RESEARCH ARTICLE

Host-similar fragments in the African swine fever virus genome: distribution, functions, and evolution

Zhaozhong Zhu[1*], Na Li[1], Qin Sun[1], Xizi Long[1], Tao Wang[2] and Hua-Ji Qiu[2*]

Abstract

African swine fever virus (ASFV) predominantly infects *Argasidae* and *suids*, resulting in high morbidity and mortality in pigs. Despite the crucial role that viral sequences resembling those of the host play in the virus's survival, there are limited comprehensive studies on the genomic similarities between ASFV and its hosts. Consequently, this study employs homology analysis to construct a similarity network between ASFV and its hosts (*Argasidae* and *suids*), investigating the distribution, function, evolution, and origins of these similar sequences in ASFV. Our findings indicate that the host-similar fragments are mainly distributed between positions 70000 and 180000 of the ASFV genome, primarily within non-coding regions. Notably, these non-coding fragments are often associated with promoter functions. Furthermore, the analysis of *suid* proteins that share similarities with ASFV proteins reveals that they predominantly exhibit RNA polymerase activity and are involved in metabolic processes. Evolutionary analysis indicates that pan-similar sequences of ASFV exist in an open state, highlighting the diversity of these analogous sequences. Additionally, a positive correlation was identified between the occurrence of recombination breakpoints and similar sequences, indicating that homologous recombination may serve as a crucial mechanism driving the formation of these analogous sequences.

Keywords African swine fever virus, host-similar fragments, similarity network, evolution, recombination

Introduction

African swine fever virus (ASFV) is a large, double-stranded DNA virus that primarily infects *Argasidae* and *suids*, resulting in significant economic losses to the global pig industry [1–3]. The virus is maintained in a sylvatic cycle between ticks and warthogs in Africa, where it remains asymptomatic in warthogs but causes severe hemorrhagic fever in domestic pigs [4, 5]. Without an effective vaccine, control relies on quarantine, culling, and biosecurity [6, 7]. ASFV has a large genome (170–194 kb) that encodes over 150 proteins, most of which are poorly understood [8, 9]. Furthermore, ASFV can be classified into approximately 24 genotypes and is primarily found in sub-Saharan Africa [10]. Its ability to evade the immune system and high mutation rate complicates vaccine development [11, 12].

All viral sequences that resemble those of the host, indicating genetic similarities between viruses and their hosts, play crucial roles in viral infection and pathogenesis [13–15]. For example, *Flaviviruses*, such as

Zhu *et al. Veterinary Research*   (2025) 56:108

Page 2 of 11

dengue and Zika viruses, exploit specific sequences similar to those of the host to manipulate the host's immune response, resulting in viral proliferation and pathogenesis [13]. Human papillomavirus (HPV), a DNA virus, uses its similarities to human DNA to infect and replicate [16]. Additionally, some HPV proteins, like E6 and E7, mimic human tumour suppressors, disrupting cell growth control and increasing the risk of cancer [16]. Over 100 sequence similarities exist between SARS-CoV-2 and the human genome, suggesting that the virus may have undergone extensive genetic exchange during its evolution [17]. Some host similarity sequences facilitate genetic exchange through horizontal gene transfer, influencing the evolution of both bacterial and bacteriophage genomes [15, 18]. Studying all virus-host-similar sequences is crucial for understanding viral pathogenesis, developing antiviral strategies, and elucidating the broader implications of viral infections on host biology [15, 17].

Several studies have examined the sequence similarities between ASFV and its hosts [19, 20], with one notably finding ASFV-like integrated (ASFLI) elements in the genome of *Ornithodoros moubata*. This discovery suggests a co-evolutionary relationship that dates back over 1.47 million years [19]. These elements, which comprise 10% of the ASFV genome, may contribute to RNA interference-based protection in *O. moubata* [19]. Additionally, ASFV shows a preference for A/U-ending codons, contrasting with its hosts' G/C preference, indicating an adaptation strategy to evade immune responses [20]. This research enhances our understanding of ASFV's evolution and host interactions, which are crucial for developing control strategies [19, 20].

Studying virus-host similarity sequences helps unravel the complex interactions between viruses and their hosts [21, 22]. For example, phages acquire host sequences to enhance infection efficiency and alter their host range [21, 22]. However, research on the similarities between ASFV and host sequences is still limited. This study addresses this gap by identifying and analysing these sequences' distribution, functions, and evolution in ASFV. The findings deepen our understanding of ASFV genetic diversity and evolution, offering new strategies for antiviral measures and vaccine development.

Materials and methods

Identification of the host-similar ASFV fragments

A total of 446 ASFV genome sequences were downloaded from the NCBI Genome database [23]. To obtain high-quality ASFV genomes, we removed all incomplete and missing genome sequences. The redundant genome sequences were eliminated at a 95% similarity threshold using CD-HIT [24]. We retained 73 complete and non-redundant ASFV genomes (Additional file 1). The *Argasidae* and *suid* genomes were downloaded from the NCBI nucleotide database [25]. To identify fragments in the ASFV genome that were similar to the host, including *Argasidae* and *suids*, BLAST (version 2.9.0) [26] was used to search the ASFV genomes. ASFV genome fragments with nucleotide sequence identities greater than 95% were considered host-similar.

Annotations of the ASFV genomes

The genes and proteins of ASFV were predicted in 73 ASFV genomes using the GeneMarkS server [27] with its default parameters. The predicted ASFV proteins were blasted against all ASFV proteins downloaded from the NCBI RefSeq database to infer protein function [28]. Predicted proteins with a sequence identity exceeding 30%, coverage greater than 40%, and e-values less than 0.001 were identified as ASFV proteins. The functional categorisation of the ASFV proteins was adapted to those outlined in Alí et al. [29].

Identification of the suid proteins similar to ASFV proteins

All *suid* proteins were obtained from the RefSeq database. A BLAST analysis was conducted to identify the *suid* proteins homologous to ASFV by comparing the annotated ASFV proteins from 73 genomes with the *suid* proteins. The *suid* proteins with a sequence identity $\geq 30\%$ and e-value ≤ 0.001 were considered potentially similar to ASFV proteins.

Functional analysis of the host-similar ASFV fragments

The promoter function of ASFV fragments was predicted using BDGP [30].

