Evaluation of un-methylated DNA enrichment in sequencing of African swine fever virus complete genome

Rémi Pereira de Oliveira, Pierrick Lucas, Amélie Chastagner, Claire de Boisseson, Laurence Vial, Marie-Frédérique Le Potier, Yannick Blanchard

Evaluation of un-methylated DNA enrichment in sequencing of African Swine Fever Virus complete genome.

Rémi Pereira De Oliveira[a, b, c], Pierrick Lucas[d], Amélie Chastagner[a], Claire De Boisseson[d], Laurence Vial[b, c], Marie-Frédérique Le Potier[a], Yannick Blanchard[d,*]

[a] Swine Virology Immunology Unit, Ploufragan-Plouzané-Niort Laboratory, ANSES, Ploufragan, France

[b] *UMR ASTRE, CIRAD, Campus International de Baillarguet, F-34398 Montpellier, France*

[c] *UMR ASTRE, Univ Montpellier, CIRAD, INRAE, Campus International de Baillarguet, F-34398, Montpellier, France*

[d] Viral Genetic and Biosecurity Unit, Ploufragan-Plouzané-Niort Laboratory, ANSES, Ploufragan, France

[*]Corresponding author. E-mail address: Yannick.BLANCHARD@anses.fr (Y. Blanchard).

Abstract

African swine fever is a febrile hemorrhagic fever disease that is caused by the African swine fever virus (ASFV) and is lethal for domestic pigs and wild boar. ASFV also infects soft ticks of the genus *Ornithodoros,* some species of which can act as a vector for ASFV. Whole genome sequencing of ASFV is a challenge because, due to the size difference of the host genome versus the viral genome, the higher proportion of host versus virus DNA fragments renders the virus sequencing poorly efficient. A novel approach of DNA enrichment, based on the separation of methylated and un-methylated DNA, has been reported but without an evaluation of its efficacy. In this study, the efficiency of the un-methylated DNA enrichment protocol was evaluated for pig and tick samples infected by ASFV. As expected, fewer reads corresponding to ASFV were found in the methylated fraction compared to the un-methylated fraction. However, the sequencing coverage of the un-methylated fraction was not improved compared to the untreated DNA. In our hands, the ASFV DNA enrichment was inefficient for tick samples and very limited for pig samples. This enrichment process

represents extra work and cost without a significant improvement of ASFV genome coverage. The efficiency of this enrichment approach and the cost/benefit ratio are discussed.

Keywords: African swine fever virus, NGS, enrichment, methylation

1-Introduction

African swine fever (ASF) is a highly contagious hemorrhagic fever in *Suidae* that is lethal for domestic pigs and wild boar. The African swine fever virus (ASFV), a large enveloped DNA virus with a genome size between 170 and 190 kb, is the etiologic agent of ASF and the only member of *Asfarviridae* family. ASFV can infect two different hosts, *Suidae* and soft ticks of the genus *Ornithodoros*, the last being a possible vector of ASFV (Boinas et al., 2011; Plowright et al., 1969). Although ASFV is a large DNA virus, with a genome of 170 to 190 kb, this genome is far smaller than the genomes of its two hosts. Pigs have a genome of 2.8 billion kb (Groenen et al., 2012) while ticks have a genome estimated at 1 billion kb (Geraci et al., 2007). Standard DNA extraction methods, like commercial kits using an affinity column or phenol chloroform, do not separate the host and viral DNA genomes, leading to a heavily unbalanced output of sequences in favor of the host's genome, even after cell culture amplification of the virus prior to sequencing. This is particularly critical for samples that have low viral titers, such as tick samples, making the sequencing of the virus from these samples highly challenging.

The first complete ASFV sequence, published in 1995, was of the non-virulent isolate BA71V, a Vero cell culture adapted strain (Yáñez et al., 1995). This first full ASFV genome sequence, 170 kb long, was obtained by Sanger sequencing after viral DNA fragmentation and cloning of the whole genome into several plasmids. Despite the subsequent emergence and use of Next Generation Sequencing (NGS), the sequencing of ASFV genomes has remained challenging, and ASFV sequences have had to be completed or confirmed by PCR amplification of some DNA fragments followed by

Sanger sequencing (Chapman et al., 2011; de Villiers et al., 2010; Olesen et al., 2018; Zani et al., 2018).

Different strategies for viral genome enrichment have been used, including centrifugation/ultracentrifugation, filtration, and DNAse and RNAse treatments (Hall et al., 2014). Genome capture, a more sophisticated technique based on a library of probes specifically designed to selectively hybridize to a target genome, also has been used for the sequencing from clinical samples of viruses that are difficult to sequence like Herpes Simplex Virus (Briese et al., 2015; Gaudin and Desnues, 2018; Greninger et al., 2018). While very efficient, capture techniques are expensive, and the preliminary knowledge of the viral sequences needed for the design of the probes also limits their use for quasispecies studies.

In eukaryotes, DNA methylation is an epigenetic modification targeting CpG dinucleotides, extremely stable DNA sequences frequently found in CpG islands that are involved in gene regulation (Dor and Cedar, 2018). DNA methylation is also present in some large DNA viruses, such as Herpesviruses and Iridoviruses, which replicate in the nucleus (Hoelzer et al., 2008). Interestingly, contrary to most DNA viruses, the viral replication of ASFV occurs exclusively in the cytoplasm of infected host cells (Dixon et al., 2013), whereas the DNA methylase is exclusively present in the nucleus of the cell. The ASFV genome thus should avoid CpG methylation, as suggested by a recent study demonstrating, on a small part of the genome, that ASFV DNA was un-methylated (Weber et al., 2018). Recently, new complete genome sequences of African ASFV strains were obtained after an enrichment step based on the separation of methylated and un-methylated DNA (Masembe et al., 2018; Ndlovu et al., 2020). However, the efficiency of this approach on the different samples used (high or low quantity of ASFV DNA) was not evaluated.

