Whole genome sequencing of porcine reproductive and respiratory syndrome virus (PRRSV) from field clinical samples improves the genomic surveillance of the virus

Christian Lalonde¹, Chantale Provost¹,², Carl A. Gagnon¹,²

¹Swine and Poultry Infectious Diseases Research Center (CRIPA) and ²Molecular diagnostic laboratory (MDL); Faculté de médecine vétérinaire, Université de Montréal, Saint-Hyacinthe, Québec, Canada.

Corresponding author: Carl A. Gagnon, Swine and Poultry Infectious Diseases Research Center (CRIPA), Faculté de médecine vétérinaire (FMV), Université de Montréal (UdeM), 3200 rue Sicotte, office 3963, St-Hyacinthe, Québec, Canada, J2S 2M2. carl.a.gagnon@umontreal.ca. Phone: 450-773-8521 (ext: 8681).

Running title: Incidence of PRRSV coinfection and recombination by WGS

Keywords: animal viral disease, swine virus, porcine reproductive and respiratory syndrome virus, PRRSV, next-generation sequencing, NGS, whole genome sequencing, WGS, recombinant, coinfection, classification
Abstract

Porcine reproductive and respiratory syndrome virus (PRRSV) is a major economic concern worldwide. There are currently large data sets available about the ORF5 gene of the virus, with thousands of sequences available, but little data is currently available on the full-length genome of PRRSV. We hypothesized that whole genome sequencing (WGS) of PRRSV genome would allow a better epidemiological monitoring compared to ORF5 gene sequencing. PRRSV PCR positive sera, oral fluids and tissue clinical samples submitted to the diagnostic laboratory for routine surveillance or diagnosis of PRRSV infection in Québec, Canada, swine herds were used. The PRRSV RT-qPCR Cq values of the processed samples varied between 11.5 and 34.34. PRRSV strain genomes were isolated using a poly(A)-tail method and were sequenced with an MiSeq Illumina sequencer. Ninety-two full length PRRSV genomes were obtained from 88 clinical samples, out of 132 tested samples, resulting in a PRRSV WGS success rate of 66.67%. Three important deletions in the ORF1a were found in most wildtype (i.e. not vaccine-like) strains. The importance of these deletions remains undetermined. Two different full-length PRRSV genomes were found in four different samples (three sera and one pool of tissues), suggesting a 4.55% PRRSV strain coinfection prevalence in swine. Moreover, six PRRSV whole genomes (6.52% of PRRSV strains) were found to cluster differently compared to ORF5 classification method. Overall, WGS of PRRSV enables a better strains classification and/or interpretation of results in 9.10% of clinical samples compared to ORF5 sequencing, as well as allowing interesting research avenues.
Porcine reproductive and respiratory syndrome (PRRS) is a major economic concern worldwide, costing 663 million dollars yearly to the American swine industry(1) and over 150 million dollars yearly to the Canadian swine industry(2). Part of this cost is due to reproductive troubles in sow (characterized by late abortion, increased incidence of stillbirth, mummified fetuses and weaker newborn piglets)(3-6) and retarded growth in piglets, mainly caused by respiratory problems (characterized by interstitial pneumonia)(5-10) and increased animal susceptibility to other pathogens(11-13). Beside this direct loss in productivity, a lot of efforts and resources are allocated to control and contain this infectious disease. Among those efforts, vaccination and epidemiological surveillance are the most prominent (14, 15).

The etiological agents of PRRS are one of two viruses of the Arteriviridae family and the Porartevirus genus that are aptly named porcine reproductive and respiratory syndrome virus (PRRSV) type 1 and type 2 (officially Betaarterivirus suid 1 and Betaarterivirus suid 2). Both PRRSV are enveloped viruses with a diameter of around 50-65 nm(16). They contain a single stranded positive-sense RNA genome, of around 15 kb in length, capped at the 5’ end and polyadenylated at the 3’ end and contains at least 11 known open reading frames (ORF)(17, 18). The first two ORFs (ORF1a and ORF1b), which constitute about 75% of the viral genomes and contain at least three ribosome shifting sites, encodes replicase polyproteins that are post-translationally cleaved into at least 16 distinct non-structural proteins (nsp) (18-20). The remaining 25% of the genomes code for at least 8 known structural proteins (GP2, GP3, GP4,
GP5, M, N, E and GP5a). Both PRRSV species share a nucleotide identity of around 60% (22, 23).

