**Summary of PhD thesis**

**THE USE OF NEXT GENERATION SEQUENCING IN THE GENOMIC EPIDEMIOLOGY OF PRRSV INFECTIONS IN HUNGARY**

Szilvia Jakab

Supervisors: Dr. Ádám Bálint and Dr. Krisztián Bányai

University of Veterinary Medicine

Doctoral School of Veterinary Science

2024

Supervisors:

…………………………………

Dr. Krisztián Bányai

HUN-REN Veterinary Medical Research Institute

…………………………………

Dr. Ádám Bálint

National Food Chain Safety Office

Veterinary Diagnostic Directorate

…………………………………..

Szilvia Jakab

**Table of contens**

[Introduction 4](#_Toc165624097)

[Objectives 5](#_Toc165624098)

[Material and methods 6](#_Toc165624099)

[Results and discussion 8](#_Toc165624100)

[New scientific results 14](#_Toc165624101)

[Publications 15](#_Toc165624102)

# Introduction

The viral pathogens of the porcine reproductive and respiratory syndrome (PRRSV-1 and PRRSV-2) are disseminated worldwide and causing significant economic losses to pig industry. The PRRS is characterized by respiratory problems which could be associated with reproductive disorders in sows. The PRRSVs belong to the family *Arteriviridae* within the order *Nidovirales*.

In Hungary, PRRSV-1 is endemic, while PRRSV-2 occurs sporadically. Hungarian PRRSV-1 strains exhibit high genetic diversity, primarily related to the importation of prefatteners. PRRSV-2 strains are mainly descendants of a modified live vaccine permitted in some European countries but not in Hungary. As a result of the PRRS eradication program initiated in 2014, the country's pig population became free of wild-type strains by spring 2022.

PRRSV is among the most variable viruses. Due to the high error-prone replication the virus population is genetically heterologous (quasispecies) within the host. The varying virulence and antigenic structure of circulating PRRSV strains complicate the application and development of vaccines and diagnostic methods.

The ORF5 encoding the GP5 envelope protein is one of the most variable regions of the PRRSV genome. Classification and molecular epidemiology of virus strains are usually performed by the sequence analysis of ORF5. These investigations are supporting the control and prevention of PRRS as well.

Sanger technology is still the gold standard method for determining the nucleotide sequence of DNA. Nonetheless, the utilisation of high-throughput next generation sequencing (NGS) methodologies in veterinary medicine is becoming increasingly prevalent, with both targeted amplicon deep sequencing and complete genome sequencing being employed in this field.

The objective of our research was to apply and expand NGS methods for the molecular genetic analysis of PRRSV. Furthermore, the origin, distribution and genetic variation of Hungarian PRRSV strains were investigated by analysing sequence data of ORF5.

# Objectives

The aims of our study were:

1. to determine the occurrence and diversity of Hungarian PRRSV-1 and PRRSV-2 strains

2. to investigate the structural region of PRRSV-2 using a traditional library construction method and NGS

3. to adapt and optimise a two-step PCR-based library preparation method coupled with amplicon deep sequencing that is specific to the ORF5 and ORF7 regions and the whole genome of PRRSV

4. to assess the potential diagnostic applications of amplicon deep sequencing

5. to analyse the resulting sequence data, focusing on the genetic diversity of the quasispecies

# Material and methods

**Origin of samples**

The samples (tissue supernatant, serum, organ, nasal swab, vaccine) were provided by the Veterinary Diagnostic Directorate of National Food Chain Safety Office (NÉBIH ÁDI), Dr. Máté Halas (Prophyl Ltd.) and Dr. Ádám Bálint (NÉBIH ÁDI). The clinical samples were obtained from Hungarian pig holdings for the purpose of routine PRRSV diagnostics.

**General molecular biology methods for sample processing**

Viral nucleic acid extraction was performed using the Nucleospin RNA virus kit (Macherey-Nagel) or the QIAmp Viral RNA Mini kit (Qiagen). The virotype PRRSV RT-PCR kit (Qiagen) specific for both PRRSV species, which is also used in routine diagnostic procedures, was used to determine the Ct values. PCR products were analyzed by agarose gel electrophoresis. The bands in the expected size were excised and purified using the Gel/PCR DNA Fragments Extraction Kit (Geneaid Biotech Ltd).