In this study, we investigated the efficiency of this un-methylated DNA enrichment strategy for ASFV sequencing. Two types of samples infected with the Georgia2007/1 ASFV strain were tested: 1) blood of viremic pigs, representing samples with a high quantity of ASFV DNA, and 2)

infected *Ornithodoros erraticus*, a European soft tick from Portugal, representing samples with a low quantity of ASFV DNA. If efficient, sample enrichment based on methylated and un-methylated DNA discrimination should separate host DNA from viral DNA. For comparison purposes, we sequenced the untreated DNA (UD), the methylated fraction (MF) and the un-methylated fraction (UMF) resulting from the enrichment process.

2-Materials and methods

2.1-Infection of SPF pigs and soft tick *Ornithodoros erraticus*

The ASFV strain Georgia2007/1 was used for this experiment. This strain, member of the ASFV genotype II group, was initially isolated from domestic pigs in Georgia in 2007 and was kindly provided by Dr. Linda Dixon (OIE reference Laboratory, Pirbright Institute, UK). This strain has a genome size of 189 344 bp coding for 269 ORFs (Forth et al., 2019). Five specific-pathogen-free (SPF) pigs (889Geo, 936Geo, 941Geo, 6547Geo, 6518Geo) were inoculated intra-muscularly with 10^4 50% hemadsorbing dose (HAD_{50}) of Georgia2007/1 ASFV strain. Heparin blood samples were collected on viremia at day 3 post-inoculation. *Ornithodoros erraticus* (E geo F1-6, E geo M1-2) were infected by feeding on Georgia2007/1 infected pigs and were frozen three months after the blood meal.

2.2-DNA extraction

DNA extractions were performed with the High Pure PCR Template Preparation Kit (Roche Diagnostics, Meylan, France). For pigs, DNA extraction was realized from 400 µL of heparin blood sample. Ticks, at 3 months post-infection, were washed 5 secs on sodium hypochlorite 2.6 % then 3 times in sterile Phosphate Buffered Saline (PBS). After washing, ticks were crushed in 200 µL of sterile PBS with two steel beads of 3 mm and 4 mm at 25 Hz during 3 min with VWR Star-Beater (VWR International bvba, Leuven, Belgium). Then 800 µL of sterile PBS were added and the whole solution was clarified by centrifugation at 5000 g during 5 min. Supernatants were filtered on 0.22 µm Sterile Millex filter (Merck Milipore, Carrigtwohill, Ireland) for E geo F1-2, E geo M1 and 0.45 µm filter for E

98 geo F3-6 and E geo M2. Filtered supernatants were centrifuged at 13000 g during 5 min and 400 µL
99 of the supernatants were used for DNA extraction.

100 **2.3-Real-time PCR**

101 Each DNA extract was tested by quantitative real-time PCR for ASFV VP-72 and beta-actin
102 genes. The primers and probe used for ASFV VP-72 gene were previously described (Tignon et al.,
103 2011). The beta-actin primers and probes used in our study were also described for swine by Tignon
104 et al.. For the tick samples, the primers described by Duron et al. (2018) were used, however, instead
105 of SYBR Green, a Taqman probe was designed for our duplex real-time PCR Hex-5'-
106 CGAGAGGAAGTACTCCGTCTGG-3'-BHQ1.

107 **2.4-Quantification of DNA, sample enrichment, library preparation and sequencing**

108 DNA samples were assayed with Qubit Fluorometer (Invitrogen, Paisley, UK). For the pig
109 samples, DNA quantification was performed with 1 µL of DNA; for the tick samples, DNA
110 quantification was performed with 10 µL of DNA. The quantities of DNA obtained for each initial
111 sample are presented in Table 1.

112 After purification, all initial DNA samples were submitted to an enrichment step by
113 separating methylated DNA and non-methylated DNA with NEBNext Microbiome DNA Enrichment Kit
114 (NEW ENGLAND Biolabs, Evry, France) according to the manufacturer's protocol. More precisely, for
115 the pig samples, 100 ng of DNA were used for the NE library and the leftover (around 550 ng) for the
116 enrichment procedure. For the tick samples, 100 ng of DNA were also used for the NE library and the
117 leftover DNA, which varied between 260 ng and 450 ng, for the enrichment procedure. After the
118 enrichment procedure, each initial sample was then split into three fractions: 1) fraction 1, the initial
119 DNA sample, corresponding to an aliquot before the enrichment process (UD), 2) fraction 2, the
120 enriched methylated DNA sample (MF), and 3) fraction 3, the depleted methylated DNA sample
121 (UMF). DNA quantification of the MF and UMF fractions was performed after the enrichment

process. For library preparation, Ion Xpress Plus Fragment Library Kit (Thermo Fischer Scientific, Frederick, Maryland, USA) was used and DNA fragments between 250 bp and 290 bp were size-selected with Ion Xpress Barcode Adapters 1-96 kit (Thermo Fischer Scientific, Frederick, Maryland, USA). For the DNA purification steps, magnetic beads from Agencourt AMPure XP Kit (Beckman Coulter, Villepinte, France) were used.

All samples were sequenced with Proton Ion Torrent technology (Thermofischer Scientific, Frederick, Maryland, USA). For the sequencing, the different libraries (UD, UMF, MF) were multiplexed and simultaneously sequenced on a P1 chip for Proton sequencing. The P1 chip allows the production of up to 100 M reads. The Phred Quality Score ≥20 was calculated for generated bases and expressed as a percentage (%≥Q20). Row data can be found in the GenBank SRA under accession no. PRJNA643370.