Many genotypes of both PRRSV exist and the virulence of each strain is highly variable. PRRSV strains found in Canadian swine herds belong to type 2 PRRSV. To our knowledge, indigenous PRRSV type 1 strains have never been reported in Canada but was found on one occasion in European imported piglets kept in quarantine. Nine distinct monophyletic lineages of PRRSV type 2 have been described worldwide. These lineages are genetically distinct, with a nucleotide (nt) identity between lineage under 89%. Of these 9 lineages, lineage 1, 2, 5 and 8 are the most prominent in Canada, with lineage 9 showing up sporadically. The vast majority of lineages 5 and 8 strains are thought to be vaccine related. These lineages have been determined using Bayesian phylogenies of the ORF5 gene encoding the membrane structural glycoprotein protein GP5 of the virion. Interestingly, recent studies have investigated the genomic diversity and relatedness of PRRSV Canadian strains, but all those reports are using PRRSV ORF5 nucleotide sequences into their genomic analyses. However, there is much less information and data available about the whole viral genome of PRRSV strains both in Canada and around the globe.

The inter-strain genetic variability of PRRSV is very high. It is believed that this enables the virus to better evade the immune system and potentially diminish vaccine efficacy. In an ongoing effort in Quebec, Canada, as many PRRSV ORF5 sequences as possible are added to an ever-growing database, with over 4695 ORF5 sequences to date. The goal of this initiative was to better understand the epidemiological links between various PRRSV strains versus outbreaks and determine the proximity of a given strain to commercially available vaccinal strains as a
strategy to fight PRRS disease. This database is also used to find genetically similar strains, enabling veterinarians to better predict the severity and outcome of a current outbreak based on the severity and outcome of past outbreaks with ORF5 genetically similar strains. The Molecular diagnostic laboratory of the Diagnostic service (Faculté de médecine vétérinaire of Université de Montréal), is the depository of that Quebec ORF5 sequences databank. The PRRSV ORF5 gene was selected for the molecular epidemiological surveillance because it was considered hypervariable and encodes the GP5 protein that acts as the main target of neutralizing antibodies.

However, the ORF5 is only 603 nucleotides long, representing only 4% of the viral genome and it has been shown that other genes are also hypervariable and that pathogenicity of PRRSV is determined by multiple genes. The GP5 antigenicity property (i.e. its recognition by neutralizing antibodies) has contributed to the selection of ORF5 for PRRSV surveillance despite the fact that other viral proteins are involved in the virion recognition by neutralizing antibodies(40). Given the high prevalence of PRRSV infections, concomitant infections with different strains are expected to occur quite frequently as it is known to happen already in other swine respiratory viruses. It was previously reported that PRRSV viral recombination occurs prominently during coinfection with two different strains. All of these facts suggest that in some clinical cases of PRRSV infections, PRRSV recombinant strains could be misclassified, affecting the interpretation of the data at hand and the subsequent intervention by veterinarian practitioners. We thus hypothesized that whole genome sequencing of PRRSV strains from clinical samples would enable a better classification of PRRSV strains compared to the current surveillance method of ORF5 sequencing. Therefore, whole genome sequencing of PRRSV strains could lead to more appropriate interventions by veterinarians and swine producers, in
addition to improving our understanding of the pathogenicity and the epidemiology of this important swine pathogen.
Swine samples. Conveniency swine samples that tested positive for PRRSV type 2 by a RT-qPCR diagnostic assay were selected for whole genome sequencing (WGS) of the entire PRRSV viral genome. The samples were submitted to and tested by the «Molecular diagnostic laboratory (MDL) of Faculté de médecine vétérinaire (FMV), Université de Montréal (UdeM)» for the identification of PRRSV after an outbreak of the disease in swine herds and to a lesser extent to conduct surveillance of the virus in swine herds. The PRRSV RT-qPCR diagnostic assay conducted by MDL (own the American Association of Veterinary Laboratory Diagnosticians, AAVLD, accreditation) was an in-house assay (protocol: PON-MOL-029). The swine samples originated from different herds and types of production throughout the province of Quebec, Canada. Those clinical samples were submitted between December 2015 and November 2018 (most of them were collected in 2017 \( n=62 \) and 2018 \( n=66 \)). A total of 132 samples were used and included: 70 sera (composed of 43 pooled sera), 2 oral fluids (OF), 32 lungs and 28 pools of tissues (PoT; mainly lungs with several other types of tissues such as lymph nodes, spleen, liver, intestine, etc.), with PRRSV RT-qPCR diagnostic assay Cq values between 11.5 and 34.34. The PRRSV viral genome was considered complete if at least 98% of the coding sequence was obtained.

