Whole-Genome Sequence of African Swine Fever Virus isolate from India provides insights into diversity and evolution

Sonalika Mahajan^{1#}, Juwar Doley^{2#}, Saravanan Subramaniam³, Ajay K. Yadav.², Seema Rani Pegu², Narayana H. Mohan², Vishal Chander¹, Karikalan Mathesh¹, Sukdeb Nandi¹, Karam Pal Singh¹, Vivek Kumar Gupta^{2*} and Gaurav Kumar Sharma^{1*}

[#]Both the authors made equal contribution.

¹ICAR-Indian Veterinary Research Institute, Bareilly, India

²ICAR-National Research Centre on Pig, Assam, India

³ICAR-Directorate of Foot-and-Mouth Disease, India

Address for Correspondence:

Gaurav Kumar Sharma, CADRAD, ICAR-Indian Veterinary Research Institute, Bareilly

243122, India; email: gaurvet@gmail.com

Vivek Kumar Gupta, Director, ICAR- National Research Centre on Pig, Assam 781015,

India; email: gupta.dr vivek@gmail.com

Abstract

African swine fever (ASF) was first reported in 1921, and since then has posed a major threat to the world pig industry and still remains a major challenge as there is no vaccine or therapy available. In May 2020, the first incidence of ASF was recorded in India, followed by a number of outbreaks in the north eastern part of India. In this study, we report the first whole genome of an Indian isolate of ASF virus (ASF/IND/20/CAD/543) using next generation sequencing and compared with the other ASFV complete genome. On phylogenetic analysis, the virus was assigned to genotype II on the basis of p72 genotyping. However, the whole genome based phylogeny distinguished it from other genotype II isolates of clade 1.1.1. This study adds to our understanding of ASFV's genetic diversity and molecular evolution.

Key words: African Swine Fever Virus, India, whole genome sequencing, p72 gene

1. Introduction

African swine fever (ASF) is a World Organization for Animal Health (OIE) notifiable disease of domestic pigs and wild boars that causes high mortality, approaching up to 100% [1]. ASF is caused by ASF virus (ASFV), a large, multi-enveloped, highly complex virus belonging to the genus *Asfivirus* of the family *Asfarviridae* [2]. The genome of ASFV is composed of a linear, double-stranded DNA molecule with a size ranging between ~170-190 kb and encodes more than 150 open reading frames (ORFs) [3]. The ASF, first described a century ago [4], remains as a challenge to control due to difficulty in development of vaccines and therapeutics [5]. The control of ASF relies on rapid diagnosis, animal stamping out, and restriction of trade and animal movements, thus leading to severe economic consequences for affected countries [6].

ASF was first reported from Kenya in 1921 and remained confined to the African continent till 1957, when it was introduced into Portugal, Europe and Cuba, Brazil, the Dominican Republic and Haiti [7]. In 2007, it was introduced into the Georgian Republic and spread to other Caucasian countries, Russia, Ukraine and Belarus in the following years [8-9]. The recent introduction of ASFV in China (2018), Myanmar (2019), and India in the year 2020 has serious economic implications and poses a major threat to the to the pork industry around the globe [10]. The North-Eastern (NE) region of India rears 47% of the Indian pig population, with an annual swine business of around US\$ 0.94 billion. It has been estimated that about 54,150 pigs have died in India up to July 2021 leading to the direct losses to a tune of INR 2.76 billion [10].

In 2020, the first laboratory confirmation of ASF from India was published, and the isolates were assigned to post-2007-p72-genotype II [11] on the basis of partial characterisation of ASF virus isolates. Until now 24 genotypes of ASFV has been described globally based on the p72 gene [3, 12, 13]. However, no full genome characterization of the ASFV isolates circulating in India has been reported yet. In the present study, complete coding region of the ASFV strain (ASF/IND/20/CAD/543) obtained directly from a field sample was determined using next generation sequencing. A comprehensive genomic comparison of ASF/IND/20/CAD/543 with other ASFV whole genome from across the world shows its distinctiveness and provides insights into the evolution of ASFV and its transboundary spread.

