EMS: Biomedical Sciences
University of Edinburgh
Honours Project

Biomedical Sciences
2023-24

B200694

Phylogeographic Analysis of the Global Spread and Evolution of African Swine Fever Virus

Analysis Project - Data
4988 words

By submitting this assignment I confirm I have adhered to the University of Edinburgh's student code of conduct and that unless otherwise indicated all this work is my own and, where applicable, full acknowledgement is made of work carried out in collaboration with, or by, other colleagues.

BIOMEDICAL TEACHING ORGANISATION

Retention of course work form

With your consent and with your name removed, your dissertation may be retained by the Deanery of Biomedical Sciences for 10 years and made available to other students. Note, however, that we cannot anonymise video recordings. This collection of work is held to assist future students and you may have used this resource in planning your own projects. If you do not consent to your work being retained, please specify this below and we will destroy the work in accordance with Taught Assessment Regulation 49.

https://www.ed.ac.uk/files/atoms/files/taughtassessmentregulations.pdf

If in future you wish to withdraw consent, please email bmto@ed.ac.uk

Please select one of the following options

I consent to my work being retained and made available to future students.	X
I do NOT consent to my work being retained and made available to future students.	
My work contains video recording and I understand that it might be retained and made available to future students. I consent to this with full understanding that it cannot be anonymized.	

Table of Contents

Acknowledgements:

I am incredibly grateful to Dr. Samantha Lycett and Dr. Gianluigi Rossi for their guidance throughout this project. Their passion for and expert understanding of the subject matter was enormously helpful, and I could not have completed this project without them.

Lay abstract:

Elucidating the spatio-temporal dynamics of pathogen spread can inform intervention strategies against African Swine Fever (ASF), a deadly viral disease of domestic pigs and wild boar. Originally endemic to Africa, the ASF virus (ASFV) has spread rapidly throughout Europe and Asia since 2007, causing severe losses to the pork industry and threatening food security. Here, we analysed the genome sequences of over 150 virus isolates from around the world to track ASFV movements and evolution over the past few decades. The available cleaned data were processed using a computational model which uses the differences between the sequences—as a result of mutations over time—to estimate the virus' genealogy. We identified many transmission events covering a vast distance in a short period of time, hinting that international trade of contaminated pork products may potentially drive dissemination. Lastly, the report looked at an ASFV variant thought to have an increased mutation rate. Though this was not evident in the model, any potential effect may only be observable in specific genes. The variant, which contains a mutation in the O174L gene, did show subtle differences in its genome structure, however, suggesting that it may be biased toward certain mutations.

Abstract:

African Swine Fever (ASF) is a deadly haemorrhagic disease of domestic pigs and wild boar. Originally endemic to Africa, it was identified in Georgia in 2007 and has since spread rapidly throughout Europe and Asia with devastating consequences for animal health and food security. In this report, we applied Bayesian phylogenetic and discrete trait analysis techniques to 153 ASF virus (ASFV) genotype II whole-genome sequences. The model established evolutionary relationships between isolates from which to reconstruct ASFV movements and estimate outbreak origins, elucidating its spatio-temporal transmission dynamics. Multiple transmission events covering vast distances in a short time frame highlight international trade in contaminated pork products as a potential driver of ASFV dissemination. East Asia was identified as a transmission hotspot, being the source of incursions to many Southeast Asian nations and Hungary. Lastly, the report looked at viruses harbouring a mutation in the ASFV polymerase X gene, O174L, which has been linked to increased rates of variant emergence in Germany. The analysis found no evidence of an increased clock rate at a genome-wide level, though potential effects may be specific to protein-coding genes. Differences in variants' dinucleotide composition were found, however, such that O174L variants may possess a substitution bias.

Introduction:

African Swine Fever (ASF), a lethal haemorrhagic disease of domestic and wild pigs, poses a severe threat to global food security and animal health. Its etiological agent is African Swine Fever Virus (ASFV), a large, unsegmented, double-stranded (ds)DNA virus comprising 24 genotypes. Of these, genotypes I and II are the only to have spread outside of Africa. Since an incursion of the highly virulent genotype II in Georgia in 2007, it has swept through much of Eurasia (Figure 1) and constitutes most of the present disease burden (reviewed by Li *et al*., 2022). Outbreaks have persisted in the European Union since its introduction in 2014, with mortality in domestic pigs approaching 100% (EFSA Panel on Animal Health and Welfare (AHAW), 2015; Blome *et al*., 2012; Gallardo *et al*., 2015). In China, the total economic loss caused by ASF outbreaks from August 2018 to July 2019 alone was estimated at 111.2 billion USD, equivalent to 0.78% of China's gross domestic product in 2019 (You *et al*., 2021).

As there is currently no effective vaccine, disease control efforts require an understanding of ASFV's transmission dynamics. Biological vectors include soft ticks of the *Ornithodoros* genus, while direct transmission occurs via animal-animal contact (World Organisation for Animal Health, 2022). However, as ASFV also persists in fomites and transmits through feeding on infected pork products, wild *Sus scrota* subspecies are regarded as an important reservoir when considering transmission dynamics in Eurasia. To this end, implementing and upholding effective biosecurity and surveillance measures are crucial to combatting further ASFV spread. Phylogenetic and phylogeographic approaches can inform such strategies by elucidating viral evolutionary and spatial histories.

The ASFV genome is highly stable and sequences can exhibit strong nucleotide sequence identity (Forth *et al*., 2023). This challenges phylogenetic studies where whole-genome data is lacking, as single locus sequences from localised samples often do not show sufficient heterogeneity (Mazloum *et al*., 2023). Recently, however, multiple genotype II variants have emerged in Germany. These variants possess novel mutations in different genes but all share a 14 base pair tandem duplication in the O174L gene encoding ASFV polymerase X (polX), which results in a frameshift and protein truncation (Forth *et al*., 2023). Structural changes to the polX enzyme induced by this insertion are thought to be linked to increased ASFV mutation rates.

This project aims to map and analyse the global spread and evolution of African Swine Fever Virus using whole-genome viral sequence data and phylodynamic modelling. We aim to

establish evolutionary relationships between isolates from which to reconstruct ASFV movements through space and time, estimate outbreak origins, and infer routes of spread. Lastly, given the links between the O174L mutation and high rates of variant emergence, we test the hypotheses that variants harbouring the mutation exhibit an increased clock rate or dinucleotide composition bias. Composition biases have also been reported in Mpox as evidence of selective pressure imposed by the host immune response (O'Toole *et al*., 2023). Thus, we also investigate whether host factors may be driving viral adaptation producing a dinucleotide composition bias in the ASFV genome.

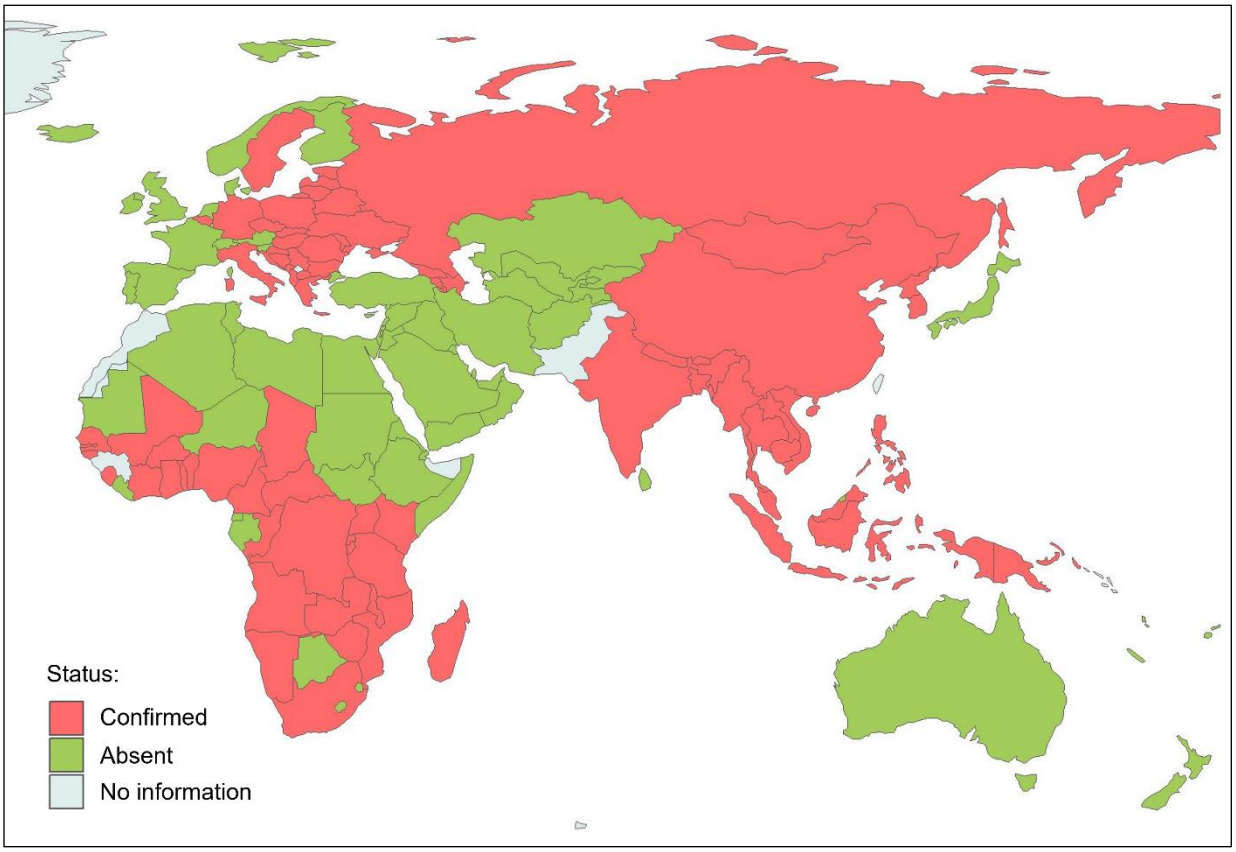


Figure 1: Infection status of African Swine Fever Virus in Africa, Europe, Asia, and Oceania (2007 – 2024). Countries highlighted red have had at least one confirmed case of ASFV between 2007 and 2024, while countries highlighted green have not detected ASFV within this period. Created by author using data obtained from the World Animal Health Information System (WAHIS, 2024).