2.5-Bioinformatic analysis

The reads were cleaned with Trimmomatic 0.36 software using the following parameters: ILLUMINACLIP:oligos.fasta: 2:30:5:1: true; LEADING: 3; TRAILING: 3; MAXINFO: 40:0.2; MINLEN: 36 (Bolger et al., 2014). An alignment then was performed using Bowtie2 (version 2.2.5) (Langmead and Salzberg, 2012) set on very fast (-D 5 –R 1 –N O –L 22 –i s,0,2.50) with cleaned down-sampled reads on a local ncbi nucleotide database. This very fast alignment detected only the highly similar sequences (>99% homology) and provided a snapshot picture of the data. The bam files were then converted to a blast output readable by Krona taxonomic viewer (Ondov et al., 2011) to visualize the gross organisms distribution in the different samples. For an accurate count of ASFV reads, the sequence of strain Georgia2007/1, Genbank accession number FR682468.1 (Chapman et al., 2011), was used as reference for a Bowtie2 alignment (command line option –al) versus all cleaned reads.

2.7-Statistical analysis

Statistical analyses were performed with RStudio software (version 1.1.463). For all data analyses, three groups were considered: 1) the pool of UD fractions, 2) the pool of UMF fractions and 3) the pool of MF fractions. The Wilxocon test was used for the statistical analysis of these three groups. Our analysis focused on the impact that the enrichment process had on the NGS results, and more specifically on ASFV genome enrichment and coverage between each group.

3-Results

3.1-Detection and relative quantification of viral and host DNA in initial samples

A duplex quantitative real-time PCR directed against ASFV and beta-actin was performed for all samples and the ASFV/beta-actin ratios of the PCR Ct were calculated for the estimation of the relative quantity of the genomes of ASFV and the hosts (Table 1). Assuming a similar efficiency of our PCR reactions, a ratio of less than 1 indicates a higher level (by genome copy number) of ASFV genomes than host genomes. Such a ratio (less than 1) was found for pig samples. For tick samples, the ratio was superior to 1, excepting the E geo F1 and E geo M1 samples for which the ratio was respectively 0.99 and 0.96.

3.2-Quality analysis of generated data

After the enrichment process, for pig samples, 158 ng to 205 ng of total DNA were obtained for UMF and 208 ng to 325 ng for MF. For tick samples, 153 ng to 391 ng of total DNA were obtained after enrichment for UMF and 9 ng to 24 ng for MF (Table 2).

The generated bases, reads and Phred quality scores are shown in Table 3. Briefly, for pig samples, a mean number of 6 635 376 reads were generated for UD, 2 510 484 for UMF and 1 963 487 for MF. For tick samples, a mean number of 4 820 610 reads were generated for UD, 7 317 952 for UMF and 1 733 646 for MF.

167 The Phred quality scores were equivalent between all of the libraries, for both pig and tick
168 samples.

169 **3.3-Metagenomic analysis**

170 Due to a high variability in the number of reads obtained by samples, from 15 000 to $21x10^6$
171 reads (Table 3), and for computing time efficiency, the metagenomic classification of the reads was
172 performed on subsets of the raw data ranging from 5 to 100% of the reads according to samples
173 (Table 4). The output of taxonomical assignment of the reads with the fast bowtie alignment varied
174 greatly according to the samples, with lower outputs for tick samples due to a poor representation of
175 this species in the databases. As far as possible, the order of magnitude in the number of reads
176 analyzed was kept in a range of hundreds of thousands of reads to be representative of the samples,
177 with the exception of the MF fraction of the 889Geo sample, which displays only 15 736 reads (Table
178 4).

179 The taxonomic classification of reads obtained from pig samples was divided into three main
180 groups: pig genome, ASFV genome and other sequences, which represent reads that are not assigned
181 to a specific taxon. For UD pig samples, 93% of detected reads belonged to the pig genome, 0.18% to
182 the ASFV genome and 6.82% to other sequences (Fig. 1A). For UMF pig samples, 83% of reads
183 belonged to the pig genome, 0.75% to the ASFV genome and 16.25% to other sequences (Fig. 1B).
184 For MF pig samples, 92.20% of reads belonged to the pig genome, 0.04% to the ASFV genome and
185 7.76% to other sequences (Fig. 1C). A statistical difference was found only for the proportion of ASFV
186 reads obtained in UD and UMF (p-value = 0.034).

187 The taxonomic classification of reads obtained from tick samples was divided into five
188 groups: tick, pig, ASFV, bacteria and other sequences. For UD tick samples, 38.63% were identified as
189 tick, 27.71% as pig, 0.03% as ASFV, 3.85% as bacteria and 32.78% as other sequences (Fig. 1D). For
190 UMF tick samples, 35.38% corresponded to tick, 25.71% to pig, 0.04% to ASFV, 4.48% to bacteria and
191 34.39% to other sequences (Fig. 1E). For MF samples, 36.63% corresponded to tick, 24.35% to pig, 0%

192 to ASFV, 1.34% to bacteria and 37.68% to other sequences (Fig. 1F). No statistical difference was
193 found between UD, UMF and MF samples.

194 **Fig. 1: Relative abundance of reads generated in the different samples classified according the**
195 **taxonomic origin**. Pig samples are presented in the left of the panel, and tick samples in the right. UD
196 samples are presented in **A** and **D**. UMF samples are presented in **B** and **E**. MF samples are presented
197 in **C** and **F**.

198

199 **3.4-Number and percentage of ASFV reads in all samples**

200 For all samples, the number of reads, the percentage and the mean coverage of reads aligned
201 on the ASFV reference are shown in Table 5.

202 For pig samples, the quantity of ASFV reads varied between 3000 and 71 286 for UD,
203 between 95 and 40 620 for UMF, and between 45 and 1220 for MF, representing 0.29%, 0.35% and
204 0.098% of total reads obtained for UD, UMF and MF of pig samples, respectively. The slight
205 difference in the number of ASFV reads between the UD and MF fractions was statistically significant
206 (p-value = 0.024).

207 For tick samples, the quantity of ASFV reads varied between 0 and 604 for UD, between 0
208 and 813 for UMF, and between 0 and 57 for MF. ASFV reads represented, for the best fraction,
209 0.002% of the total reads in tick samples (Table 5). No statistical difference was found between the
210 UD, UMF and MF samples.

