SHORT TERM SCIENTIFIC MISSION (STSM) – SCIENTIFIC REPORT

The STSM applicant submits this report for approval to the STSM coordinator

Action number: CA15116, STSM reference number: 38765
STSM title: Improvement of molecular typing methods for use in molecular epidemiology studies of African swine fever (ASF) virus
STSM start and end date: 20/11/2017 to 01/12/2017
Grantee name: Annika Vilem

PURPOSE OF THE STSM

After an extensive spread of the ASF virus (ASFV) genotype II in Eastern Europa, the first case of ASF in Estonia was diagnosed in September 2014. By the end of November 2017, 3540 ASF-positive wild boar have been found as well as 26 domestic pig outbreaks have been confirmed. The disease has spread all over Estonia and only one county out of 15 is still free of the disease.

Although ASFV genotype II is circulating in Eastern Europa already for ten years, we know about the virus a lot less than we would like. Therefore, it is crucial to contribute as much as possible to the improvement of molecular typing methods, especially in the affected countries. Up to now, the EURL for ASF (CISA-INIA, Spain) has supported Estonian Veterinary and Food Laboratory (acting also as NRL for ASF in Estonia), sequencing a part of ASFV positive samples. As the result of these analysis, one new CVR variant (CVR2) has been found in Estonia and also, one SNP (single nucleotide polymorphism) within CVR variant 1. Furthermore, has been the great co-operation with the Friedrich-Loeffler-Institut (Germany) with the aim to characterize different virus strains circulating in different regions in Estonia. Up to now, this co-operation has been very fruitful. Based on the results of animal experiments and next generation sequencing (NGS), the association between deletion at the 5'- end of Estonian ASFV strains and attenuated phenotype has raised (Zani et al., 2018-in preparation).

The main goal of the short term scientific mission to the Friedrich-Loeffler-Institut was to strengthen and improve the knowledges of ASFV sequencing and data analysis. Furthermore, to get an overview and basic experience of the NGS approaches.

During two weeks of short term scientific mission genetic characterization of 6 ASFV isolates was performed by PCR and further sequencing throughout the initial analysis of different regions of ASF genome. The regions were i) the C-terminal end of the VP-72 protein coding gene, which differentiates up to 23 distinct genotypes ii) the central variable region (CVR) within the B602L-gene iii) the full E183L gene encoding the p54 protein as a valuable additional genotyping method for molecular epidemiological studies of p72 genotype I viruses, particularly in West Africa where this genotype predominates and iv) the gene K196R coding for the ASFV thymidine kinase. Samples 1 to 5 were obtained from ASFV positive domestic pig samples from Ukraine

and sample 6 was an ASFV isolate obtained from an African *Ornithodoros porcinus* tick.

Secondly, metagenomic analyses of four bovine swab samples were performed using IonTorrent S5 XL NGS platform to get an overview of the important steps and workflow of the next generation sequencing, which can be successfully used for ASF sequencing also.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

Material and Methods for molecular characterization of 6 ASFV isolates

Amplification of the regions of ASFV genome

The C-terminal end of VP-72 coding protein gene, the CVR within gene B602L, the genes E183L and K196R were amplified using Phusion Green Hot Start II High-Fidelity PCR Master Mix (Thermo Scientific) according to the manufacturer's recommendations and primers listed in table 1. After amplification, the PCR product sizes were evaluated by gel electrophoresis.

