From the Department for Pathobiology

at the University for Veterinary Medicine Vienna

Institute for Virology

(Institute Leader: Univ.Prof. Dr.med.vet. Till Rümenapf)

Development of an improved method of fast sequence determination of PRRSV field isolates

Bachelor thesis

University of Veterinary Medicine Vienna

Submitted by

Gloria Scheikl

Vienna, June 2023

Supervisor: Univ.Prof. Dr.med.vet. Till Rümenapf

Institute for Virology

Department for Pathobiology

University for Veterinary Medicine Vienna

Reviewer: Priv.-Doz. Dr.med.vet. Dipl.ECVM Joachim Spergser

Declaration of independence:

I hereby declare that I have written the submitted work independently and have not used any sources or aids other than those specified. All text passages taken from external sources have been marked.

I did the crucial work by myself and listed all the contributors with their contribution to the work.

This work has not been submitted or published elsewhere.

Vienna, at 02.06.23 Gloria Scheikl

ZUSAMMENFASSUNG

Die Themenstellung des durchgeführten Projektes befasst sich mit dem Porcinen reproduktiven und respiratorischen Syndrom Virus (PRRSV) und dessen Diagnostik. Dieses stellt nicht nur gesundheitlich, sondern auch wirtschaftlich eine große Herausforderung für die Schweine-Industrie dar. Zusätzlich ist auch der Nachweis des Virus erschwert, was vor allem an der hohen genetischen Variabilität zwischen verschiedenen Stämmen liegt und die Detektion mittels PCR verkompliziert.

Der Fokus des Projekts liegt daher auf der Etablierung eines optimierten Protokolls, zur Erzeugung von sogenannten Amplikons, die das gesamte Genom möglichst vieler PRRSV-1 Stämme abdecken und die anschließend mittels Next Generation Sequencing (NGS) analysiert werden. Ziel ist es, ein universelles Protokoll zu entwickeln, mit dem das vollständige Genom von PRRSV-1 Stämmen, mittels NGS ermittelt werden kann.

Eine Schlüsselstellung hat dabei das Design der verwendeten Primerpaare, die möglichst viele der heterogenen PRRSV-1 Stämme amplifizieren können. Die Amplikons wurden mit Längen von etwa 1,5 kb konzipiert, so dass das Genom mit 11 jeweils um ca. 100 Nukleotide überlappenden Amplikons abgedeckt wird. Eine wichtige Voraussetzung für das Primerdesign ist die zunehmende Zahl an verfügbaren Gesamtsequenzen in der Genbank.

.

Im Zuge der Projektarbeit wurden die Überstände im Labor kultivierter Viren zur RNA-Extraktion und weiteren Bearbeitung verwendet. Eine ebenfalls mitgeführte cDNA aus Gewebematerial einer diagnostischen Probe erwies sich als problematisch. Hierfür müsste die Methode weiterentwickelt werden.

ABSTRACT

The topic of the implemented project deals with the porcine reproductive and respiratory syndrome virus (PRRSV) and its diagnostics. This virus poses a major challenge for the pig industry, not only in terms of health, but also economically. Another difficulty is the detection of the virus. This is mainly due to the high genetic variability between different strains, which complicates the detection by PCR.

The focus of the project is therefore to establish an optimized protocol for generating so-called amplicons that cover the entire genome of as many PRRSV-1 strains as possible and can then be analysed using next generation sequencing (NGS). The aim is to develop a universal protocol that can be used to determine the complete genome sequence of PRRSV-1 strains using NGS.

The design of the primer pairs, which can amplify as many of the heterogenous PRRSV-1 strains as possible, has a key position here. The amplicons were designed to have a length of about 1,5 kb, so that the genome is covered with 11 amplicons, each overlapping by about 100 nucleotides. An important requirement for the primer design is the increasing number of total sequences available in the gene library.

In the course of the project, supernatants of cultivated virus were used for RNA extraction and further processing. cDNA from tissue material, form a diagnostic sample that was also carried along proved to be problematic. For this, the metho would have to be further developed.

Table of Contents

1. Introduction

1.1. PRRSV

Since being discovered, the porcine reproductive and respiratory syndrome virus constitutes a high cost to the swine industry. Attempts to regulate the virus are complicated, because of the persistent infections, as well as the high heterogeneity of the virus (1). The induced disease is characterized by respiratory problems, as well as weight loss, poor growth performance and reproductive failure in pregnant individuals (2).

The porcine reproductive and respiratory syndrome virus is an enveloped, positive-sense single-stranded RNA virus, with a genome length of approximately 15 kilobases, containing at least ten ORFs. ORF 1a and 1b near the 5' end of the genome encode for two polyproteins, which result in 14 non-structural proteins. ORF 2-7 near the 3' end, encode for eight known structural proteins (3). The virus belongs to the Arteriviridae family, which belongs to the order of Nidovirales, which additionally includes the Coronaviridae and Roniviridae families. Within the family of Arteriviridae, PRRSV belongs to the subfamily of Variarterividae and to the genus of Betaarterivirus. The order of Nidovirales share replication and transcription strategies, as well as similarities in genomic organisation, but differ in host species and range, as well as disease phenotype. The Arteriviridae family contains 5 viruses, which have similar genomic organization and content and similar cellular tropism for the macrophage lineage. The viruses belonging to this family are PRRSV, simian haemorrhagic fever virus, lactate-dehydrogenase elevating virus, newly recognized wobbly possum disease virus and equine arteritis virus. (4)

Originally, PRRSV was divided into two genotypes, the European virus (type 1), which is also known as Lelystad virus and the American genotype (type 2). Although the disease phenotype, the clinical symptoms, genomic organization and temporal emergence are mostly similar, the two types share only about 60 % homology on nucleotide level. (4) Therefore, the two types of PRRSV both received virus status, with PRRSV-1 being the European virus and PRRSV-2 being the American virus.(5)

PRRSV is characterized by a high genetic variability. As typical for most plus stranded RNA viruses PRRSV tends to mutate at a very high rate, due to an absent -proofreading activity of the RNA dependent RNA polymerase, which allows a constant generation of new mutations. (3). In addition, the discontinuous transcription strategy employs RNA recombination that upon coinfection can result in the generation of chimeric viruses.

1.2. Current diagnostic method

The mentioned high diversity of PRRSV impacts, not only virulence and prevention through vaccines, but also the diagnosis of the virus (3).

Currently reverse transcription (RT) PCR is the most common technique for detecting templates from RNA viruses. After the RNA has been reverse transcribed into cDNA the polymerase chain reaction can be applied. In essence two primers hybridise to the template and the region between the primers will be amplified exponentially.(6)

The current method to detect PRRSV in samples, at the institute of virology at the university of veterinary medicine in Vienna is RT-PCR and RT-qPCR, which are conducted at ORF1, ORF5 and ORF7.

RT-PCR is a technique, which can increase the amount of RNA in samples. Another advantage is that, with RT-PCR it is possible to diagnose pathogens in secretions, tissue samples, blood or cell cultures, as well as a very quick diagnostic time. This kind of PCR includes a reverse transcription, which generates complementary DNA transcripts. All these steps, from RNA, to cDNA synthesis, to PCR amplification can be done in a single tube, or as a two-step process in two separate tubes. qPCR is a technique for detecting and quantifying gene expression in real time. It continuously measures the accumulation or reduction of fluorescent signal while the sample is amplified. One advantage of qPCR is a very high sensitivity. Additionally, the two mentioned methods can be combined, in order to combine reverse transcription with a continuous quantification of the sample. (7)

The heterogeneity between different PRRSV strains complicates the design of primers, as divergence in the genomic sequence impairs the binding of the primer and hence no amplicon is generated.