Functional analysis of the suid proteins

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and gene ontology (GO) enrichment analysis were performed using the enrichKEGG and enrichGO functions from the clusterProfiler package (version 4.6.2) [31] in R (version 4.2.2) [32]. The pathways and GO terms with *q*-values below 0.05 were considered significantly enriched.

Phylogenetic trees and ASFV genotyping

To determine ASFV genotypes, a maximum likelihood phylogenetic tree was constructed using the C-terminal sequence (415 bp) of the p72 gene from 73 ASFV isolates [10]. The tree was generated using MEGA (version X) [33] with default parameters. To assess the reliability of the phylogenetic tree, we performed bootstrap analysis with 100 replicates. Genotypes were assigned based on established criteria from previous studies, and the

Zhu *et al. Veterinary Research*   (2025) 56:108

Page 3 of 11

resulting phylogeny was visualised using Dendroscope (version 3.8.8) [34].

Pan-genomic analysis of the host-similar ASFV fragments

All host-similar ASFV fragments were grouped into clusters based on 100% sequence identity. PanGP [35], employing DG sampling algorithms, was used to visualise the characteristic curves of both pan-host-similar ASFV fragments and core-host-similar ASFV fragments. In R software, the 'plotrix' package (version 4.2.2) [36] generated a flower plot illustrating the distribution of core, dispensable, and unique host-similar ASFV fragment clusters.

Identification of the recombination events in the ASFV genome

To detect recombination events in the aligned ASFV genomes, we employed RDP (version 5) [37], including GENECONV, Bootscan, Maxchi, Chimaera, SiSscan, PhylPro, and 3Seq. These methods were applied with default parameters, and only recombination events with a significant *p*-value < 0.05 were considered. For each detected event, RDP provided information on the recombination region, recombination breakpoints, the recombinant virus, the potential major and minor parent viruses, and the support from each method. Only events detected by at least two methods were used for subsequent analysis to ensure robustness (Additional file 2).

Statistical analysis

All statistical analyses were conducted in R (version 4.2.2) [32]. The Wilcoxon rank-sum test was performed using the wilcox.test function; linear regression was fitted with the lm function, and the correlation coefficient was calculated using the cor.test function.

Results
Identification of similar fragments between the ASFV and host genomes

Seventy-three high-quality ASFV genome sequences were obtained from the NCBI Genome database. The ASFV genomes were found to contain a large number of fragments similar to those of the hosts (host-similar fragments), including *Argasidae* and *suids*. A similarity network was constructed using a 95% sequence identity threshold (Figure 1A). In total, 147 *Argasidae*-similar fragments and 578 *suid*-similar fragments were identified. Notably, two ASFV fragments exhibited similarity to sequences in both *suid* and *Argasidae* genomes.

The analysis examined the length, number, and distribution ratio of the host-similar fragments in the ASFV genome. The lengths of *Argasidae*-similar fragments ranged from 23 to 42 (median: 26; Figure 1B). In contrast, the lengths of *suid*-similar fragments ranged from 24 to 34 (median: 27; Figure 1C). The number of *Argasidae*-similar fragments per genome varied from 8 to 24 (median: 19; Figure 1D). *Suid*-similar fragments ranged from 15 to 84 (median: 59; Figure 1D). The ratio of *Argasidae*-similar fragments ranged from 0.0011 to 0.0034 (median: 0.0026; Figure 1E), whereas *suid*-similar fragments ranged from 0.0021 to 0.0122 (median: 0.0084; Figure 1E). *Argasidae*-similar fragments were shorter, fewer, and had lower ratios, suggesting that ASFV may have acquired more *suid*-similar sequences during its adaptation from *Argasidae* to *suids*.

Distribution of the host-similar fragments in the ASFV genome

We analysed the distribution within the viral genome to determine the location of host-similar ASFV fragments and subsequently calculated their occurrences per 10 kb window of ASFV genomes. Specifically, the median of *Argasidae*-similar ASFV coding region fragments, *Argasidae*-similar ASFV non-coding fragments, *suid*-similar ASFV coding region fragments, and *suid*-similar ASFV non-coding region fragments per 10 kb window of the ASFV genome were 36, 6.5, 102.5, and 116, respectively. Interestingly, two prominent hotspots were identified at window positions 70 kb and 180 kb (Figure 2A, red arrows). Furthermore, among the ten ASFV genotypes we identified, some host-similar ASFV fragments were identified as genotype-specific (Figure 2B, black box). This finding enriches our understanding of ASFV genetic diversity and reveals specific adaptive mechanisms involved in its evolutionary process.

Functional and sequence characteristics of the host-similar ASFV fragments

These fragments were found more frequently in non-coding regions than in ASFV gene sequences (Figure 3A). Approximately 50% of *Argasidae*-similar fragments and 26% of *suid*-similar fragments were located in non-coding regions. In contrast, only 12% of the entire ASFV genome was found in non-coding regions. Given the abundance of host-similar ASFV fragments in non-coding regions, we explored their potential role in transcriptional regulation. Previous research indicates that viral fragments resembling host genes frequently contain promoter sequences, aiding immune evasion and adaptation to the host [38].