**DNA library preparation and next generation sequencing**

Two distinct methodologies were applied for the construction of the DNA libraries.

In some cases, PCR products were used as samples for DNA library preparation using conventional commercial kits. The sequencing was then performed on either the IonTorrent PGM or the Illumina NextSeq 500 platforms.

We also adapted a novel approach: a two-step PCR-based library construction method compatible with Illumina systems. In this case the sequencing was conducted on Illumina ISeq100 or MiSeq600 platforms. In this method fusion primers are utilized in PCR-1, which contain the Illumina sequencing primer sequences adjacent to target region specific sequence segments. In the second PCR step, Illumina specific adapters and barcodes are coupled to the PCR-1 product using another fusion primer containing these sequences. Once the second PCR has been completed, the resulting amplicon DNA libraries can be sequenced directly on Illumina instruments.

**Bioinformatics methods**

The "raw" sequence reads obtained during NGS were processed in Geneious Prime. After qualitative screening of the reads, the consensus sequence of the genomic regions or whole genomes was generated by reference mapping. Single nucleotide variants (SNVs) were detected using the built-in algorithm of the Geneious program.

The nucleotide and amino acid sequence alignments (nt and aa) were assembled using the MUSCLE or MAFFT algorithms. Pairwise nucleotide identity values between sequences were calculated using the p-distance method in MEGA X. N-glycosylation positions were estimated using NetGlyc 1.0 Server. Selection pressure was estimated using the FEL, SLAC and FUBAR algorithms available on the Datamonkey web server. Phylogenetic relationships were reconstructed using MEGA X software or IQ-TREE web server.

**Sequence data**

We analysed ORF5 sequence datasets (2003-2020) provided by the NÉBIH ÁDI to describe the distribution and molecular epidemiology of PRRSV strains in Hungary. In case of the PRRSV-1 a representative dataset was created, comprising those sequences (n=301, full ORF5) that were considered unique per pig holding, i.e. showing more than 2% nucleotide divergence from previously identified sequence variation. For PRRSV-2, all sequence data (n=44, partial ORF5) were used.

# Results and discussion

**Sequence data analysis**

**ORF5-based analysis of PRRSV-1 strains**

The Hungarian PRRSV-1 strains investigated in this study were classified into 16 clades (1A, 1B, 1C, 1E, 1F, 1G, 2, 3C, 3D, 3F, Porcilis, Porcilis-like, Spanish, Reprocyc, X1 and X2). Of these, four clades consisted of previously unclassified sequences (Porcilis, Porcilis-like, Spanish and Reprocyc), while two were new clades (X1, X2). The clades 3D and X2 showed the highest genetic diversity, while 1F, Porcilis and Reprocyc were the most conservative. With regard to their origin, the following clades were identified as being associated with pigs imported from abroad: 1A, 1F, 1G, X2, 2, 3C, 3D, 3F and Spanish. In addition, three clades consisting of vaccine-derived strains (Porcilis, Reprocyc, Spanish - Unistrain or commonly known as Amervac) were identified. This is not unexpected given the common use of these PRRSV-1 MLV vaccines in our country. The genetic analyses indicated that the Porcilis vaccine strain is genetically stable under field conditions, whereas the Reprocyc and Amervac vaccine strains exhibit less stability.

The comparative analysis of the ORF5 amino acid sequences revealed high degree of genetic variation. In contrast to the data from Europe, one amino acid position (as32) within the generally conserved neutralising epitope, several antigenic regions and T-cell epitopes were found to be diverse.

We identified six potential N-glycosylation sites in the GP5. Although N46 and N53 are typically conserved, we observed mutations in the N46 position, with the majority occurring in the Spanish clade. In parallel, the absence of N46 was accompanied by the presence of N37. The predominant glycosylation pattern of the Hungarian PRRSV-1 strains was the N37-N46-N53, which is the most prevalent in Europe and Asia.

A total of 25 amino acid positions showed significant positive selective pressure in the GP5. The majority were located in the N-terminal signal sequence and ectodomain protein regions. Analysing the selection pressure in each clade we found that the number of unique codons under diversifying selection was highest in the Spanish clade (6 codons). This finding also supports the high diversity of this clade.