1.3. Universal primer pairs for amplicons

The PRRSV genome is, as already mentioned, prone to mutations. Therefore, even when looking at constant regions in the sequence, there are quite a lot of single nucleotide polymorphisms. Because of these differences, universal, degenerated primers were designed for the experiment. Degenerated primers are like normal primers, short single stranded oligonucleotides used in nature, as a starting point for replicating enzymes and in PCR to amplify input DNA sequences. However, unlike standard primers, degenerated primers have

positions where several bases are possible (8). These primers are defined by a so-called ambiguity code, for alternative nucleotides. This allows the amplification of a heterogenous group, as the degenerate primers offers a combination of slightly different primers, for each alternative input sequence. However, there are also some difficulties, when designing degenerated primers. One would be the number of degenerated bases possible in a primer sequence, as the likelihood of the primer binding to unwanted regions increases with the number of degenerated nucleotides. (9) For, designing primers, it is also important to ensure, that the primer pairs have annealing temperatures (T_A), which fit together, for the PCR reaction to work.

1.4. Illumina sequencing

When introduced, next generation sequencing resulted in a great increase of data output from sequencing. This increase is a result of a much higher throughput, achieved by parallel sequencing of single DNA molecules.

Illumina sequencing belongs to the category of second-generation sequencing, also called short-read sequencing. These technologies generally start by fragmenting the input DNA and then sequencing short, amplified DNA molecules, which are sequenced in parallel. With these methods millions of sequencing reactions occur at the same time, producing DNA fragments, which than must be reassembled. For sequencing Illumina uses the “sequencing by synthesis” method with a fluorescent-labelled reversible terminator technology, which is based on the optical signal of incorporating fluorescent nucleotides coupled to a reversible terminator by a DNA polymerase. When a fluorescently labelled nucleotide is incorporated into the new forming nucleic acid chain a fluorescent signal occurs, which is then detected by the sequencing instrument. The terminator and the fluorescent label are then cleaved from the incorporated nucleotide, allowing the next fluorescently labelled nucleotide to be incorporated. Additionally, to single-end sequencing, Illumina also allows paired-end sequencing, where the sequencing reaction occurs on both ends of the input DNA fragment. The workflow for Illumina sequencing usually contains the preparation of the library, which involves fragmentation, end-repair and adapter ligation, the sequencing reaction and data analysis. The analysis of the resulting data involves different steps. At first the base-calling for each amplified DNA fragment, which is performed by the instrument software. After that quality control procedures, such as read filtering and trimming are done. This is followed by aligning the reads. The reads

can either be aligned using a reference genome, or de-novo. This results in contigs, with different lengths, which ultimately can be assembled by mapping them to a reference genome. (10)

2. Material and Methods

2.1. Primer-Design

For establishing the new universal primers, the complete genome sequence from the PRRSV strain Aut15_33 (MT000052.1) was used. The sequence was aligned with other similar sequences using the NCBI Blast tool (11), to find conserved regions, with no mutations between the compared sequences. These regions were used to design primer pairs with melting temperatures between 59 °C and 61 °C. The primers were designed to produce amplicons with a size around 1000 – 1500 bp. Therefore, eleven primer-pairs covering most of the genome sequence of PRRSV were designed.

2.2. RNA isolation of cultured viral strains

The cultivated viral strains used for this experiment were grown in immortalized porcine immune cells. For isolating the viral RNA, the supernatant of the cells, which showed a positive cytopathic effect were centrifuged at 35000 x g for two hours. To extract the viral RNA from the pelleted cells the Monarch Total RNA Miniprep Kit (New England BioLabs GmbH, Ipswich, Massachusetts) was used. For this the supernatant was discarded and the pellet was resuspended in 400 µl RNA lysis buffer. Then 400 µl ethanol (95 %) were added to the lysate. The mixture was transferred to the RNA purification column and spined at 16 000 x g for 30 s. After that 500 µl RNA priming buffer were added to the column and the column was spined for 30 s. After discarding the flow-through, 500 µl RNA wash buffer was added, and the column was centrifuged for 30 s at 16000 x g. 500 µl RNA wash buffer was added a second time. Then the column was spined at 16000 x g for 2 min. To elute the RNA from the column, 50 µl nuclease-free water were added directly to the column matrix, and the column was centrifuged at 16000 x g for 30 s.

Cultivated strains used or RNA extraction	
1522_22	1099_22
1510_22	Eber 2
Aut21 2201	77_22
1445_22	Best_22
1245_22	

Table 1: Cultivated strains used for RNA extraction

2.3. Reverse transcription

In addition to the strains where RNA extraction was performed, an additional Strain (2857_22) was used to carry out reverse transcription. The isolated RNA from 1245_22 however, was not used to generate cDNA, the reason being a weak cytopathic effect and a low viral concentration.

For reverse transcription five of the established universal reverse primers, covering the whole sequence were used. For the reaction 8 µl Luna Script Reverse Transcriptase 5x Master Mix (New England BioLabs GmbH, Ipswich, Massachusetts), 5 µl of the isolated RNA, 17 µl water and 2 µl of each Primer were combined in an PCR Tube, vortexed and spined. The tubes were then put in a Thermo cycler and incubated at 55 °C for 12 min, before cooling it down to 4 °C. The generated cDNA was then cleaded up by adding 10 µl water, transferring the mixture on a Quantum Prep PCR Kleen Spin Column (Bio-Rad Laboratories Inc., Hercules, California) and then spin at 735 x g.

Reverse Primer
PRS 734
PRS 738
PRS 742
PRS 746
PRS 752

Table 2: Reverse Primer used for reverse transcription

Strains used for reverse transcription	
2857_22	1445_22
1522_22	Eber 2
1510_22	77_22
Aut21 2201	Best_22
1099_22	

Table 3: Strains used for reverse transcription

2.4. PCR and gel-electrophoresis

For every generated cDNA eleven PCRs were performed, one reaction per primer-pair. For each reaction 10 µl 2x Phanta Max Master Mix (Vazyme Biotech, Nanjing, China), 2 µl cDNA, 2 µl forward reverse primer of the primer-pair and 4 µl water are combined, resulting in a reaction volume of 20 µl.

PCR programme		
95 °C	2 min	1 cycle
95 °C	11 s	40 cycles
T_A amplicon	20 s	
72 °C	45 s	
72 °C	5 min	1 cycle
6 °C	∞	

Table 4: PCR programme

The annealing temperature depends on the primer-pair used in the PCR reaction. The PCR tubes were placed in the thermocycler (Veriti Thermal Cycler, Applied Biosystems, Waltham, USA) accordingly.

Primer Pair	Primer	Annealing temperature
1	PRS731 PRS732	59 °C
2	PRS733 PRS759	55 °C
3	PRS760 PRS761	55 °C
4	PRS737 PRS738	59 °C
5	PRS757 PRS758	61 °C
6	PRS741 PRS742	55 °C
7	PRS743 PRS744	55 °C
8	PRS745 PRS762	59 °C

9	PRS747 PRS748	55 °C
10	PRS749 PRS750	58 °C
11	PRS751 PRS752	58 °C

Table 5: Annealing temperatures for each primer-pair

For each PCR reaction gel electrophoresis was performed, to check if every amplicon has the expected length. For this a 0,8 % agarose gel was used. A mixture of 2 µl of the respective PCR reaction and 1 µl of Loading dye (New England Biolabs, Massachusetts, USA) was applied to the gel and the electrophoresis was carried out, at 100 V, for 30 min. For Aut21 2201 not all primers were used, as it was seen, that the fourth primer-pair did not work for that strain. Therefore, instead of using primer-pair four, the forward primer of pair three (PRS737) and the reverse primer of primer-pair five (PRS758) were used to produce a bigger amplicon.

Lastly the amplicons of each viral strain were pooled, and the concentration of each pool was determined.