We calculated the ratio of promoters within the host-similar sequences in the ASFV non-coding regions using randomly selected non-coding sequences as a control (see Materials and methods). The results showed significantly higher promoter ratios in *Argasidae*- and *suid*-similar non-coding fragments compared with the control
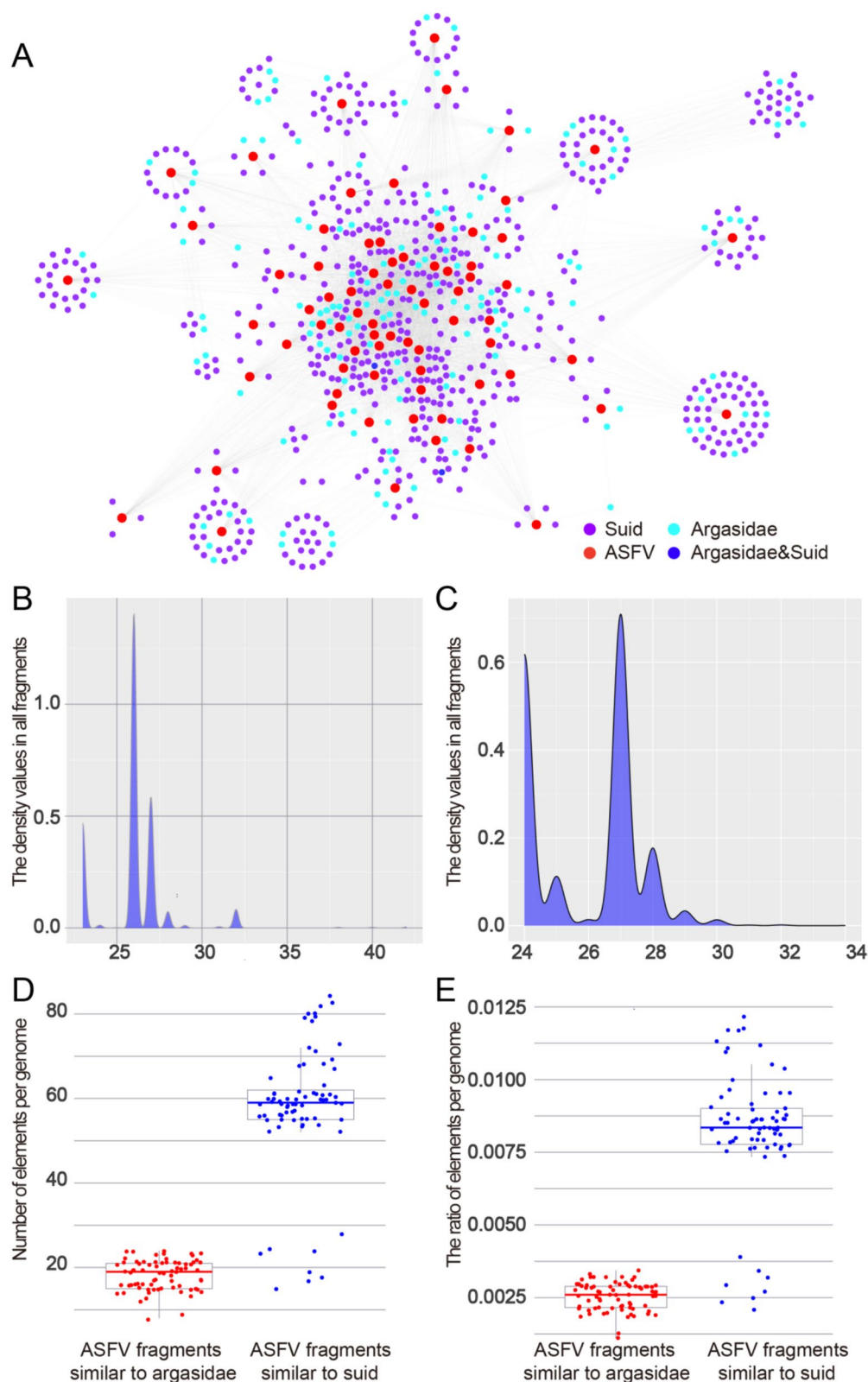
Zhu *et al. Veterinary Research*   (2025) 56:108

Page 4 of 11

Figure 1 **Overview of the similarity network between ASFV and its hosts (Argasidae and suids). A** ASFV similarity network with *Argasidae* and *suids*. **B** The length of the ASFV fragments similar to *Argasidae*. **C** The length of the ASFV fragments similar to *suids*. **D** Number of the ASFV fragments similar to *Argasidae* and *suids* per genome. **E** The ratio of the ASFV fragments similar to *Argasidae* and *suids* per genome.
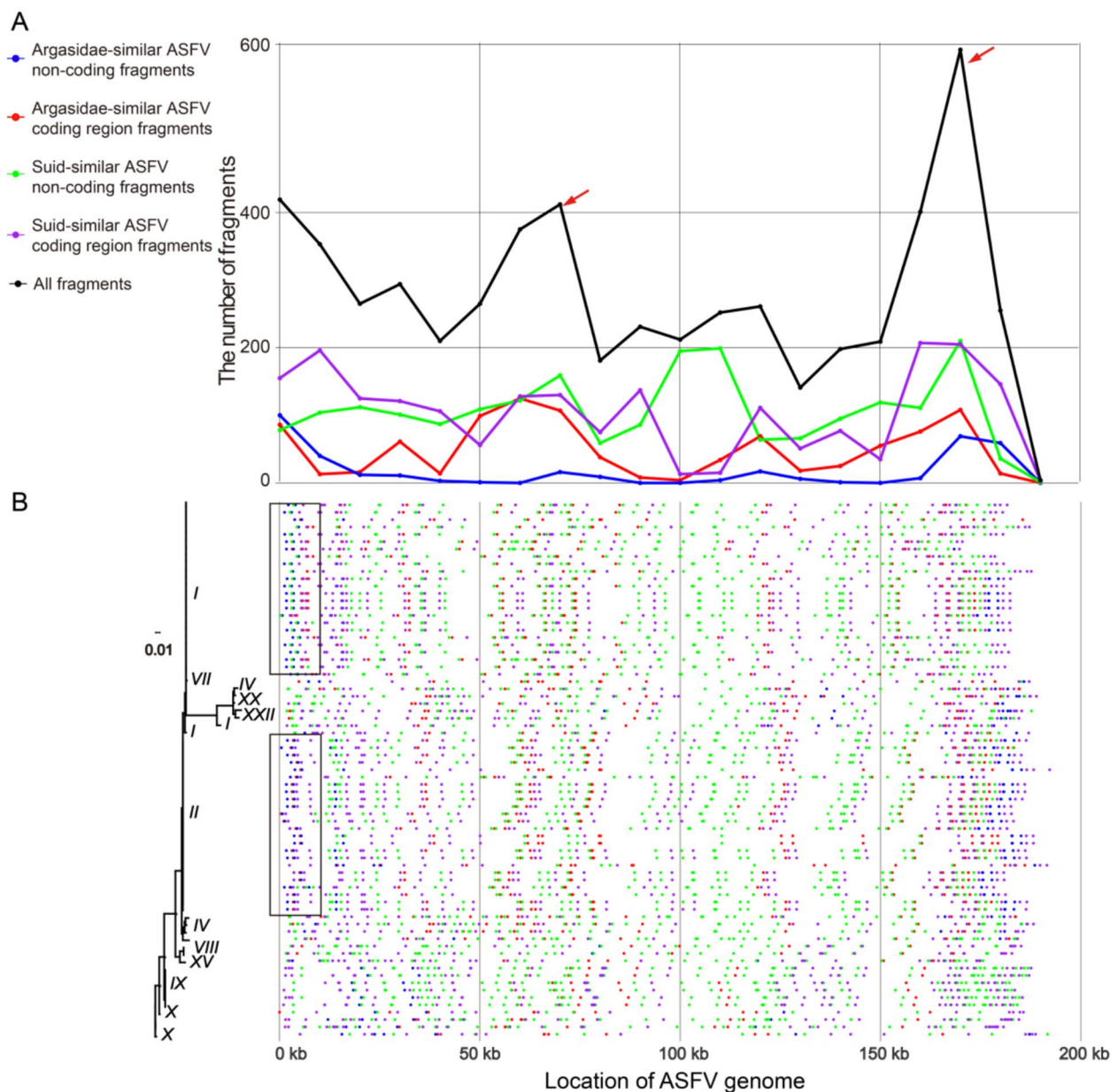
Zhu *et al. Veterinary Research*   (2025) 56:108