**ORF5-based analysis of PRRSV-2 strains**

The number of Hungarian pig farms infected with PRRSV-2 was approximately four or five per year during the period 2005-2021, and these holdings distributed throughout the country. Our investigations revealed that only PRRSV-2 strains belonging to the L5 and L1 lineages were present.

The L5 includes the wild-type VR-2332 strain, its attenuated vaccine strain (Ingelvac MLV) and its derivatives. The Hungarian L5 sequences (n=39) showed the highest similarity to the Ingelvac MLV strain at both nt and aa levels. Furthermore, the assumption of vaccine origin is supported by the fact that most of the L5 strains were isolated from imported pig herds or from pig farms associated with such herds. It is of note that Ingelvac MLV is not authorised in our country, in contrast to some other European countries.

The Hungarian L1 sequences (n=5) showed the closest relationship to a Slovakian sequence, in addition they formed a separate group within L1 in the phylogenetic tree. As regard to the exact origin of the Hungarian L1 strains, no information is available. A recent study updated the classification of PRRSV-2 in which a Hungarian L1 strain belonged to a clade that was one of the dominant clades in the USA from the 1990s to 2003. In Hungary, the presence of the L1 strains was first detected in 2005. All L1 strains were associated with different pig farms of a single integration.

When analysing the amino acid sequence of GP5, we found a high degree of similarity within both lineages. However, several substitutions were observed at aa151 within the L5 sequences. This position is different between VR-2332 and Ingelvac MLV, and in about half of the Hungarian strains it was identical to VR‑2332. This aa position is characterised by high mutation rate and positive selection pressure.

**Two-step PCR-based library preparation**

**ORF7 region (both PRRSV species)**

The Illumina NGS system compatible, two-step PCR-based library construction method was initially designed specifically for the ORF7, as this is one of the most conserved regions of the PRRSV genome and thus several diagnostic methods and phylogenetic analyses are based on it. Sequencing of the ORF7 amplicon DNA libraries was performed on Illumina Miseq600 platform.

The PRRSV specific region of the fusion primer used in the PCR-1 was designed to be universal, i.e. capable of amplifying the entire ORF7 region of both species. Our preliminary studies have shown that the PCR system works well and its sensitivity is similar for both virus species in term of viral load (Ct 34).

The developed library construction method was tested on serum samples gathered from Hungarian pig farms (n=15, belonging to 10 different clades according to the ORF5 sequence analysis). Although the sequencing depth varied over a relatively wide range between the samples (202×-14986×), ORF7 consensus sequences were generated in all but one sample. The observed genetic divergence of ORF7 sequences was similar to that described worldwide. We identified two sequence variants in five samples by analysing deep sequencing data. Four minor variants showed the closest relationship to the Porcilis vaccine strain (nt identity: >99%). Our results suggest that deep sequencing of ORF7 amplicons may contribute to PRRS diagnostics, as it is widely applicable as well as may detect the simultaneous circulation of multiple PRRSV variants in a monitored pig population.

We identified heterogeneous viral populations in eight serum samples. A total of 33 SNVs were detected. These were unevenly distributed among the serum samples and along the ORF7 region, and their prevalence was varied from 10.2% to 46.7%. According to the literature, the variation observed in the quasispecies can be influenced by several factors, such as the type of sample, the time elapsed since infection or the method of SNV analysis.

**Whole genome (two PRRSV-1 vaccine strains)**

In this study, we developed a multiplex PCR system specific for Porcilis and Unistrain vaccine strains and their descendants, which was integrated into the two-step PCR-based library construction process. For our experiments, we directly extracted nucleic acid from three Porcilis® and one Unistrain® vaccine vials representing different batches and performed PCR-based library construction, which was sequenced on an Illumina Miseq600 platform.

First, we designed primers to amplify 34 overlapping amplicons (size 544-583 bp) that provide 100% coverage along the genome. Second, we convert them into fusion primers for the PCR-1. Afterwards, we divided the primer set into two pools (odd and even amplicons) and prepared three pooling schemes where the relative proportions of the primers were adjusted. Finally, after optimising the pools, we reached a more balanced sequencing depth along the entire genome and completely eliminate regions of low depth.