Genome extraction and purification. One hundred mg of lung or PoT were grinded using a Beadbeater apparatus (model: Mini-Beadbeater-96; BioSpec Products Inc., OK, USA) in PBS, then centrifuged at full speed for one minute and supernatant was use for viral extraction. Two hundred ul of sera and OF were centrifuged for 5 minutes at 10 000 g and supernatant was used
for viral extraction. Viral RNA was extracted using Quick-RNA Viral Kit (#R1035, Zymo Research, CA, USA) as described in the company's protocol. Thereafter, RNA was eluted using 50 ul of nuclease free water (Corning, NY, USA). Total elution volume was used to isolate RNA with poly(A)-tails with a magnetic beads purification method, the NEBNext® Poly(A) mRNA Magnetic Isolation Module, as described by the company’s protocol, and poly(A)-tails RNA was resuspended in 15 ul Tris Buffer (#E7490; New England BioLabs, ON, Canada). Then, first strand cDNA was synthesised using the Non-directional Reaction Step Up protocol of the NEBNext® RNA First Strand Synthesis Module (#E7525) (New England BioLabs, ON, Canada), starting with 10 ul of isolated poly(A)-tails RNA. Immediately after first strand cDNA synthesis, the second DNA strand was synthesized using NEBNext® Ultra II Non-Directional RNA Second Strand Synthesis Module, as described by the manufacturer protocol with a minor modification at the incubation step in the thermal cycler for 2 hours (instead of 1.5 hours) at 16⁰C (#E6111; New England BioLabs, ON, Canada). dsDNA was then purified using AxyPrep Mag™ PCR Clean-up Kits (Axygen – Corning, NY, USA) using 1.8X of beads and 70% ethanol. The purified dsDNA was diluted in 30 ul of 10mM Tris-HCl at pH 8.0 and stored at -20⁰C for later use.

**PRRSV whole genome sequencing.** dsDNA was quantified using Qubit™ dsDNA HS Assay Kit and a Qubit™ fluorometer (ThermoFisher Scientific, MA, USA). Sequencing libraries were prepared using Nextera XT DNA Library Preparation Kit (Illumina, CA, USA). Sequencing libraries construction were still performed even if the dsDNA quantification results were below the Qubit™ dsDNA HS Assay Kit threshold of detection (0.2 ng). Briefly, 0.2 to 0.3 ng of dsDNA were tagmented with 10 µl Tagment DNA Buffer (TD) and 5 µl Amplicon Tagment Mix
(ATM) at 55°C for 5 minutes, using a Thermocycler TProfessional Basic 96 (Biometra GmbH, 37079 Göttingen, Germany). The transposomes were inactivated with 5 µl of Neutralize Tagment Buffer (NT). Sequencing libraries were then amplified using index adapter and Nextera PCR Master Mix (NMP) following those PCR steps: 72°C for 3 minutes, 95°C for 30 seconds; then 14 cycles of 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 seconds; a final elongation step at 72°C for 5 minutes; and a hold at 10°C until the next step. Sequencing libraries were then purified using AxyPrep Mag™ PCR Clean-up Kits (Axygen – Corning, NY, USA) as described by the Nextera XT protocol. Sequencing libraries quality were assessed using Agilent High Sensitivity DNA Kit with a Bioanalyzer (Agilent, CA, USA). Sequencing libraries were normalised using LNB1 beads (Nextera XT protocol) or the manual normalisation protocol if the concentrations of the libraries were on the lower end of the Bioanalyzer curves. Sequencing libraries were sequenced in a v3 600-cycle cartridge (#MS-102-3003; Illumina) using an MiSeq Illumina benchtop sequencer (located at the “Veterinary High Throughput Sequencing Laboratory (VHTSL)”, at FMV) and PhiX was included at around 1% of the total sequencing libraries, as a control to establish the sequencing run efficacy (Illumina, CA, USA).