Page 5 of 11

Figure 2 The number and positions of the host-similar fragments in the ASFV genome. **A** The number of host-similar ASFV fragments per kb in the 73 ASFV genomes. **B** Phylogenetic tree (left) based on the C-terminal sequence of p72 (415 bp) and the genotype labels (Greek numerals). The positions of the host-similar ASFV fragments (right) with genotype-specific examples are highlighted (black frames).

group (Figure 3B). This finding suggests that these fragments may contribute to ASFV transcriptional regulation by enriching promoter elements, thereby influencing its life cycle and pathogenicity. Additionally, we analysed the GC content of host-similar sequences in coding and non-coding regions (Additional file 3). These fragments exhibited significantly lower GC content than the overall genome, suggesting they may have undergone higher mutation rates during viral evolution.

Functions of the suid proteins similar to the ASFV proteins

We conducted GO and KEGG pathway enrichment analyses on *suid* proteins similar to ASFV proteins to investigate the biological processes in *suid* cells that ASFV may disrupt by mimicking these proteins (see Materials and methods). We identified 42 *suid* proteins that share sequence similarity with ASFV proteins. Functional enrichment analysis revealed that most enriched GO terms were closely associated with the RNA polymerase
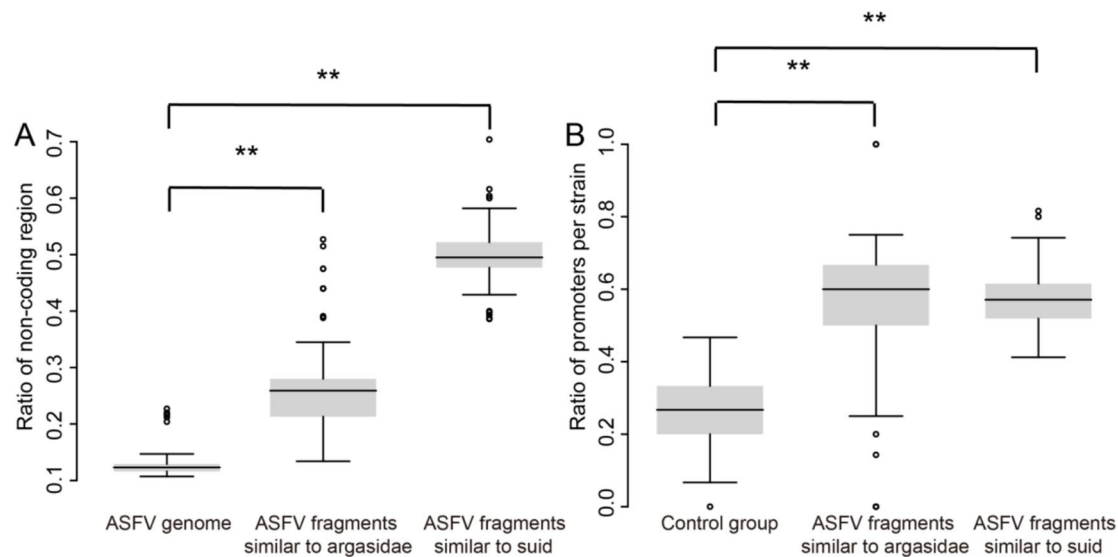
Zhu *et al. Veterinary Research*   (2025) 56:108

Page 6 of 11

Figure 3 The functions of the host-similar fragments in the non-coding regions in the ASFV genome. A Distribution of the non-coding regions in the ASFV genome, *Argasidae*-similar, and *suid*-similar fragments. **B** The ratio of the promoters in the host-similar (*Argasidae* and *suids*) vs the non-host-similar fragments of the ASFV genome. '**', $p < 0.01$.

complex in the cellular component category (Figure 4). Moreover, most enriched GO terms were related to RNA polymerase activity in the molecular function category, including 5'−3'RNA polymerase, RNA polymerase