2. Materials and Methods

Samples were collected from 12 pigs died of acute ASF from Dhemaji district of Assam, India in February 2021 during an outbreak and spleen tissues were collected during post-

69 mortem examination. Nucleic acid extraction, followed by World Organization for Animal
70 Health recommended qPCR confirmed the presence of ASFV in all samples.

71 *2.1 Screening of samples for ASFV*

72 Viral nucleic acid was extracted from the tissue homogenates using QIAamp DNA mini kit
73 (Qiagen, Germany) as per the manufacturer's protocol. The samples were tested for ASFV
74 using OIE recommended protocol [14]. In brief, a 20µl reaction was assembled consisting of
75 mastermix, probes (reference) and amplification was carried out in ABI 7500 Real-Time PCR
76 system (Applied Biosystems, USA) with following thermal profile: 1 cycle at 95°C for 5 min,
77 followed by 45 cycles at 95°C for 15 s and 58°C for 1 min.

78 *2.2. Genome Sequencing*

79 Amongst the samples, one representative isolate (ASF/IND/20/CAD/543) was selected for
80 next generation sequencing. The genome sequencing was generated at a commercial facility
81 (Eurofins India). Briefly, nucleic acid was extracted from the tissue sample using QIAamp
82 DNA extraction kit (Qiagen, Germany) as per the recommended protocol. Sequencing library
83 was prepared with TruSeq library preparation kit (Illumina, USA) followed by sequencing in
84 a NextSeq 500 system (Illumina, USA) to generate 2 X 150bp paired reads. The sequenced
85 raw data was processed to obtain high quality clean reads using Trimmomatic v0.38 [15] to
86 remove adapter sequences, ambiguous reads (reads with unknown nucleotides "N" larger
87 than 5%), and low-quality sequences (reads with more than 10% quality threshold (QV) < 20
88 phred score). The high quality reads of the samples were aligned to the reference ASF
89 genome of strain BAV17 available in the NCBI database (NC_001659.2) [16] followed by
90 annotation and variant calling. Annotation was performed by using Genome annotation
91 transfer utility (GATU) software [17], The high quality reads of the samples were aligned to
92 the reference sequence using Burrows–Wheeler Algorithm (software version 0.7.17) [18].
93 The sequence data is available in BioSample database (Accession numbers; XXXXXXXX).
94 Consensus sequence of 170,095 bp was extracted using mpileup utility of Samtools v 0.1.18
95 [19].

96 *2.3 Phylogenetic analysis*

97 A total of 35 genome sequences representing viral strains from different endemic countries
98 were obtained from GenBank and the details are listed (Table 1). The pair-wise sequence
99 alignment was performed using MAFFT [20] and nucleotide identity matrix was created
100 using BioEdit 7.2.5 [21]. The phylogenetic analysis was performed on the basis of whole
101 genome and three different genes namely B646L, KP177R and A238L using maximum
102 likelihood model available in MEGA X with 1000 bootstrap replications [22]. The data set

103 was best fitted by the nucleotide substitution model HKY+G, which was used to infer
104 phylogenetic relationships.

105 **3. Results and Discussion**

106 The ASFV complete genome from the Indian isolate consisted of 170,095 nucleotides with
107 144 Open Reading Frames (ORFs), 2,806 SNPs and had a mean GC content of 37.6%. The
108 genome sequence of the ASFV isolate ASF/IND/20/CAD/543 has been deposited in
109 GenBank under the accession OK236383. The pair-wise sequence alignment performed to
110 compare ASF/IND/20/CAD/543 with other ASFV whole genomes revealed 73.3% to 86.3%
111 identity with other sequences (Table S1). The Indian isolates sequenced in the study belonged
112 to Genotype II and shared 100% sequence homology with isolates collected earlier from
113 outbreaks in the Indian states of Arunachal Pradesh and Assam between January and April
114 2020 [11]. Till date, only genotypes I and II have been reported outside of Africa, with
115 genotype II being responsible for the ongoing epidemic and having the most global
116 significance. ASF genotype II viruses resulting from the Georgia introduction have been
117 identified in a number of European countries since 2007, recently in China and Southeast
118 Asia [3], and most recently in India [11].

