

Université de Montréal

**Amélioration des services de génomiques et de surveillance du virus du syndrome  
reproducteur et respiratoire porcin**

Présenté par

**Christian Lalonde**

Département de pathologie et microbiologie

Faculté de médecine vétérinaire

Mémoire présenté à la Faculté de médecine vétérinaire  
en vue de l'obtention du grade de *Maîtrise ès sciences* (M. Sc.)  
en sciences vétérinaires, option microbiologie

Avril 2020

© Christian Lalonde, 2020

Université de Montréal

Département de pathologie et microbiologie, Faculté de médecine vétérinaire

*Ce mémoire intitulé*

**Amélioration des services de génomiques et de surveillance du virus du syndrome reproducteur et respiratoire porcine**

*Présenté par*

**Christian Lalonde**

*A été évalué par un jury composé des personnes suivantes*

**Josée Harel**

Présidente-rapporteuse

**Carl A. Gagnon**

Directeur

**Neda Barjesteh**

Membre du jury

## Résumé

Le virus du syndrome reproducteur et respiratoire porcin (VSRRP) est un pathogène important, entraînant des pertes économiques de 130 millions de dollars annuellement au Canada. La surveillance est effectuée par séquençage Sanger du gène ORF5 mais nous croyons que le séquençage du génome entier (SGE) du VSRRP permettrait une meilleure surveillance épidémiologique comparé au séquençage du gène ORF5.

Pour développer une méthode efficace de SGE du VSRRP, 149 échantillons (sérums, poumons, tissus, autres) d'animaux malades ou récoltés pour fin de surveillance ont été analysés. L'ARN viral a été concentré par enrichissement d'ARN à queue poly (A) et le séquençage effectué sur une plateforme Illumina. Le SGE a été efficace dans 67,11% des échantillons, réussissant dans certains échantillons de poumons et de sérums possédant une valeur de quantification (Cq) du virus par RT-qPCR jusqu'à 26,50 et 34.13, respectivement. La méthodologie développée de SGE du VSRRP a été 4650 fois plus sensible que les méthodes décrites précédemment.

Pour quantifier l'impact du SGE, 88 échantillons (dont le SGE a réussi) ont été utilisés pour comparer le SGE au séquençage ORF5. Deux génomes de VSRRP différents ont été trouvés dans quatre échantillons différents (taux de coinfection de 4,55%). Six génomes de VSRRP (6,52% des souches) ont été classés différemment par rapport à la classification ORF5. Ainsi, le SGE du VSRRP a permis une meilleure caractérisation de 9,10% des échantillons VSRRP positifs comparé au séquençage ORF5.

Donc, le SGE du VSRRP est à la fois sensible et plus précis que la classification par l'ORF5.

**Mots-clés** : Virus du syndrome reproducteur et respiratoire porcin, virus, coinfection, recombinant, classification, séquençage de génome entier, porc, MiSeq, séquençage à haut débit

## Abstract

Porcine reproductive and respiratory syndrome virus (PRRSV) is an important pathogen, costing over 130 million dollars annually in Canada. Surveillance is done by Sanger sequencing of the ORF5 gene, but we hypothesized that whole genome sequencing (WGS) of PRRSV genome will allow a better epidemiological monitoring of PRRSV compared to ORF5 gene sequencing.

To develop an efficient method of PRRSV WGS, 149 PRRSV samples (sera, lungs, pool of tissues and others) collected for surveillance or from sick animals were tested. Viral RNA was concentrated using a poly(A) tailed RNA enrichment method, and sequencing was done on an Illumina platform. WGS was successful in 67.11% of cases. WGS was successful in some tissues and lungs samples with RT-qPCR cycle quantification (Cq) values up to 26.50, and in some sera with Cq value up to 34.13. The developed WGS methodology was 4650 times more sensitive for PRRSV WGS than previously described methods.

To quantify the impact of WGS, 88 successful samples for the WGS of PRRSV were used to compare efficiency of WGS and ORF5 sequencing. Two different full-length genomes of PRRSV were found in four of those samples (coinfection rate of 4.55%). Six full-length PRRSV genomes (6.52% of PRRSV strains) were found to cluster differently compared to ORF5 sequencing. WGS of PRRSV also enabled a better classification or characterisation of 9.10% of the PRRSV infected samples compared to ORF5 sequencing.

Thus, WGS can be both sensitive and more accurate than ORF5 classification for the characterisation of PRRSV strains.