I activity, and DNA-directed 5'−3'RNA polymerase activity. Four of the ten enriched KEGG pathways were also linked to metabolic processes, such as glutathione and purine metabolism. These results emphasise the
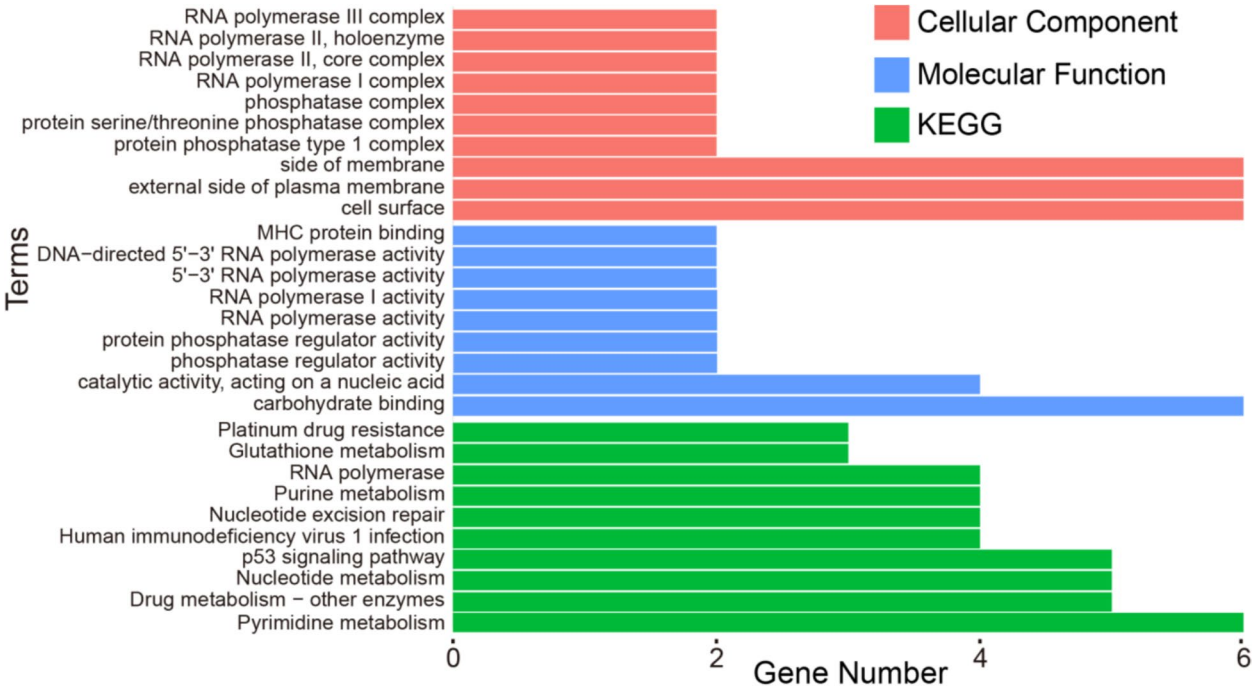

Figure 4 Functional analysis of the suid proteins similar to the ASFV proteins. The top 10 enriched terms in the *suid* proteins within the domains of biological processes, cellular components, molecular functions, and KEGG pathways.

significance of RNA polymerase complexes and metabolic pathways in the functions of these *suid* proteins.

We also performed a functional analysis of ASFV proteins similar to *suid* proteins, revealing that most ASFV proteins had unknown functions, followed by those involved in structure and morphogenesis (Additional file 4). These findings suggest that ASFV may disrupt host cell processes by mimicking proteins involved in host structure and morphogenesis, thereby supporting viral survival.

Pan-genomic analysis of the host-similar ASFV fragments

A pan-genomic analysis of the host-similar ASFV fragments was carried out, during which all host-similar ASFV fragments in all ASFV genomes were collectively defined as 'pan-ASFV fragments'. These were categorised into two types: core ASFV fragment clusters, which appear in every genome, and non-core ASFV fragments, which were found in one or more but not all genomes. The non-core fragments appearing only in one genome are considered unique ASFV fragments. Analysis of the pan-ASFV fragments similar to *Argasidae* revealed no core fragment clusters but identified 147 non-core fragment clusters, including 55 unique fragments (Figure 5A). Similarly, analysis of the *suid*-similar pan-fragments identified two core fragment clusters and 576 non-core fragment clusters, including 268 unique fragments (Figure 5C).

A power-law regression model was applied to fit the number of dispensable ASFV fragment clusters analogous to those found in *Argasidae* (Figure 5B) and *suids* (Figure 5D). The models, $y = 32.22x^{0.37} - 15.17$ and $y = 87.16x^{0.45} - 31.7$, accurately captured this relationship, respectively. As more ASFV genomes were analysed, dispensable host-similar fragment clusters steadily increased. This trend indicates ASFV's adaptability, likely due to the expansion of pan-fragments with each additional genome, which can generate new host-similar fragment clusters.

An exponential model was used to fit the number of core ASFV fragment clusters similar to *Argasidae* (Figure 5B) and *suids* (Figure 5D). The models, $y = 28.04e^{(-0.5x)} + 0.23$ and $y = 90.39e^{(-0.54x)} + 1.53$, accurately described this trend. As the number of the genomes analysed increased, the core ASFV fragment clusters significantly declined. This outcome denotes the diverse and dynamic nature of the host-similar fragment within the ASFV genome.

The relationship between the host-similar ASFV fragments and the ASFV homologous recombination

Studies suggest that homologous recombination may facilitate the exchange of nucleic acid sequences between viruses and their hosts. Therefore, we examined the effect of homologous recombination on host similarity fragments in ASFV. Recombination events were observed across the 73 ASFV genomes, with each genome undergoing between 1 and 17 recombination events (median: 3; Figure 6A). Interestingly, the number of recombination breakpoints identified in each 20 kb ASFV genome window was positively correlated with the number of *Argasidae*-similar ASFV fragments (Figure 6B) and *suid*-similar ASFV fragments (Figure 6D) identified in the same region, with Pearson correlation coefficients of 0.19 and 0.21, respectively. Specifically, the median number of *Argasidae*-similar fragments in windows containing recombination breakpoints was 1, while *suid*-similar fragments had a median of 5. In contrast, windows without recombination breakpoints had medians of 2 for the *Argasidae*-similar fragments (Figure 6C) and 6 for the *suid*-similar fragments (Figure 6E).

Discussion

The research highlights a significant discovery of fragments in ASFV genomes that closely resemble those of their hosts, *Argasidae and suids*. This similarity suggests a complex evolutionary history involving potential genetic exchange or recombination. These host-similar fragments were primarily distributed in non-coding regions and were more inclined to include promoter regions. Additionally, *suid* proteins share similarities with ASFV proteins, mainly functioning as RNA polymerases and participating in metabolic processes. Such findings may prompt further investigation into the virus's adaptability and ability to influence the host's immune response, providing valuable insights into ASFV's pathogenicity and host range.

Our research has identified numerous similar sequences between the ASFV and its host, consistent with previous studies [19]. These sequences, likely resulting from nucleic acid exchanges between the virus and host, provide the virus with advantages in infection and survival, enhancing its adaptability and pathogenicity. Moreover, some ASFV protein fragments are similar to RNA polymerase and metabolism-related proteins in pigs. Therefore, these fragments may interfere with essential functions such as transcription and metabolism, providing a viral advantage. They may also help viruses evade the host's immune system by mimicking host proteins and facilitating replication and transmission within host cells. This process underscores the complexity of host-virus interactions and the necessity for a thorough understanding to develop effective countermeasures against ASFV.

