Virology Journal

RESEARCH   **Open Access**

CrossMark

Intra-epidemic genome variation in highly pathogenic African swine fever virus (ASFV) from the country of Georgia

Jason Farlow[3*] , Marina Donduashvili[1], Maka Kokhreidze[1], Adam Kotorashvili[2], Nino G. Vepkhvadze[1], Nato Kotaria[2] and Ana Gulbani[1]

Abstract

Background: African swine fever virus (ASFV) causes an acute hemorrhagic infection in suids with a mortality rate of up to 100%. No vaccine is available and the potential for catastrophic disease in Europe remains elevated due to the ongoing ASF epidemic in Russia and Baltic countries. To date, intra-epidemic whole-genome variation for ASFV has not been reported. To provide a more comprehensive baseline for genetic variation early in the ASF outbreak, we sequenced two Georgian ASFV samples, G-2008/1 and G-2008/2, derived from domestic porcine blood collected in 2008.

Methods: Genomic DNA was extracted directly from low-volume ASFV PCR-positive porcine blood samples and subjected to next generation sequencing on the Illumina Miseq platform. De novo and mapped sequence assemblies were performed using CLCBio software. Genomic illustrations, sequence alignments and assembly figures were generated using Geneious v10.2.4. Sequence repeat architecture was analyzed using DNASTAR GeneQuest 14.1.0.

Results: The G-2008/1 and G-2008/2 genomes were distinguished from each other by coding changes in seven genes, including MGF 110-1 L, X69R, MGF 505-10R, EP364R, H233R, E199L, and MGF 360-21R in addition to eight homopolymer tract variations. The 2008/2 genome possessed a novel allele state at a previously undescribed intergenic repeat locus between genes C315R and C147L. The C315R/C147L locus represents the earliest observed variable repeat sequence polymorphism reported among isolates from this epidemic. No sequence variation was observed in conventional ASFV subtyping markers. The two genomes exhibited complete collinearity and identical gene content with the Georgia 2007/1 reference genome. Approximately 56 unique homopolymer A/T-tract variations were identified that were unique to the Georgia 2007/1 genome. In both 2008 genomes, within-sample sequence read heterogeneity was evident at six homopolymeric G/C-tracts confined to the known hypervariable ~ 7 kb region in the left terminal region of the genome.

Conclusions: This is the first intra-epidemic comparative genomic analysis reported for ASFV and provides insight into the intra-epidemic microevolution of ASFV. The genomes reported here, in addition to the G-2007/1 genome, provide an early baseline for future genome-level comparisons and epidemiological tracing efforts.

---

* Correspondence: farlowscience@gmail.com
[3]Farlow Scientific Consulting Company, LLC, Lewiston, UT 84320, USA
Full list of author information is available at the end of the article

Farlow *et al. Virology Journal* (2018) 15:190

Page 2 of 8

Introduction

African Swine Fever Virus (ASFV) is a large, icosahedral, dsDNA virus and sole member of the *Asfarviridae* family [1]. ASFV is the only known DNA arbovirus [2]. ASFV is highly adapted for both sylvatic (enzootic) and domestic transmission cycles and infects wild boars (*Sus scrofa scrofa*), warthogs (*Phacochoerus aethiopicus*), bushpigs (*Potamochoerus*spp.), domestic pigs and *Ornithodoros* ticks [1, 3–5]. Following the 2007 ASF outbreak in the Caucasus, ASFV has been detected in Russia, Ukraine and Baltic countries [6–8].

Conventional genetic subtyping strategies for ASFV utilize tandem-repeat sequences (TRSs) within p72 (B646L), p54 (E183L), the central variable region (CVR) of the B602L gene and an intergenic region between *I73R* and *I329L* [9–14]. More recently, loci within p30 and EP402R (CD2v) have also shown discriminatory utility [15]. Previous studies have identified at least 22 distinct p72 genotypes among East and South African isolates with genotype I representing the major genotype present in West Africa and genotype II predominating in East Africa [15, 16]. Thus far, only p72 genotype II has been detected in eastern Europe and Russia following a putative single introduction of this genotype into Georgia in 2007 [17]. An insertion observed at the *I73R/I329L* TRS marker also confirmed allele variation in Russia or elsewhere in the Caucasus prior to the detection of the virus in Baltic countries [7, 8]. While existing ASFV marker-based assays provide genetic discrimination among isolates, higher-resolution genomic comparisons may facilitate more robust epidemiological tracing [18]. Lack of whole genome data for both closely related endemic strains as well as those generated during intra-epidemic transmission remain unavailable. To date, only a single ASFV genome sequence has been reported from the current epidemic (G-2007/1) [19].