Table 1 Primer sequences used in amplification of different regions of ASFV genome

Name of the primer	Sequence 5'-3'	Amplicon size
B646L-P72_for_2	TGGAAGCCCACAGATCGAAC	1100 bp
B646L-P72_ rev_2	GCGCTCTGGATTAAGTTGCG	
B602L2For	ACTTTGAAACAGGAAACATAATGATG	450 bp
B602L2rev	ATATTTTGTAATATGTGGGCTGCTG	
ASFV_P54_for1	AGA ATG GTG CAC ACG TAG CT	762 bp
ASFV_P54_rev1	TCT GTA ATT TCA TTG CGC CAC A	
ASFV_TK_for1	TCT TCG TCC ATG CAT TCC GG	795 bp
ASFV_TK_rev1	GAT GAC CCT CGA GCT TCG TA	

Purification of the PCR products and amplification for sequencing

PCR products with expected amplicon size were cut out of the gel and purified with QIAquick Gel extraction kit (Qiagen) according to the manufacturer's instructions. The amplification of the DNA for both strands was performed separately using BigDye Terminator v3.1 Cycle Sequencing kit (Thermo Scientific) and primers listed in table 1. PCR amplicons were purified from residual enzyme, dNTPs and primers using NucleoSEQ® Columns and sequenced with ABI 3130 Genetic Analyzer.

Material and methods for the metagenomic analysis of four bovine swab samples by NGS

RNA extraction from swabs and quantification

RNA was extracted using TRIZOL LS Reagent and QIAamp RNeasy Mini kit (Qiagen). Some minor derivations from manufacturer's instructions were made. Extracted RNA was quantified with Agilent 2100 Bioanalyzer using Agilent RNA 6000 Pico chip according to the

manufacturer's recommendations.

cDNA syntesis

Before cDNA synthesis, RNA was concentrated using Agencourt RNAClean XP beads, which utilizes solid-phase paramagnetic bead technology for purification and concentration of RNA. The magnetic beads were washed with ethanol and total RNA was finally eluted in 20 µl of nuclease free water. The synthesis of double stranded (ds) cDNA from concentrated RNA was performed using the cDNA Synthesis System from Roche. The ds cDNA is required for the subsequent construction of non-directional cDNA libraries for next-generation sequencing.

cDNA fragmentation and DNA library construction

dsDNA was fragmented to 500bp fragments using the Covaris M220 Focused ultrasonicator. As the volume of DNA after fragmentation is too high, the DNA needs to be concentrated using Agencourt AMPure XP beads. After washing the beads with ethanol twice, the DNA was eluted with 20,5 µl of nuclease free water. Library preparation including end-repair and barcode ligation for sequencing with IonTorrent S5 XL was performed using GeneRead DNA Library L Core Kit (Qiagen) and Ion Xpress Barcode Adapters (Life Technologies) according to the manufacturer's instructions. End-repaired and adapter ligated IonTorrent library was subsequently purified and concentrated using Agencourt AMPure XP beads. Library was eluted in 50 µl nuclease free water.

Size exclusion

After DNA fragmentation the obtained fragments have different sizes. For optimal sequencing results by IonTorrent S5 XL platform, the peak size 500bp is recommended. DNA fragments sized >750 bp were removed using diluted Agencourt AMPure XP beads. After binding the large fragments to the beads, the cleared solution was transferred to a new tube, where binding the average size of DNA fragments was occurred using undiluted Agencourt AMPure XP beads. Magnetic beads were washed twice with ethanol and library was eluted in 30 µl EB Buffer.

DNA Library quantification and pooling

DNA library was quantified using Agilent 2100 Bioanalyzer and high sensitivity DNA chip from Agilent Technologies. After receiving the fragment sizes the KAPA PCR was performed to quantify the exact molarity of the libraries. Dilutions of the libraries were analyzed in duplicate using the KAPA Library Quantification Ion Torrent kit. Results received from PCR run were inserted to the excel file where the concentrations of the libraries were calculated. Libraries were pooled subsequently according to the concentrations received.