211 ASFV genome sequencing coverages are also shown in Table 5. For pig samples, the mean
212 coverage was 15.11 for UD, 9.63 for UMF and 0.25 for MF fractions. For tick samples, the mean
213 coverage was 0.11 for UD, 0.12 for UMF and 0.005 for MF fractions.

214 **4-Discussion and conclusion**

In this study, we evaluated a new strategy of DNA enrichment for high-throughput sequencing of the ASFV genome. The strategy, already used in two studies (Masembe et al., 2018; Ndlovu et al., 2020) but without an evaluation of its efficacy, is based on a specificity of the ASFV DNA genome which is not methylated, contrary to the host genome. Separation of methylated and un-methylated DNA should result in an enrichment in ASFV reads in un-methylated samples. This strategy was tested against samples corresponding to the different ASFV hosts, i.e., pig and tick samples.

Before the enrichment process, ratios in genome copy number of ASFV and hosts were roughly estimated (by PCR ratio) to be 1 in tick samples, and around 0.7 in pig samples. After mapping the reads on the different genomes, the ASFV DNA indeed represented less than 0.4% and 0.01% of the reads obtained in pig and tick samples respectively.

For all pig samples, in the UMF fractions (depleted methylated DNA), the percentage of ASFV genome reads increased from 0.18% to 0.75% and, conversely, the percentage of pig genome reads decreased by 10% compared with the UD fractions. For tick samples, the proportion of reads corresponding to the host genome decreased by 3% between UMF and UD, indicating a low efficiency of the enrichment process. The decrease between the UD and UMF factions was more extensive for pig genome reads (10%) than the decrease for tick genome reads (3%). These observations can be explained by the difference in DNA methylation between invertebrates and vertebrates. Indeed, the DNA of invertebrates is less methylated than the DNA of vertebrates (Bird and Taggart, 1980; Tweedie et al., 1997). However, the sequence of the whole *Ornithodoros erraticus* genome is unknown, so the percentage of tick reads is probably underestimated. Moreover, the proportion of ASFV reads in tick samples only increased in two samples out of eight and remained lower than 0.01%, which in our case is insufficient to obtain the whole sequence of the ASFV genome, with a mean coverage of less than 0.61.

In comparison with another enrichment approach for ASFV whole genome sequencing, DNA extraction from erythrocytes of a viremic pig showed 0.84% of ASFV reads in their total reads with a mean coverage of 103 reads per nucleotide (Olesen et al., 2018). This approach could be a good strategy for fresh isolated erythrocytes but is not suitable for frozen whole blood.

In 2019, a workflow for efficient ASFV sequencing was proposed. This consisted of a target enrichment approach by ASFV hybridization with probes fixed on magnetic beads followed by the use of NGS technologies (either Illumina alone, or a combination of Illumina and Nanopore sequencing). This target enrichment approach allows better coverage compared to the no enrichment approach, and a smaller amount of data is generated during sequencing (Forth et al., 2019). However, this approach has only been tested on the reference genome used for the design of the probes and a strain close to this reference. Knowing that the core genome was constituted of 102 ortholog genes on the 301 ortholog genes described in ASFV pan-genome (Wang et al., 2019), the efficiency of enrichment by ASFV hybridization should be validated on other genotypes before being applied to unknown strains.

Our study is the first to investigate the efficiency and the utility of the separation of methylated and un-methylated DNA for the whole genome sequencing of the ASFV genome. Our results suggest a low enrichment of ASFV DNA using this approach. The cost of the enrichment process by separation of methylated and un-methylated DNA is around $35 per sample. An increase in the deepness of sequencing would most probably give similar results as EMD samples, with less extra work on the samples and may be cheaper than the enrichment process. In comparison, the target enrichment strategy, which costs approximately $225 per sample, is most effective, resulting in a significant reduction of data generated. However, the reduction of the sequencing costs does not compensate for the extra cost of the enrichment process; the real advantage of this strategy is the high quality of the sequence obtained.

In our hands, the enrichment of samples in un-methylated DNA for the sequencing of the ASFV genome represented extra work and cost without a significant improvement in the final results for very low ASFV load samples.

Author contributions

Rémi Pereira De Oliveira: Investigation, writing original draft, writing-review & editing. **Pierrick Lucas**: Investigation. **Amélie Chastagner**: Writing-review & editing. **Claire De Boissesson**: Investigation. **Laurence Vial**: Resources, writing-review & editing. **Marie-Frédérique Le Potier**: Writing-review & editing. **Yannick Blanchard**: Conceptualization, writing-review & editing, validation supervision.

Funding

This work was partially funded by the European Union Horizon 2020 program 503 COMPARE (Collaborative Management Platform for detection and Analyses of [Re-]emerging and foodborne outbreak in Europe; no. 643476).

Declaration of competing Interest

The authors declare that they have no competing interests.

Acknowledgements

The authors are grateful to Jean-Marie Guionnet and Angélique Moro for animal care and sampling at ANSES. Thanks to the Vectopole Sud for funding the insectary where the *Ornithodoros erraticus* were raised, the *Direction Générale de l'Alimentation* for the financial support provided CIRAD, and the European Union Horizon 2020 program 503 COMPARE (no. 643476) for partially financing this work. We are thankful to CIRAD and ANSES for funding the PhD grant of Rémi Pereira De Oliveira. Thanks to Grace Delobel for English editing.