Bioinformatic analyses. At first, reads were trimmed for adaptors and quality by the MiSeq software during FastQ generation. Using CLC Genomic Workbench software (version 12.0.3, Qiagen, CA, USA), reads from a sample were assembled using the software’s built-in de novo assembly module, including another step of adaptor and quality trimming. Contigs obtained were checked against a database made using reference type 2 PRRSV genomes available online to identify the contigs that corresponded to full-length PRRSV sequences. These full-length genomes obtained by de novo assembly were then added to our PRRSV reference database.
Secondly, reads from each sample were mapped against the same database. For each sample, the reference genome onto which the highest number of reads mapped and offered a uniform coverage was used as the reference genome for resequencing analysis using the software’s map to reference module. When more than one reference genome seemed to offer an appropriate depth and coverage, both were used separately for resequencing analysis, and the resulting sequences compared to one another. Samples from which two PRRSV genomes could be assembled and possessed a nucleotide (nt) identity of < 96% were considered to contain more then one PRRSV strain. The < 96% nt identity cut-off value was determined based on the sequencing error rate of < 1% and the fact that we have ambiguity in less then 1% of the positions for each complete sequence. Thus, the worst case scenario where both PRRSV coinfecting strains could each have 1% error rate and 1% uncertainty could lead to identical strains with a nt identity of ≥ 96%. Therefore, this nt identity value was established as the cut-off value. PRRSV recombinant status was determined by comparing the location of a given strain’s individual ORFs on their respective phylogenetic trees. Bioinformatics and data analyses were performed using the CLC Genomic Workbench software (version 12.0.3, Qiagen, CA, USA). Phylogenetic trees were constructed using a maximum likelihood model with a type 1 PRRSV reference strain (Lelystad virus) used as an outgroup and with a bootstrap setting of 1000. The trees were constructed with the Geneious Prime software (Biomatter, version 2019).

Accession Numbers

All entire viral genomes of successfully sequenced PRRSV strains were submitted to GenBank databank, with accession numbers MN865482 to MN865573 (Suppl. Table 1).
Results

PRRSV WGS were obtained from a total of 88 swine clinical samples (an overall 66.67% PRRSV WGS success rate). In those clinical samples, the PRRSV WGS was successful in two OF, 13 PoT, 21 lungs and 52 sera. WGS made a difference over ORF5 sequencing in eight of those samples (Table 1). Ninety-two PRRSV entire genomes were obtained from those clinical samples (Suppl. Table 1). The additional 4 PRRSV genomes were the consequences of the identification of 4 PRRSV coinfections, samples in which two different PRRSV strains with a nt identity below 96% were identified (listed in Table 1). Therefore, a total of 4.55% PRRSV coinfections were found among PRRSV WGS successful cases. The inter-strains nt identities of the 92 PRRSV entire viral genome sequences were found to vary between 79.57% and 99.94% (data not shown). At the viral gene level, the nsp2 and nsp7 encoded genes nt identities were found to be the least conserved between the 92 PRRSV Quebec strains, with 66.04% and 64.38% nt identities, respectively, indicating that ORF1a genetic diversity seemingly a little higher compared to the others ORF. Noteworthy, the overall genetic diversity was scattered throughout all the ORFs (Fig. 1, Table 2). The highest nt identities were found for the nsp4, E, N and M gene encoded proteins with 92.41%, 86.94%, 86.13% and 86.10%, respectively.

Three deletions were identified by WGS in the ORF1a of most wild-type strains (Fig. 1). The first one, located in the nsp2 coding region at nt positions 2159 to 2491 and is 333 nt long. The second one, also in the nsp2 coding region, was found around nt positions 2717 to 2773 with a span of 57 nt. Both deletions occurred together and were found in 63 of the 92 PRRSV Québec sequenced strains. The third deletion was found in the nsp7 coding region, at position 7106 to 7228 and was 123 nt long. The nt position of the deletions are given relative to the coding
sequence of the MLV vaccine strain used as a reference. All the strains that harbor the nsp7
deletion also have the nsp2 deletions but some strains that do not possess the nsp7 deletion have
the nsp2 deletions. Only 47 of the 92 sequenced strains have the nsp7 deletions.

Based on ORF5 sequencing classification method, all 92 PRRSV strains were found to
belong to lineage 1, 5 and 8 (Fig 2b). Interestingly, 18 PRRSV strains were considered to be
MLV vaccine-like strains (i.e. with nt identities > 96%) if the tree was built using ORF5 while
17 strains were considered to be MLV vaccine-like strains while using the whole genome.
Noteworthy, one discrepancy was found in regard to MLV vaccine-like strain relatedness
classification. In fact, PRRSV strain 2072533 was classified as a MLV vaccine-like strain based
on the ORF5 method whereas it was not genetically related to MLV vaccine-like strain based on
WGS results, with nt identities of 98.51% and 84.01%, respectively (Fig. 2). Moreover, PRRSV
strain 2072533 was identified to be a recombinant virus which explains the discrepancy between
the ORF5 and WGS classification methods (Table 1). No difference was found for the ATP
vaccine-like and Fostera vaccine-like classification status (Fig 2a compared to 2b).
Consequently, misclassification by ORF5 sequencing of PRRSV vaccine-like strains was
revealed by WGS in 3.23% of samples.