We report here a comparative genomic analysis of two ASFV isolates (G-2008/1 and G-2008/2) sequenced directly from PCR-positive porcine blood isolates taken in 2008 during the initial phase of the outbreak in Georgia. These data represent the first intra-epidemic comparative genomic analysis of ASFV, and to our knowledge, represent the first ASFV genome-level sequences derived directly from naturally-infected suids without prior cell culture propagation. These data also provide insight into the wild-type genome microevolution of ASFV and provide an additional baseline for genetic variation early in the ASF outbreak.
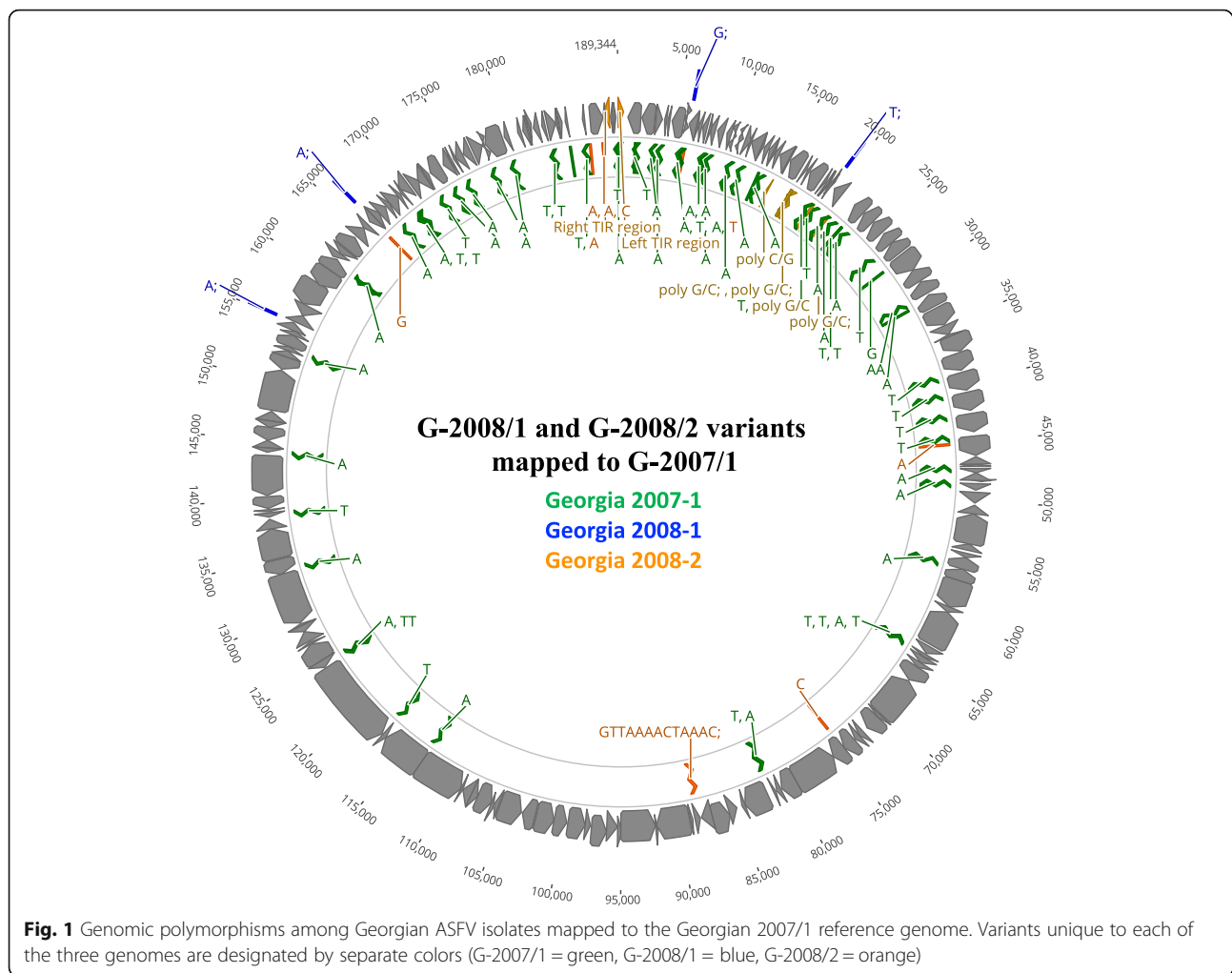
Materials and methods

Genomic DNA was extracted directly from low-volume ASFV PCR-positive porcine blood samples, designated G-2008/1 and G-2008/2, using the DNeasy Blood & Tissue kit (QIAGEN, Hilden, Germany). Village and district-level geographical locations were not available for the isolates. Genomic libraries were quantified using the Qbit (Qbit 2.0 Fluorometer, Life technologies) and sheared using the Covaris M220 focused ultrasonicator (Covaris, Inc.) Genome fragmentation was performed on the BioAnalyzer 2100 (Agilent Technologies, Inc.). Libraries were subsequently modified using the Illumina Paired End Sample Prep Kit and sequenced on the Illumina MiSeq platform (Illumina, Inc.) per manufacturer's instructions. Raw fastq files for each virus sample were subjected to de novo assembly using CLCBio software (CLC Bio, Aarhus, Denmark). Mapped assemblies were also performed using the previously reported ASFV Georgia 2007/1 sequence (FR682468) as a reference. Variant calling, genome alignments and sequence illustrations were generated using Geneious software version 10.2.4. The C315R/C147L locus from global representatives were aligned using available nucleic acid sequence data from Genbank (Fig. 2c).