References

286 Bird, A.P., Taggart, M.H., 1980. Variable patterns of total DNA and rDNA methylation in animals.
287 Nucleic Acids Res. 8, 1485–1497. https://doi.org/10.1093/nar/8.7.1485
288 Boinas, F.S., Wilson, A.J., Hutchings, G.H., Martins, C., Dixon, L.J., 2011. The Persistence of African
289 Swine Fever Virus in Field-Infected Ornithodoros erraticus during the ASF Endemic Period in
290 Portugal. PLoS ONE 6, e20383. https://doi.org/10.1371/journal.pone.0020383
291 Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence
292 data. Bioinforma. Oxf. Engl. 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170
293 Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O.J., Lipkin, W.I., 2015. Virome Capture
294 Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. mBio 6,
295 e01491-01415. https://doi.org/10.1128/mBio.01491-15
296 Chapman, D.A.G., Darby, A.C., Da Silva, M., Upton, C., Radford, A.D., Dixon, L.K., 2011. Genomic
297 analysis of highly virulent Georgia 2007/1 isolate of African swine fever virus. Emerg. Infect.
298 Dis. 17, 599–605. https://doi.org/10.3201/eid1704.101283
299 de Villiers, E.P., Gallardo, C., Arias, M., da Silva, M., Upton, C., Martin, R., Bishop, R.P., 2010.
300 Phylogenomic analysis of 11 complete African swine fever virus genome sequences. Virology
301 400, 128–136. https://doi.org/10.1016/j.virol.2010.01.019
302 Dixon, L.K., Chapman, D.A.G., Netherton, C.L., Upton, C., 2013. African swine fever virus replication
303 and genomics. Virus Res. 173, 3–14. https://doi.org/10.1016/j.virusres.2012.10.020
304 Dor, Y., Cedar, H., 2018. Principles of DNA methylation and their implications for biology and
305 medicine. Lancet Lond. Engl. 392, 777–786. https://doi.org/10.1016/S0140-6736(18)31268-6
306 Duron, O., Morel, O., Noël, V., Buysse, M., Binetruy, F., Lancelot, R., Loire, E., Ménard, C., Bouchez,
307 O., Vavre, F., Vial, L., 2018. Tick-Bacteria Mutualism Depends on B Vitamin Synthesis
308 Pathways. Curr. Biol. CB 28, 1896-1902.e5. https://doi.org/10.1016/j.cub.2018.04.038
309 Forth, J., Forth, L., King, J., Groza, O., Hübner, A., Olesen, A., Höper, D., Dixon, L., Netherton, C.,
310 Rasmussen, T., Blome, S., Pohlmann, A., Beer, M., 2019. A Deep-Sequencing Workflow for
311 the Fast and Efficient Generation of High-Quality African Swine Fever Virus Whole-Genome
312 Sequences. Viruses 11, 846. https://doi.org/10.3390/v11090846
313 Gaudin, M., Desnues, C., 2018. Hybrid Capture-Based Next Generation Sequencing and Its Application
314 to Human Infectious Diseases. Front. Microbiol. 9, 2924.
315 https://doi.org/10.3389/fmicb.2018.02924
316 Geraci, N.S., Spencer Johnston, J., Paul Robinson, J., Wikel, S.K., Hill, C.A., 2007. Variation in genome
317 size of argasid and ixodid ticks. Insect Biochem. Mol. Biol. 37, 399–408.
318 https://doi.org/10.1016/j.ibmb.2006.12.007
319 Greninger, A.L., Roychoudhury, P., Xie, H., Casto, A., Cent, A., Pepper, G., Koelle, D.M., Huang, M.-L.,
320 Wald, A., Johnston, C., Jerome, K.R., 2018. Ultrasensitive Capture of Human Herpes Simplex
321 Virus Genomes Directly from Clinical Samples Reveals Extraordinarily Limited Evolution in
322 Cell Culture. mSphere 3. https://doi.org/10.1128/mSphereDirect.00283-18
323 Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-
324 Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D.M., Kim, H., Frantz, L.A.F.,
325 Caccamo, M., Ahn, H., Aken, B.L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie,
326 C.W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S., Bujie,
327 Z., Bystrom, M., Capitanu, B., Carvalho-Silva, D., Chardon, P., Chen, C., Cheng, R., Choi, S.-H.,
328 Chow, W., Clark, R.C., Clee, C., Crooijmans, R.P.M.A., Dawson, H.D., Dehais, P., De Sapio, F.,
329 Dibbits, B., Drou, N., Du, Z.-Q., Eversole, K., Fadista, J., Fairley, S., Faraut, T., Faulkner, G.J.,
330 Fowler, K.E., Fredholm, M., Fritz, E., Gilbert, J.G.R., Giuffra, E., Gorodkin, J., Griffin, D.K.,
331 Harrow, J.L., Hayward, A., Howe, K., Hu, Z.-L., Humphray, S.J., Hunt, T., Hornshøj, H., Jeon, J.-
332 T., Jern, P., Jones, M., Jurka, J., Kanamori, H., Kapetanovic, R., Kim, J., Kim, J.-H., Kim, K.-W.,
333 Kim, T.-H., Larson, G., Lee, K., Lee, K.-T., Leggett, R., Lewin, H.A., Li, Y., Liu, W., Loveland, J.E.,
334 Lu, Y., Lunney, J.K., Ma, J., Madsen, O., Mann, K., Matthews, L., McLaren, S., Morozumi, T.,
335 Murtaugh, M.P., Narayan, J., Truong Nguyen, D., Ni, P., Oh, S.-J., Onteru, S., Panitz, F., Park,
336 E.-W., Park, H.-S., Pascal, G., Paudel, Y., Perez-Enciso, M., Ramirez-Gonzalez, R., Reecy, J.M.,
337 Rodriguez-Zas, S., Rohrer, G.A., Rund, L., Sang, Y., Schachtschneider, K., Schraiber, J.G.,

Schwartz, J., Scobie, L., Scott, C., Searle, S., Servin, B., Southey, B.R., Sperber, G., Stadler, P., Sweedler, J.V., Tafer, H., Thomsen, B., Wali, R., Wang, Jian, Wang, Jun, White, S., Xu, X., Yerle, M., Zhang, G., Zhang, Jianguo, Zhang, Jie, Zhao, S., Rogers, J., Churcher, C., Schook, L.B., 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491, 393–398. https://doi.org/10.1038/nature11622