Analysis of the NGS data revealed high genetic complexity in the virus population of Porcilis vaccines and some differences between the different batches as well. Recently, Eclercy et al. reported that Porcilis vaccines are composed of four different genome sequence variants (FULL-LENGTH and deletion variants: LONG-DEL, SHORT-DEL, SHIFT-DEL), which we also detected in the three tested vaccine samples. The detected quasispecies in the vaccine samples showed a consistent pattern, i.e. the identified SNVs occurred in similar locations and at similar frequencies.

By examining our and three additional whole genome sequences downloaded from the genebank, we found that the Porcilis vaccine strain is genetically stable at the consensus sequence level, with only 0.1% genetic divergence between them.

**Conventional DNA library preparation**

**Detection of mixed infections**

We utilized conventional DNA library preparation and Ion Torrent platform to analyse the PCR products for a six month period, that were generated in the routine PRRSV monitoring during the eradication program.

We detected minor variants in some of the processed samples (182 ORF5 and 10 ORF7): two in two ORF5s, one in one ORF5 and two ORF7s. Nucleotide identity values suggested co-occurrence of the Porcilis and Amervac vaccine strains or a wild-type and vaccine strain in some mixed samples.

**Investigation of PRRSV-2 outbreaks**

In the period from autumn 2020 to summer 2021, we contributed to the inspection of two PRRSV-2 outbreaks (case I - 5 outbreaks, case II - 2 outbreaks in affected pig farms) with the genetic analyses of the ORF2-7 region (encoding structural proteins). In this study, conventional DNA library preparation was applied from ORF2-7 PCR products, and the sequencing was performed on Illumina NextSeq 500 platform.

The examined ORF2-7 consensus sequences (failed in case of two pig farms) showed high nucleotide identity with the VR-2332 wild-type strain and the Ingelvac MLV vaccine strain. However, amino acid sequence analysis identified several amino acids identical to the Ingelvac MLV strain.

Unexpectedly high genetic divergence (up to 2.5%) was observed among the PRRSV-2 strains tested in these outbreaks. Based on literature data, the maximum nucleotide divergence in the whole genome sequence of Ingelvac MLV vaccine strains could reach as high as 6.4% within one year.

We gained insight into the diversity of the quasispecies of the vaccine-derived PRRSV-2 strains circulating in our country by analysing the NGS data. We identified a total of 77 SNVs in four samples. No patterns were observed in the distribution of SNVs within or between ORF regions and in the position of samples in the time course of infection. In response to our question whether a trend towards reversion in the quasispecies was observed in the vaccine-like strains, we found that only a few SNVs matched the wild-type VR-2332 strain.

Sequencing, phylogenetic and epidemiological studies of the ORF2-7 region have successfully reconstructed the possible transmission pathways of the PRRSV-2 vaccine strain. In both cases, we determined that the origin of the Ingelvac MLV vaccine strain could be the imported pigs transferred through collecting stations from Denmark. Within Hungary, the vaccine strain spread between the pig farms via live animal transport and other service vehicles.

# New scientific results

1. We established the distribution map of both PRRSV species in Hungary and assessed their genetic diversity by examining sequence data of ORF5.
2. We adapted a two-step PCR-based DNA library construction method compatible with Illumina systems for PRRSV and performed amplicon deep sequencing.

Two different PCR systems have been developed and optimised:

i.) for the ORF7 region of both PRRSV species;

ii.) for the whole genome of the Porcilis and Unistrains PRRSV-1 vaccine strains and their descendants.

1. The applicability of ORF7 amplicon deep sequencing has been successfully evaluated with clinical samples collected from Hungarian pig farms. In addition, we have found that whole genome sequencing of the Porcilis and Unistrain vaccine strains is also feasible. Based on our results, both approaches could be used to support both diagnostics and research of PRRSV, including the identification of circulating PRRSV strains, their genetic diversity and the assessment of the complexity of the viral population. We showed that amplicon deep sequencing is a fast, simple and cost-effective alternative to the introduction of NGS in the diagnostic field.

# Publications

**Publications in peer-reviewed journals related to the thesis**

Jakab, S., Marton, S., Szabó, I., Kecskeméti, S., Bálint, Á., Bányai, K.: **Kezdeti tapasztalatok az amplikon-mélyszekvenálás PRRS-mentesítési programba történő integrálásával kapcsolatban.** Magyar Állatorvosok Lapja. 144. 115–128. 2022.