2.5. Library Preparation

For preparing the different Libraries, the Illumina DNA Prep Kit (Illumina Inc., San Diego, California), containing the Tagmentation Kit, the Index Set and the Purification Beads was used.

The library preparation was conducted separately for each pool. The total Input of DNA of each Pool was 500 ng. Nuclease-free water was added, to bring the sample-volume to 30 µl. After vortexing the Bead-linked transposomes a mixture of 11 µl Bead-linked transposomes and 11 µl Tagmentation Buffer was prepared. 20 µl of the mixture were transferred to the sample and resuspended. The Tubes were placed in the thermocycler (Veriti Thermal Cycler, Applied Biosystems, Waltham, USA) and the TAG Program was started.

TAG programme	
55 °C	5 min
10 °C	∞

Table 6: TAG programme for library prep

After running the programme 10 µl Tagment Stop Buffer were added to the reaction and slowly resuspended. Following this the PTC programme was started.

PTC programme	
37 °C	15 min
10 °C	∞

Table 7: PTC programme for library prep

After the programme finished, the tubes were placed on a magnetic stand until the liquid cleared up and the supernatant was discarded. For washing the tubes were removed from the magnetic stand and 100 µl Tagment Wash Buffer were added onto the beads. The mixture was resuspended and again placed on the magnetic stand to clear the liquid. The supernatant was discarded, and the tubes again removed from the magnet. Tagment Wash Buffer was added a second time and the tubes were placed on the magnetic rack again.

A mixture of 22 µl Enhanced Primer Mix and 22 µl Nuclease-free water was prepared. The supernatant was removed from the tubes, on the magnetic stand and 40 µl of the prepared mixture were added to the sample. After that 10 µl of the appropriate index were added to the sample and the BLT PCR programme was started.

BLT PCR programme		
68 °C	3 min	1 cycle
98 °C	3 min	1 cycle
98 °C	45 s	5 cycles
62 °C	30 s	
68 °C	2 min	
68 °C	1 min	1 cycle
10 °C	∞	

Table 8: BLT PCR programme for library prep

45 µl supernatant of each sample were transferred to a new tube and 81 µl Illumina Purification Beads were added. The mixture was resuspended and incubated at room temperature for 5 min. To clear the liquid the tubes were placed on the magnetic stand. After that, the supernatant was discarded. To wash the beads, 200 µl 80 % Ethanol was added and the mixture was incubated at room temperature for 30 s. The Ethanol was removed from the beads and the wash-step was repeated. After air drying the beads, the tubes were removed from the magnet and 32 µl Resuspension Buffer were added onto the beads. The mixture was again incubated at room temperature for 2 min and after that 30 µl supernatant were transferred into a new tube.

2.6. Illumina Sequencing

For Sequencing the prepared libraries, the Illumina MiniSeq system (Illumina Inc., San Diego, California) was used. To generate one library-pool, 5 µl of each prepared library were transferred in a new tube and vortexed. The pool was than quantified using Qubit (Thermo Fisher Scientific Inc., Waltham, Massachusetts). For dilution, 2,8 µl Library pool were combined with 47,2 µl Resuspension Buffer and mixed. For further diluting 50 µl Resuspension buffer were added. The library-pool then had a concentration of 1 nM. Following that dilution, the library-pool was denatured by adding 5 µl 0,1 N NaOH to 5 µl of the 1 nM library-pool, incubating the mixture for 5 min and then adding 5 µl of Resuspension Buffer. To further dilute the library-pool to the final concentration of 1,3 pM, 985 µl Hybridization Buffer were added to the denatured library-pool. Then 370 µl Hybridization Buffer were added to 130 µl of the library-pool. The diluted and denatured library-pool was then loaded onto the Cartridge and the paired-end sequencing was performed.

2.7. Sequence Analysis

Before analysing the Sequences regarding their similarity with each other or possible recombination, the Illumina generated data has to be combined to one single sequence per strain. For this the Galaxy-Tool Shovill was used (12). This tool performs a de-novo assembly, which results in different contigs, of various lengths. These contigs, were then mapped and subsequently combined to longer sequences, by aligning them with the reference genome Aut15_33 (GenBank: MT000052.1) with the NCBI tool BLAST. In cases where the de-novo assembly produced more contigs, which aligned at the same position, the contig with the higher

coverage was used to generate the sequence. For the combination of the contigs, to longer sequences, the software Serial Cloner (13).

For further analysis of the different sequences, again NCBI BLAST was used. With that, the obtained sequences were aligned with the reference genome Aut15_33, to determine the identity of these alignments. Additionally, similar sequences to the reference genome were collected, with the BLAST search. Together with these sequences and the 9 obtained Illumina sequences a recombination analysis was done. For this a proportion test was conducted, using SimPlot++ (14). Furthermore, a phylogenetic tree of the obtained sequences was created using the CLC Sequence Viewer 8.0 (15).

3. Results

3.1. Universal primer-pairs

The Blast of the reference sequence Aut15_33 (GenBank: MT000052.1) resulted in eleven primer pairs, which all together cover a section from approximate 80 – 15100 bp, as seen in the table below. The PCR-amplicons resulting from these primer pairs each overlap with the following fragment with about 100 - 200 bp. Which not only allows Illumina sequencing of the whole genome of the respective strain, but also sanger sequencing, without leaving any gaps.

Fragment	Primer		Position according to Aut15_33
1	TTGGCYCCTGTTCTAGCC	forward	86-110
	TTCAACATCCTCTCAACAGCAT	reverse	2053-2074
2	ACAAAATGYTGACCTYYCCG	forward	1942-1961
	CCAGARRTAACGTGACC	reverse	3958-3974
3	GTTCTTCTTGTGACCACG	forward	3817-3834
	GTACTGRTGRATRTCATAAG	reverse	5146-5165
4	GGCYCTYARRGCRGAYATGTT	forward	4901-4921
	AGRRCYCTRGGYARGAAGM	reverse	6247-6265
5	CTYTGGCGYATGATGGGC	forward	5959-5976
	CCYAARGTRAAAGTTCTGCTGTG	reverse	7484-7506
6	ARGCTTTAAACTGYTAGCCGC	forward	7393-7413
	AGRTCYTCRAAYTTRAGCTG	reverse	8870-8889
7	GYCATGAYCTYGTGGC	forward	8694-8709
	YCTYGCCCGRGTRATRGC	reverse	10403-10420
8	AYCCARCCTYGTYACAGRG	forward	10184-10203
	TGGGARACYCGCATRTGC	reverse	12079-12097
9	GGTTTGCTCCGCGMTTCT	forward	11944-11961
	CMAKCGRCAAAGGACCA	reverse	13575-13591
10	CAACCATYGCTWGYTTGTTSGC	forward	13454-13475
	GCCAGAATGTATYGCCGG	reverse	14395-14412
11	CTTCCAGATGCAGATTGTGTTG	forward	14368-14388
	TAATTACGGTCACATGGTTY	reverse	15075-15094