Results

Illumina deep sequencing recovered near full length draft genomes for the two isolates. Average read coverage obtained for G-2008/2 (118X) was significantly higher than G-2008/1 (8.5X). A 547 bp region spanning the 3′ end of the annotated left terminal inverted repeat (TIR) region and the proximal end of the first open reading frame, multigene family (MGF) protein 360-1 L, was not recovered in the 2008/2 genome, despite significant read depth across the remainder of the genome. The architecture of the ASFV genome and genetic polymorphisms are illustrated in Fig. 1 and Table 1. The two genomes exhibited complete collinearity and identical gene content with the Georgia 2007/1 reference genome. The G-2008/1 and G-2008/2 genomes were distinguished from each other by coding changes in seven genes, including MGF 110-1 L, X69R, MGF 505-10R, EP364R, H233R, E199L, and MGF 360-21R in addition to eight homopolymer tract variations in the left ($N = 6$) and right ($N = 2$) terminal regions (Table 1). Molecular targets used in conventional ASFV subtyping assays, including p72 (B646L), P54 (E183L), CVR (pB602L), and the *I73R/I329L* TRS locus exhibited no sequence variation. Genome comparisons also revealed 56 unique homopolymer A/T-tract variations unique to the Georgia 2007/1 reference genome (Table 1).

A novel intergenic sequence insertion (GTTAAAACT AAAC) was observed in the G-2008/2 genome between genes C315R (TFIIB-like protein) and C147L (RNApol subunit 6) (Fig. 2a, b, and c). The short tandem repeat array (Fig. 2b) forms a portion of a larger indirect repeat structure illustrated in Additional File 2 (Additional file 1: Figure S1D and E). This locus displays significant copy unit diversity among global ASFV strains (Fig. 2c). The

Fig. 1 Genomic polymorphisms among Georgian ASFV isolates mapped to the Georgian 2007/1 reference genome. Variants unique to each of the three genomes are designated by separate colors (G-2007/1 = green, G-2008/1 = blue, G-2008/2 = orange)

length of the intergenic region among global strains ranged from 20 bp (Ken05-Tk1) to 93 bp (OURT and NHV) (Fig. 2b). The limits of the indel boundaries extended to the terminal end of C147L gene.

Within-sample sequence read heterogeneity occurred at six homopolymeric G/C-tracts confined to a ~ 7 kb region in the LTR (13,267-20,834) (Table 1, Fig. 1). This region is illustrated in further detail in Additional File 1 (Additional file 1: Figure S1A-C). These loci were found to possess approximately 10 or more residues and represented the only G/C homopolymer tracts of this length within the genome. Variable sites included a homopolymeric C-tract in MGF_110_13L, two homopolymeric G-tracts in genes ACD_00290 and ACD_00350, respectively, and two intergenic homopolymeric tracts (Table 1). This variability may represent sequencing artifact generated during library amplification or derives from within-host genome variation during natural infection.

Genome sequences of the Georgia 2008/1 and Georgia 2008/2 strains were deposited in GenBank under accession numbers MH910495 and MH910496, respectively.

Discussion

The genomes reported here provide additional references that may inform epidemiological tracing of related intra-epidemic isolates and provides insight into viral microevolution during acute wild-type infection. The novel allele at the C315R/C147L locus currently represents the earliest observed repeat variation among isolates derived from the current epidemic lineage. ASFV genomes exhibit elevated diversity in the multi-gene family (MGF) genes located in the left and right variable regions (LVR and RVR, respectively) [19–23]. Approximately 11 of the fifteen total SNPs that differentiated the two 2008 genomes were non-synonymous. The remaining four SNPs represent intergenic homopolymer tract variation. Additional genomic data from this epidemic lineage may allow an expanded assessment of constraints on synonymous variation.

Homopolymer tract variation

Resolution of correct homopolymer tract allele states is warranted due to their ability to generate frameshift

Table 1 Genome-wide genetic polymorphisms identified between ASFV isolates 2007/1, 2008/1, and 2008/2