The pan-genomic analysis of host-similar ASFV fragments provides valuable insights into the virus's
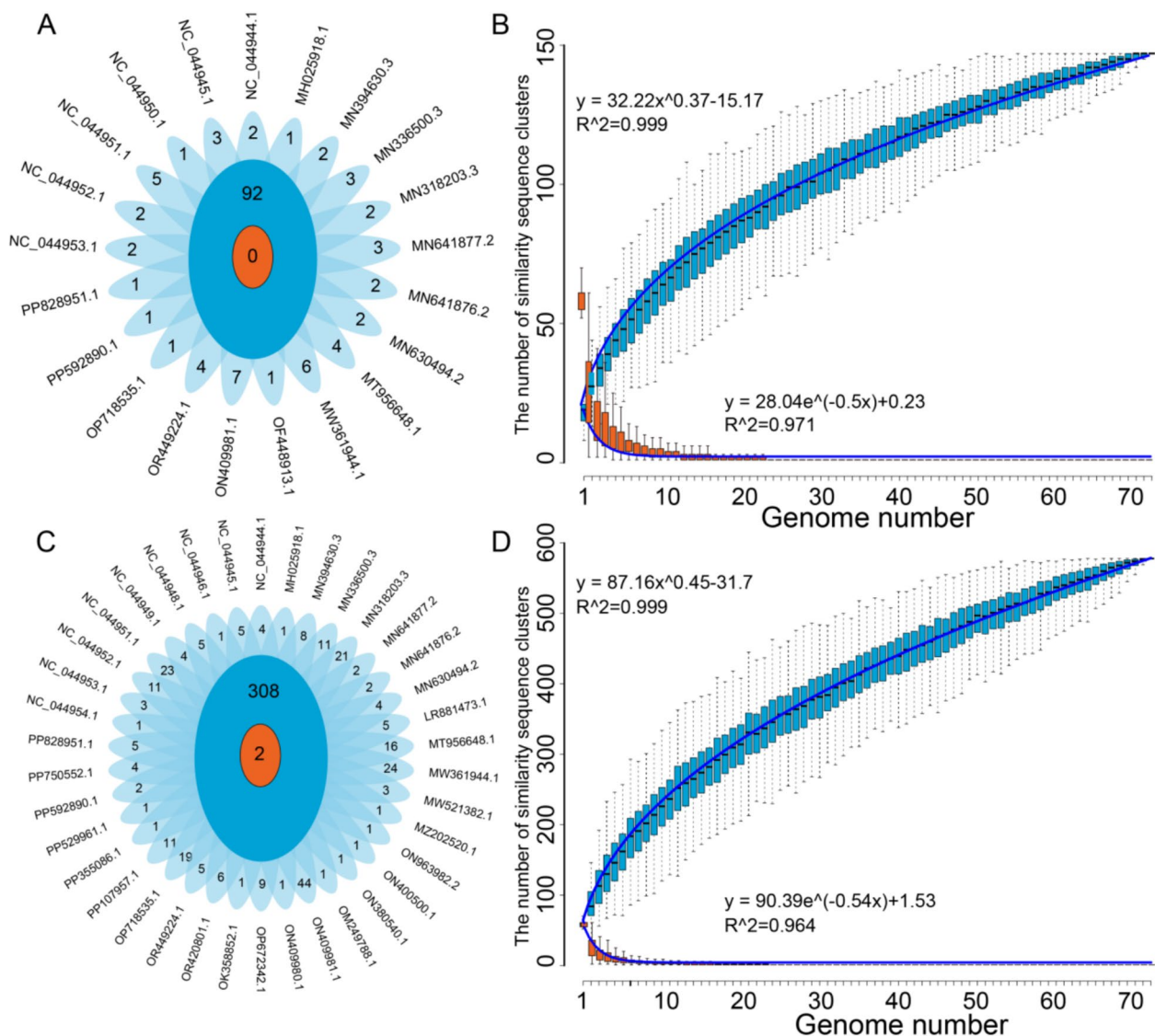
Zhu *et al. Veterinary Research*   (2025) 56:108

Page 8 of 11

Figure 5 Pan-genome analysis of the host-similar fragments in the ASFV genome. A A flower plot depicting the core, dispensable, and strain-specific ASFV fragments similar to *Argasidae* across the 73 ASFV strains. The diagram depicts the core (in the centre), dispensable (in the annulus), and strain-specific (in the petals) fragment number for the 73 ASFV strains. **B** The profile of the pan-ASFV and core fragments similar to *Argasidae*. The exponent (0.37) of the pan-genome curve is greater than zero, indicating an open pan-genome. **C** A flower plot depicting the core, dispensable, and strain-specific ASFV fragments similar to *suids* across 73 ASFV strains. The diagram depicts the core fragments number (in the centre), the dispensable fragments number (in the annulus), and the strain-specific fragments number (in the petals) for the 73 ASFV strains. **D** The profile of the pan-ASFV fragments and core fragments similar to *suids*. The exponent (0.45) of the pan-genome curve is greater than zero, indicating an open pan-genome.

genetic diversity and adaptability. Identifying core and non-core fragment clusters, including unique ones, emphasises the genome's dynamic nature. The power-law regression model shows a steady increase in dispensable fragment clusters with each genome analysed, reflecting the virus's capability to expand its genetic repertoire and adapt to new hosts [39, 40]. In contrast, the exponential model suggests a decline in essential core fragment clusters as more genomes are studied,

underscoring the virus's flexibility and potential for rapid evolution.