As illustrated in Figure 2, PRRSV whole genome sequences phylogenetic tree has
allowed to construct a more robust classification, based on the bootstrap values, compared to the
ORF5 sequencing method. More interestingly, six PRRSV strains (identified by colored arrows
in Fig. 2) were found to cluster differently within the WGS phylogenetic tree compared to the
ORF5 phylogenetic tree. A summary of those classification differences is presented in Table 1.
One of the reasons that could explain this difference is recombination events, at some point in time, between two PRRSV strains during coinfection in swine. As an example of this phenomenon, the 2072533 PRRSV strain nucleotide identity with a presumed parental PRRSV strain (the MLV vaccine strain) is shown at Figure 3 in which the recombination position was identified around nt position 12,100. The nt identity between these two PRRSV strains for the first 12100 nt of the 5’ end (i.e. encompassing ORF1a, ORF1b, ORF2a and ORF2, is 80.46% whereas it is 98.63% (data not shown) for the 3’ end region of the genome (the last 3,000 nt of the genome). For strain 1952821 #1, the first half of the genome (mainly the ORF1a) shares a 85.53% nt identity with the MLV vaccine strain, whereas the second half (mostly ORF1b to ORF7) shares a 96.55% nt identity with the MLV vaccine strain (data not shown), suggesting again a recombination between a wild-type and vaccine-like strain. For the other possible recombinant strains, the specific recombination point could not be determined, either because the parental strains are unknown or because those strains may have undergone several recombination events over time. Overall, 6.52% possible recombinant viruses over a total of 92 PRRSV strains were found among PRRSV WGS successful cases. Noteworthy, 2 of the 6 possible recombinant viruses were found in PRRSV coinfected samples. This is however to be expected given that coinfection is a prerequisite for recombination events to occur, taking into account that more then one third of produced virions during coinfections can be PRRSV recombinant viruses(42).

In four of the 88 successfully sequenced PRRSV positive samples, two different PRRSV strains were found (Table 1), representing 4.55% coinfections among PRRSV WGS successful cases as indicated above. The nt identities of these four pairs of PRRSV were 88.8%, 92.36% 91.57% and 81.83% for samples 1890826, 1952821, 1968868 and 2153073, respectively (data
not shown). A graph of the nucleotide conservation status of the PRRSV strains in sample 2153073 is shown in Figure 4 as an example. Interestingly, the 2153073 #2 strain possesses the nsp2 and nsp7 deletions like Quebec wildtype strains whereas its coinfecting strain, 2153073 #1, is more related to ATP vaccine-like strains (sharing a 99.02% nt identity). All coinfections were identified from one individual animal (i.e. sample coming from a single animal and not from a pool of different animals). Interestingly, coinfections were identified in three sera samples (5.77% of WGS successful sera) and one PoT sample (7.14% of WGS successful PoT) (Suppl. Table 1 and Table 1).
Discussion