Position	2007/1	2008/1	2008/2	Annotation	aa_chg	Codon	Position	2007/1	2008/1	2008/2	Annotation	aa_chg	Codon
105	(A)6	UND	(A)7	TIR region	NA	NA	63,868	(T)5	(T)6	(T)6	IG	NA	NA
1605	TT	delTT	delTT	MGF_360-1L	Frame Shift	2	63,877	(T)7	(T)8	(T)8	IG	NA	NA
1,623	(T)3	(T)4	(T)4	MGF_360-1L	Frame Shift	1	74,591	T	T	C	EP364R	M24T	2
3,247	(A)8	(A)9	(A)9	KP177R	Frame Shift	3	81,526	(T)8	(T)9	(T)9	C84L	Frame Shift	2
3,862	(A)8	(A)9	(A)9	IG	NA	NA	81,620	(A)8	(A)9	(A)9	IG	NA	NA
5,671	(A)8	(A)9	(A)9	ASFV_G_ACD_00070	Frame Shift	2	88,181	minus 1	minus 1	plus 1	IG	NA	NA
5,821	(T)8	(A)9	(A)9	IG	NA	NA	113,167	(A)5	(A)6	(A)6	IG	NA	NA
5,870	(A)7	(A)8	(A)8	IG	NA	NA	117,182	(T)8	(T)9	(T)9	IG	NA	NA
6,049	(G)5	(G)6	(G)5	MGF_110-1L	Frame Shift	1	124,705	(A)8	(A)9	(A)9	IG	NA	NA
6,095	C	C	T	MGF_110-1L	Truncation	2	124,804	(T)8	(T)10	(T)10	CP204L	Frame Shift	3
7,824	(A)8	(A)9	(A)9	ASFV_G_ACD_00120	Frame Shift	3	133,464	(A)8	(A)9	(A)9	IG	NA	NA
7,941	(A)7	(A)8	(A)8	ASFV_G_ACD_00120	Frame Shift	3	138,422	(T)8	(T)9	(T)9	D129L	Frame Shift	2
8,506	(A)7	(A)8	(A)8	IG	NA	NA	143,758	(T)7	(A)8	(A)8	IG	NA	NA
10,466	(A)6	(A)7	(A)7	IG	NA	NA	152,793	(A)5	(A)6	(A)6	IG	NA	NA
11,610	(A)9	(A)10	(A)10	ASFV_G_ACD_00190	Frame Shift	2	155,165	G	A	G	H233R	A215T	1
13,267	C14	UND	C12/13	MGF_110-14L	delG112	2	161,252	(A)8	(A)7	(A)7	QP383R	Frame Shift	2
13,666	(A)8	(A)9	(A)9	ASFV_G_ACD_00240	Frame Shift	2	161,293	(A)6	(A)7	(A)7	QP383R	Frame Shift	1
14,711	C16	C12-13	C11,14-17	MGF_110-13L	Frame Shift	2	166,192	C	C	G	E199L	A85P	1
16,663	(G)12	(G)11-12	(G)9-11	ASFV_G_ACD_00290	Frame Shift	3	166,200	G	A	G	E199L	T82I	2
16,880	(G)10	(G)8	(G)10-12	IG	NA	NA	167,764	(A)9	(A)10	(A)10	IG	NA	NA
18,737	(T)8	(T)9	(T)9	IG	NA	NA	169,455	(T)7	(T)8	(T)8	IG	NA	NA
19,033	(G)15	(G)10,13,15	(G)8,9,10	ASFV_G_ACD_00350	delGly-Gly	1	169,463	(A)7	(A)8	(A)8	IG	NA	NA
19,384	C	T	C	X69R	T53I	2	169,742	(T)9	(T)10	(T)10	I267L	Frame Shift	2
19,462	(T)8	(T)9	(T)9	IG	NA	NA	172,018	(T)9	(T)10	(T)10	IG	NA	NA
20,670	(A)7	(A)8	(A)8	IG	NA	NA	173,566	(A)7	(A)8	(A)8	IG	NA	NA
20,834	(G)11	(G)15,17	(G)9,10	IG	NA	NA	173,625	(T)9	(T)8	(T)9	IG	NA	NA
21,589	(A)8	(A)9	(A)9	IG	NA	NA	174,955	(A)5	(A)6	(A)6	V_G_ACD_01	Frame Shift	2
23,028	(A)8	(A)9	(A)9	IG	NA	NA	175,633	(A)9	(A)10	(A)9	IG	NA	NA
26,456	(T)9	(T)10	(T)10	IG	NA	NA	177,494	(A)7	(A)8	(A)8	MGF_360-16R	Frame Shift	2
31,679	(AA)4	(AA)5	(AA)5	IG	NA	NA	179,380	(A)8	(A)9	(A)9	IG	NA	NA
39,025	(T)8	(T)9	(T)9	IG	NA	NA	183,306	(T)3	(T)4	(T)4	IG	NA	NA
40,731	(T)8	(T)9	(T)9	IG	NA	NA	183,315	(T)8	(T)9	(T)9	IG	NA	NA
42,617	(T)8	(T)9	(T)9	IG	NA	NA	184,968	(T)8	(T)7	(T)7	V_G_ACD_01	Frame Shift	2
44,548	(T)8	(T)9	(T)9	IG	NA	NA	186,353	(T)8	(T)9	(T)9	IG	NA	NA
45,077	G	G	A	MGF_505-10R	E172K	1	186,802	G	G	A	IG	NA	NA
47,322	(A)6	(A)7	(A)7	IG	NA	NA	187,981	T	T	C	MGF_360-21R	silent	1
48,664	(A)7	(A)8	(A)8	A151R	Frame Shift	2	187,999	G	G	A	MGF_360-21R	E344K	1
55,975	(A)8	(A)9	(A)9	IG	NA	NA	188,019	G	G	A	MGF_360-21R	M350I	3
63,846	(T)6	(T)7	(T)7	IG	NA	NA	189,283	(T)4	(T)5	(T)5	IG	NA	NA
63,861	(A)5	(A)6	(A)6	IG	NA	NA							

mutations, alter amino acid repeat number and protein structure, produce gene fusions, and alter putative promoter and termination signals. Substitution mutations and homopolymer tract variations among the Georgian ASFV genomes were elevated in the terminal regions of the genome, particularly in the LVR (Fig. 1). It is unclear whether the homopolymer tract diversity present in the Georgia 2007/1 isolate genome reflects the original wild-type sequence, sequencing error, or mutations arising in cell culture prior to harvesting. Homopolymer tract sequences in the Georgia 2007/1 genome were reported to have been validated using Sanger sequencing [19]. The G-2008/1 and G-2008/2 genomes were sequenced directly from porcine blood samples. In contrast, the Georgia 2007/1 isolate was obtained from infected porcine organ tissue, cultured, and harvested from primary porcine bone marrow cells [19].