119 The whole genome based phylogeny indicates that this virus belongs to clade 1.1.1/beta (Fig
120 1 b), which is consistent with the previous reports [23-24]. Previously, based on full genome
121 analysis [23] genotype II ASFV isolates were clustered within clade 1.1.1, which also
122 comprised genotype I. Such discrepancies between the phylogeny based on the p72 genomic
123 region and the entire genome was reported earlier also [24]. Unlike the ASFV genotype
124 classification based on p72 gene, presently there is no universal system for phylogeny based
125 on complete genome analysis. Five clades (alpha, beta, gamma, delta, and epsilon), with all
126 genotype II ASFV belonging to the beta clade was reported [25].

127 The genome size of the Indian isolate of ASF was smaller than the recently reported
128 Tanzania/Rukwa/2017/1 isolate [3] (170095 against 183186) with less number of ORFs (144
129 vs 188) and more SNPs than Tanzanian or Polish or Chinese isolates. The relevance of
130 missing ORFs is presently not known, but presumably not essential for viral replication.
131 Furthermore, there is insertion of TRS between the I73R and I329L (Fig S1).

132 We also performed phylogenetic analysis based on KP177R and A238L genes. KP177R
133 encodes p22, a structural viral protein that is early transcribed and found in the ASFV

134 particle. Although its particular function is uncertain, p22 has been speculated to be an
 135 interaction partner of multiple host proteins [26]. The tree reconstructed using the KP177R
 136 gene exhibited a complete lack of clarity, with isolates of different genotypes clustering
 137 together and genotype II isolates collected from Estonia in 2014 constituting the basal node
 138 (data not shown). In case of A238L gene based phylogeny, the p72 genotype clustering
 139 pattern was preserved to a greater extent, with the exception of slight aberrations in genotype
 140 placement (Fig 1c). There were three primary clades in the p72 clustering pattern, with
 141 genotypes I, II, and III forming one clade, genotypes IX and X forming one clade, and
 142 genotype VIII forming one clade alone. However, in the A238L gene-based phylogeny, only
 143 two primary clades were detected, with clustered genotypes I, II, and III retained in one
 144 clade. But, the genotype VIII, which previously remained as a distinct clade in the p72-based
 145 classification, was grouped with genotypes IX and X in a single clade. The above findings
 146 suggest that the A238L gene can be used as a supplement for p72-based phylogenetic
 147 analysis. The observed differences in clustering patterns, especially with single genes, could
 148 be due to a variety of factors, including evolutionary processes that result in evolution of viral
 149 genome [27].

150 In the present study, we report the first whole genome based characterization of ASFV
 151 belonging to genotype II isolated from India. The isolate was assigned to genotype II and
 152 clade 1.1.1. The distinct placement of the present ASFV isolate on the basis of whole genome
 153 phylogeny indicates the genome evolution as the virus spread across the different countries
 154 with varying climate. The study reemphasizes the relevance of whole genome based
 155 characterization and classification in understanding molecular evolution of ASFV and
 156 epidemiological tracing of the disease spread.

157

158 **Funding Information**

159 This work was funded by the Department of Animal Husbandry Dairying and Fisheries under
 160 Central Disease Diagnostic laboratory (CDDL) scheme and ICAR.

161 **Acknowledgements**

162 The authors are grateful to the Indian Council of Agricultural Research.

163 **Conflict of Interest Statement**

164 The authors declare no conflict of interest.

165 **Ethics Statement**

166 The authors confirm that the ethical policies have been adhered to.

167 **Author Contributions:**

168 SM was involved in methodology and writing. JD, AKY, SRP, NHM, VC, KM, SN in
169 sample procurement and data analysis. SS and GKS in conceptualization, sequencing analysis
170 and manuscript draft. KPS and VKG in supervision and manuscript writing.