Jakab, S., Kaszab, E., Marton, S., Bányai, K., Bálint, Á., Nemes, I., Szabó, I.: **Genetic diversity of imported PRRSV-2 strains, 2005–2020, Hungary.** Front. Vet. Sci. 9. 1581. 2022.

Jakab, S., Bányai, K., Bali, K., Nemes, I., Bálint, Á., Szabó, I.: **Transmission dynamics of imported vaccine-origin PRRSV-2 within and between commercial swine Integrations in Hungary**. Animals. 13 (19). 3080. 2023.

Jakab, S., Bali, K., Freytag, C., Pataki, A., Fehér, E., Halas, M., Jerzsele, Á., Szabó, I., Szarka, K., Bálint, Á., Bányai, K.: **Deep sequencing of porcine reproductive and respiratory syndrome virus ORF7: A promising tool for diagnostics and epidemiologic surveillance.** Animals. 13(20). 3223. 2023.

Jakab, S., Bálint, Á., Cseri, K., Bali, K., Kaszab, E., Domán, M., Halas, M., Szarka, K., Bányai, K.: **Genome stability assessment of PRRS vaccine strain with new ARTIC-style sequencing protocol.** Front. Vet. Sci. 10. 1327725. 2024.

Bálint, Á., Jakab, S., Kaszab, E., Marton, S., Bányai, K., Kecskeméti, S., Szabó, I.: **Spatiotemporal distribution of PRRSV-1 clades in Hungary with a focus on the era of disease eradication.** Animals. 14(1). 175. 2024.

**Publications that are not closely related to the topic of the thesis**

Homonnay, Z., Jakab, S., Bali, K., Kaszab, E., Mató, T., Kiss, I., Palya, V., Bányai, K.: **Genome sequencing of a novel variant of fowl adenovirus B reveals mosaicism in the pattern of homologous recombination events.** Arch. Virol. 166(5). 1477–1480. 2021.

Fehér, E., Jakab, S., Bali, K., Kaszab, E., Nagy, B., Ihász, K., Bálint, Á., Palya, V., Bányai, K.: **Genomic epidemiology and evolution of duck hepatitis A virus**. Viruses. 13(8). 1592. 2021.

Fehér, E., Bali, K., Kaszab, E., Ihász, K., Jakab, S., Nagy, B., Ursu, K., Farkas, S.L., Bányai, K.: **A novel gyrovirus in a common pheasant (*Phasianus colchicus*) with poult enteritis and mortality syndrome.** Arch. Virol. 167(5). 1349–1353. 2022.

Fornyos, K., Búza, L., Makkai, I., Polyák, F., Pogácsás, I., Savoia, L., Szegedi, L., Bálint, Á., Jakab, S., Bányai, K., Szabó, I.: **Sampling strategies in PRRS elimination in Hungary: An observational study involving four farrow-to-finish swine herds.** Vet. Sci. 10. 546. 2023.

Jakab, S., Bali, K., Homonnay, Z., Kaszab, E., Ihász, K., Fehér, E., Mató, T., Kiss, I., Palya, V., Bányai, K.: **Genomic epidemiology and evolution of fowl adenovirus 1.** Animals. 13(18). 2819. 2023.

**Conferences**

Jakab, S., Bálint, Á., Szabó, I., Bányai, K.: **Current situation of the eradication program of porcine reproductive and respiratory syndrome (PRRS) in Hungary: The molecular epidemiological perspective.** 10th Jubilee Interdisciplinary Doctoral Conference, Pécs, 12-13th of November 2021, Book of Abstracts.

Jakab, S., Bálint, Á., Bányai, K.: **PRRSV törzsek kimutatása és jellemzése NGS-alapú amplikon szekvenálással.** Akadémia Beszámolók, Budapest, 2023. január 30-31.

**Book chapter that are not closely related to the topic of the thesis**

Vlasova, A.N., Deol, P., Sircar, S., Ghosh, S., Jakab, S., Bányai, K., Dhama, K, Amimo, J.O., Saif, L.J., Malik, Y.S.: **Animal Rotaviruses.** In: Malik, Y.S., Singh, R.K., Dhama, K. (eds). Springer, Singapore, 163–202. 2020.