Under a scenario where the 2007/1 genome represents the ancestral sequence, the subsequent emergence and maintenance of homopolymeric A/T variations fixed in the G-2008/1 and G-2008/2 genomes would suggest a rapid and stable adaptation to the local host population. Previously published whole genome data gathered following extensive in vitro passage of individual strains revealed only minimal poly A/T tract variability. Whole genome comparisons of the Vero cell culture-adapted BA71V virus and the highly virulent BA71 parent strain revealed only five variable homopolymeric A/T loci and no G-tract polymorphisms [24]. Furthermore, no homopolymeric tract variations were reported in Vero-adapted Georgia 2007/1 strains (ASFV-G passages 30, 60, 80, and 110) although these genomes are not publicly available for independent verification [25]. Compared to such in vitro data, the extent of homopolymeric A/T-tract variability observed between the G-2007/1 genome and G-2008/1 and G-2008/2 appears pronounced, although the context of genome evolution may be expected to differ between published in vitro studies and wild-type genome variation reported here. We found that select homopolymeric A/T-tract loci unique to the 2007/1 genome in our analysis also displayed variation in previous studies, such as the terminal poly-T tract the CP204L gene and the homopolymer tract responsible for fusion of MGF 110-13 L homologs in L60, E75, and NHV28. Homopolymer-based variability has also previously been described in 14 additional ASFV
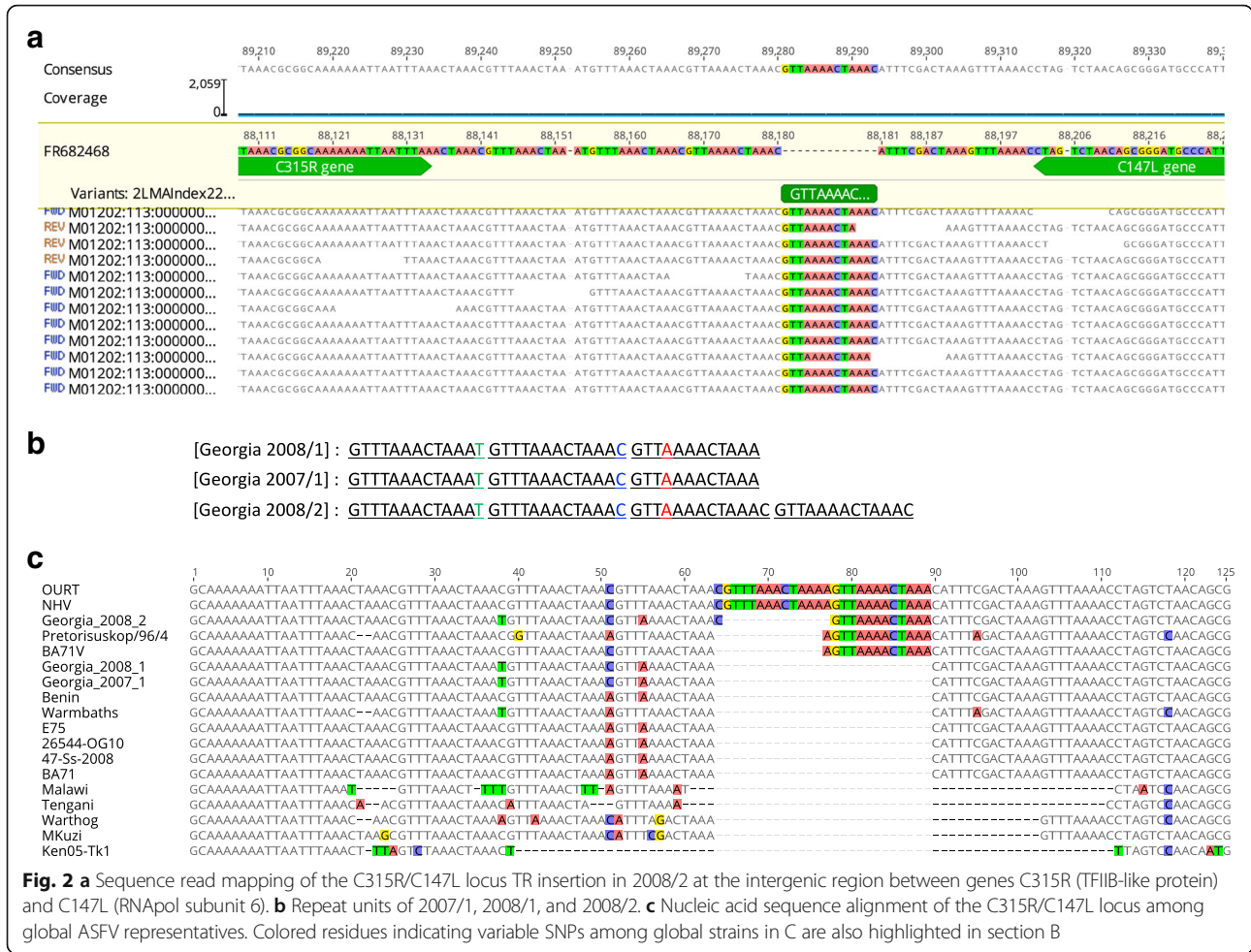
Farlow *et al. Virology Journal* (2018) 15:190

Page 5 of 8

Fig. 2 a Sequence read mapping of the C315R/C147L locus TR insertion in 2008/2 at the intergenic region between genes C315R (TFIIB-like protein) and C147L (RNApol subunit 6). **b** Repeat units of 2007/1, 2008/1, and 2008/2. **c** Nucleic acid sequence alignment of the C315R/C147L locus among global ASFV representatives. Colored residues indicating variable SNPs among global strains in C are also highlighted in section B

proteins, including p54 and E183L [26, 27]. Homopolymeric tracts that displayed notable tract length variability in other strains, such as that within the M1249 L gene of strains L60 and E7528 showed single-allele state in the three Georgian ASFV genomes here.