Table 9: established universal primer pairs

3.2. Gel electrophoresis

3.2.1. Strain 2857_22

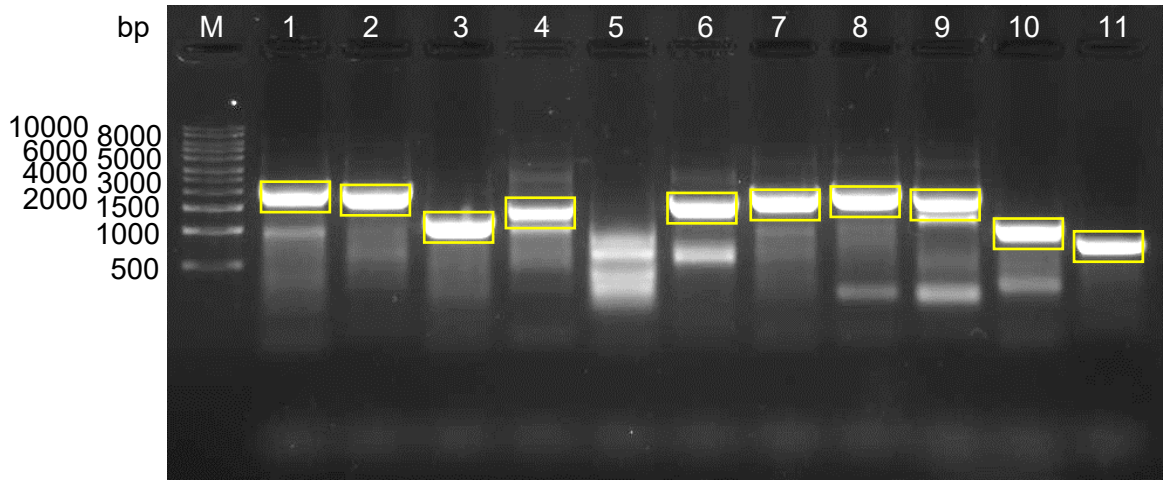

Figure 1: Gel electrophoresis 2857_22

The figure above shows the gel electrophoresis, of strain 2857_22. This was the first PCR, done while still working on the universal primer set. Therefore, some of the primers used in this PCR were afterwards optimized, with regard to overlapping with the previous or the following amplicon.

On the left side of the gel is the 1 kb marker, labelled with “M”. The bands of the marker are labelled with the according length in bp. As seen in the picture, besides fragment 5, all the fragments show a band with the expected length of the amplicon. The expected bands of each column are marked in yellow. As the picture clearly shows, lane five has no band in the correct size. However, also after repeating the PCR for fragment 5, there was no amplicon in the correct size. Because a second PCR of fragment 5 also showed no fragment, it was assumed that the primer pair is not working and a new primer pair in slightly different parts of the reference sequence was designed, PRS757 and PRS758.

3.2.2. Strain 77_22

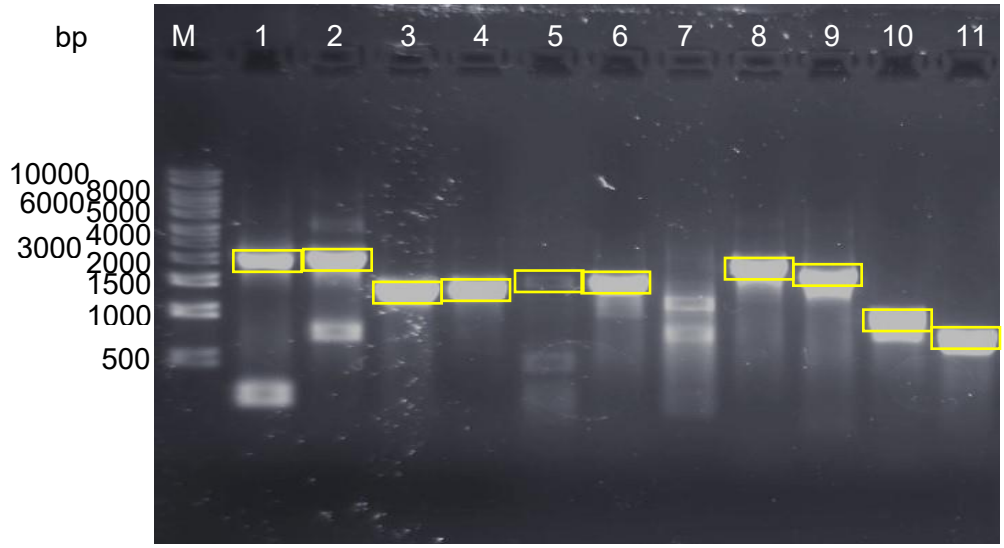

Figure 2: Gel electrophoresis of 77_22

Shown in figure number 2 is the result of the gel electrophoresis conducted after the PCR. From left to right the picture shows first the column containing the reference ladder (M) and after that the obtained PCR fragments from 1 – 11. On the left side of the picture the bands from the reference ladder are labelled with their lengths in bp. The right band from each column was marked in yellow. As seen in the picture there is additional bands in some of the lanes. In these lanes the right fragment was isolated by excising out the expected fragment and purifying it with the Monarch DNA Gel Extraction Kit (New England BioLabs Inc., Massachusetts). Because the correct band from fragment number 7 was missing and the right fragment from column 5 was not clearly visible, the PCR of these two fragments was repeated. This produced the correct fragments. Therefore, all eleven fragments were ready for library preparation.

3.2.3. Strain 136_22

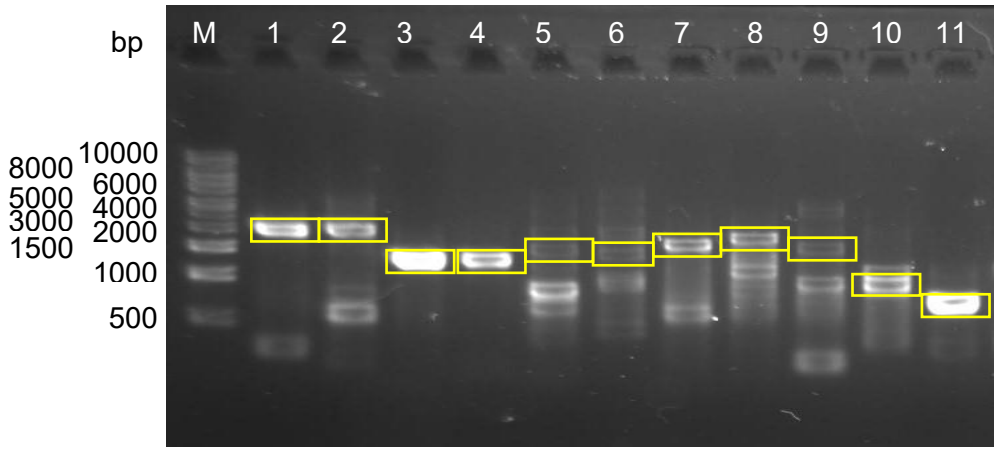

Figure 3: Gel electrophoresis 136_22

Figure number 3 shows the gel electrophoresis of strain 136_22 after the conducted PCR. As shown in the picture every fragment produced a band with the expected length. On the left side of the picture are the lengths corresponding to the bands of the 1 kb ladder. For fragment 2, 5, 6, 8 and 9 gel electrophoresis was repeated, as there were additional fragments in the columns. For the repetition electrophoresis the remaining PCR products were applied to the gel and the correct bands were excised and purified, before further steps of the protocol were taken.