Methods:

Sequence and metadata retrieval:

146 sequences were obtained from GenBank and the European Bioinformatics Institute (EBI), and a further 10 pre-processed sequences were sent to us by collaborators at the National Reference Center for the Study of Pestivirus and Asfivirus Diseases (CEREP) in Italy. All 156 sequences were aligned in MAFFT. Aligned sequences were each ~187 kb in length. For each sequence, metadata including sampling location, world region, sampling date, host type and O174L gene identity was obtained. Sampling location was represented as paired latitude and longitude values, pinpointed to the geographic centroid of the most specific location found in Genbank or associated publications. In two cases, both in China, we could not find any obvious region associated to the sequence: as the country was deemed too large, the province of the institute to have sequenced the isolate was used. Depending on their country or region of origin, sequences were assigned to one of six world macro-areas based on an aggregation of UN sub-regions (Supplementary Figure A). Sampling date was taken as the midpoint of the smallest time interval given (e.g mid-month, or mid-year), in cases where a specific date was not provided. Host identities are specified in Genbank as either domestic pig or wild boar. Presence of the 14 bp tandem duplication mutation in the O174L gene was determined from genome sequences in MEGA (Supplementary Figure B).

Model selection:

Bayesian Evolutionary Analysis Sampling Trees (BEAST) is a program for Bayesian analysis of molecular sequences. BEAST models use Markov Chain Monte Carlo (MCMC) algorithms to infer values taken by multiple evolutionary parameters, such as divergence times and evolutionary rates, by repeated sampling of probability distributions (Suchard et al., 2018). Obtaining the most useful output requires a process of model selection. Models were constructed in BEAUTi v1.10.4 with sampling dates and date uncertainty values assigned to each sequence and run in BEAST v1.10.4 using MCMC. Preliminary model selection was conducted for a variety of substitution models, clock models and models of population size (tree priors). Their performance was evaluated based on marginal likelihood estimation values, effective sample sizes, and log file exploration in Tracer v1.7.2. At this stage, three sequences (from Armenia, Estonia, and China) were removed from the analysis as outliers in a root-to-tip plot of divergence against sampling date, which was constructed in TempEst v1.5.3 (Rambaut

et al., 2016). The selected models included a General Time Reversible (GTR, Gamma + Invariant sites = 4) substitution model, an uncorrelated relaxed lognormal clock model, and a Bayesian Skygrid (35/35) tree prior.

Tree construction:

The origin and inter-regional spread of ASFV lineages were estimated by combining Bayesian phylogeny with discrete trait analysis (DTA) (Lemey et al., 2009). Using the chosen models, we ran BEAST with four traits of interest: Location (latitude and longitude, continuous), O174L gene identity (binary), host type (binary) and world region (discrete). Upon further inspection in Tracer v1.7.2, we found that including spatial and discrete model components in the initial run dramatically reduced model performance. We therefore used subsets of 1,001 trees from the original posterior distributions of trees to generate empirical tree distributions with trait partitions (Duchatel, Bronsvoort and Lycett, 2019). In total, we generated 10,001 post-burnin posterior trees (i.e. an average of 10 trees with traits to every one original tree). A maximum clade credibility (MCC) tree was created in TreeAnnotator v1.10.4.

Data analysis and visualization:

All data analysis and visualization was performed in R studio (R 4.3.1). Transition and clock rates were obtained from branch information extracted from the log output file of the BEAST run. When comparing clock rates, significance was determined by randomly sampling 10,000 posterior observations (of the rate, normalized to its tree's mean rate weighted by branch length, and then log-transformed) for branches in Europe with and without the O174L tandem duplication mutation, and comparing the distribution of differences against the Region of Practical Equivalence (ROPE) as appropriate for data generated by Bayesian inference (Makowski, Ben-Shachar and Lüdecke, 2019). We defined ROPE values as one-tenth the standard deviation of the log-transformed data, according to convention. Genome composition was analysed using the original MAFFT alignment of all 153 sequences. Dinucleotide count was estimated by multiplying the relative proportion of dinucleotides by the length of the alignment minus 1. As the data violated assumptions of normality which could not be rectified by data transformation, the Wilcoxon rank sum test for non-parametric data was used to test differences in dinucleotide count among O174L gene identities and host types. Bonferroni correction was applied for n=16 dinucleotides. In the absence of appropriate methods for calculating confidence intervals (CIs) for data following varied distributions, transition rate and dinucleotide count 95% CIs were calculated as the range between the 2.5^{th} and 97.5^{th} percentiles.

Results:

Phylogeographic analysis of global and inter-regional spread:

In order to infer evolutionary relationships between isolates, we constructed a time-scaled maximum clade credibility (MCC) tree with discrete trait information for 153 ASFV genotype II WGS (Figure 2). To conduct phylogeographic analysis, spatial data corresponding to each branch, node and tip were then plotted onto a world map at three different time points (Figure 3A-C). Overall, we observed a mean substitution rate of 1.297×10^{-5} substitutions per site per year (95% highest posterior density (HPD) interval $9.334 \times 10^{-6} – 1.733 \times 10^{-5}$). All sequences included in the analysis can be traced back to East Africa (root node posterior probability: 100%; region probability: 98%; most recent common ancestor (MRCA) median coordinates: 13.579°S, 41.294°E) in the late 1980s (median time to MRCA (tMRCA): 1989; 95% HPDs 1960.7 – 1998.5). The genetic diversity of African sequences, sampled between 1998 and 2022, hints to endemic circulation of the virus in the continent. As of 2015 (Figure 3A), ASFV Genotype II had spread locally to the East African countries of Madagascar, Mauritius, Malawi, Mozambique, and Tanzania, as well as Eastern Europe and the Caucasus, representing the first transcontinental incursion of the virus. ASFV may have reached West Africa (Ghana and Nigeria) more recently as a consequence of a single long-distance transmission event; West African sequences form a clade with a median tMRCA in 2018 (95% HPDs 2014.5 – 2020.1).

All sequences outside Africa comprise a monophyletic clade and originate in Eastern Europe and the Caucasus (node A posterior probability: 97%; region probability: 97%; median tMRCA: 2005; 95% HPDs 2003.1 – 2007.2). The introduction of ASFV from Africa to Eastern Europe and the Caucasus was estimated as in 2005 (node B 95% HPDs 2001.8 – 2007.0). The European epidemic is paraphyletic, likely caused by multiple introductions from the Trans-Caucasus region. At some point in the early 2010s (node C median tMRCA: 2012; 95% HPDs 2009.7 – 2014.9), a divergence occurred of which one lineage would eventually reach Poland, Ukraine, Germany, and the Russian exclave of Kaliningrad, and the other the larger geographical area encompassing Moldova, the Czech Republic, Serbia, Belgium, Italy, as well as Asia (node C posterior probability: 100%). The 14bp tandem duplication mutation in the O174L gene was found in all German and four Polish sequences (highlighted in pink). As of 2019 (Figure 3B), ASFV had spread west through Europe and east into Asia, where it was first identified in China in 2018 (Li et al., 2022).

All Asian sequences are monophyletic and originate in East Asia (node D posterior probability:

95%; region probability: 96%; median tMRCA: 2016; 95% HPDs 2014.8 – 2017.6), migrating from Eastern Europe and the Caucasus in 2015 (node E 95% HPDs 2012.7 – 2017.1). Within the East Asian lineage, there are few clades in which sequences share a single country of origin and represent the majority of sequences from that country, indicating much transboundary spread between non-contiguous areas. Two notable exceptions include a clade comprising all sequences from India (posterior probability: 99%; median tMRCA: 2019; 95% HPDs 2018.4 – 2020.0) and a clade comprising six sequences from South Korea (posterior probability: 99%; tMRCA: 2019; 95% HPDs 2018.6 – 2019.5). Between 2019 and 2023, ASFV spread further to multiple Southeast Asian nations, including the Philippines, Singapore, Vietnam, Indonesia, and Timor-Leste (Figure 3C). Notably, the results suggest two instances of ASFV transmission from East Asia back to Europe. A single sequence from Hungary, sampled in 2018, is descended from the East Asian lineage. So too are four Serbian sequences, sampled between 2019 and 2022. Neither of these two intercontinental transmissions are estimated to have spread outside the countries where they were first identified.

Inter-regional ASFV transmission was further defined using a transition matrix reporting the mean number of transitions across 10,001 trees (Figure 3D). The results support the notion that ASFV arrived in Eurasia by a single incursion from Africa (mean: 1.0; median: 1; 95% CIs 1 – 2). East Asia was the dominant source of inter-regional spread, primarily to Southeast Asia (mean: 13.6; median: 14; 95% CIs 7 – 18) but also to Eastern Europe and the Caucasus (mean: 1.0; median: 1; 95% CIs 0 – 2) and Southern Europe (mean: 1.1; median: 1; 95% CIs 1 – 2). Eastern Europe may also be considered a hotspot of ASFV dissemination, being the source of transmission events to Western Europe (mean: 3.2; median: 3; 95% CIs 2 – 5) and East Asia (mean: 1.0; median: 1; 95% CIs 0 – 2).