Emotion PCR for library amplification and the chip loading

Library was amplified using Ion S5 Reagent Mix and Ion OneTouch 2 instrument. The library enrichment was subsequently followed with Ion OneTouch ES instrument using magnetic beads to get rid of the empty ISPs, which will take up wells on the chip and therefore need to be removed. The cleaned and amplified library was transferred to the ION530 chip according to the manufacturer's instructions and subsequently sequenced with IonTorrent S5 XL platform.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

Molecular characterization of six ASF isolates

All six ASFV positive samples were successfully amplified and all gene amplicons corresponded to

the sizes that are shown in table 1. PCR amplicons were sequenced in order to place them in one of the 23 defined p72 genotypes. Samples 1-5 were clustered in genotype 2 and showed 100% nucleotide identity with all compared ASFV isolates from Georgia, Estonia, Lithuania and Poland. Sample 6 had 98% identity to the Kenya 2009 isolate from a tick and clustered in genotype 10. Five nucleotide changes were observed in sample 6 and only one nucleotide change resulted in an amino acid change (S instead of T).

According to the full gene of the E183L gene encoding the p54 protein, samples 1-5 showed 100% nucleotide identity with all compared genotype 2 ASFV isolates from Ukraine, Poland, Lithuania, Belarus and Russian Federation. Sample 6 revealed insertions of three amino acid tandem repeat compared to the closest Kenya 2009 tick isolate.

CVR subtyping placed the samples 1-5 to the genotype 2 CVR variant 1 with the presence of 10 copies of amino acid tetramer repeats, which are identical to the ASFV isolates circulating in Estonia, Lithuania and Ukraine, except of that circulating in certain areas of Tartu Estonia characterized by the deletion of three amino acid tetramer repeats and western part of Estonia characterized by the SNP within CVR variant 1. CVR subtyping of sample 6 revealed one insertion of a tetrameric repeat (NVDT) and 16 nucleotide changes which resulted in six amino acid change compared to the closest Kenya 2009 and 2005 sequences isolated from ticks (see Fig 1).

Fig. 1 Illustration of amino acid alignment of sample six (kam17-4668 and kam17-4662)

Sample 1-5 thymidine kinase sequence of the K196R gene were compared to the Georgia 2007 full genome and showed 100% nucleotide identity. Sample 6 could not be compared because of the bad sequence quality.

Description of NGS results

Library concentrations and expected number of the reads are shown in table 2.

Table 2 Library concentrations, expected and received number of the reads

Name of the library	Concentration pM	Expected number of reads	Received number of reads
Lib02396	0,5	350000	15176
Lib02397	0,6	437500	71590
Lib02398	0,6	437500	68511
Lib02399	1,0	822500	267953

Additional five libraries were sequenced in the same sequencing run and are not presented in table 2. According to the run report 83% of the chip loading was successful. 72% of the ISPs were clonal and 54% (ca 12 mln reads) of the ISPs were in final library. 24% of the ISPs had low quality and 20% were adapter dimers. The amount of adapter dimers was high probably due to low concentration of the library. Mean read length was 252bp. Received number of reads were lower than expected because of the primer dimers.

Reads were classified using the software pipeline RIEMS (Scheuch, M et al 2015), which was designed for the analysis of metagenomics datasets, comprising software of Newbler 3.0 (454 Life Sciences-A Roche Company), blast 2.6.0 and Emboss 6.3.1. No substantial results were observed in lib02396, lib02397, lib02398 and are therefore not presented further. The run report of the lib02399 was conducted using RIEMS 4.0 software (Scheuch, M et al 2015). Reads of the lib02399 classified into superkingdoms are shown in figure 2.