3.2.4. Strain 1099_22

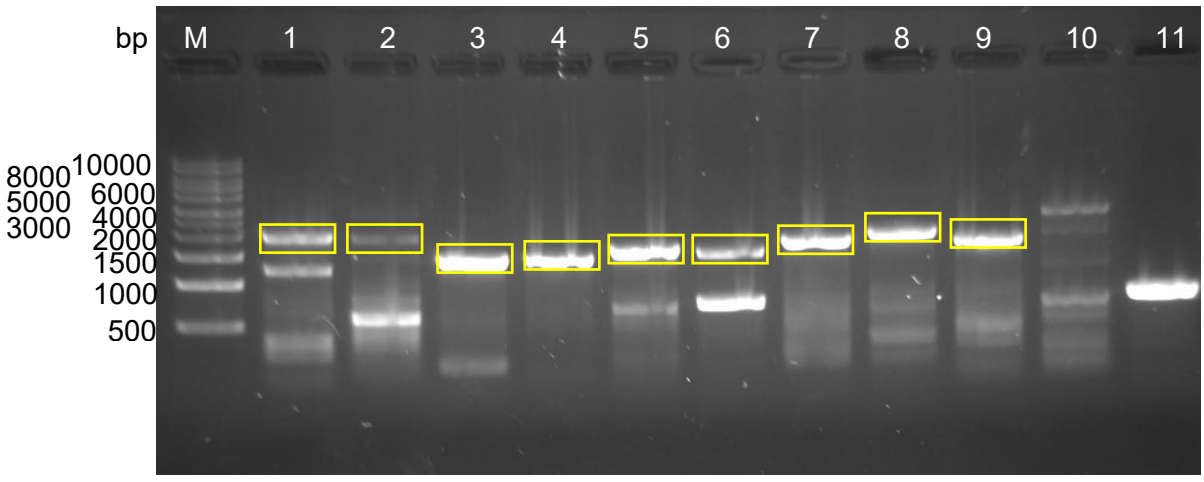

Figure 4: Gel electrophoresis 1099_22

Again, on the left side of the picture the bands of the 1 kb ladder are labelled with the corresponding lengths in bp. The lane labelled “M” contains the 1 kb ladder. From that lane on, from left to right are the fragments from 1 to 11. As seen in the picture, marked in yellow, fragment 1 to 9 contain bands with the expected length. Lane 10 and 11 on the other hand had to be repeated, as they right band was missing. Additionally, the PCR of fragment 2 was repeated, as the amount of the fragment seemed to be quite low, to ensure enough input material for Illumina sequencing.

3.2.5. Strain 1445_22

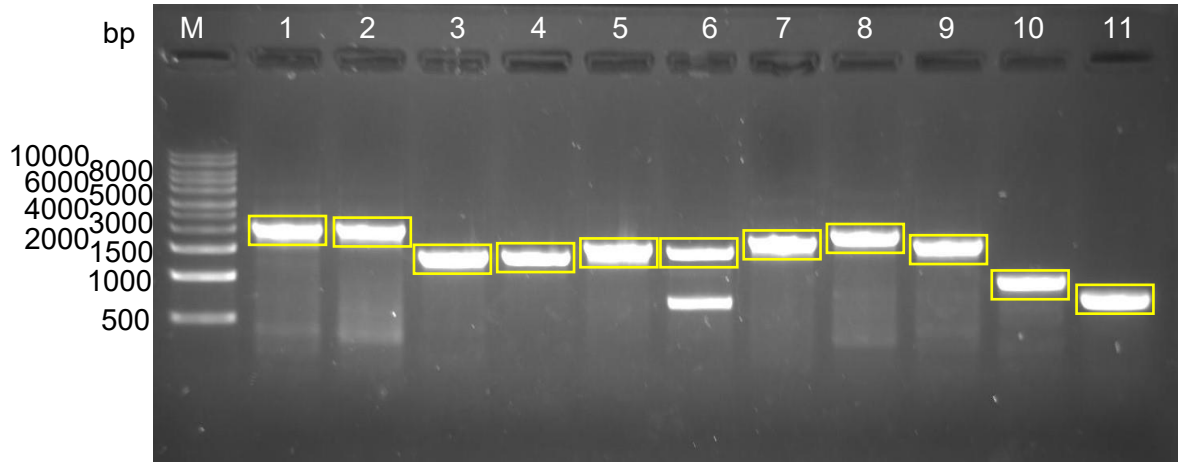

Figure 5: Gel electrophoresis 1445_22

Figure 5 shows the gel electrophoresis of strain 1445_22. As seen in the picture all the lanes containing PCR product (1 to 11) contain a fragment with the correct length, when compared to the 1 kb ladder, which is seen in the first lane, on the left side of the figure. The right fragments in each lane are circled in yellow. The ladder is labelled with the right lengths in bp on the left side of the picture. Additionally, the figure shows almost no other bands than the expected fragments, so no repetition of the PCR was necessary, and all Fragments could be used directly for library preparation.

3.2.6. Strain 1510_22

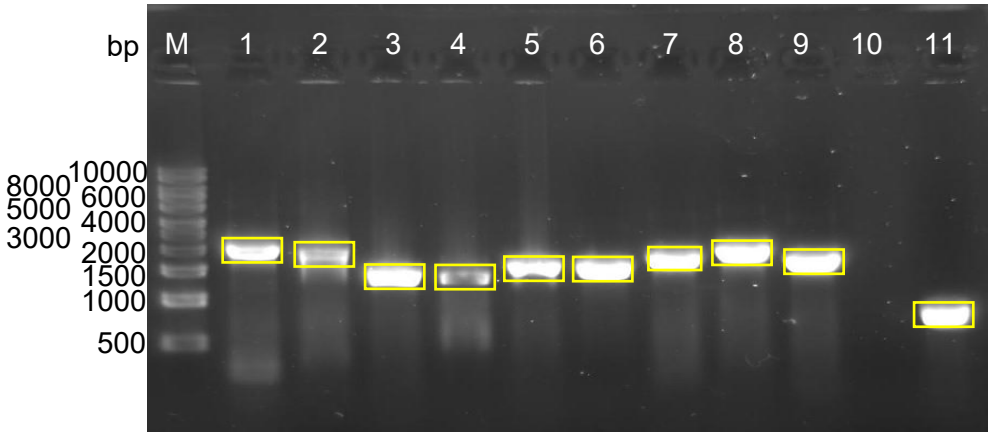

Figure 6: Gel electrophoresis 1510_22

In figure 6, shown from left to right are first the 1 kb ladder, which is labelled with the right lengths on the left side of the picture and after that from 1 to 11 the different fragments, which resulted from the PCR. The picture shows, that all the lanes except lane 10 contain a band with the expected length of the fragment. The correct bands are marked in yellow. The PCR on fragment 10 was repeated, as the first gel electrophoresis showed no band, of any length. After the repeated PCR, the finished PCR product was again analysed by gel electrophoresis, which showed that the repeated PCR produced the fragment with the expected length.

3.2.7. Strain 1522_22

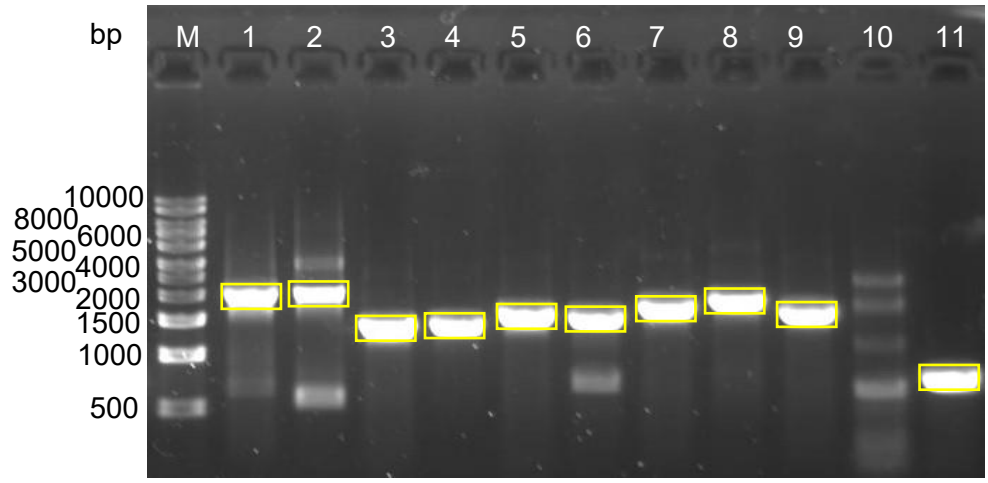

Figure 7: Gel electrophoresis 1522_22

Figure 7 shows, on the left side of the picture, the 1 kb ladder, labelled with the according lengths, which was used to compare the results of the PCR. As seen in the picture all lanes except for lane 11 each show a band with the expected length. The PCR for fragment 10 was repeated, as there were a lot of different bands in the column, but none with the correct length. This repeated PCR resulted in a fragment, with the right length, which was then used for further work, together with the other conducted fragments.