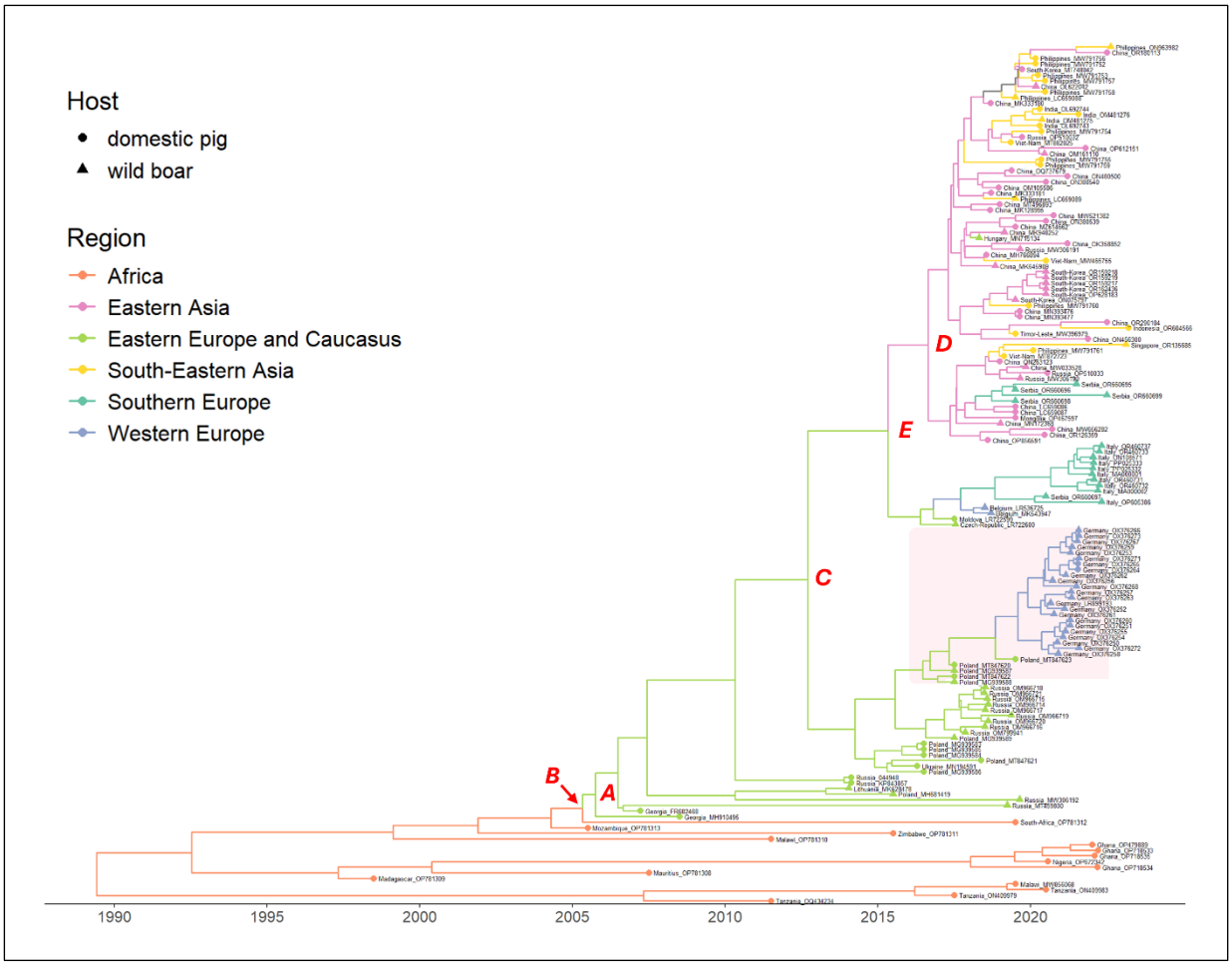


Figure 2: Time-scaled phylogenetic tree of African Swine Fever Virus whole-genome sequences: All 153 sequences that were included in the study are displayed. Branch colours represent ancestral regions, and tip shapes represent host types. Sequences highlighted in pink all contain an identical tandem duplication mutation in the O174L gene. Tip labels refer to individual sequences' country of origin and accession number. Branches coloured grey represent those whose region could not be distinguished between East and Southeast Asia. Nodes A – E are referred to in the text.

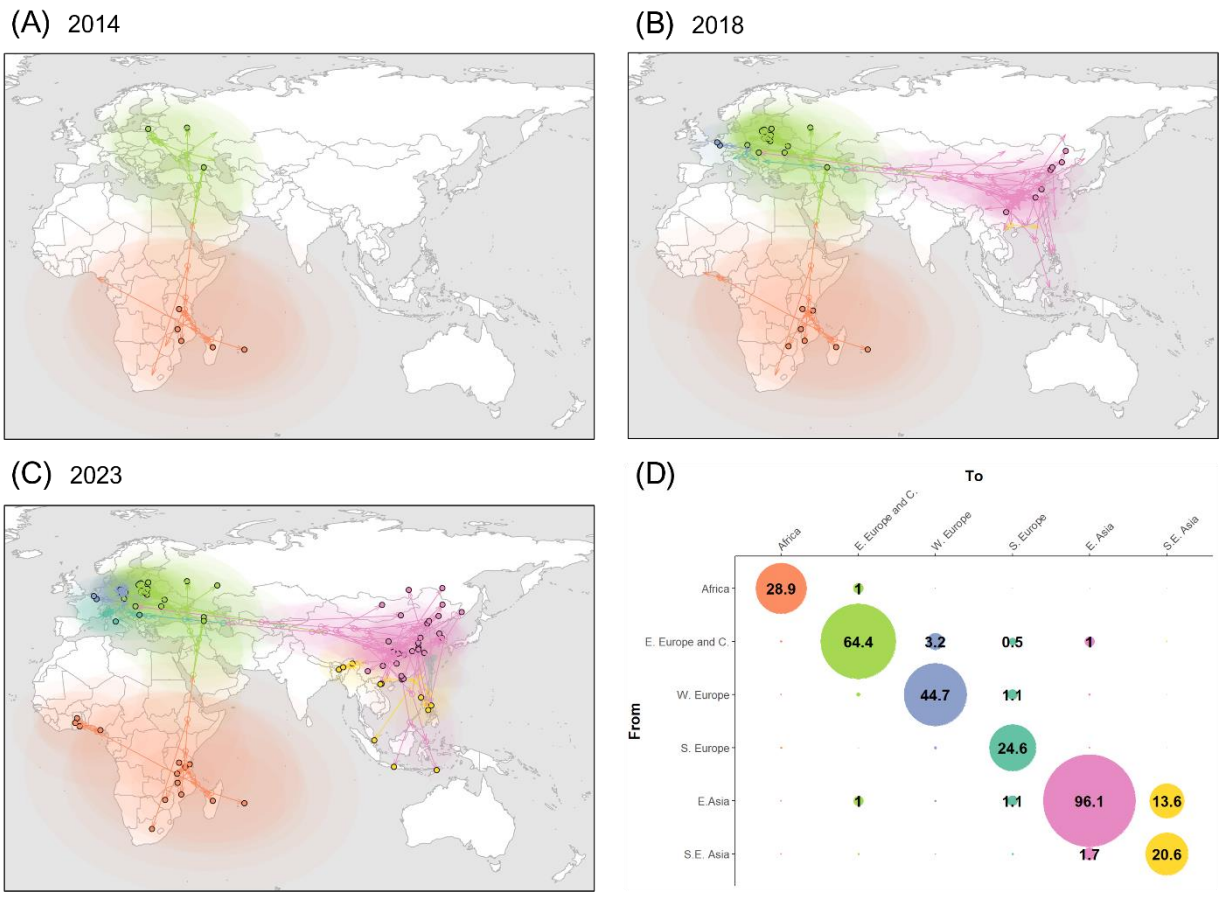


Figure 3: Global and inter-regional spread of African Swine Fever Virus: (A-C) Spatial tree of ASFV spread up to and including the years **(A)** 2014, **(B)** 2018 and **(C)** 2023. Solid circles represent ASFV sequence locations. Hollow circles represent node locations, with semitransparent clouds representing their respective 80% HPDs. Arrows represent branches pointed in the direction of spread. Arrows and nodes coloured grey represent those whose region could not be distinguished between East and Southeast Asia. **(D)** Transition matrix showing number of transitions from a node of one region to a node of another region, averaged over 10,001 trees. Values of transition frequencies of > 0.25 are displayed in text on the graph. Colours correspond to region: Africa = orange; Eastern Europe and Caucasus = green; Western Europe = blue; Southern Europe = mint; East Asia = pink; Southeast Asia = yellow.

European and Asian theatres of ASFV dissemination:

ASFV dissemination in Europe and Asia is reported in Figure 4. International dissemination events of ASFV within mainland Asia (Figure 4A) include transmissions to Eastern Russia, India, South Korea and Mongolia. The map suggests at least four introductions of ASFV to Eastern Russia. All Indian sequences are located in the country's Northeastern region, with an MRCA in China but close to its border with India (median coordinates: 27.751°N, 99.199°E), such that a single incursion from China is plausible. The clade containing six South Korean sequences may represent one of two inferred viral incursions to the country, with the other first identified in

the northern Gyeonggi Province. Figure 4A also suggests that there have been three introductions of ASFV to Vietnam, all appearing to have originated in China.

Long-distance transmissions of ASFV to Southeast Asia include those to Singapore, Indonesia, Timor-Leste, and the Phillippines (to which multiple introductions of ASFV are shown). However, their origins are difficult to infer precisely. Given that the earliest ASFV isolates in East Asia were all identified in China and node 80% HPDs are generally confined to within China's borders, it can be inferred that much of the reported transmission events originate in China. However, it cannot be ruled out that South Korea may also have exported ASFV internationally.

In the European theatre, transmissions are inferred across a range of distances (Figure 4B). All German sequences, which form a unique clade comprised mostly of sequences isolated from wild boar, share an MRCA in Western Poland (median coordinates: 51.900°N, 15.520°E) and cluster with a Polish sequence near the German-Polish border, indicating Poland as the source of this incursion. The Kaliningrad cluster, which is entirely composed of wild boar sequences, also share an MRCA located in Poland (median coordinates: 53.776°N, 21.433°E). Taken together, the results show that international ASFV transmission events to Germany and Kaliningrad may have been partly mediated by wild boar migration from Poland.

The results do not provide sufficient evidence to determine the origins of outbreaks in Belgium and Italy. There is an absence of non-Belgian sequences near the Belgian border, and ASFV has not yet been reported in the neighbouring countries of France, the Netherlands, or Luxembourg (World Animal Health Information System, 2024). While ASFV is present in Germany, these sequences are localised along its eastern border, and Figure 4B indicates that ASFV was introduced to Belgium via a different route than to Germany. Figure 4B suggests that ASFV may have been introduced to Italy twice, with incursions identified in the Lazio and Piedmont regions. No country with which Italy shares a border has reported ASFV infection (World Animal Health Information System, 2024). It is plausible, therefore, that human activity mediated these transmission events, though wild boar migration from Croatia to Italy may be possible.