**Keywords** : Porcine reproductive and respiratory syndrome virus, virus, coinfection, recombinant, classification, whole genome sequencing, swine, MiSeq, high-throughput sequencing

# Table des matières

Résumé .....	3
Abstract .....	4
Table des matières .....	5
Liste des tableaux .....	9
Liste des figures .....	10
Liste des sigles et abréviations .....	11
Remerciements .....	14
1 - Revue de littérature .....	15
1.1 - Le syndrome reproducteur et respiratoire porcin.....	15
1.1.1 - La découverte du syndrome reproducteur et respiratoire porcin .....	15
1.1.2 - Les manifestations cliniques du SRRP .....	16
1.1.3 - L'agent étiologique du SRRP .....	17
1.1.4 - Le génome du VSRRP .....	20
1.1.5 - Les protéines structurales du VSRRP .....	20
1.1.6 - La réplication du VSRRP.....	22
1.1.7 La pathogenèse du VSRRP .....	22
1.1.8 - L'immunité au VSRRP .....	24
1.1.9 - Les vaccins commerciaux contre le VSRRP .....	25
1.1.10 - La recombinaison du génome virale .....	26
1.1.11 - La variabilité génétique du VSRRP .....	27
1.2.0 - Les méthodes de séquençage de l'acides nucléiques .....	30
1.2.1 - L'histoire du séquençage .....	30
1.2.2 - Le séquençage à haut débit .....	31
1.2.3 - Le séquençage par synthèse .....	32

1.2.4 - Séquençage <i>de novo</i> .....	33
1.2.5 - Séquençage en temps réel d'une seule molécule (SMRTS) .....	33
1.2.6 – Oxford Nanopore.....	34
1.3.0 - La surveillance épidémiologique moléculaire du VSRRP.....	35
1.3.1 - Le séquençage du VSRRP .....	35
1.3.2 - La classification des souches de VSRRP.....	35
1.3.4 - La banque centralisée de séquence Québécoise.....	36
1.3.5 - Le diagnostic moléculaire du VSRRP au Québec .....	36
2 - Problématique, hypothèse et objectifs .....	36
2.1 - Problématique .....	37
2.2 - Hypothèse .....	38
2.3 - Objectifs.....	38
2.3.1 - Objectif 1 .....	38
2.3.2 - Objectif 2 .....	38
2.3.3 - Objectif 3 .....	38
2.3.3.1 - Objectif 3.1 .....	38
2.3.3.2 - Objectif 3.2 .....	38
3 - Premier article.....	39
Porcine reproductive and respiratory syndrome virus whole genome sequencing efficacy with field clinical samples using a poly(A)-tail viral genome purification method.....	40
3.1 - Abstract.....	41
3.2 - Introduction.....	42
3.3 - Materials and methods .....	44
3.3.1 - Swine clinical samples.....	44
3.3.2 - PRRSV RT-qPCR assay .....	45

3.3.3 - Genome extraction and purification.....	45
3.3.4 - PRRSV whole genome sequencing .....	46
3.3.5 - Highly qualified personnel for PRRSV WGS efficiency .....	47
3.3.6 - Bioinformatic analyses.....	48
3.3.7 - Reads dispersion .....	48
3.3.8 - Amount of PRRSV-specific reads .....	49
3.3.9 - Statistical analyses .....	49
3.4 - Results.....	49
3.5 - Discussion .....	52
3.6 - Acknowledgments.....	56
3.7 - Declaration of conflicting interests.....	56
3.8 - Funding .....	57
3.9 - Supplemental data.....	66
4 - Deuxième article .....	69
Whole genome sequencing of porcine reproductive and respiratory syndrome virus (PRRSV) from field clinical samples improves the genomic surveillance of the virus .....	70
4.1 - Abstract.....	71
4.2 - Introduction.....	72
4.3 - Material and Methods .....	75
4.3.1 - Swine samples.....	75
4.3.2 - Genome extraction and purification.....	75
4.3.3 - PRRSV whole genome sequencing. ....	76
4.3.4 - Bioinformatic analyses.....	77
4.3.5 - Accession Numbers .....	78
4.4 - Results.....	79

4.5 - Discussion.....	83
4.6 - Acknowledgments.....	87
4.7 - Supplemental data.....	97
5 - Discussion .....	100
5.1 - Mise au point du protocole .....	100
5.2 - Le type d'échantillons.....	102
5.3 - Diversité génétique .....	103
5.4 - Délétions .....	104
5.5 - Coinfections .....	104
5.6 - Classification.....	105
5.7 - Implication pour l'industrie .....	107
5.8 - Implication pour la recherche .....	108
Références bibliographiques .....	110
Annexes .....	129
Liste des réalisations en cours de Maîtrise.....	129
Bourses et prix:.....	129
Implications sociales et académiques.....	129
Conférences et présentations:.....	129
Publications .....	130
En révision: .....	130