The first observation that can be easily made from our type 2 PRRSV complete genome analyses is that many of the PRRSV strains currently circulating in Quebec appear to have three relatively large deletions (Fig. 1). The first two of these occur in the region coding for the nsp2 protein and the third one in the region coding for the nsp7 protein. While the exact implication of these deletions is unknown, several investigations have been conducted on type 2 PRRSV strains possessing various nsp2 deletions, especially with Asian PRRSV strains. The nsp2 protein is known to play a role in both immune system modulation and pathogenicity. For example, it is believed that some amino acids in the nsp2 protein could play a role in evading the ubiquitin-ISG15 system of the innate immune response. There are also many reports of highly pathogenic or epidemic strains recognizable by nsp2 deletions. The importance of nsp2 in the pathogenicity of PRRSV is what makes these deletions so interesting, and why they warrant further investigations. Noteworthy, these specific nsp2 deletions can also be found in some PRRSV strains that were reported in the United States (i.e.: GenBank #KT258000 and #EF532803), meaning that they are not markers of a specific North American country but are probably specific to lineage 1 or at least to a subset of lineage 1 PRRSV strains. The involvement of the third deletion, in the region coding for the nsp7 protein into the pathogenicity of PRRSV, is much more elusive. PRRSV pathogenicity studies involving nsp7 deletions and mutations experiment are scarce, and usually involve pinpoint mutations or very short deletions(50). However, at least one study has reported that nucleotide mutations at some nsp7 specific site could lead to changes in PRRSV replication efficiency in vitro. Noteworthy, to our knowledge, this specific nsp7 deletion does not seem to have been previously reported by other investigators. Therefore, this deletion certainly requires further investigation.
Overall, the whole genome diversity of PRRSV type 2 strains in Quebec is quite high, as expected based on ORF5 diversity, with up to 20.43% difference in nucleotide identities between strains. This diversity is higher in regions coding for proteins considered more or less accessory and lower for those considered essential, like the polymerase nsp9 or the helicase nsp10, as well as the structural proteins (Table 2). Interestingly, the most conserved nt encoded protein gene was found to be the serine protease nsp4 (Table 2), which is the main PRRSV protease responsible to cleave most of the replicase polyprotein into it’s smaller functional parts (52).

Type 2 PRRSV WGS classification stays similar for most PRRSV strains compared to ORF5 classification method which includes 9 lineages. Although, few PRRSV strains appeared to be of an uncertain lineage when looking at the whole genome sequences. This could be because WGS may have improve the depth of the genomic analyses compared to ORF5 method (i.e. 15,000 nt based analyses compared to 603 nt based analyses, respectively). Two examples of this are the strains 2108698 and 2137004. While both clustered inside the lineage 8 based on ORF5 phylogenetic tree, they could just as easily be grouped with the reference strain MN30100 of lineage 9 based on whole genome sequencing. Another example of this phenomenon is the reference strain IAF-klop. We know it belongs to lineage 1, but the whole genome seems to cluster more closely to the lineages 5-9. Another possible explanation of this is that this specific strain has been passaged repeatedly in vitro in MARC-145 cell line. Lastly, based on whole genome sequencing, both strains from the 1890896 sample seem to belong to an uncertain lineage. The limited availability of PRRSV whole genome sequences data and the relatively frequent occurrence of viral genome recombination events raise the need to come up with a
 genomic classification system that would take into account the higher variability of some non-
structural proteins and the incidence of viral genome recombination events. Such recombination
events pose a real challenge to the current monophyletic lineage classification system.

A more adequate classification of the strains could be made using both the nsp2 and
ORF5 gene sequences in a manner akin to influenza subtyping(54). The nsp2 coding region
seems to routinely contain deletions which can serve as markers of sub-lineage as well as being
implicated in pathogenicity, making it a good candidate for typing of different strains (45, 46,
49). WGS remains however the method of choice compared to nsp2/ORF5 sequencing for a few
reasons. Firstly, the nsp2 coding-region is over 3000 nt long and highly variable, making it’s
amplification by PCR for Sanger sequencing quite difficult. Secondly, while adding a second
gene from the 5’ end of the genome would greatly improve odds of catching recombinants,
simply doing a second Sanger sequencing instead of whole genome sequencing would not detect
coinfections in swine samples.

The 4.55% occurrence of PRRSV coinfections amongst PRRSV WGS positive cases was
in line with what has been previously reported for swine influenza virus (SIV). In fact, more then
one subtype (such as H1N1 and H3N2) of SIV are found in at least 2% of the tested samples, not
accounting at all for coinfections with strains from the same subtype(41).

The six possible recombinant PRRSV strains (a prevalence of 6.52%) that cluster
differently, based on WGS analyses, are also of high importance. In the field, veterinarian
practitioners and swine producers are mostly interested in knowing if a strain within a PRRSV
positive sample is genetically related to a PRRSV vaccine-like or wild-type strains, and if it is
related to previous infections on that farm or to other known outbreaks. When a PRRSV strain
changes clustering with WGS, due to a possible recombination event, this can have an important
impact on the clinical interpretation as it can change its relatedness to a known high or low pathogenic strain.

From a clinical point of view, the 2072533 strain would have been identified by veterinarian practitioners, as a MLV vaccine-like strain based on the ORF5 gene method whereas in fact, it is not a MLV vaccine-like strain based on WGS. Overall, PRRSV WGS has changed the vaccine-like strain status of one PRRSV strain, i.e. 3.23% of the vaccine-like related cases (data not shown).