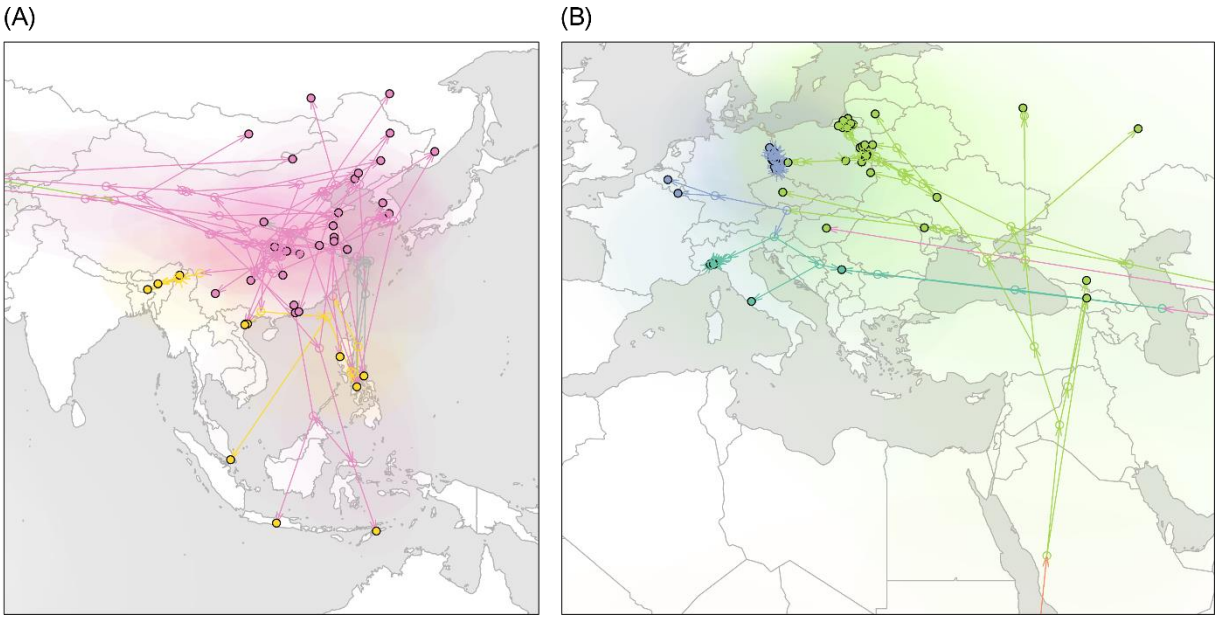


Figure 4: Spread of African Swine Fever Virus within (A) Asia and (B) Europe: Spatial tree of ASFV spread including all 153 ASFV WGS. Solid circles represent ASFV sequence locations. Hollow circles represent node locations, with semitransparent clouds indicating their respective 80% HPDs. Arrows represent tree branches, pointed in the direction of spread. Arrows and nodes coloured grey represent those whose region could not be distinguished between East and Southeast Asia. Colours correspond to region: Africa = orange; Eastern Europe and Caucasus = green; Western Europe = blue; Southern Europe = mint; East Asia = pink; Southeast Asia = yellow.

The O174L tandem duplication mutation is not conferring an increased molecular clock rate:

Recently, a uniquely high rate of variant emergence was reported in Germany, with all novel variants sharing a 14 bp tandem duplication in the O174L gene encoding the ASFV polX enzyme (Forth *et al*., 2023). Estimations of mutation origin are complicated by the presence of a Polish sequence (accession number: MG939588) lacking the O174L gene mutation in the same clade as those with it. By taking only posterior probabilities of over 95%, its emergence can be traced to Eastern Poland (98% posterior probability; median location: 52.355°N, 22.659°E) sometime in 2016 (80% HPDs 2015 – 2017). The O174L mutation has since spread to western Poland and eastern Germany (Figure 5A).

Given the increased rate of variant emergence in viruses harbouring the O174L gene insertion, it is hypothesised to have increased viral mutation rates (Forth *et al*., 2023). When inferring the original time-scaled trees, we used an uncorrelated relaxed lognormal clock model allowing each branch to have its own clock rate drawn from a log normal distribution. Therefore, we

tested this hypothesis by comparing model clock rates among branches within Europe according to the models' estimation of O174L mutation presence. Figure 5B shows the log-transformed relative clock rates from the posterior time-scaled trees of branches with (pink) and without (blue) the mutation. This analysis shows a lower relative clock rate for mutated branches (mean: 0.79; mean and 95% HPDs of log-transformed data: -0.38, -1.378 – 0.594) than for unmutated branches (mean: 1.23; mean and 95% HPDs of log-transformed data: -0.28, -1.362 – 0.862) but found the effect to be of undecided significance, as 6.21% of the 95% CIs of the posterior distribution of the effect overlaps with the region of practical equivalence (ROPE).

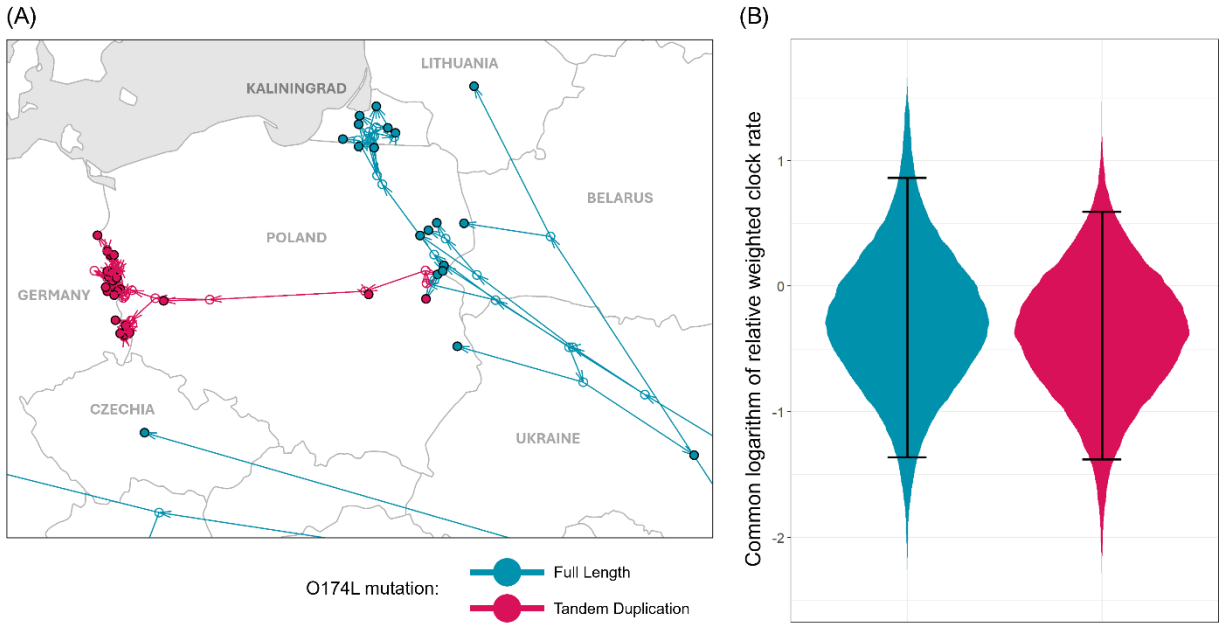


Figure 5: The O174L tandem duplication mutation is not conferring an increased molecular clock rate: (A) Spatial tree of ASFV spread in Eastern Europe, coloured according to the presence (pink) or absence (blue) of the O174L tandem duplication mutation. Solid circles represent ASFV sequence locations. Hollow circles represent node locations. Arrows represent tree branches, pointed in the direction of spread. **(B)** Violin plot of the common logarithm of clock rate, relative to its original tree's mean (with the mean weighted by the sum of branch lengths) among branches within Europe estimated to lack (blue) or have (pink) the O174L tandem duplication mutation, plotted for all 10,001 generated trees. Branch clock rates were normalized to the mean rate of the tree, where the mean was weighted by the sum of branch lengths. Error bars represent 95% highest posterior densities (HPDs) of log-transformed data.

ASFV genomes may be subject to substitution biases:

We also hypothesised that the mutated polX enzyme possesses a substitution bias which may account for observed rate of variant emergence. To test this, we compared genome-wide dinucleotide compositions of ASFV genomes with and without the O174L tandem duplication mutation (Figure 6A). The Wilcoxon rank sum test revealed a number of significant differences

between dinucleotide counts when Bonferroni correction (n=16 dinucleotides) was applied. The distribution of AG, CA, GT, and TA dinucleotide counts was significantly different between sequences with and without the O174L mutation, with O174L-mutated isolates having higher mean counts than unmutated isolates (p<0.001 for all tests). The distributions of GC and GG dinucleotides were also significantly different, but with O174L-mutated sequences having a lower mean count than unmutated ones (p<0.001 for both tests). CG dinucleotide content and methylation in viral genomes is widely researched due to its effects on viral life cycles and immune recognition (Hoelzer, Shackelton and Parrish, 2008). The analysis reported that the distribution of CG count was significantly different between groups (p<0.001), with O174L-mutated isolates having a lower mean count (mean: 6190.6; 95% CIs 6190.12 – 6191.39) than isolates lacking the mutation (mean: 6201.48; 95% CIs 6190.73 – 6306.17). All means, medians and 95% CIs are given in Supplementary Table A. Though the test is robust to outliers, the means are not, and differences between means are small considering the large size of ASFV genomes. Thus, we conclude that these results support the hypothesis but warrant further investigation at the level of individual gene loci.

Host factors can also drive dinucleotide composition biases. This has been described for Mpox virus, whereby a higher-than-expected rate of TC>TT and GA>AA mutations was shown to be driven by the human antiviral enzyme APOBEC3 (O'Toole *et al*., 2023). Thus, we also tested whether dinucleotide composition differed by host type (Figure 6B). After Bonferroni correction (n=16), the analysis identified three dinucleotides reporting a difference between count distributions. This includes a significant difference in CG count distribution (p<0.001), with domestic pig isolates (mean: 6206.7; 95% CIs 6190.3 – 6310.8) having a higher mean count than wild boar isolates (mean: 6192.2; median: 6191.8; 95% CIs 6190.3 – 6199.2). This was also the case for TC (p<0.01), while mean counts of dinucleotide AA were higher in wild boar (p<0.01). All means, medians and 95% CIs are given in Supplementary Table B. As with Figure 6A, the test is robust to outliers but differences in means are relatively small. The results support the hypothesis that a difference in host factors is conferring a substitution bias, but further investigation is required.