Hall, R.J., Wang, J., Todd, A.K., Bissielo, A.B., Yen, S., Strydom, H., Moore, N.E., Ren, X., Huang, Q.S., Carter, P.E., Peacey, M., 2014. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. J. Virol. Methods 195, 194–204. https://doi.org/10.1016/j.jviromet.2013.08.035

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

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Masembe, C., Sreenu, V.B., Da Silva Filipe, A., Wilkie, G.S., Ogweng, P., Mayega, F.J., Muwanika, V.B., Biek, R., Palmarini, M., Davison, A.J., 2018. Genome Sequences of Five African Swine Fever Virus Genotype IX Isolates from Domestic Pigs in Uganda. Microbiol. Resour. Announc. 7. https://doi.org/10.1128/MRA.01018-18

Ndlovu, S., Williamson, A.-L., Malesa, R., van Heerden, J., Boshoff, C.I., Bastos, A.D.S., Heath, L., Carulei, O., 2020. Genome Sequences of Three African Swine Fever Viruses of Genotypes I, III, and XXII from South Africa and Zambia, Isolated from *Ornithodoros* Soft Ticks. Microbiol. Resour. Announc. 9, e01376-19, /mra/9/10/MRA.01376-19.atom. https://doi.org/10.1128/MRA.01376-19

Ondov, B.D., Bergman, N.H., Phillippy, A.M., 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12, 385. https://doi.org/10.1186/1471-2105-12-385

Olesen, A.S., Lohse, L., Dalgaard, M.D., Woźniakowski, G., Belsham, G.J., Bøtner, A., Rasmussen, T.B., 2018. Complete genome sequence of an African swine fever virus (ASFV POL/2015/Podlaskie) determined directly from pig erythrocyte-associated nucleic acid. J. Virol. Methods 261, 14–16. https://doi.org/10.1016/j.jviromet.2018.07.015

Plowright, W., Parker, J., Pierce, M.A., 1969. The epizootiology of African swine fever in Africa. Vet. Rec. 85, 668–674.

Tignon, M., Gallardo, C., Iscaro, C., Hutet, E., Van der Stede, Y., Kolbasov, D., De Mia, G.M., Le Potier, M.-F., Bishop, R.P., Arias, M., Koenen, F., 2011. Development and inter-laboratory validation study of an improved new real-time PCR assay with internal control for detection and laboratory diagnosis of African swine fever virus. J. Virol. Methods 178, 161–170. https://doi.org/10.1016/j.jviromet.2011.09.007

Tweedie, S., Charlton, J., Clark, V., Bird, A., 1997. Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol. Cell. Biol. 17, 1469–1475. https://doi.org/10.1128/MCB.17.3.1469

Wang, Z., Jia, L., Li, J., Liu, H., Liu, D., 2019. Pan-Genomic Analysis of African Swine Fever Virus. Virol. Sin. https://doi.org/10.1007/s12250-019-00173-6

Weber, S., Hakobyan, A., Zakaryan, H., Doerfler, W., 2018. Intracellular African swine fever virus DNA remains unmethylated in infected Vero cells. Epigenomics 10, 289–299. https://doi.org/10.2217/epi-2017-0131

Yáñez, R.J., Rodríguez, J.M., Nogal, M.L., Yuste, L., Enríquez, C., Rodriguez, J.F., Viñuela, E., 1995. Analysis of the complete nucleotide sequence of African swine fever virus. Virology 208, 249–278. https://doi.org/10.1006/viro.1995.1149

Zani, L., Forth, J.H., Forth, L., Nurmoja, I., Leidenberger, S., Henke, J., Carlson, J., Breidenstein, C., Viltrop, A., Höper, D., Sauter-Louis, C., Beer, M., Blome, S., 2018. Deletion at the 5'-end of Estonian ASFV strains associated with an attenuated phenotype. Sci. Rep. 8, 6510. https://doi.org/10.1038/s41598-018-24740-1

389 **Table 1: Relative quantification of ASFV genome, host beta-actin and total DNA in pig and tick**

390 **samples.**

Sample	Host	VP-72 ASFV PCR (Ct*)	Beta-actin PCR (Ct*)	DNA quantity (ng)	ASFV PCR/Beta-actin PCR ratio
889Geo	Pig	17.07	22.64	2610	0.74
936Geo		17.07	22.27	2180	0.77
941Geo		16.12	22.43	2130	0.72
6518Geo		15.88	22.11	2060	0.72
6547Geo		16.39	22.66	1540	0.72
E geo F1	Tick	27.2	27.1	554.4	0.99
E geo F2		33.95	29.33	392.4	1.16
E geo M1		27.13	28.35	594	0.96
E geo F3		35.51	28.5	374.2	1.24
E geo F4		38.77	28.85	357.5	1.34
E geo F5		33.79	28.39	330.8	1.19
E geo F6		31.29	28.26	410.8	1.10
E geo M2		31.11	28.66	407.3	1.08

391 *Ct = Cycle Threshold.

392 **Table 2: DNA quantity used for enrichment process and DNA quantity obtained after enrichment.**

Sample	Host	DNA for UD fraction (ng)	DNA for enrichment (ng)	DNA after enrichment for UMF fraction (ng)(% input)	DNA after enrichment for MF fraction (ng)(% input)	% DNA recovery (UMF+MF)
889Geo	Pig	99	550	178 (32)	282 (51)	84
936Geo		100	550	203 (37)	209 (38)	75
941Geo		98	550	162 (29)	275 (50)	79
6518Geo		99	550	205 (30)	290 (53)	90
6547Geo		98	550	158 (29)	326 (59)	88
E geo F1	Tick	100	454	368 (81)	24 (5)	86
E geo F2		100	292	210 (72)	9 (3)	75
E geo M1		100	494	347 (71)	23(4)	75
E geo F3		100	274	202 (74)	14(5)	79
E geo F4		100	257	172 (67)	9 (3)	70
E geo F5		100	230	153 (67)	9 (4)	70
E geo F6		100	310	392 (126)	13 (4)	131
E geo M2		100	307	246 (80)	10 (3)	83

393

Table 3: Quantity of total bases generated, quality score of bases, quantity of total reads and mean read length per sample. Quality of bases was analyzed by Phred Quality Scores ≥20 (%≥Q20) and expressed in percentage.