171 **References:**

- 172 1. OIE, World Organisation for Animal Health. (Global situation of African swine fever.
173 African Swine Fever (ASF). 2021 Report 47: 2016 – 2020.
174 <https://www.oie.int/app/uploads/2021/03/report-47-global-situation-asf.pdf> (2021).
- 175 2. Dixon LK, Abrams CC, Bowick G, Goatley LC, Kay-Jackson PC, Chapman D, Liv-
176 erani E, Nix R, Silk R, Zhang F. African swine fever virus proteins involved in evad-
177 ing host defence systems. *Vet Immunol Immunopathol*. 2004 Aug;100(3-4):117-34.
178 doi: 10.1016/j.vetimm.2004.04.002. PMID: 15207450.
- 179 3. Njau EP, DomelevoEntfellner JB, Machuka EM, Bochere EN, Cleaveland S, Shirima
180 GM, Kusiluka LJ, Upton C, Bishop RP, Pelle R, Okoth EA. The first genotype II Af-
181 rican swine fever virus isolated in Africa provides insight into the current Eurasian
182 pandemic. *Sci Rep*. 2021 Jun 22;11(1):13081. doi: 10.1038/s41598-021-92593-2.
- 183 4. Penrith ML, Kivaria F M, Masembe C. One hundred years of African swine fever: A
184 tribute to R. Eustace Montgomery. *TransboundEmerg Dis*. 2021 Sep;68(5):2640-
185 2642. doi:10.1111/tbed.14183
- 186 5. Gavier-Widen D, Stahl K, Dixon L. No hasty solutions for African swine fever. *Sci-*
187 *ence*. 2021 367(6478), 622-624. doi:10.1126/science.aaz8590
- 188 6. Costard S, Mur L, Lubroth J, Sanchez-Vizcaino JM, Pfeiffer DU. Epidemiology of
189 African swine fever virus. *Virus Res*. 2013 Apr;173(1):191-7. doi:
190 10.1016/j.virusres.2012.10.030. Epub 2012 Nov 1. PMID: 23123296.
- 191 7. Sánchez-Vizcaíno JM, Mur L, Martínez-López B. African swine fever: an epidemiol-
192 ogical update. *Transbound Emerg Dis*. 2012 Mar;59 Suppl 1:27-35. doi:
193 10.1111/j.1865-1682.2011.01293.x. Epub 2012 Jan 9. PMID: 22225967.
- 194 8. Gabriel C, Blome S, Malogolovkin A, Parilov S, Kolbasov D, Teifke JP, Beer M.
195 Characterization of African swine fever virus Caucasus isolate in European wild
196 boars. *Emerg Infect Dis*. 2011 Dec;17(12):2342-5. doi: 10.3201/eid1712.110430.
197 PMID: 22172247; PMCID: PMC3311204.