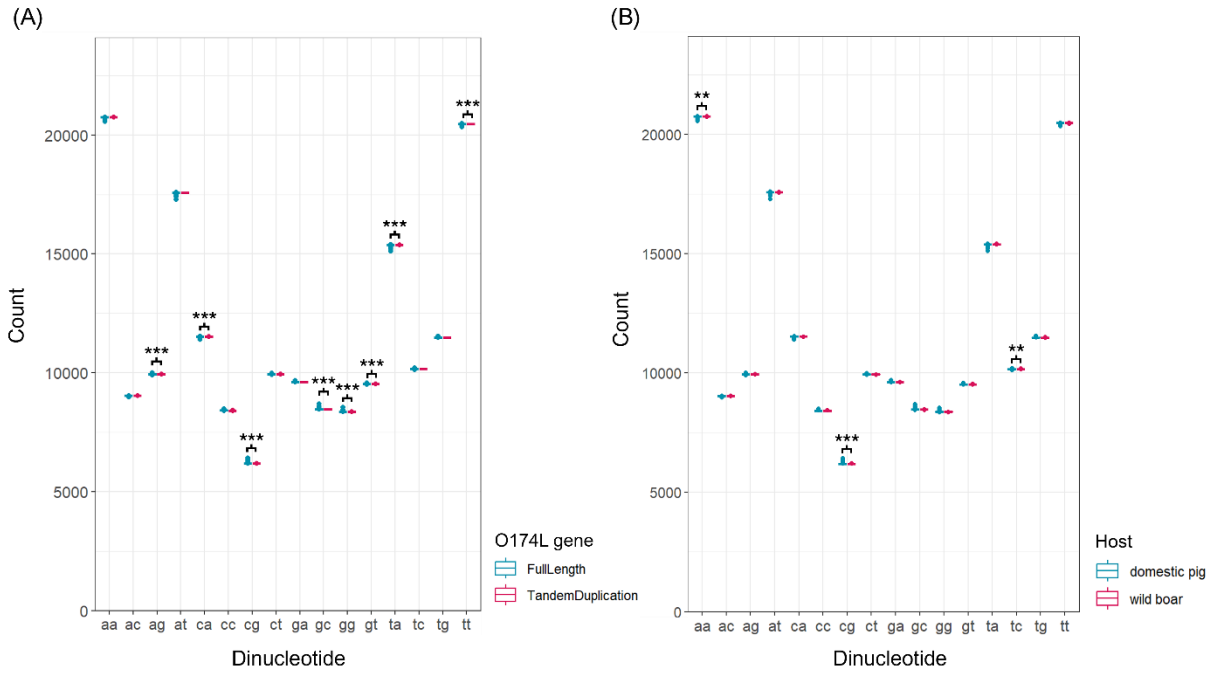


Figure 6: ASFV may be subject to substitution biases: Boxplots showing dinucleotide counts in ASFV genomes based on **(A)** O174L gene identity or **(B)** host type. Dinucleotide counts were estimated by multiplying dinucleotide proportions by sequence length *n* – 1. Differences in dinucleotide count were tested using the Wilcoxon rank sum test for non-parametric data. Bonferroni correction was applied for n = 16 dinucleotides. Asterisks indicate: * = p < 0.05; ** = p < 0.01; *** = p < 0.001.

Discussion:

Merits and limitations:

In this report, we have applied Bayesian phylogenetic analysis methods in BEAST v1.10.4 to 153 ASFV genotype II whole-genome sequences (WGS) from across Africa, Europe, and Asia. This methodology is upheld by three main assumptions: (i) There are enough differences between sequences to infer evolutionary relationships; (ii) the molecular clock signal is sufficient to estimate divergence times; (iii) the virus does not undergo genetic recombination.

ASFV's high genetic stability hinders the definition of reliable genetic markers; single locus sequences often do not demonstrate sufficient heterogeneity to infer meaningful tree topologies or a robust evolutionary rate (Mazloum *et al*., 2023). This obstacle is partly overcome by capturing genetic differences distributed across whole genomes, thus improving the resolution and accuracy of the analysis and satisfying the first assumption (Kao *et al*., 2014). However, sequences are still highly similar and the spatial and temporal resolution, though sufficient, is restricted accordingly. This is a barrier imposed by ASFV itself and underlines the traditional application of phylogenetic analysis to RNA viruses with a higher evolutionary rate.

One limitation of our analysis is that the phylogenetic model assumes no sequence recombination. When not accounted for, recombination can precipitate excessively long terminal branches, an underestimation of the time to most recent common ancestor, an overestimation of the number of mutations, or a loss of a molecular clock signal (Schierup and Hein, 2000). Whilst recombination in ASFV is unlikely in Europe/Asia in the study data set (Li *et al*. 2020), limited recombination has been recently reported for ASFV Genotype II (Kwon *et al*., 2024; Zhao *et al*., 2023), and in Genotype I/II African sequences (Li *et al*., 2020). However, model evaluation in TempEst confirmed that our clock signal was not lost and may have identified recombinant sequences. Recombination-aware phylogenetic analysis can elucidate the movement of genetic material under all evolutionary scenarios and has been applied to ASFV (Li *et al*., 2020), but is beyond the scope of this report.

Another limitation of the analysis is the scarcity of available sequences from multiple ASFV-affected countries which may offer useful insight into ASFV global transmission dynamics. The Balkan states shoulder a significant proportion of Europe's current ASF burden, but our only sequences in this region come from Serbia. 1,660 outbreaks were reported in Romania in 2021 alone (Balmoș *et al*., 2023). Between May 2023 and January 2024, there were 1,510 reports of

ASF in domestic pigs in Bosnia and Herzegovina and 293 in Croatia (Department for Environment, Food & Rural Affairs, 2024). Ukrainian sequences, of which only one was included in the analysis, may lend important insight to our understanding of ASFV's historical spread, having been prevalent there since 2012 (Omelchenko *et al*., 2022). Our data also does not represent the extent of ASFV infection in Southeast Asia (Hsu et al., 2023). Despite these gaps, Bayesian phylogeographic methods are robust to unbalanced and sparse sampling, which is further helped by including data from a wide timespan. Nonetheless, access to sequences from these and other nations would improve spatial and temporal estimations of ASFV movement.

Outbreak origins and mechanisms of spread:

Wild boar represent an important viral reservoir in Eurasia and a formidable challenge to ASF containment. This is often attributed to transmission via contaminated boar carcasses, in which ASFV can persist for several weeks, and the high reproductive capacity of wild boar, which maintains susceptible populations (Chenais *et al*., 2019; Schulz *et al*., 2019). Results from our phylogeographic analysis point to wild boar as having a visible role in the international transmission of ASFV in Poland, Germany, and Kaliningrad. Sauter-Louis *et al*. (2020) concluded that an introduction of ASFV to Germany by wild boar migrating from Poland is likely, while Mazloum *et al*. (2023) corroborate the close evolutionary link between sequences from Poland, Germany, and Kaliningrad isolated from wild boar. Without direct evidence, however, we cannot conclude that these transmission events were mediated by wild boar alone. Wild boar are considered relatively sedentary (Miettinen et al., 2023), such that they may play a greater role in sustaining local ASFV transmission than facilitating its movement into new regions.

Much ASFV transmission has been attributed to the consumption of contaminated pork products, feed, or contact with fomites as a product of human activity. ASFV is highly durable outside the host; in meat stored at 4–8°C, a viable virus could be detected for up to ~150 days and may remain infectious in frozen meat for several years (EFSA Panel on Animal Health and Welfare (AHAW), 2014). Feed contaminated by infectious blood preserved viable ASFV for 30 days at 4°C, and in water for 50 days at room temperature (Sindryakova *et al*., 2016). This survivability greatly elevates ASFV's potential for long-distance transmission. We report many incursion events for which anthropogenic factors may be responsible. In Europe, these include introductions to Italy and Belgium, and in agreement with Mazur-Panasiuk *et al* (2020), the jump of ASFV from eastern to western Poland. Human activity likely facilitated virus migration from Europe to East Asia, and in turn, the island nations of the Philippines, Singapore, Indonesia, and Timor-Leste. Mazur-Panasiuk, Żmudzki and Woźniakowski (2019) contend that the spread

of ASFV within China is associated primarily with contaminated feed, which is supported by the many long-distance branches and widely scattered sequences seen in Figure 4A. Lastly, the first outbreaks of the virus in Georgia are thought to have been triggered by improper disposal of contaminated pork products by international ships docked in the port city of Poti (Beltrán-Alcrudo *et al*., 2008). It is important to acknowledge, however, that our phylogeographic and discrete trait analysis is more conducive to inference of viral geographic origin than mechanisms of spread.

Our estimations of inter-regional sources of spread are largely supported by the literature. In concordance with Figure 3, the Georgia 2007 outbreak and subsequent dissemination through Europe is widely accepted as being a result of virus migration from Southeast Africa (Rowlands *et al*., 2008; Quembo *et al*., 2018). ASF outbreaks in East Asia are believed to be of European origin, though more specific estimates are debated. In direct comparison with our report, two studies applying Bayesian inference of phylogeny to ASFV WGS in BEAST concluded ASFV to have been imported from central Europe and central or eastern Europe, respectively (Shen *et al*., 2022; Zhang *et al*., 2023). A third study using PhyML suggested a Western European origin (Xin *et al*., 2023). Our results suggest that the East Asian lineage is of Eastern European origin, but given the aforementioned scarcity of sequences across the region, importation of ASFV from central Europe cannot be ruled out. Notably, two of these studies corroborate the finding that the Hungarian sequence (accession number: MN715134) is descended from the East Asian lineage (Figure 2) (Xin *et al*., 2023; Zhang *et al*., 2023), though the other places it within the European transmission network (Shen *et al*., 2022). No studies offer support for an East Asian origin of ASF in Serbia.