Fraction	Samples		Generated bases	%≥Q20	Quantity of total reads	Mean read length (bp)
UD	pigs	889Geo	229 542 433	80.79	1 682 871	136
		936Geo	385 414 722	80.86	2 749 366	140
		941Geo	410 776 191	76.63	3 124 738	131
		6518Geo	452 263 244	81.64	3 792 151	119
		6547Geo	2 935 297 256	78.98	21 827 756	134
		mean for pigs	**882 658 769**	**79.78**	**6 635 376**	**132**
	ticks	E geo F1	1 213 297 206	82.27	9 285 185	131
		E geo F2	1 013 734 715	84.43	8 726 201	116
		E geo M1	855 764 828	87.38	7 243 049	118
		E geo F3	404 865 960	82.73	3 132 828	129
		E geo F4	468 405 245	83.33	3 618 451	129
		E geo F5	319 338 365	84.07	2 578 862	124
		E geo F6	177 777 191	84.00	1 437 348	124
		E geo M2	305 292 664	83.29	2 542 952	120
		mean for ticks	**594 809 522**	**83.94**	**4 820 610**	**124**
		global mean	**705 520 771**	**82.34**	**5 518 597**	**127**
UMF	pigs	889Geo	285 321 178	80.59	2 246 476	127
		936Geo	222 820 790	80.98	1 751 451	127
		941Geo	17 393 456	82.28	150 956	115
		6518Geo	986 698 698	83.10	6 468 942	153
		6547Geo	248 717 004	80.28	1 934 593	129
		mean for pigs	**352 190 225**	**81.45**	**2 510 484**	**130**
	ticks	E geo F1	963 624 225	83.76	7 817 144	123
		E geo F2	940 491 921	83.99	8 008 483	117
		E geo M1	1 118 167 936	86.58	9 220 704	121
		E geo F3	152 407 597	82.92	1 203 993	127
		E geo F4	244 449 864	84.03	2 013 848	121
		E geo F5	324 824 629	84.41	2 728 169	119
		E geo F6	287 311 527	82.70	2 170 620	132
		E geo M2	148 976 803	84.68	1 380 652	108
		mean for ticks	**522 531 813**	**84.13**	**4 317 952**	**121**
		global mean	**457 015 818**	**83.10**	**3 622 772**	**125**
MF	pigs	889Geo	2 151 395	80.31	15 736	137
		936Geo	56 192 863	76.39	414 846	135
		941Geo	104 536 025	79.83	754 918	138
		6518Geo	1 139 572 253	82.56	7 757 735	147
		6547Geo	126 458 347	80.45	874 201	145
		mean for pigs	**285 782 177**	**79.91**	**1 963 487**	**140**
	ticks	E geo F1	885 423 136	85.64	8 165 392	108
		E geo F2	254 288 416	84.29	2 169 858	117
		E geo M1	109 382 507	84.16	977 897	112
		E geo F3	53 307 894	85.23	445 062	120
		E geo F4	16 326 955	83.85	149 147	109
		E geo F5	17 295 136	82.75	135 737	127
		E geo F6	127 530 447	84.14	1 048 408	122
		E geo M2	80 531 098	84.62	777 670	104
		mean for ticks	**193 010 699**	**84.34**	**1 733 646**	**115**
		global mean	**228 692 036**	**82.63**	**1 822 047**	**125**

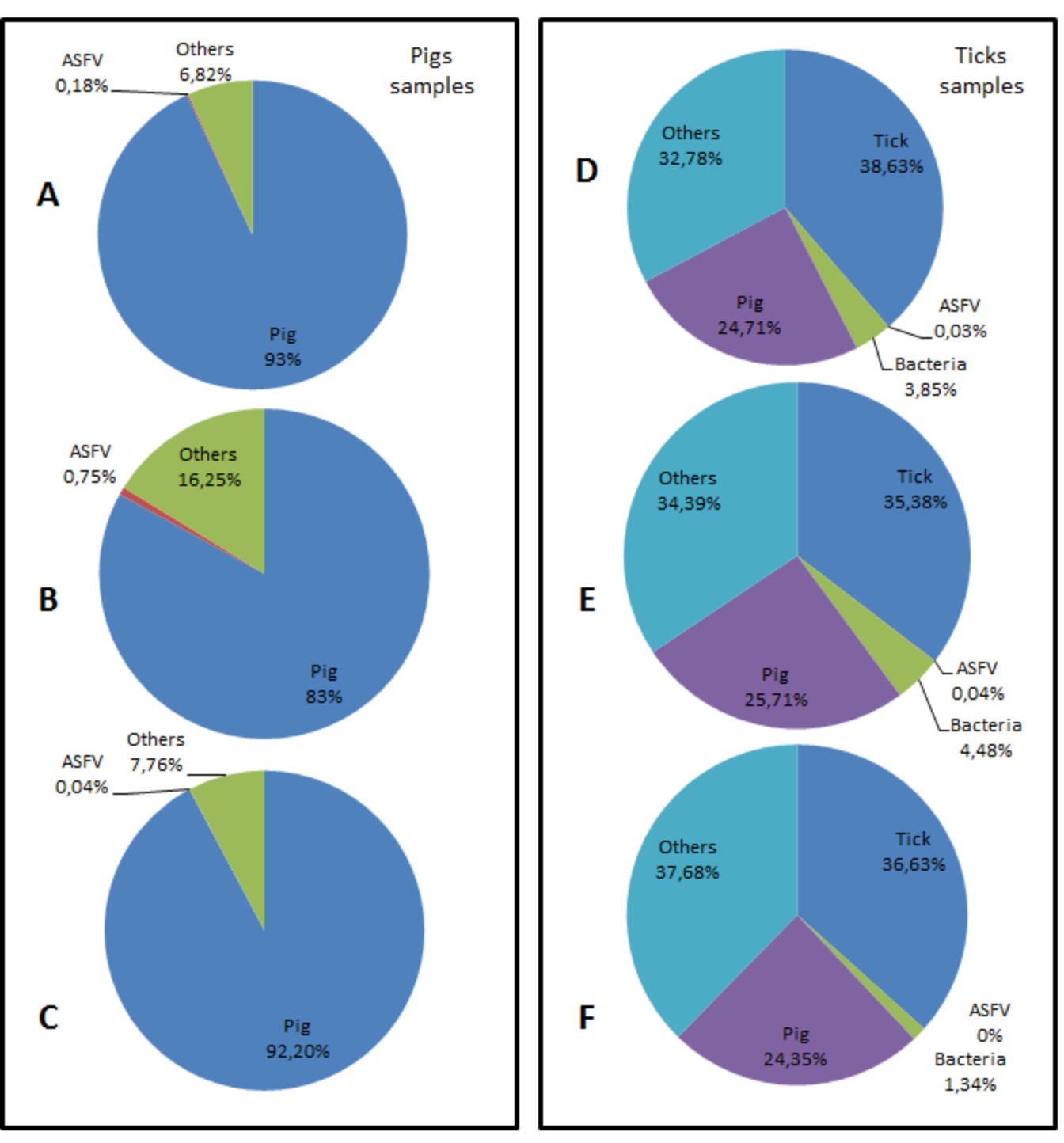
Table 4: Number and percentage of the total reads used for metagenomics analysis.