Exploring the relationship between host-similar ASFV fragments and homologous recombination events provides a compelling narrative of the virus's genetic plasticity. The positive correlation between recombination breakpoints and the presence of host-similar sequences suggests a mechanism through which the virus may acquire and integrate host genetic
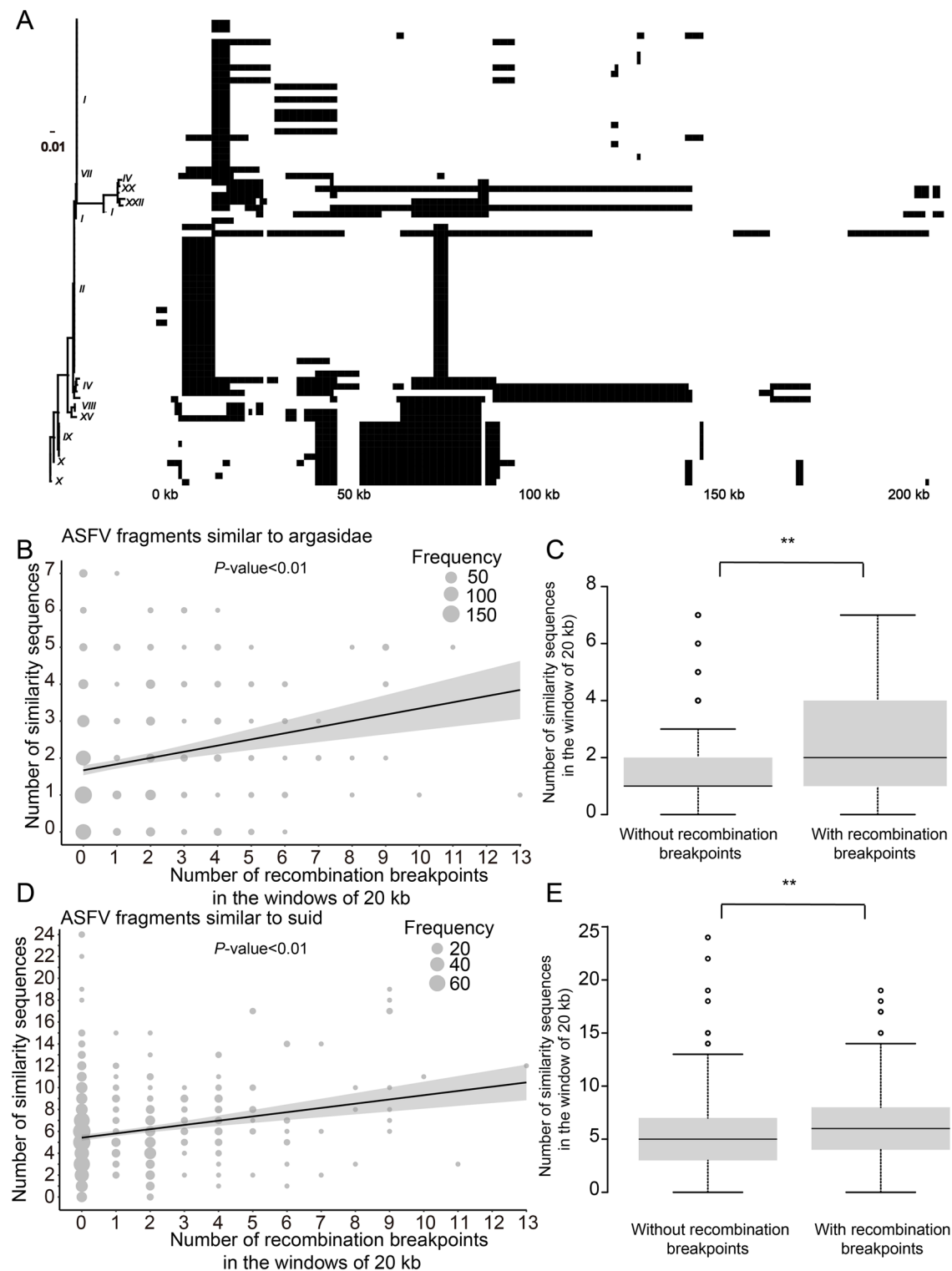
Figure 6 The relationship between the recombination breakpoints and the host-similar fragments in the ASFV genome. A The recombination regions (in black) in the ASFV genome. **B** The number of the recombination breakpoints in a sliding window of 20 kb *versus* the number of *Argasidae*-similar ASFV fragments. **C** Comparison of the number of the *Argasidae*-similar ASFV fragments in a sliding window of 20 kb with and without the recombination breakpoints in the ASFV genome. **D** The number of recombination breakpoints in a sliding window of 20 kb versus the number of *suid*-similar ASFV fragments. **E** Comparison of the number of *suid*-similar ASFV fragments in a sliding window of 20 kb with and without the recombination breakpoints in the ASFV genome. '******', $p < 0.01$.

Zhu *et al. Veterinary Research*   (2025) 56:108

Page 10 of 11

material. This finding is particularly significant as it implies that recombination may contribute to the virus's ability to adapt and evolve, potentially leading to new strains with altered virulence or host range. The study's insights into the role of recombination in shaping the ASFV genome contribute to a more nuanced understanding of viral evolution and the complexities of host-virus co-evolution [41].

However, it is important to note that this study has two limitations. First, the size and number of similar sequences in the ASFV genome may be influenced by the chosen similarity threshold. We applied a 95% similarity threshold to ensure the authenticity of the sequences, as using a lower threshold could identify more sequences but might also increase false positives. Second, the predicted promoter functions within the ASFV-similar fragments need further validation. In future studies, experimental approaches, such as reporter assays or mutagenesis analyses, could be employed to validate the predicted promoter activities.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13567-025-01539-3.

Additional file 1: The ASFV strains used in this study.

Additional file 2: The recombination events detected by at least two methods in RDP5. The results detected by RDP5 listed the recombination regions, recombinant viruses, potential major and minor parents, and the methods supporting the events.

Additional file 3: The GC content of host-similar fragments in the coding and non-coding regions of the ASFV genome.

Additional file 4: The functions of the ASFV proteins similar to those of the Argasidaeand suidproteins.

Authors' contributions
HJQ, ZZ, and TW conducted the experiments and drafted the manuscript. HJQ, ZZ, NL, QS, and XL critically revised the manuscript and the experimental design. HJQ, ZZ, and NL helped with the experiments. All authors read and approved the final manuscript.

Availability of data and materials
Datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests
The authors declare that they have no competing interests.