Sources of intra-regional transmission are inferred with varying degrees of certainty. Phylogenetic analysis of South Korean ASFV sequences corroborate our finding that ASFV was likely introduced to the country more than once. However, the authors suggest that these repeated introductions were caused by wild boar migrating across the demilitarized zone (DMZ) rather than importation from China as suggested in Figure 4A (Kwon *et al*., 2024). An independent whole-genome analysis of two ASFV sequences (accession numbers: OL692743, OL692744) included in this study, both sampled during initial ASF outbreaks in India, substantiate their close evolutionary association with sequences in China (Senthilkumar *et al*., 2022). Lastly, comparative analyses of a Belgian WGS (accession number: LR536725) and two sequences from Northwest Italy were also unable to define a geographical origin for these outbreaks (Forth *et al*., 2019; Giammarioli *et al*., 2023).

Clock rates and dinucleotide composition:

Overall, we observed a mean substitution rate of 1.297×10^{-5} substitutions per site per year, similar to that estimated by other Bayesian WGS studies (Shen *et al*., 2022; Zhang *et al*., 2023).

ASFV PolX, encoded by the O174L gene, is a low-fidelity polymerase and key driver of the virus' genetic drift (Showalter *et al*., 2001). Thus, the sudden emergence of multiple variants carrying a tandem insertion in the O174L gene—which leads to a frameshift and protein truncation—has raised questions regarding its potential impact on ASFV phylodynamics. We estimate that the mutation emerged sometime between 2015 and 2016 (Figure 2) in Eastern Poland (Figure 5A) but emphasize caution due to the uncertain topology of our discrete trait MCC tree. A phylogenetic study based on concatenated O174L, K145R and IGR I73R/I329L genomic regions, by contrast, suggested that it was introduced into Poland via natural migration of wild boars from Belarus in 2015–2016 (Mazur-Panasiuk *et al*., 2020).

We hypothesised that the increased rate of variant emergence in viruses harbouring the mutation would be reflected by an increased genome-wide molecular clock rate (Figure 5B). However, upon comparing clock rates among European branches of our trees, we found the mean clock rate to be lower in branches with mutation than those without, though this effect was of undecided significance. To determine whether mutated polX enzymes and host factors are driving a substitution bias, we compared the dinucleotide composition of ASFV whole genomes. We found several dinucleotides for which a significant difference in distribution was reported, but differences in means appear to be driven by outliers and are small relative to the size of the ASFV genome. Interestingly, CG dinucleotides were slightly repressed in wild boar and in variants with the O174L mutation, perhaps reflecting a mechanism for avoiding the host immune response (Mordstein *et al*., 2021). Taken together, our results warrant further investigation of the O174L tandem duplication and its potential impact. In particular, analysis of the ratio of non-synonymous to synonymous substitutions (dN/dS) in protein-coding genes would help determine whether variant genes are positively selected and potentially advantageous (Jeffares *et al*., 2015).

Conclusion:

The global spread of African Swine Fever Virus genotype II since its arrival on the Eurasian continent has been rapid and damaging. In this report, we reconstruct the virus' evolutionary and spatial history using phylogeographic and discrete trait analysis of 153 whole-genome sequences. ASFV's environmental survivability begets a worrying capacity for transboundary spread mediated by anthropogenic factors. Our results corroborate the literature regarding the inter-continental sources of ASFV transmission, highlight East Asia as a platform for ASFV dissemination and possible source of an ASF outbreak in Hungary, and reinforce the need for heightened surveillance and sharing of sequence data. Lastly, a novel mutation in the ASFV polymerase X gene appears to be influencing ASFV evolutionary dynamics and warrants further research.

References:

Balmoș, O.M., Supeanu, A., Tamba, P., Horváth, C., Panait, L.C., Sándor, A.D., Cazan, C.D., Ungur, A., Motiu, M., Manita, F.A., Ancuceanu, B.C., Bărbuceanu, F., Dhollander, S. and Mihalca, A.D. (2023). African Swine Fever Virus Load in Hematophagous Dipterans Collected in Outbreaks from Romania: Risk Factors and Implications. *Transboundary and Emerging Diseases*, 2023, pp.1–9. doi:https://doi.org/10.1155/2023/3548109.

Beltrán-Alcrudo, D., Lubroth, J., Depner, K. and De La Rocque, S. (2008). *EMPRES Watch: African swine fever in the Caucasus*. [online] *FAO.org*, Food and Agriculture Organisation of the United Nations (FAO), p.2. Available at: https://www.fao.org/publications/card/en/c/1111d08d-5e2a-5819-b49d-39bcfc121fed/ [Accessed 18 Apr. 2024].

Blome, S., Gabriel, C., Dietze, K., Breithaupt, A. and Beer, M. (2012). High Virulence of African Swine Fever Virus Caucasus Isolate in European Wild Boars of All Ages. *Emerging Infectious Diseases*, 18(4). doi:https://doi.org/10.3201/eid1804.111813.

Chenais, E., Depner, K., Guberti, V., Dietze, K., Viltrop, A. and Ståhl, K. (2019). Epidemiological considerations on African swine fever in Europe 2014–2018. Porcine Health Management, 5(1). doi:https://doi.org/10.1186/s40813-018-0109-2.

Department for Environment, Food & Rural Affairs and Animal and Plant Health Agency (2024). *African swine fever in Europe*. [online] *WWW.GOV.UK*. GOV.UK. Available at: https://www.gov.uk/government/publications/african-swine-fever-in-pigs-and-boars-in-europe [Accessed 17 Apr. 2024].

Duchatel, F., Bronsvoort, B.M. and Lycett, S. (2019). Phylogeographic Analysis and Identification of Factors Impacting the Diffusion of Foot-and-Mouth Disease Virus in Africa. Frontiers in Ecology and Evolution, 7. doi:https://doi.org/10.3389/fevo.2019.00371.

EFSA Panel on Animal Health and Welfare (AHAW) (2014). Scientific Opinion on African swine fever. *EFSA Journal*, 12(4). doi:https://doi.org/10.2903/j.efsa.2014.3628.

EFSA Panel on Animal Health and Welfare (AHAW) (2015). African swine fever. *EFSA Journal*, 13(7). doi:https://doi.org/10.2903/j.efsa.2015.4163.

Forth, J.H., Calvelage, S., Fischer, M., Hellert, J., Sehl-Ewert, J., Roszyk, H., Deutschmann, P., Reichold, A., Lange, M., Thulke, H.-H., Sauter-Louis, C., Höper, D., Mandyhra, S., Sapachova, M., Beer, M. and Blome, S. (2023). African swine fever virus – variants on the rise. Emerging Microbes & Infections, 12(1). doi:https://doi.org/10.1080/22221751.2022.2146537.

Forth, J.H., Tignon, M., Cay, A.B., Forth, L.F., Höper, D., Blome, S. and Beer, M. (2019). Comparative Analysis of Whole-Genome Sequence of African Swine Fever Virus Belgium 2018/1. Emerging Infectious Diseases, 25(6), pp.1249–1252. doi:https://doi.org/10.3201/eid2506.190286.

Giammarioli, M., Alessandro, D., Cammà, C., Masoero, L., Torresi, C., Marcacci, M., Zoppi, S., Curini, V., Rinaldi, A., Rossi, E., Casciari, C., Pela, M., Pellegrini, C., Iscaro, C. and Feliziani, F. (2023). Molecular Characterization of the First African Swine Fever Virus Genotype II Strains Identified from Mainland Italy, 2022. Pathogens, [online] 12(3), p.372. doi:https://doi.org/10.3390/pathogens12030372.

Gallardo, C., Soler, A., Nieto, R., Cano, C., Pelayo, V., Sánchez, M.A., Pridotkas, G., Fernandez-Pinero, J., Briones, V. and Arias, M. (2015). Experimental Infection of Domestic Pigs with African Swine Fever Virus Lithuania 2014 Genotype II Field Isolate. *Transboundary and Emerging Diseases*, 64(1), pp.300–304. doi:https://doi.org/10.1111/tbed.12346.

Hoelzer, K., Shackelton, L.A. and Parrish, C.R. (2008). Presence and role of cytosine methylation in DNA viruses of animals. *Nucleic Acids Research*, 36(9), pp.2825–2837. doi:https://doi.org/10.1093/nar/gkn121.

Hsu, C.-H., Schambow, R., Montenegro, M., Miclat-Sonaco, R. and Perez, A. (2023). Factors Affecting the Spread, Diagnosis, and Control of African Swine Fever in the Philippines. Pathogens, [online] 12(8), pp.1068–1068. doi:https://doi.org/10.3390/pathogens12081068.

Jeffares, D.C., Tomiczek, B., Sojo, V. and dos Reis, M. (2015). A Beginners Guide to Estimating the Non-synonymous to Synonymous Rate Ratio of all Protein-Coding Genes in a Genome. In: *Parasite Genomics Protocols*. New York, NY: Humana Press.

Kao, R.R., Haydon, D.T., Lycett, S.J. and Murcia, P.R. (2014). Supersize me: how whole-genome sequencing and big data are transforming epidemiology. Trends in Microbiology, 22(5), pp.282–291. doi:https://doi.org/10.1016/j.tim.2014.02.011.

Kwon, O.-K., Kim, D.-W., Heo, J.-H., Kim, J.-Y., Nah, J.-J., Choi, J.-D., Lee, D.-W., Cho, K.-H., Hong, S.-K., Kim, Y.-H., Kang, H.-E., Kwon, J.-H. and Shin, Y.-K. (2024). Genomic Epidemiology of African Swine Fever Virus Identified in Domestic Pig Farms in South Korea during 2019–2021. *Transboundary and emerging diseases*, 2024, pp.1–11. doi:https://doi.org/10.1155/2024/9077791.

Lemey, P., Rambaut, A., Drummond, A.J. and Suchard, M.A. (2009). Bayesian Phylogeography Finds Its Roots. PLoS Computational Biology, [online] 5(9), p.e1000520. doi:https://doi.org/10.1371/journal.pcbi.1000520.

Li, X., Xiao, K., Zhang, Z., Yang, J., Wang, R., Shen, X., Pan, J., Irwin, D.M., Chen, R.-A. and Shen, Y. (2020). The recombination hot spots and genetic diversity of the genomes of African swine fever viruses. *Journal of Infection*, 80(1), pp.121–142. doi:https://doi.org/10.1016/j.jinf.2019.08.007.