While these analyses have been conducted on Quebec PRRSV strains, there is nothing to suggest that these findings could not extend to most if not all regions currently affected by PRRSV type 2 outbreaks as both intra and inter-lineage recombination events have been described and as coinfections can also occur with type 2 PRRSV strains from the same or different lineages. These types of recombination events are known to have occurred elsewhere.

The main limit of our current study is that since Canada is exempt of type 1 PRRSV, we cannot say if these finding also extends to type 1 PRRSV strains. While type 2 PRRSV is the most frequent PRRSV found in both China and the US, type 1 is also known to occur and data on the latest is currently limited(57). It is also currently unknown if recombination between both viral species causing PRRS is possible in a clinical setting. As both species are now becoming endemic to regions where the other one used to predominate, this could be something to investigate in the future. It should be noted however that the lack of type 1 PRRSV strains do not seem to have hindered the Quebec diversity of type 2 PRRSV strains.

To our knowledge, it is the first time that a study has investigated the entire viral genome of several Canadian PRRSV type 2 strains. In addition to allowing the description of 3 substantial deletions in the ORF1a region of many contemporary lineage 1 strains, WGS appears
to enable a better interpretation of clinical cases in about 9.10% of PRRSV cases due to either PRRSV coinfecting strains or the presence of PRRSV recombinant strains. Again, to our knowledge, it is the first time that both the proportion of misclassifications and the incidence of coinfections of type 2 PRRSV strains using whole genome sequencing data have been quantified.

In those cases, this additional information (that more then 1 strain was present in a sample or that the strain is not related to the same strains than those we would have thought with ORF5 sequencing method) could lead directly to more appropriate preventive and palliative measures being applied by the veterinarians and swine producers, which could have a positive impact in the overall health of swine herds around the world.
Acknowledgments

This research was financially supported by the Ministère de l’Agriculture, des Pêcheries et de l’Alimentation du Québec (MAPAQ) Innov’Action program. Carl A. Gagnon was financially supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant and a Canadian Swine Research and Development Cluster (CSRDC) grant. Christian Lalonde was a recipient of a scholarship from the CRIPA, a research network financially supported by the Fonds de recherche du Québec – Nature et technologies (FRQNT). The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


aerosols from experimentally infected pigs. Canadian Journal of Veterinary Research 70:297.


Table 1: List of clinical samples containing more than one PRRSV strain (coinfection) and of strains that are classified differently by WGS compared to ORF5 method.

<table>
<thead>
<tr>
<th>Sample id number</th>
<th>Type of sample</th>
<th>Viruses whole genome sequencing characteristics</th>
<th>Impact on PRRSV classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1898026</td>
<td>Serum</td>
<td>Coinfection&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Both strains clustered in different lineages</td>
</tr>
<tr>
<td>1952821</td>
<td>Pool of Tissues</td>
<td>Coinfection</td>
<td>Both strains clustered differently in the same lineage</td>
</tr>
<tr>
<td>1968868</td>
<td>Serum</td>
<td>Coinfection</td>
<td>Both strains clustered differently in the same lineage</td>
</tr>
<tr>
<td>2153073</td>
<td>Serum</td>
<td>Coinfection</td>
<td>Both strains clustered in different lineages</td>
</tr>
<tr>
<td>1425619</td>
<td>Lung</td>
<td>Recombinant&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Strain clustered differently in the same lineage compared to ORF5 method</td>
</tr>
<tr>
<td>1890826 #2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Serum</td>
<td>Recombinant</td>
<td>Strain clustered in a different lineage compared to ORF5 method</td>
</tr>
<tr>
<td>1927781</td>
<td>Serum</td>
<td>Recombinant</td>
<td>Strain clustered differently in the same lineage compared to ORF5 method</td>
</tr>
<tr>
<td>1952821 #1</td>
<td>Pool of Tissues</td>
<td>Recombinant</td>
<td>Strain clustered in a different lineage compared to ORF5 method</td>
</tr>
<tr>
<td>2072533</td>
<td>Serum</td>
<td>Recombinant</td>
<td>Strain clustered in a different lineage compared to ORF5 method</td>
</tr>
<tr>
<td>2087409</td>
<td>Lung</td>
<td>Recombinant</td>
<td>Strain clustered differently in the same lineage compared to ORF5 method</td>
</tr>
</tbody>
</table>

<sup>a</sup>At least two PRRSV different strains were identified in the same clinical sample and were possessing a nucleotide identity of < 96%.