3.2.8. Strain Aut21 2201

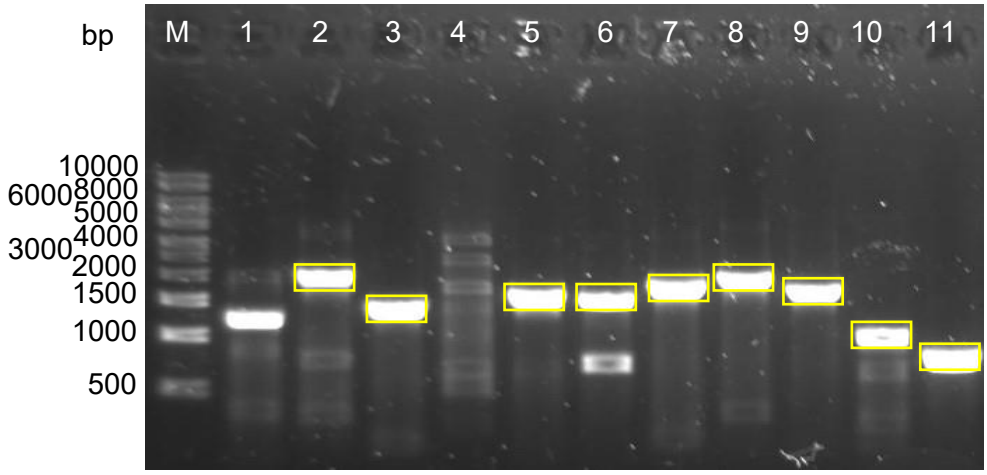

Figure 8: Gel electrophoresis Aut21 2201

The gel electrophoresis of the PCR products of Aut21 2201 showed, that except for lane 1 and lane 4 all lanes contained a band with the right length. For fragment 1 and 4 the PCR was repeated. After doing another gel electrophoresis with the products of the repeated PCR it was seen that fragment one had the correct length, but the PCR for fragment four again did not work. Therefore, another PCR was done, this time using the forward primer from fragment 3 and the reverse primer from fragment 5. The result was a larger fragment, which was a combination of fragment three, four and five, which was also confirmed by gel electrophoresis.

3.2.9. Strain Best_22

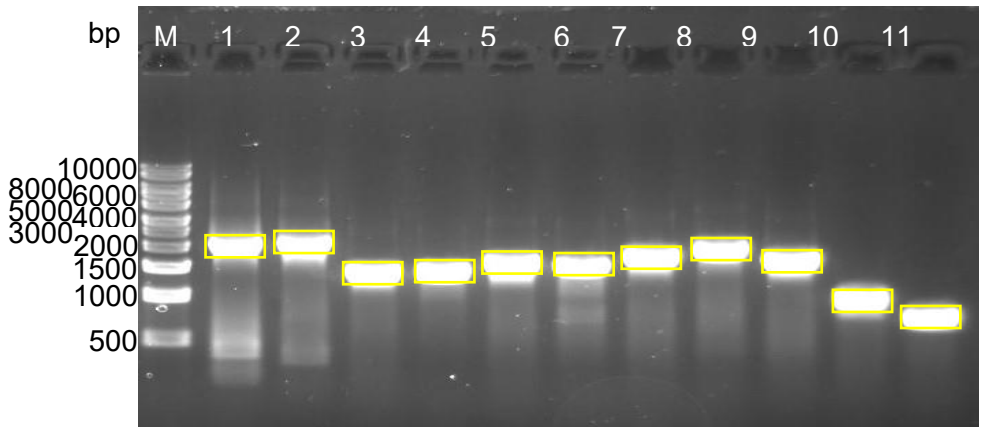

Figure 9: Gel electrophoresis Best

In figure 9 the gel electrophoresis following the PCR of Best is shown. On the left side labelled "M" is the 1 kb ladder, which was used to determine the length of the bands of the different PCR amplicons. As seen in the lanes labelled 1 to 11 all the produced PCR fragments had the expected length and no PCR had to be repeated. There were also no considerable additional bands seen on the gel. Therefore, no isolation of the PCR products was necessary, and all fragments could be pooled directly after the gel electrophoresis.

3.2.10. Strain Eber 2

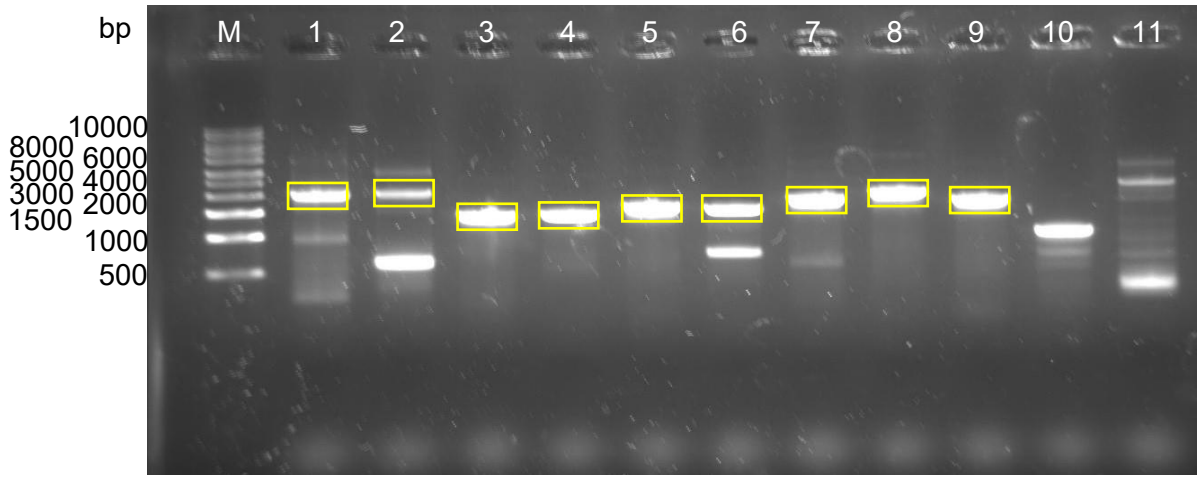

Figure 10: Gel electrophoresis Eber 2

Figure 10 shows the gel electrophoresis conducted after the PCR of the amplicons of the strain Eber 2. Shown on the left side, as in the other gels, is the 1 kb ladder used to determine the lengths of the different amplicons. The lanes 1 to 9 all contain a band with the expected length of the right amplicon. Column 10 and 11 on the other hand do not contain the respective band with the right length. Therefore, PCR was repeated for these two fragments. Additionally, the PCR for fragment two was repeated, as the gel showed a second, very strong band. Because the second fragment again produced an additional band, the corrected band was extracted from the gel, to ensure, that only the correct amplicon is used for sequencing.

3.3. Illumina Sequencing

Illumina Sequencing produced a large amount of sequence data for each strain. As already mentioned, the obtained data, was further analysed with the Galaxy tool Shovill and subsequently mapped to the genome of the PRRSV strain Aut15_33, which was also used to design the primers.

The assembly of the different strains produced sequences, which except for 1522_22 all cover the whole genome of the respective strains. Of course, the assembled sequences of the different strains slightly differ in their lengths, but all the strains where a complete sequence was obtained, can be aligned to the reference genome Aut15_33 over its entire length.