Li, Z., Chen, W., Qiu, Z., Li, Y., Fan, J., Wu, K., Li, X., Zhao, M., Ding, H., Fan, S. and Chen, J. (2022). African Swine Fever Virus: A Review. *Life*, [online] 12(8), p.1255. doi:https://doi.org/10.3390/life12081255.

Mazloum, A., van Schalkwyk, A., Shotin, A.R., Zinyakov, N., Igolkin, A., Chernishev, R., Debeljak, Z., Korennoy, F.I. and Sprygin, A.V. (2023). Whole-genome sequencing of African swine fever virus from wild boars in the Kaliningrad region reveals unique and distinguishing genomic mutations. *Frontiers in Veterinary Science*, 9. doi:https://doi.org/10.3389/fvets.2022.1019808.

Makowski, D., Ben-Shachar, M. and Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. Journal of Open Source Software, 4(40), p.1541. doi:https://doi.org/10.21105/joss.01541.

Mazur-Panasiuk, N., Walczak, M., Juszkiewicz, M. and Woźniakowski, 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 IGR I73R/I329L Genomic Sequences. Viruses, 12(10), p.1094. doi:https://doi.org/10.3390/v12101094.

Mazur-Panasiuk, N., Żmudzki, J. and Woźniakowski, G. (2019). African Swine Fever Virus – Persistence in Different Environmental Conditions and the Possibility of its Indirect Transmission. Journal of Veterinary Research, [online] 63(3), pp.303–310. doi:https://doi.org/10.2478/jvetres-2019-0058.

Miettinen, E., Melin, M., Holmala, K., Meller, A., Väänänen, V.M., Huitu, O. and Kunnasranta, M. (2023). Home ranges and movement patterns of wild boars (Sus scrofa) at the northern edge of the species' distribution range. *Mammal Research*, 68(4), pp.611–623. doi:https://doi.org/10.1007/s13364-023-00710-5.

Mordstein, C., Cano, L., Morales, A.C., Young, B., Ho, A.T., Rice, A.M., Liss, M., Hurst, L.D. and Kudla, G. (2021). Transcription, mRNA Export, and Immune Evasion Shape the Codon Usage of Viruses. *Genome Biology and Evolution*, 13(9). doi:https://doi.org/10.1093/gbe/evab106.

O'Toole, Á., Neher, R.A., Ndodo, N., Borges, V., Gannon, B., Gomes, J.P., Groves, N., King, D.J., Maloney, D., Lemey, P., Lewandowski, K., Loman, N.J., Myers, R., Omah, I., Suchard, M.A., Worobey, M., Chand, M., Ihekweazu, C., Ulaeto, D.O. and Adetifa, I. (2023). APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016. *Science*, 382(6670), pp.595–600. doi:https://doi.org/10.1126/science.adg8116.

Omelchenko, H., Avramenko, N.O., Petrenko, M.O., Wojciechowski, J., Pejsak, Z. and Woźniakowski, G. (2022). Ten Years of African Swine Fever in Ukraine: An Endemic Form of the Disease in the Wild Boar Population as a Threat to Domestic Pig Production. *Pathogens*, [online] 11(12), pp.1459–1459. doi:https://doi.org/10.3390/pathogens11121459.

Quembo, C.J., Jori, F., Vosloo, W. and Heath, L. (2018). Genetic characterization of African swine fever virus isolates from soft ticks at the wildlife/domestic interface in Mozambique and identification of a novel genotype. *Transboundary and Emerging Diseases*, 65(2), pp.420–431. doi:https://doi.org/10.1111/tbed.12700.

Rambaut, A., Lam, T.T., Max Carvalho, L. and Pybus, O.G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution, 2(1), p.vew007. doi:https://doi.org/10.1093/ve/vew007.

Rowlands, R.J., Michaud, V., Heath, L., Hutchings, G., Oura, C., Vosloo, W., Dwarka, R., Onashvili, T., Albina, E. and Dixon, L.K. (2008). African Swine Fever Virus Isolate, Georgia, 2007. *Emerging Infectious Diseases*, [online] 14(12), pp.1870–1874. doi:https://doi.org/10.3201/eid1412.080591.

Sauter-Louis, C., Forth, J.H., Probst, C., Staubach, C., Hlinak, A., Rudovsky, A., Holland, D., Schlieben, P., Göldner, M., Schatz, J., Bock, S., Fischer, M., Schulz, K., Homeier-Bachmann, T., Plagemann, R., Klaaß, U., Marquart, R., Mettenleiter, T.C., Beer, M. and Conraths, F.J. (2021). Joining the club: First detection of African swine fever in wild boar in Germany. *Transboundary and Emerging Diseases*, [online] 68(4), pp.1744–1752. doi:https://doi.org/10.1111/tbed.13890.

Schierup, M.H. and Hein, J. (2000). Consequences of Recombination on Traditional Phylogenetic Analysis. *Genetics*, [online] 156(2), pp.879–891. doi:https://doi.org/10.1093/genetics/156.2.879.

Schulz, K., Staubach, C., Blome, S., Viltrop, A., Nurmoja, I., Conraths, F.J. and Sauter-Louis, C. (2019). Analysis of Estonian surveillance in wild boar suggests a decline in the incidence of African swine fever. Scientific Reports, [online] 9(1), p.8490. doi:https://doi.org/10.1038/s41598-019-44890-0.

Senthilkumar, D., Rajukumar, K., Venkatesh, G., Singh, F., Tosh, C., Kombiah, S., Dubey, C.K., Chakravarty, A., Barman, N.N. and Singh, V.P. (2022). Complete genome analysis of African swine fever virus isolated from domestic pigs during the first ASF outbreaks in India. *Transboundary and Emerging Diseases*, 69(5). doi:https://doi.org/10.1111/tbed.14536.

Shen, Z.-J., Jia, H., Xie, C.-D., Shagainar, J., Feng, Z., Zhang, X., Li, K. and Zhou, R. (2022). Bayesian Phylodynamic Analysis Reveals the Dispersal Patterns of African Swine Fever Virus. *Viruses*, 14(5), pp.889–889. doi:https://doi.org/10.3390/v14050889.

Showalter, A.K., Byeon, I.J., Su, M.I. and Tsai, M.D. (2001). Solution structure of a viral DNA polymerase X and evidence for a mutagenic function. *Nature Structural & Molecular Biology*, 8(11), pp.942–946. doi:https://doi.org/10.1038/nsb1101-942.

Sindryakova, I.P., Morgunov, Yu.P., Chichikin, A.Yu., Gazaev, I.Kh., Kudryashov, D.A. and Tsybanov, S.Zh. (2016). THE INFLUENCE OF TEMPERATURE ON THE RUSSIAN ISOLATE OF AFRICAN SWINE FEVER VIRUS IN PORK PRODUCTS AND FEED WITH EXTRAPOLATION TO NATURAL CONDITIONS. *Sel'skokhozyaistvennaya Biologiya*, 51(4), pp.467–474. doi:https://doi.org/10.15389/agrobiology.2016.4.467eng.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J. and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evolution, 4(1). doi:https://doi.org/10.1093/ve/vey016.

World Organisation for Animal Health (WOAH) (2022). AFRICAN SWINE FEVER: WOAH Technical Disease Card. [online] woah.org. Available at: https://www.woah.org/en/document/african_swine_fever/ [Accessed 23 Apr. 2024].

World Organization for Animal Health (WOAH) (2024). World Animal Health Information System (WAHIS). [online] wahis.woah.org. Available at: https://wahis.woah.org/#/home [Accessed Apr. 2024].

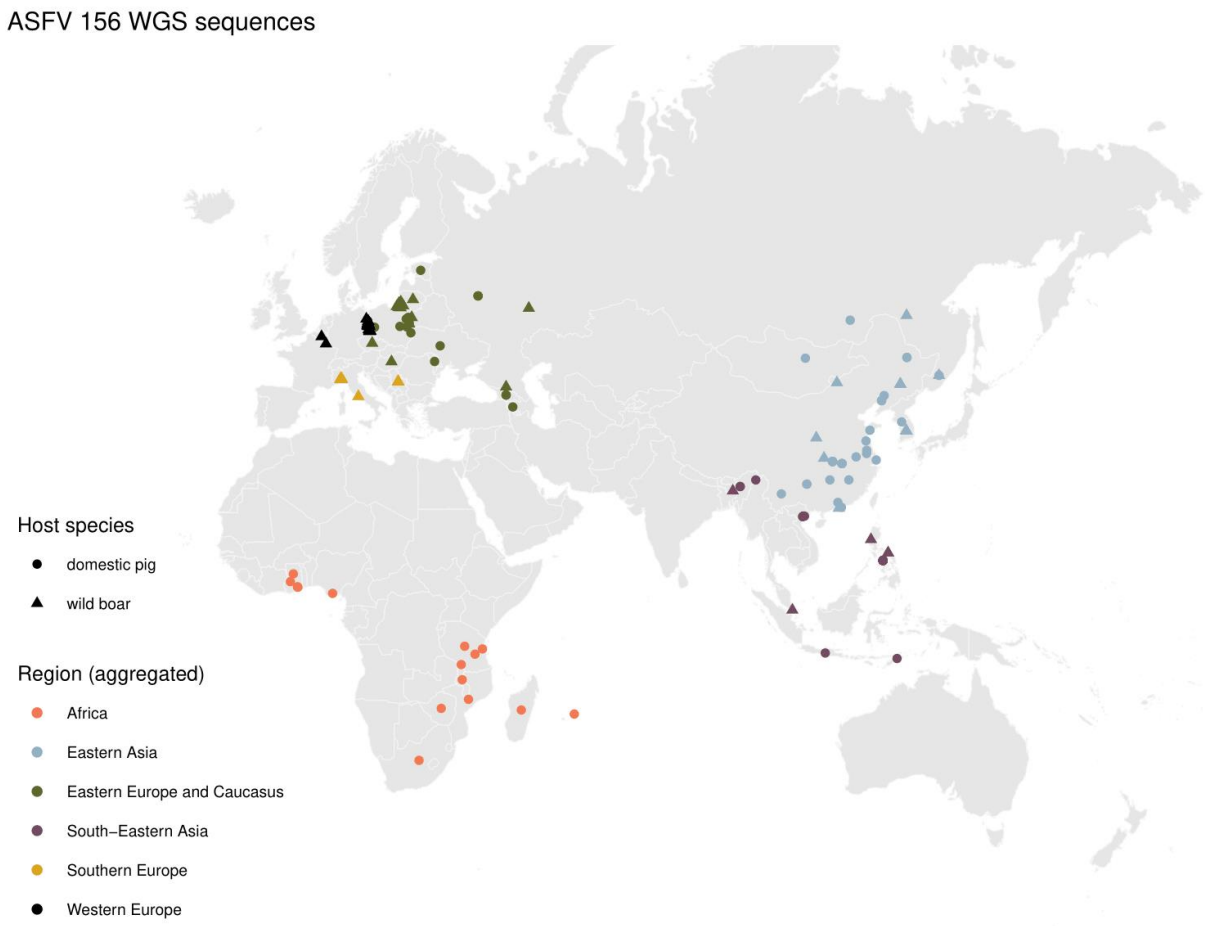
Xin, G., Kuang, Q., Le, S., Wu, W., Gao, Q., Gao, H., Xu, Z., Zheng, Z., Lu, G., Gong, L., Wang, H., Zhang, G., Shi, M. and Sun, Y. (2023). Origin, genomic diversity and evolution of African swine fever virus in East Asia. *Virus Evolution*, [online] 9(2). doi:https://doi.org/10.1093/ve/vead060.