Molecular marker stability
Cell culture propagation is frequently performed prior to ASFV genome sequencing. The genetic stability of ASFV epi-markers in traditional ASFV in vitro cell culture systems remains understudied. Limiting errors in phylogenetic inference, such as those arising from homoplasious variation and evolutionary selection are a crucial component of molecular assay design strategy. The extent of in vitro passage or natural transmission events needed to mediate homoplasious variation in current ASFV molecular markers remains unknown and warrants further investigation. We find previous data indicate in vitro propagation may alter genetic loci currently used for ASFV subtyping. Our analysis of genome sequence data generated in a previous study [24] describing adaption of strain BA71 to Vero cells revealed that the adapted

strain BA71V uniquely possessed a new CVR (pB602L) allele that is identical to that of the Portuguese strain Por 63 (Supp. 1F). Homoplasious variation was also found to occur at the CVR (B602L) marker in select global strains during replication in pig macrophages and Vero cells [28].

In addition to their use in ASFV subtyping, tandem repeat (TR) markers are also used for dissecting molecular variation in other dsDNA viruses including Hepatitis C virus [29], Adenoviruses [30], human Cytomegaloviruses [31], and Herpesviruses [32–34]. The approach used by Deback et al. is similar to that commonly used in developing microbial variable-number tandem-repeat (VNTR) assays. This approach includes selection of candidate TR loci that are first informed by comparative genome analyses and subsequently tested for genetic stability in vitro [32]. This approach represents a viable strategy for testing existing ASFV marker stability as well as future markers relevant for tracing current intra-epidemic ASFV isolates. Cataloguing genome-wide variation among a longitudinal set of geographically diverse isolates from the current epidemic would facilitate the

Farlow *et al. Virology Journal* (2018) 15:190

Page 6 of 8

identification of new more epidemiologically-informative marker loci. The genomes reported here, in addition to the G-2007/1 genome, provide an early baseline for such efforts.

Variability in the genetic architecture of the genomic region spanning the six hypervariable G/C-tract loci in our study has been reported previously [35, 36]. The repeat tracts are positioned at the core of homopolymeric A + T rich regions within and adjacent to select MGF 110 and MGF 300 genes that together form diverse larger internal repeat sequences in the LVR (38). These regions were previously shown to possess architectural similarity to eukaryotic scaffold-associated regions (SARs) and satellite III repeats of *Drosophila* [36]. Whether the sequence read heterogeneity at these loci observed in our data represent artifact or a natural mutant spectrum is unclear. Lack of sample volume prevented further molecular validation. Each homopolymeric locus uniquely contained ≥10 Gs or Cs and was confined to the same 7 kb region in the LVR and is illustrated in Additional file 1: Figure S1A. All remaining homopolymeric G/C-tracts in the G-2008/1 and G-2008/2 genomes possessed less than 10 residues and displayed no within-sample copy number variation. Homopolymer A/T tract loci in the genome, by comparison, exhibited no detectable intra-sample variation. In contrast to the G-2007/1 genome, the G-2008/1 and G-2008/2 genomes exhibited only two homopolymeric A/T-tracts that differed and displayed no within-sample heterogeneity in their sequence read populations. The basis for the occurrence of within sample homopolymer tract variation confined only to this region of the genome in both isolates remains unknown.

Wild-type genomic homogeneity

To date, this study is the first analysis of intra-epidemic ASFV genome sequences and provides insight into genome stability during wild-type infection. DNA viruses generally display low mutation rates largely due their encoded high fidelity DNA proofreading repair enzymes. ASFV genomes uniquely encode the only known X-type polymerase (X Pol) and a DNA ligase that each exhibit low fidelity [37, 38]. The in vitro fidelity of the ASFV DNA ligase is currently the lowest known. The genetic and phenotypic basis for the extreme antigenic diversity displayed by ASFV strains has not been determined [39, 40]. Elevated antigenic diversity in field isolates has been speculated to arise from the activity of the X Pol/DNA ligase. Previous authors have hypothesized that the Pol X/ligase system functions as a strategic DNA mutator in the base excision repair (BER) pathway where rapid genetic drift occurs due to elevated DNA repair-based error [37, 41, 42]. Conflicting data for Pol X fidelity have been reported [37, 41, 43]. It has been speculated that the incongruous reports resulted from enzyme redox state differences between the

published studies [44], wherein greater fidelity is exhibited by the reduced form of the enzyme [45]. We speculate realization of the DNA mutator model proposed previously for ASFV would be reflected by greater intra-sample genome-wide variation and elevated genomic polymorphisms between the wild-type isolates studied here. Our data suggest homogeneity of within-host and inter-genome sequences. These data are consistent with other studies describing factors supporting DNA fidelity in ASFV genome replication, including the activity of the viral class II apurinic/apyrimidinic (AP) endonuclease (E296R protein) [46] and the recently reported reparative role for Pol X in vivo [47, 48].

Conclusions

These data provide insight into the intra-epidemic micro-evolution of ASFV and provides an early baseline for additional epidemiological tracing efforts in the region and elsewhere. While our effort was performed with limited resource availability in Tbilisi, Georgia, genomic data obtained directly from porcine blood was sufficient to yield genome-level data. Our results confirm the utility of deep sequencing as a highly informative culture-free method for ASFV genomic interrogation directly from PCR-positive blood. Future efforts may utilize additional sample processing methods to increase viral DNA yield prior to library preparation. Additional genome data from this epidemic will better facilitate targeted assay development and provide more robust epidemiological tracing as this epidemic continues.