Strain 1522_22 differs from the other sequenced strains, as it was not possible to obtain one complete sequence with Illumina sequencing, because of a previous mistake made in library preparation. This mistake resulted in fragment 5 of the strain not being used in the pool for library preparation. Therefore, for further analysis regarding the similarities and differences of the strains the sequence of the missing fragment was added to the Illumina sequence by sanger sequencing.

Another strain that led to complications was 136_22. Sequencing of this strain led to 1,5 kb long Sequence. However, the additionally obtained Sanger sequences of these amplicons did not coincide with the Illumina sequences.

3.4. Sequence analysis

A BLAST alignment of all obtained strains with the reference genome Aut15_33 showed, that strain Eber 2 had with 96,42 % the highest identity with the reference genome. This was followed by strains 77_22 (identity: 95,71 %), 1099_22 (identity: 94.17 %) and 1522_22 (identity: 92,66 %). The other five strains all only showed an identity bellow 90 % when aligned with the reference genome. While 1510_22, 1445_22 and Best_22 had an identity higher than 85 %, 2857_22 and Aut21 2201 had the lowest similarities with 84,58 % and 83,98 %. The exact values are listed in the table below.

Identities with Aut15_33	
Strain	**Identity [%]**
Eber2	96,42
77_22	95,71
1099_22	94,17
1522_22	92,66
1510_22	87,63
1445_22	85,15
Best_22	85,08
2857_22	84,58
Aut21 2201	83,98

Table 10: Identity of the sequenced strains with Aut15_33

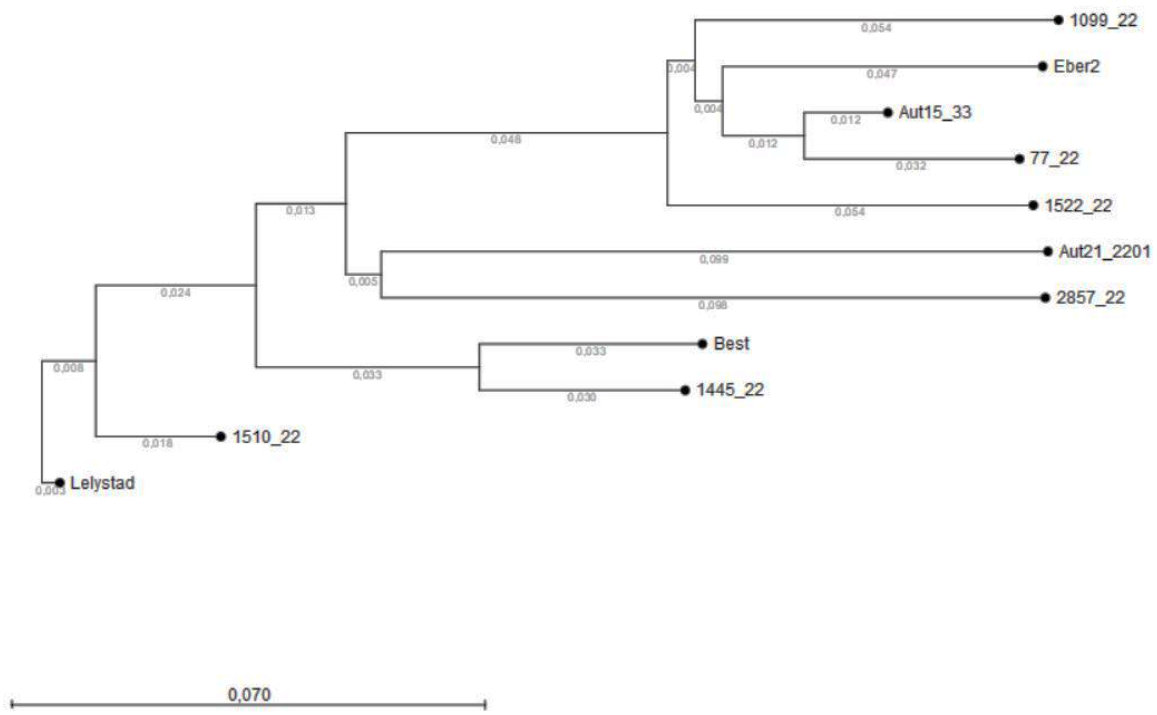

Figure 11: Phylogenetic tree of obtained sequences with Lelystad as root

The figure above shows a phylogenetic tree, of the nine obtained illumina sequences, with two reference sequences. The reference sequences are Lelystad (Genbank: NC_043487) and Aut15_33 (Genbank: MT000052.1). In this graphic, the Sequence of Lelystad was used as root for the phylogenetic tree. The branches of the phylogenetic tree, are labled with their lengths, wich shows the nucleotide substitutions per site. The tree clearly shows, that 1510_22 has the highest similarity to the reference sequence Lelystad, with about 0,029 nucleotide substitutions per site, between the two sequences. In case of 1510_22, which has a length of 14910 bp this would be about 432 nucleotide substitutions. Interestingly, the strain which is the second most similar to Lelystad, 1445_22 already has a difference of 0,098 nucleotide substitutions per site compared to the reference. As 1445_22 has a sequence length of 15007 bp, this would mean that 1445_22 differs in about 1470 nucleotide substitutions from Lelystad. Another information, the phylogenetic tree provides, is that the second reference sequence used for the analysis, Aut15_33 already shows 1933 nucleotide substitutions compared to Lelystad. 1099_22 shows the least similarity to Lelystad, as the phylogenetic tree shows 0,154 nucleotide substitutions per site, which would be about 2295 nucleotide substitutions in the sequence of 1099_22.