Fraction	Samples		Number of reads used for metagenomics analysis	Percentage of reads used for metagenomics analysis
UD	pigs	889Geo	504 861	30%
		936Geo	824 810	30%
		941Geo	937 421	30%
		6518Geo	1 137 645	30%
		6547Geo	1 091 388	5%
		mean for pigs	**899 225**	**25%**
	ticks	E geo F1	464 256	5%
		E geo F2	436 310	5%
		E geo M1	724 305	10%
		E geo F3	939 848	30%
		E geo F4	1 085 535	30%
		E geo F5	773 659	30%
		E geo F6	431 204	30%
		E geo M2	762 886	30%
		mean for ticks	**702 250**	**21.25%**
		global mean	**778 010**	**23.12%**
UMF	pigs	889Geo	673 943	30%
		936Geo	525 435	30%
		941Geo	150 956	100%
		6518Geo	646 894	10%
		6547Geo	580 378	30%
		mean for pigs	**515 521**	**40%**
	ticks	E geo F1	390 857	5%
		E geo F2	400 424	5%
		E geo M1	922 070	10%
		E geo F3	361 198	30%
		E geo F4	604 154	30%
		E geo F5	818 451	30%
		E geo F6	651 186	30%
		E geo M2	414 196	30%
		mean for ticks	**551 199**	**21.25%**
		global mean	**534 982**	**30.62%**
MF	pigs	889Geo	15 736	100%
		936Geo	414 846	100%
		941Geo	452 951	60%
		6518Geo	775 774	10%
		6547Geo	524 521	60%
		mean for pigs	**436 765**	**66%**
	ticks	E geo F1	408 270	5%
		E geo F2	216 986	10%
		E geo M1	391 159	40%
		E geo F3	356 050	80%
		E geo F4	149 147	100%
		E geo F5	135 737	100%
		E geo F6	314 522	30%
		E geo M2	466 602	60%
		mean for ticks	**304 809**	**53.12%**
		global mean	**355 561**	**59.56%**

Table 5: Number and percentage of ASFV reads per sample and the sequencing coverage of the reference strain Georgia2007/1.

Samples	UD			UMF			MF		
	ASFV reads	%	Cov.	ASFV reads	%	Cov.	ASFV reads	%	Cov.
889Geo	3000	0.18	2.20	5866	0.26	3.42	67	0.43	0.05
936Geo	5544	0.2	4.25	4594	0.26	3.25	45	0.01	0.03
941Geo	10760	0.34	7.68	95	0.06	0.06	72	0.01	0.06
6518Geo	14222	0.38	9.19	40620	0.63	33.96	1220	0.02	0.99
6547Geo	71286	0.33	52.21	10491	0.54	7.46	132	0.02	0.1
Mean	**20962**	**0.29**	**15.11**	**12333**	**0.35**	**9.63**	**307**	**0.098**	**0.25**
E geo F1	604	0.007	0.47	391	0.005	0.29	57	0.001	0.04
E geo F2	12	0	0.01	6	0	0	0	0	0
E geo M1	546	0.008	0.39	813	0.009	0.61	6	0.001	0
E geo F3	1	0	0	0	0	0	0	0	0
E geo F4	0	0	0	0	0	0	0	0	0
E geo F5	2	0	0	0	0	0	0	0	0
E geo F6	5	0	0	26	0.001	0.02	1	0	0
E geo M2	28	0.001	0.02	10	0.001	0.01	3	0	0
Mean	**150**	**0.002**	**0.11**	**156**	**0.002**	**0.12**	**8**	**0.00025**	**0.005**

Cov. = sequencing coverage in average number of reads per nucleotide position.


Pigs samples
A
ASFV 0,18%
Others 6,82%
Pig 93%
B
ASFV 0,75%
Others 16,25%
Pig 83%
C
ASFV 0,04%
Others 7,76%
Pig 92,20%
Ticks samples
D
Others 32,78%
Tick 38,63%
Pig 24,71%
ASFV 0,03%
Bacteria 3,85%
E
Others 34,39%
Tick 35,38%
Pig 25,71%
ASFV 0,04%
Bacteria 4,48%
F
Others 37,68%
Tick 36,63%
Pig 24,35%
ASFV 0%
Bacteria 1,34%