Additional file

Additional file 1: Figure S1. Genetic architecture of genomic region containing hypervariable G/C-tract loci, C315R/C147L locus repeat array structures, and BA71/BA71V CVR (pB602L) locus diversity. A) Location and organization of the six hypervariable G/C-tract loci. Box B) and C) illustrate examples of within-sample sequence read diversity at hypervariable G/C-tract loci with the MGF 110-14 L gene (B) and ASFV_G_ACD_00290 (C). Direct and indirect repeat structure of Georgia 2008/1 (D) and Georgia 2008/2 (E) calculated using DNASTAR GeneQuest repeat analysis. (F) Nucleic acid sequence alignment of CVR (pB602L) locus of BA71, BA71V and Por63. (PDF 15068 kb)

Abbreviations
AP: apurinic/apyrimidinic; ASFV: African swine fever virus; BER: base excision repair; CVR: central variable region; DNA: Deoxyribonucleic acid; MGF: multi-gene family; PCR: Polymerase chain reaction; SARs: scaffold-associated regions; SNP: Single –nucleotide polymorphism; TRSs: tandem-repeat sequences

Farlow *et al. Virology Journal*   (2018) 15:190

Page 7 of 8

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Laboratory of the Ministry of Agriculture, Tbilisi, Georgia. [2]Richard G. Lugar Center for Public Health Research at the National Center for Disease Control (NCDC), Tbilisi, Georgia. [3]Farlow Scientific Consulting Company, LLC, Lewiston, UT 84320, USA.