References

1. Gaudreault NN, Madden DW, Wilson WC, Trujillo JD, Richt JA (2020) African swine fever virus: an emerging DNA arbovirus. Front Vet Sci 7:215
2. Wang T, Sun Y, Qiu HJ (2018) African swine fever: an unprecedented disaster and challenge to China. Infect Dis Poverty 7:66–70
3. 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, Bu Z (2023) Highly lethal genotype I and II recombinant African swine fever viruses detected in pigs. Nat Commun 14:3096
4. Jori F, Bastos AD (2009) Role of wild suids in the epidemiology of African swine fever. EcoHealth 6:296–310
5. Sun Y, Xing J, Hong SL, Bollen N, Xu S, Li Y, Zhong J, Gao X, Zhu D, Liu J, Gong L, Zhou L, An T, Shi M, Wang H, Baele G, Zhang G (2024) Untangling lineage introductions, persistence and transmission drivers of HP-PRRSV sublineage 8.7. Nat Commun 15:8842
6. Mutua F, Dione M (2021) The context of application of biosecurity for control of African swine fever in smallholder pig systems: current gaps and recommendations. Front Vet Sci 8:689811
7. Chathuranga K, Lee JS (2023) African swine fever virus (ASFV): immunity and vaccine development. Vaccines 11(2):199
8. Dixon LK, Chapman DA, Netherton CL, Upton C (2013) African swine fever virus replication and genomics. Virus Res 173:3–14
9. Cackett G, Matelska D, Sýkora M, Portugal R, Malecki M, Bähler J, Dixon L, Werner F (2020) The African swine fever virus transcriptome. J Virol 94:e00119-20
10. Bastos AD, Penrith ML, Cruciere C, Edrich J, Hutchings G, Roger F, Couacy-Hymann E, Thomson GR (2003) Genotyping field strains of African swine fever virus by partial p72 gene characterisation. Arch Virol 148:693–706
11. Wang T, Sun Y, Huang S, Qiu HJ (2020) Multifaceted immune responses to African swine fever virus: implications for vaccine development. Vet Microbiol 249:108832
12. Li Z, Chen W, Qiu Z, Li Y, Fan J, Wu K, Li X, Zhao M, Ding H, Fan S, Chen J (2022) African swine fever virus: a review. Life 12:1255
13. Lobo FP, Mota BE, Pena SD, Azevedo V, Macedo AM, Tauch A, Machado CR, Franco GR (2009) Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. PLoS One 4:e6282
14. Kaján GL, Doszpoly A, Tarján ZL, Vidovszky MZ, Papp T (2020) Virus-host coevolution with a focus on animal and human DNA viruses. J Mol Evol 88:41–56
15. Hendrix RW, Hatfull GF, Ford ME, Smith MC, Burns RN (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci USA 96:2192–2197
16. McBride AA (2022) Human papillomaviruses: diversity, infection and host interactions. Nat Rev Microbiol 20:95–108
17. Li H, Hong X, Ding L, Meng S, Liao R, Jiang Z, Liu D (2022) Sequence similarity of SARS-CoV-2 and humans: Implications for SARS-CoV-2 detection. Front Genet 13:946359
18. Chen DS, Wu YQ, Zhang W, Jiang SJ, Chen SZ (2016) Horizontal gene transfer events reshape the global landscape of arm race between viruses and homo sapiens. Sci Rep 6:26934
19. Forth JH, Forth LF, Lycett S, Bell-Sakyi L, Keil GM, Blome S, Calvignac-Spencer S, Wissgott A, Krause J, Höper D, Kampen H, Beer M (2020) Identification of African swine fever virus-like elements in the soft tick genome provides insights into the virus' evolution. BMC Biol 18:136
20. Wang Y, Chi C, Zhang J, Zhang K, Deng D, Zheng W, Chen N, Meurens F, Zhu J (2024) Systematic analysis of the codon usage patterns of African swine fever virus genome coding sequences reveals its host adaptation phenotype. Microb Genom 10:001186
21. Hyman P, Abedon ST (2010) Bacteriophage host range and bacterial resistance. Adv Appl Microbiol 70:217–248
22. Iuchi H, Kawasaki J, Kubo K, Fukunaga T, Hokao K, Yokoyama G, Ichinose A, Suga K, Hamada M (2023) Bioinformatics approaches for unveiling virus-host interactions. Comput Struct Biotechnol J 21:1774–1784
23. Brister JR, Ako-Adjei D, Bao Y, Blinkova O (2015) NCBI viral genomes resource. Nucl Acid Res 43:D571–D577

Zhu *et al. Veterinary Research*   (2025) 56:108

Page 11 of 11

24. Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152
25. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2012) GenBank. Nucl Acid Res 41:D36–D42
26. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinform 10:421
27. Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucl Acid Res 33:W451–W454
28. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P et al (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucl Acid Res 44:D733–D745
29. Alejo A, Matamoros T, Guerra M, Andrés G (2018) A proteomic atlas of the African swine fever virus particle. J Virol 92:e01293-18
30. Reese MG (2001) Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput Chem 26:51–56
31. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (Camb) 2(3):100141
32. Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comput Graph Stat 5:299–314
33. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549
34. Huson DH, Scornavacca C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol 61:1061–1067
35. Zhao Y, Jia X, Yang J, Ling Y, Zhang Z, Yu J, Wu J, Xiao J (2014) PanGP: a tool for quickly analyzing bacterial pan-genome profile. Bioinformatics 30:1297–1299
36. Package 'plotrix' (2015) https://cran.r-project.org/web/packages/plotrix/index.html. Accessed 10 Dec 2024
37. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, Kelz Z, Kumar V, Murrell B (2021) RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. Virus Evol 7:veaa087
38. Bohn-Wippert K, Tevonian EN, Megaridis MR, Dar RD (2017) Similarity in viral and host promoters couples viral reactivation with host cell migration. Nat Commun 8:15006
39. Wang Z, Jia L, Li J, Liu H, Liu D (2020) Pan-genomic analysis of African swine fever virus. Virol Sin 35:662–665
40. Zhu Z, Ge S, Cai Z, Wu Y, Lu C, Zhang Z, Fu P, Mao L, Wu X, Peng Y (2022) Systematic identification and characterization of repeat sequences in African swine fever virus genomes. Vet Res 53:101
41. Zhu Z, Xiao CT, Fan Y, Cai Z, Lu C, Zhang G, Jiang T, Tan Y, Peng Y (2019) Homologous recombination shapes the genetic diversity of African swine fever viruses. Vet Microbiol 236:108380

Publisher's Note