<sup>b</sup>A PRRSV recombinant status was assigned when a strain clustered more than two nodes away in phylogenetic trees based on PRRSV whole genome analysis compared to ORF5 classification method.

<sup>c</sup>The # after the sample identification (id) number represent one of the two PRRSV strains that was identified within the same clinical sample.
Table 2: Nucleotide conservation status of the main PRRSV protein encoded genes. The lowest inter-strains nucleotide identities between major PRRSV proteins are indicated.

<table>
<thead>
<tr>
<th>Corresponding proteins</th>
<th>Lowest inter-strain identities(^a)</th>
<th>Corresponding proteins</th>
<th>Lowest inter-strain identities</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsp1</td>
<td>76.15%</td>
<td>nsp11</td>
<td>84.01%</td>
</tr>
<tr>
<td>nsp2</td>
<td>66.04%</td>
<td>nsp12</td>
<td>82.47%</td>
</tr>
<tr>
<td>nsp3</td>
<td>80.09%</td>
<td>GP2</td>
<td>84.70%</td>
</tr>
<tr>
<td>nsp4</td>
<td>92.41%</td>
<td>E</td>
<td>86.94%</td>
</tr>
<tr>
<td>nsp5</td>
<td>78.82%</td>
<td>GP3</td>
<td>82.35%</td>
</tr>
<tr>
<td>nsp6</td>
<td>77.08%</td>
<td>GP4</td>
<td>84.73%</td>
</tr>
<tr>
<td>nsp7</td>
<td>64.38%</td>
<td>GP5a</td>
<td>82.05%</td>
</tr>
<tr>
<td>nsp8</td>
<td>82.61%</td>
<td>GP5</td>
<td>82.95%</td>
</tr>
<tr>
<td>nsp9</td>
<td>85.75%</td>
<td>M</td>
<td>86.10%</td>
</tr>
<tr>
<td>nsp10</td>
<td>84.50%</td>
<td>N</td>
<td>86.13%</td>
</tr>
</tbody>
</table>

\(^a\)The highest nucleotide identities obtained between strains were \(\geq 99.8\%\) for all protein encoded genes.
Figure 1: Nucleotide conservation status between Québec PRRSV strains. All 92 PRRSV sequenced strains are depicted here with the main ORFs shown. Three important and recurrent deletions in some of the strains are indicated with an *. The scale to the left of the bar indicates the percentage of strains that shared identity at a given location, with a full blue bar signifying that all strains are identical at this position. The numbers on top of the blue bar indicate the nucleotide position.
**Figure 2: PRRSV whole genome and ORF5 gene nucleotide phylogenetic trees comparison.**

A) PRRSV WGS nucleotide phylogenetic tree and B) PRRSV ORF5 nucleotide phylogenetic tree. The brackets indicate the PRRSV type 2 lineages. MLV, ATP, Fostera, PrimePac and Prevacent corresponds to commercially available live attenuated vaccines. Strains 97-7895, MN30100, VR2385 and MN184 are American reference strains. Strain PA8 is the first Canadian PRRSV whole genome sequenced strain and strain IAF-Klop is a Québec reference strain. The branches for the Lelystad virus strain (type 1 PRRSV reference strain which was used as an outgroup) should be three times longer but were shortened to fit the graph. PRRSV coinfecting strains, i.e. two different strains that were identified in the same clinical samples, are indicated with a #1 or #2. Colored arrows identify potentially PRRSV recombinant strains that are clustering differently when WGS and ORF5 phylogenetic trees are compared. The scale bars represent a 5% difference in nucleotide identity.
Figure 3: Comparison of the PRRSV recombinant strain 2072533 with one of its possible parental strain, the MLV vaccine strain. The red arrow points to the region of the genome where the recombination probably occurred. A full blue line indicates that both strains are identical in that given nucleotide position and gray/white indicate that they differ at that position. The numbers on top of the blue bar indicate the nucleotide position.
Figure 4: Nucleotide conservation status of two PRRSV coinfecting strains found in sample 2153073. Nucleotide identity between the two strains found in sample 2153073. A full blue line indicates that both strains are identical in that given nucleotide position and gray/white indicate that they differ at that position. The nucleotide identity between both PRRSV strains is 81.83%. The numbers on top of the blue bar indicate the nucleotide position.