You, S., Liu, T., Zhang, M., Zhao, X., Dong, Y., Wu, B., Wang, Y., Li, J., Wei, X. and Shi, B. (2021). African Swine Fever Outbreaks in China Led to Gross Domestic Product and Economic Losses. *Nature Food*, [online] 2, pp.1–7. doi:https://doi.org/10.1038/s43016-021-00362-1.

Zhang, Y., Wang, Q., Zhu, Z., Wang, S., Tu, S., Zhang, Y., Zou, Y., Liu, Y., Liu, C., Ren, W., Zheng, D., Zhao, Y., Hu, Y., Li, L., Shi, C., Ge, S., Lin, P., Xu, F., Ma, J. and Wu, X. (2023). Tracing the Origin of Genotype II African Swine Fever Virus in China by Genomic Epidemiology Analysis. *Transboundary and emerging diseases*, 2023, pp.1–14. doi:https://doi.org/10.1155/2023/4820809.

Zhao, D., Sun, E., Huang, L., Ding, L., Zhu, Y., Zhang, J., Shen, D., Zhang, X., Zhang, Z., Ren, T., Wang, W., Li, F., He, X. and Bu, Z. (2023). Highly lethal genotype I and II recombinant African swine fever viruses detected in pigs. *Nature Communications*, 14(1). doi:https://doi.org/10.1038/s41467-023-38868-w.

Supplementary material:



Supplementary Figure A: Spatial distribution of ASFV WGS and assigned regions. 156 WGS were collected in the sequence retrieval stage of the analysis, and assigned an aggregated region based on UN sub-regions. Figure by Dr. Gianluigi Rossi.

ID	O174L Sequence (sites 1 - 82)		
2007-03-20 Georgia domestic-pig FR682468	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	_____	TTTTTCAGTAGTGA
2017-06-30 Poland wild-boar MG939588	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	_____	TTTTTCAGTAGTGA
2016-06-30 Poland domestic-pig MG939583	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	_____	TTTTTCAGTAGTGA
2016-06-30 Poland domestic-pig MG939585	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	_____	TTTTTCAGTAGTGA
2016-06-30 Poland domestic-pig MG939586	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	_____	TTTTTCAGTAGTGA
2017-06-30 Poland wild-boar MG939587	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2017-06-30 Poland wild-boar MG939589	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	_____	TTTTTCAGTAGTGA
2017-06-30 Poland domestic-pig MT847620	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2017-06-30 Poland domestic-pig MT847622	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2019-06-30 Poland domestic-pig MT847623	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2020-11-16 Germany wild-boar OX376250	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2020-11-26 Germany wild-boar OX376258	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-04-19 Germany wild-boar OX376251	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-01-21 Germany wild-boar OX376254	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-02-18 Germany wild-boar OX376255	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-04-19 Germany wild-boar OX376260	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-07-29 Germany wild-boar OX376272	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-04-20 Germany wild-boar OX376253	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-02-22 Germany wild-boar OX376262	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-05-11 Germany wild-boar OX376259	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-07-28 Germany wild-boar OX376273	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-07-26 Germany wild-boar OX376266	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-07-27 Germany wild-boar OX376271	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-07-12 Germany wild-boar OX376267	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTG
2021-07-16 Germany domestic-pig OX376264	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-06-29 Germany wild-boar OX376268	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-02-09 Germany wild-boar OX376252	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2020-08-24 Germany wild-boar LR899193	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-04-29 Germany wild-boar OX376257	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-05-06 Germany wild-boar OX376263	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2020-09-20 Germany wild-boar OX376256	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2020-10-07 Germany wild-boar OX376261	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA
2021-07-16 Germany domestic-pig OX376265	TTATAAACGTTTCTTAGGTATGCGATACGTAAATCCTAATTCTTTAATAAGTTC	TTTTTCAGTAGTGA	TTTTTCAGTAGTGA

Supplementary Figure B: Structure and presence of the 14 base pair tandem duplication mutation in the ASFV O174L gene of Polish and German sequences, as well as the Georgia/2007 sequence. The insertion and sequences containing it are highlighted in pink.

O174L	Dinucleotide	Mean	Quantile_2.5th	Median	Quantile_97.5th
FL	aa	20738.76	20638.72	20745.54	20754.12
TD	aa	20746.46	20744.66	20746.22	20749.14
FL	ac	9032.48	9003.28	9034.57	9039.82
TD	ac	9034.75	9033.34	9034.76	9035.79
FL	ag	9947.36	9938.70	9946.74	9966.02
TD	ag	9948.77	9946.63	9948.76	9949.92
FL	at	17557.43	17416.34	17566.72	17579.23
TD	at	17568.02	17566.18	17568.09	17569.87
FL	ca	11520.15	11491.28	11523.43	11527.08
TD	ca	11525.54	11523.68	11525.73	11527.26
FL	cc	8417.90	8405.23	8415.77	8463.53
TD	cc	8412.57	8406.88	8411.77	8419.88
FL	cg	6201.48	6190.73	6192.78	6306.17
TD	cg	6190.59	6190.12	6190.56	6191.39
FL	ct	9936.56	9929.78	9935.29	9972.02
TD	ct	9935.06	9933.67	9935.23	9935.83
FL	ga	9623.34	9618.00	9620.84	9651.41
TD	ga	9621.27	9619.59	9621.39	9622.59
FL	gc	8475.19	8463.69	8467.49	8590.79
TD	gc	8466.22	8465.42	8466.39	8466.90
FL	gg	8373.74	8361.27	8368.44	8451.58
TD	gg	8362.28	8356.81	8361.83	8366.23
FL	gt	9525.53	9519.71	9524.01	9541.52
TD	gt	9525.56	9524.86	9525.58	9526.16
FL	ta	15373.43	15221.33	15383.21	15393.14
TD	ta	15385.11	15383.00	15385.16	15386.39
FL	tc	10167.42	10161.68	10166.18	10191.52
TD	tc	10165.77	10164.40	10165.86	10167.28
FL	tg	11485.55	11478.85	11482.64	11528.54
TD	tg	11482.17	11480.89	11482.33	11483.46
FL	tt	20465.68	20413.97	20469.38	20479.18
TD	tt	20471.85	20470.09	20471.66	20473.68

Supplementary Table A: Means, medians, 2.5th and 97.5th quantiles for dinucleotide count data by O174L mutation presence (FL = full length, unmutated; TD = tandem duplication, mutated).

Host	Dinucleotide	Mean	Quantile_2.5th	Median	Quantile_97.5th
domestic pig	aa	20733.92	20638.33	20745.09	20753.16
wild boar	aa	20746.56	20743.04	20746.22	20754.12
domestic pig	ac	9030.98	9001.97	9034.65	9039.34
wild boar	ac	9034.85	9032.94	9034.61	9038.07
domestic pig	ag	9948.23	9935.64	9947.19	9967.92
wild boar	ag	9946.96	9942.27	9946.84	9949.78
domestic pig	at	17551.60	17412.46	17566.98	17583.39
wild boar	at	17567.30	17563.07	17567.36	17571.02
domestic pig	ca	11518.22	11487.46	11523.43	11531.98
wild boar	ca	11524.10	11520.92	11524.51	11526.52
domestic pig	cc	8419.11	8405.06	8415.77	8463.82
wild boar	cc	8414.72	8406.72	8414.45	8422.80
domestic pig	cg	6206.68	6190.29	6193.11	6310.78
wild boar	cg	6192.16	6190.35	6191.78	6199.19
domestic pig	ct	9938.16	9930.28	9935.50	9972.76
wild boar	ct	9934.35	9929.05	9935.12	9936.77
domestic pig	ga	9624.90	9618.00	9620.81	9654.67
wild boar	ga	9620.97	9617.84	9621.23	9622.95
domestic pig	gc	8480.07	8463.77	8467.52	8591.26
wild boar	gc	8466.90	8464.72	8466.80	8470.68
domestic pig	gg	8376.99	8360.93	8367.99	8452.19
wild boar	gg	8366.22	8358.31	8367.14	8370.98
domestic pig	gt	9526.66	9519.07	9524.31	9543.86
wild boar	gt	9524.37	9520.87	9524.66	9527.04
domestic pig	ta	15367.02	15220.21	15383.38	15393.71
wild boar	ta	15384.30	15380.13	15384.31	15387.68
domestic pig	tc	10168.67	10163.70	10166.36	10195.16
wild boar	tc	10165.52	10161.21	10165.84	10167.50
domestic pig	tg	11487.53	11479.17	11482.73	11530.47
wild boar	tg	11482.27	11478.99	11482.33	11484.49
domestic pig	tt	20463.25	20412.08	20469.37	20477.62
wild boar	tt	20470.43	20459.57	20470.72	20479.42

Supplementary Table B: Means, medians, 2.5th and 97.5th quantiles for dinucleotide count data by host type.