References
1. Dixon, L. K., et al. Family *Asfarviridae*. In Virus taxonomy: seventh report of the International Committee on Taxonomy of Viruses. (Academic Press, 2000).
2. Brown F. The classification and nomenclature of viruses: summary of results of meetings of the international committee on taxonomy of viruses in Sendai, September 1984. Intervirology. 1986;25:141–3.
3. Plowright W, Parker J, Pierce MA. The epizootiology of African swine fever in Africa. Vet Rec. 1969;85:668–74.
4. Plowright W, Parker J, Pierce MA. African swine fever virus in ticks (*Ornithodoros moubata*, Murray) collected from animal burrows in Tanzania. Nature. 1969;221:1071–3.
5. Thomson, G. R., et al. The relationship between ASFV, the warthog and Ornithodoros species in southern Africa. In Proceedings of CEC'FAO (1983).
6. Rowlands RJ, et al. African swine fever virus isolate, Georgia. Emerg Infect Dis. 2008;14:1870–4.
7. Gallardo C, et al. Genetic variation among African swine fever genotype II viruses, eastern and Central Europe. Emerg Infect Dis. 2014;20:1544–7.
8. Goller KV, et al. Tandem Repeat Insertion in African Swine Fever Virus, Russia, 2012. Emerg Infect Dis. 2015;21:731–2.
9. Rodriguez JM, et al. Characterization and molecular basis of heterogeneity of the African swine fever virus envelope protein p54. J Virol. 1994;68:7244–52.
10. Bastos BAS, et al. Genotyping field strains of African swine fever virus by partial p72 gene characterisation. Arch of Virol. 2003;148:693–706.
11. Gallardo C, et al. Enhanced discrimination of African swine fever virus isolates through nucleotide sequencing of the p54, p72, and pB602L (CVR) genes. Virus Genes. 2009;38:85–95.
12. Nix RJ, Gallardo C, Hutchings G, Blanco E, Dixon LK. Molecular epidemiology of African swine fever virus studied by analysis of four variable genome regions. Arch Virol. 2006;151:2475–94.
13. Gallardo C., et al. African Swine Fever Virus p72 Genotype IX in Domestic Pigs, Congo, 2009. Emerg Infect Dis. 2011;17,1556–17,1558.
14. Sanna G, et al. Improved strategy for molecular characterization of African swine fever viruses from Sardinia, based on analysis of p30, CD2V and I73R/I329L variable regions. Transbound Emerg Dis. 2017;64:1280–6.
15. Boshoff CI, Bastos AD, Gerber LJ, Vosloo W. Genetic characterisation of African swine fever viruses from outbreaks in southern Africa (1973-1999). Vet Microbiol. 2007;121:45–55.
16. Lubisi BA, Bastos AD, Dwarka RM, Vosloo W. Molecular epidemiology of African swine fever in East Africa. Arch Virol. 2005:2439–52.
17. Malogolovkin A, Yelsukova A, Gallardo C, Tsybanov S, Kolbasov D. Molecular characterization of African swine fever virus isolates originating from outbreaks in the Russian Federation between 2007 and 2011. Vet Microbiol. 2012;158:415–9.
18. Onzere CK, et al. Virus Genes. 2018;54:111–23.
19. Chapman DAG, et al. Genomic analysis of highly virulent Georgia 2007/1 isolate of African swine fever virus. Emerg Infect Dis. 2011;17:599–605.
20. Neilan JG, Zsak L, Lu Z, Kutish GF, Afonso CL, Rock DL. Novel swine virulence determinant in the left variable region of the African swine fever virus genome. J Virol. 2002;76:3095–104.
21. Zsak L, et al. swine fever virus multigene family 360 and 530 genes are novel macrophage host range determinants. J Virol. 2001;**75**:3066–76.
22. Chapman DA, Tcherepanov V, Upton C, Dixon LK. Comparison of the genome sequences of non-pathogenic and pathogenic African swine fever virus isolates. J Gen Virol. 2008;89:397–408.
23. Yáñez RJ, et al. Analysis of the complete nucleotide sequence of African swine fever virus. Virology. 1995;208:249–78.
24. Rodríguez JM, et al. Genome sequence of African swine fever virus BA71, the virulent parental strain of the nonpathogenic and tissue-culture adapted BA71V. PLoS One. 2015;10:e0142889.
25. Krug PW, et al. The progressive adaptation of a Georgian isolate of African swine fever virus to Vero cells leads to a gradual attenuation of virulence in swine corresponding to major modifications of the viral genome. J Virol. 2015;89:2324–32.
26. Michaud V, Randriamparany T, Albina E. Comprehensive Phylogenetic Reconstructions of African Swine Fever Virus: Proposal for a New Classification and Molecular Dating of the Virus. Donlin MJ, ed. PLoS ONE. 2013;8:e69662.
27. Sun H, Jacobs SC, Smith GL, Dixon LK, Parkhouse RM. African swine fever virus gene j13L encodes a 25–27 kDa virion protein with variable numbers of amino acid repeats. J Gen Virol. 1995;76:1117–27.
28. Irusta PM, et al. Amino acid tandem repeats within a late viral gene define the central variable region of African swine fever virus. Virology. 1996;220:20–7.
29. Edwards VC, et al. Use of short tandem repeat fingerprinting to validate sample origins in hepatitis C virus molecular epidemiology studies. J Gen Virol. 2014;95:66–70.
30. Houng H-SH, et al. Adenovirus microsatellite reveals dynamics of transmission during a recent epidemic of human adenovirus serotype 14 infection. J Clin Micro. 2009;47:2243–8.
31. Picone O, Ville Y, Costa JM, Rouzioux C, Leruez-Ville M. Human cytomegalovirus (HCMV) short tandem repeats analysis in congenital infection. J Clin Virol. 2005;32:254–6.
32. Deback C, et al. Utilization of microsatellite polymorphism for differentiating herpes simplex virus type 1 strains. J Clin Virol. 2009;47:533–40.
33. Renault T, et al. Genotyping of a microsatellite locus to differentiate clinical Ostreid herpesvirus 1 specimens. Vet Res. 2014;45:3.
34. Avarre JC, Madeira JP, et al. Investigation of cyprinid herpesvirus-3 genetic diversity by a multi-locus variable number of tandem repeats analysis. J Virol Meth. 2011;173:320–7.
35. Portugal R, et al. Related strains of African swine fever virus with different virulence: genome comparison and analysis. J Gen Virol. 2015;96:408–19.
36. Almazán F, Murguía JR, Rodríguez JM, de la Vega I, Viñuela E. A set of African swine fever virus tandem repeats shares similarities with SAR-like sequences. J Gen Virol. 1995;76:729–40.
37. Showalter AK, Byeon IJL, Su MI, Tsai MD. Solution structure of a viral DNA polymerase X and evidence for a mutagenic function. Nat Struct Bio. 2001;8:942–6.
38. Benítez BAS, Arora K, Balistreri L, Schlick T. Mismatched base pair simulations for ASFV pol X/DNA complexes help interpret frequent G•G misincorporation. J Mol Bio. 2008;384:1086–97.
39. Malogolovkin A, et al. Comparative analysis of African swine fever virus genotypes and serogroups. Emer Inf Dis. 2015;21:312–5.
40. García-Barreno B, Sanz A, Nogal ML, Viñuela E, Enjuanes L. Monoclonal antibodies of African swine fever virus: antigenic differences among field virus isolates and viruses passaged in cell culture. J Virol. 1986;58:385–92.
41. Lamarche BJ, Showalter AK, Tsai MD. An error-prone viral DNA ligase. Biochemistry. 2005;44:8408–17.
42. Lamarche BJ, Tsai MD. Contributions of an endonuclease IV homologue to DNA repair in the African swine fever virus. Biochemistry. 2006;45:2790–803.
43. Garcia-Escudero R, Garcia-Diaz M, Salas ML, Blanco L, Salas J. DNA polymerase X of African swine fever virus: insertion fidelity on gapped DNA substrates and AP lyase activity support a role in base excision repair of viral DNA. J Mol Bio. 2003;326:1403–12.
44. Dixon LK, Chapman DA, Netherton CL, Upton C. African swine fever virus replication and genomics. Virus Res. 2012;173:3–14.

Farlow *et al. Virology Journal*   (2018) 15:190

Page 8 of 8

45. Voehler MW, Eoff RL, McDonald WH, Guengerich FP, Stone MP. Modulation of the structure, catalytic activity, and fidelity of African swine fever virus DNA polymerase X by a reversible disulfide switch. J Biol Chem. 2009;284: 18434–44.

46. Redrejo-Rodríguez M, García-Escudero R, Yáñez-Muñoz RJ, Salas ML, Salas J. African swine fever virus protein pE296R is a DNA repair Apurinic/ Apyrimidinic endonuclease required for virus growth in swine macrophages. J Virol. 2006;80:4847–57.

47. Redrejo-Rodríguez M, Rodríguez JM, Suárez C, Salas J, Salas ML. Involvement of the reparative DNA polymerase pol X of African swine fever virus in the maintenance of viral genome stability in vivo. J Virol. 2013;87:9780–7.

48. Kearse M, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.