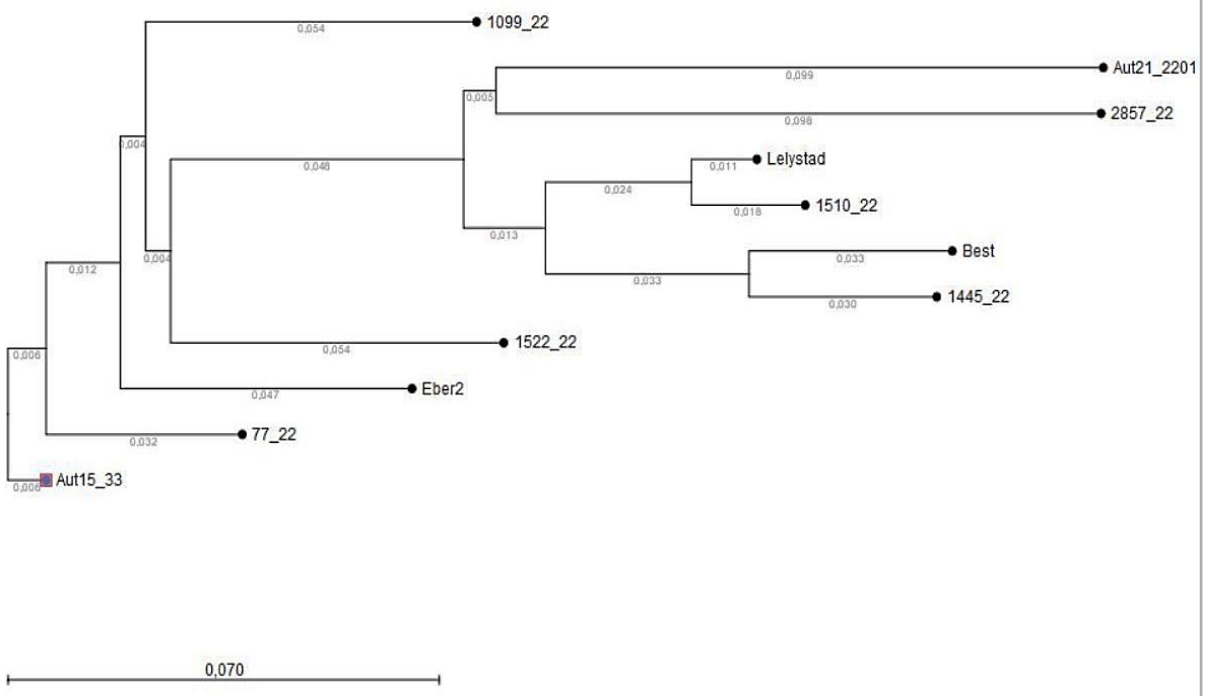

Figure 12: Phylogenetic tree of obtained sequences with Aut15_33 as root

Figure 12 shows the phylogenetic tree of the nine obtained sequences, with the reference sequences Aut15_33 and Lelystad. In this tree Aut15_33 is the root. The lengths of the branches are labelled underneath the respective branch. As seen in the picture, strain 77_22 has the least nucleotide substitutions compared to Aut15_33, with only about 660 nucleotide substitutions in 14996 bp. Aut21 2201 on the other hand, is according to the phylogenetic tree, the least similar to Aut15_22, with about 0,184 nucleotide substitutions per site. With a sequence length of 14958 bp, this would be around 2752 nucleotide substitutions compared to the reference. Interestingly, as seen in both figure 11 and figure 12, Aut21 2201 and 2857_22, Lelystad and 1510_22, as well as Best_22 seem to respectively share a common ancestor. Although, it has to be said, that the branch lengths of Aut21 2201 and 2857_22 to their common ancestor are quit long compared to the others, with 0,099 and 0,098 nucleotide substitutions per site.

Additionally, the recombination analysis, a proportion test, which was also with the Simplot++ software, showed that there is a possible mosaic region between the strain 1099_22 and PRRSV LV4.2.1 (Genbank: AY588319.1). This possible recombination would be in the area

4500 to 4740 bp. However, other than this mosaic region, no other possible recombinated sequences were found by this analysis.

4. Discussion

This execution of this bachelor thesis led to different results along the way. The first result is the list of universal, degenerate primers. These not only have the advantage of working with different PRRSV strains, because of their ability to bind alternative sequences, they also offer a way to amplify the whole genome of the input PRRSV strain, as the eleven primer pairs cover nearly all 15 kb of the viral genome. Additionally, the fragments these primer pairs produce, have enough nucleotides overlapping with each other, to allow sanger sequencing of the whole genome, if Illumina sequencing is not the desired method for sequencing.

To consider, when planning to do Illumina sequencing, is that the protocol implemented in this bachelor thesis involves several steps, therefore time management is important in this protocol. Overall, the protocol tested during the conduct of this bachelor thesis, worked very well and quite efficient, despite the few obstacles it provides. One great benefit of this protocol is that it produces full length sequences of the target strains. Of course, for only detecting PRRSV in samples, it would be enough to only sequence certain parts of the viral genome, however, full length sequences not only allow the comparison of different strains on a whole-genome basis, but it also allows better understanding of evolutionary events, that happened to PRRSV over time, for example recombination or nucleotide substitution. Another advantage Illumina Sequencing offers, is that the necessary amount of input sequence that is required for sequencing is quite low in comparison to sanger sequencing, which allows sequencing, even if the viral load in a sample or the viral concentration of cultivated virus is low.

The sequences obtained by Illumina sequencing where overall a good quality. The assembly of the sequences of course is more time consuming and more complicated, than an average sanger sequencing. Therefore, an important thing to consider, when analysing the results of the Illumina sequencing, is the choice of tool or software used for assembly, as every tool works differently in the background, which is rarely known to the user. The offered tools are also often very complex to use, and the user must define a lot of parameters, before running the analysis.

One other thing to consider, with this protocol is the cost that comes with it. The main cost-factor of the implemented protocol is the Illumina sequencing and the materials necessary for the library prep and the sequencing itself. Therefore, it has to be evaluated, if the cost of the analysis is profitable, as Illumina sequencing offers a certain capacity, which should be used. However, in case there is only a small number of samples it might be more profitable to conduct the PCR with the universal primers, but afterwards perform sanger sequencing instead of Illumina sequencing.

Overall, the implemented method works efficiently with cultivated viral material. In case of clinical samples, an improvement of the established method would be needed. One option of improvement would be an additional purification step, as there are additional sequences in tissue samples, for example the host DNA. And even if it might not be profitable, if only a few samples are to be analysed, considering that an Illumina MiniSeq instrument is available to the department and next generation sequencing methods are getting more and more important, this protocol presents a detailed step by step instruction, for future analysis of PRRSV samples.

5. List of Abbreviations

NGS	Next generation sequencing
Blast	Basic Local Alignment Search Tool
ORF	Open reading frame
RT-PCR	reverse transcription PCR
RT-qPCR	reverse transcription quantitative PCR
T_A	Annealing temperature

6. References

1. Dokland T. The structural biology of PRRSV. Virus Res. 2010 Dec;154(1–2):86–97.
2. Rossow KD. Porcine Reproductive and Respiratory Syndrome. Vet Pathol. 1998 Jan 26;35(1):1–20.
3. Cortey M, Díaz I, Martín-Valls GE, Mateu E. Next-generation sequencing as a tool for the study of Porcine reproductive and respiratory syndrome virus (PRRSV) macro- and micro- molecular epidemiology. Vet Microbiol. 2017 Sep;209:5–12.
4. Kappes MA, Faaberg KS. PRRSV structure, replication and recombination: Origin of phenotype and genotype diversity. Virology. 2015 May;479–480:475–86.
5. Kuhn JH, Lauck M, Bailey AL, Shchetinin AM, Vishnevskaya T V., Bào Y, et al. Reorganization and expansion of the nidoviral family Arteriviridae. Arch Virol. 2016 Mar 25;161(3):755–68.
6. Mackay IM, Arden KE, Nitsche A. Real-time PCR in virology. Nucleic Acids Res. 2002 Mar 15;30(6):1292–305.
7. Singh C, Roy-Chowdhuri S. Quantitative Real-Time PCR: Recent Advances. 2016. p. 161–76.
8. Linhart C, Shamir R. The Degenerate Primer Design Problem: Theory and Applications. Journal of Computational Biology. 2005 May;12(4):431–56.
9. Li K, Shrivastava S, Stockwell TB. Degenerate Primer Design for Highly Variable Genomes. 2015. p. 103–15.
10. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. Hum Immunol. 2021 Nov;82(11):801–11.
11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct;215(3):403–10.
12. Afgan E, Nekrutenko A, Grüning BA, Blankenberg D, Goecks J, Schatz MC, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Res. 2022 Jul 5;50(W1):W345–51.
13. SerialBasics. Serial Cloner 2.6.1, 2022.
14. Samson S, Lord E, Makarenkov V. Simplot++. 2022.
15. Qiagen. CLC Sequence Viewer 8.0. Hilden, Germany; 2017.

7. Tables and Figures

ACKNOWLEDGEMENT

At the end of this bachelor project, I want to thank everybody, that supported me while working on this bachelor thesis.

At first, I want to thank Univ.Prof. Dr.med.vet. Till Rümenapf, who not only supervised this bachelor thesis, and allowed me to work with his team for a few months, but also introduced me to this topic and over the course of my project always offered helpful advice and interesting ideas.

Additionally, I want to thank the team of the institute of virology, for not only supporting me during my project, but also for including me in their team in those few months I worked for the institute. Rarely have I met a team nicer or more welcome than this one.

I also want to thank my parents for their never-ending support and for always having an open ear, especially during the tough times of this degree. Without their support studying in Vienna would have not been possible for me.

Additionally, I want to thank all my family and friends, who not only listened to my ideas for this project, but who also motivated me during the tough times of this project and never doubted that I could make it.