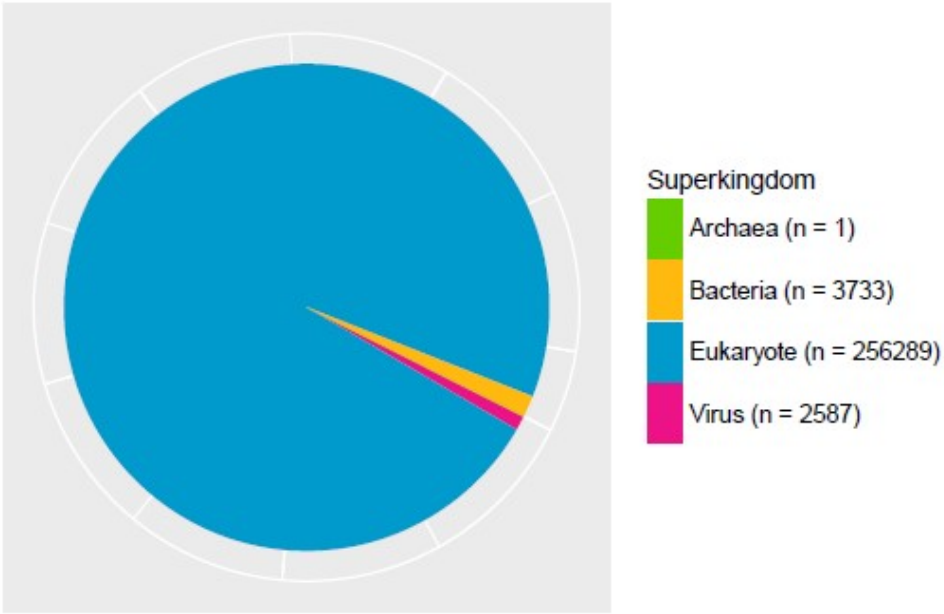

Fig. 2 Pie chart of the lib02399 reads classified into superkingdoms. In the legend, the absolute number of reads classified into the respective taxa is given (lib02399.IonXpress-029.R-2017-11-24-11-50-11-user-S5XL-0003-17-lauf0017-resultprotocol.pdf, Scheuch, M et al 2015)

Although the library concentrations were significantly low, the sequencing results of lib02399 were promising. Based on the count of reads assigned to *Betacoronavirus 1* it can be concluded that viral load in the sample was significantly high and this virus may be the cause of the illness in bovine herd. Further investigations of this virus should be considered. Read assignments to the superkingdom Virus is shown in figure 3

Family	*Species*	SkTax	FamTax	Tax	counts
Coronaviridae	*Betacoronavirus 1*	10239	11118	694003	2400
Coronaviridae	*Camel coronavirus HKU23*	10239	11118	1699096	145
Unknown	*Xenohaliotis phage pCXc HC2016*	10239		1933103	23
Retroviridae	*Avian myeloblastosis associated virus*	10239	11632	11960	4
Arenaviridae	*Guanarito mammarenavirus*	10239	11617	45219	4

Fig 3 Read assignments to the superkingdom Virus (lib02399.IonXpress-029.R-2017-11-24-11-50-11-user-S5XL-0003-17-lauf0017-resultprotocol.pdf, Scheuch, M et al 2015)

ACKNOWLEDGEMENTS

I would like to thank Dr. Sandra Blome for the invitation to the Friedrich-Loeffler-Institut and sharing her work with me. Further, I would like to thank Jan Hendrik Forth for sharing his good knowledge of African swine fever sequencing and giving me all his support during my visit to the FLI. I acknowledge the whole staff from labor Höper, especially Leonie Forth and Patrick Zitzow for giving me a good introduction to next generation sequencing and sharing their best knowledge.

Furthermore, I would like to thank the European Cooperation in Science and Technology (COST) for partially funding the Short Term Scientific Mission (COST Action: CA15116, STSM reference number: 38765) within the COST Action Understanding and Combating African Swine Fever in Europe (ASF-STOP).

References

Scheuch, M, D.Höper and M.Beer (2015)
RIEMS: a software pipeline for sensitive and comprehensive taxonomic classification of reads from metagnomics datasets. BMC Bioinformatics 16(1):69

L. Zani, L. Kauer, J. H. Forth, S. Leidenberger,J. Carlson,I. Nurmoja, A. Viltrop, M.Beer, D.Höper, S.Blome (2018, in preparation)
Deletion at the 5'-end of Estonian ASFV strains associated with an attenuated phenotype?

Result Protocol- lib02399.IonXpress-029.R-2017-11-24-11-50-11-user-S5XL-0003-17-lauf0017-resultprotocol.pdf