- 198 9. Rowlands RJ, Michaud V, Heath L, Hutchings G, Oura C, Vosloo W, Dwarka R,
199 Onashvili T, Albina E, Dixon LK. African swine fever virus isolate, Georgia, 2007.
200 Emerg Infect Dis. 2008 Dec;14(12):1870-4. doi: 10.3201/eid1412.080591. PMID:
201 19046509; PMCID: PMC2634662.
- 202 10. Mohan NH, Misha MM, Gupta VK. Consequences of African swine fever in India:
203 Beyond economic implications. Transbound Emerg Dis. 2021 Nov;68(6):3009-3011.
204 doi: 10.1111/tbed.14318. Epub 2021 Sep 19. PMID: 34498419.
- 205 11. Rajukumar K, Senthilkumar D, Venkatesh G, Singh F, Patil VP, Kombiah S, Tosh C,
206 Dubey CK, Sen A, Barman NN, Chakravarty A, Dutta B, Pegu SR, Bharali A, Singh
207 VP. Genetic characterization of African swine fever virus from domestic pigs in India.
208 Transbound Emerg Dis. 2021 Sep;68(5):2687-2692. doi: 10.1111/tbed.13986. Epub
209 2021 Jan 29. PMID: 33415828.
- 210 12. Bastos AD, Penrith ML, Crucire C, Edrich JL, Hutchings G, Roger F, Couacy-
211 Hymann E, R Thomson G. 2003. Genotyping field strains of African swine fever vi-
212 rus by partial p72 gene characterisation. Arch Virol 148:693–706.
213 <http://dx.doi.org/10.1007/s00705-002-0946-8>.
- 214 13. Achenbach JE, Gallardo C, Nieto-Pelegrn E, Rivera-Arroyo B, Degefa-Negi T, Arias
215 M, Jenberie S, Mulisa DD, Gizaw D, Gelaye E, Chibssa TR, Belaye A, Loitsch A,
216 Forsa M, Yami M, Diallo A, Soler A, Lamien CE, Snchez-Vizcano JM. Identifica-
217 tion of a New Genotype of African Swine Fever Virus in Domestic Pigs from Ethio-
218 pia. TransboundEmerg Dis. 2017 Oct;64(5):1393-1404. doi: 10.1111/tbed.12511.
219 Epub 2016 May 22. PMID: 27211823
- 220 14. Fernndez-Pinero J, Gallardo C, Elizalde M, Robles A, Gmez C, Bishop R, Heath L,
221 Couacy-Hymann E, Fasina FO, Pelayo V, Soler A, Arias M. Molecular diagnosis of
222 African Swine Fever by a new real-time PCR using universal probe library. Trans-
223 bound Emerg Dis. 2013 Feb;60(1):48-58. doi: 10.1111/j.1865-1682.2012.01317.x.
224 Epub 2012 Mar 7. PMID: 22394449.
- 225 15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina se-
226 quence data. Bioinformatics. 2014 Aug 1;30(15):2114-20. doi:
227 10.1093/bioinformatics/btu170.
- 228 16. Yanez RJ, Rodriguez JM, Nogal ML, Yuste L, Enriquez C, Rodriguez JF, Vinuela E.
229 Analysis of the complete nucleotide sequence of African swine fever virus. Virology
230 1995 208 (1), 249-278.

- 231 17. Tcherepanov V, Ehlers A, Upton C. 2006. Genome annotation transfer utility
232 (GATU): rapid annotation of viral genomes using a closely related reference genome.
233 BMC Genomics 7:150. <http://dx.doi.org/10.1186/1471-2164-7-150>.
- 234 18. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–
235 Wheeler transform. *Bioinformatics*, 25, 1754-1760.
236 <https://doi.org/10.1093/bioinformatics/btp324>.
- 237 19. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., & Durbin, R.
238 (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25,
239 2078-2079. doi: 10.1093/bioinformatics/btp352.
- 240 20. Katoh, Kazutaka; Misawa, Kazuharu; Kuma, Kei-ichi; Miyata, Takashi (2002).
241 "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier
242 transform". *Nucleic Acids Research*. 30 (14): 3059–66.
- 243 21. Hall TA. (1999) BioEdit: a user-friendly biological sequence alignment editor and
244 analysis program for Window 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.
- 245 22. Kumar, S., Stecher, G., Michael Li, M., Knyaz, C., and Koichiro Tamura, K (2018)
246 MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms.
247 *Molecular Biology and Evolution* 35:1547-1549
- 248 23. Bao J, Wang Q, Lin P, et al. Genome comparison of African swine fever virus
249 China/2018/AnhuiXCGQ strain and related European p72 Genotype II strains. *Trans-*
250 *bound Emerg Dis*. 2019;66(3):1167-1176. doi:10.1111/tbed.13124
- 251 24. de Villiers EP, Gallardo C, Arias M, et al. Phylogenomic analysis of 11 complete Af-
252 rican swine fever virus genome sequences. *Virology*. 2010;400(1):128-136.
253 doi:10.1016/j.virol.2010.01.019
- 254 25. Aslanyan L, Avagyan H, Karalyan Z. Whole-genome-based phylogeny of African
255 swine fever virus. *Vet World*. 2020;13(10):2118-2125.
256 doi:10.14202/vetworld.2020.2118-2125
- 257 26. Vuono EA, Ramirez-Medina E, Pruitt S, Rai A, Espinoza N, Velazquez-Salinas L,
258 Gladue DP, Borca MV. Evaluation of the Function of the ASFV KP177R Gene, En-
259 coding for Structural Protein p22, in the Process of Virus Replication and in Swine
260 Virulence. *Viruses*. 2021 May 26;13(6):986. doi: 10.3390/v13060986. PMID:
261 34073222; PMCID: PMC8227490.
- 262 27. Domingo, Esteban. Long-term virus evolution in nature. *Virus as Populations* (2020):
263 225–261. doi:10.1016/B978-0-12-816331-3.00007-6

264 **Figure legends**

265 Fig 1: Evolutionary relationships of representative strains of African swine fever virus
 266 inferred using maximum likelihood method based on B646L(p72) gene (1a), full genome
 267 (1b) and A238L gene (1c). The uncertainty was evaluated using 1,000 bootstrap replications,
 268 and the node values above 70% are shown. The isolate sequenced in this study is indicated by
 269 filled circle.

270 Fig S1: Nucleotide sequences alignment of Indian isolate with representative strains of
 271 African swine fever virus

272

273 **Table 1**

274 ASFV sequences used in this study

Name	GenBank Accession No.	Origin	Year	Host
Kenya 1950	AY261360	Kenya	1950	Domestic Pig
Ken05/Tk1	KM111294	Kenya	2005	Tick
Ken06.Bus	KM111295	Kenya	2006	Pig
R35	MH025920	Uganda	2015	Domestic Pig
N10	MH025919	Uganda	2015	Domestic Pig
R25	MH025918	Uganda	2015	Domestic Pig
R7	MH025917	Uganda	2015	Domestic Pig
R8	MH025916	Uganda	2015	Domestic Pig
Malawi Lil-20/1 (1983)	AY261361	Malawi	1983	Pig
E75	FN557520	Spain	1975	Domestic Pig
OURT 88/3	AM712240	Portugal	1988	Ornithodoros sp.
Benin97/1	AM712239	Benin	1997	Domestic Pig
L60	KM262844	Portugal	1960	Domestic Pig
NHV	KM262845	Portugal	1986	Domestic pig
47/Ss/2008	KX354450	Italy	2008	Domestic Pig
BA71	KP055815	Spain	1971	Pig
26544/OG10	KM102979	Italy	2010	Pig
Mkuzi 1979	AY261362	South Africa	1979	Ornithodoros sp.
Warthog	AY261366	Namibia	N.A.	Warthog
Pretoriuskop/96/4	AY261363	South Africa	1996	Ornithodoros sp.
Warmbaths	AY261365	South Africa	N.A.	Tick
Tengani 62	AY261364	Malawi	1962	Pig
IND/AS/SD-13/2020	MT612962	India	2020	Domestic Pig
IND/AR/SD-61/2020	MT612964	India	2020	Domestic Pig
IND/AR/SD-59/2020	MT612963	India	2020	Domestic Pig
IND/AS/SD-02/2020	MT612961	India	2020	Domestic Pig
Germany 2020/1	LR899193	Germany	2020	Wild Boar
Wuhan 2019-2	MN393477	China	2019	Pig
Timor-Leste/2019/1	MW396979	Timor-Leste	2019	Pig
China/2018/AnhuiXCGQ	MK128995	China	2018	Pig
CN/2019/InnerMongolia-AES01	MK940252	China	2019	Wild Boar
POL/2015/Podlaskie	MH681419	Poland	2015	Wild Boar
Estonia 2014	LS478113	Estonia	2014	Wild Boar
Georgia 2007/1	FR682468	Georgia	2007	Pig
Odintsovo_02/14	KP843857	Russia	2014	Wild Boar

275 ^{N.A.} - Not available