## Liste des tableaux

Tableau 1. –	Protéines structurales du VSRRP .....	19
Tableau 2. –	Vaccins commerciaux contre le VSRRP de type 2 .....	26
Tableau 3. –	PRRSV RT-qPCR positive sample description.....	58
Tableau 4. –	PRRSV samples information and sequencing coverage.....	66
Tableau 5. –	List of clinical samples containing more than one PRRSV strain (coinfection) or from which PRRSV strains were genetically misclassified by the ORF5 method.	88
Tableau 6. –	Nucleotide conservation status of the main PRRSV protein encoded genes. ....	90
Tableau 7. –	PRRSV whole genome sequenced strains description .....	97

## Liste des figures

Figure 1. – Problèmes reproducteurs engendrés par le VSRRP .....	16
Figure 2. – Structure d’un virion du VSRRP par cryo-microscopie électronique.....	18
Figure 3. – Schématisation de la structure d’un virion du VSRRP .....	18
Figure 4. – Organisation génomique du VSRRP.....	19
Figure 5. – Protéines transmembranaires de l’enveloppe du VSRRP .....	21
Figure 6. – Réseau d’ARN sous-génomique du VSRRP.....	22
Figure 7. – Chronologie de l’infection au VSRRP .....	24
Figure 8. – Recombinaison du génome viral.....	27
Figure 9. – Schématisation de la diversité génétique des virus causant le SRRP .....	29
Figure 10. – Schématisation du séquençage de type Sanger. ....	31
Figure 11. – Séquençage par la technologie PacBio (Escalante <i>et al</i> , 2014).....	34
Figure 12. – Number of PRRSV whole genome sequencing positive cases per type of clinical samples. ....	59
Figure 13. – Percentage of PRRSV whole genome sequencing success compared to samples viral load. ....	60
Figure 14. – PRRSV whole genome sequencing success compared to samples viral load. ....	61
Figure 15. – Amount of PRRSV high throughput sequencing specific reads in clinical samples. ....	62
Figure 16. – Highly qualified personnel (HQP) whole genome sequencing efficiency. ....	63
Figure 17. – PRRSV specific reads dispersion throughout the viral genome.....	65
Figure 18. – Nucleotide conservation status between Québec PRRSV strains. ....	91
Figure 19. – PRRSV whole genome and ORF5 gene nucleotide phylogenetic trees comparison. ....	94
Figure 20. – Comparison of the PRRSV recombinant strain 2072533 with one of its possible parental strain, the MLV vaccine strain.....	95
Figure 21. – Nucleotide conservation status of two PRRSV coinfecting strains found in sample 2153073. ....	96

## Liste des sigles et abréviations

ADN : Acide désoxyribonucléique

ADNc : ADN complémentaire

ADNdb : Acide désoxyribonucléique double brin

ARN : Acide ribonucléique

ATP : Vaccin commercial contre le virus du syndrome reproducteur et respiratoire porcin.

BCSQ : Banque centralisée de séquences Québécoises

CD25 : Chaîne alpha du récepteur d'interleukin-2

Cq : Cycle de quantification (quantitation cycle)

ddNTPs : Didésoxynucléotides

dNTPs : Désoxynucléotides

E : Protéine d'enveloppe mineure du VSRRP

FMV : Faculté de médecine vétérinaire de l'université de Montréal

GC : Guanine et cytosine

GP : Glycoprotéine

INF- $\gamma$  : Interféron gamma

IL: Interleukine

M : Protéine transmembranaire majeure du VSRRP

MLV : Modified live vaccine. Vaccin commercial contre le virus du syndrome reproducteur et respiratoire porcin.

MDL : Laboratoire de diagnostic moléculaire (Molecular diagnostic laboratory)

MSD : Maladie mystérieuse du porc (mystery swine disease)

N : Protéine de la nucléocapside du VSRRP

NSP : Protéine non-structural (non-structural protein)

OF : Fluide oral (oral fluid)

ORF : Cadre de lecture ouvert (open reading frame)

PAM : Macrophage alvéolaire pulmonaire (pulmonary alveolar macrophage)

PF : Processing fluid (liquide obtenu lors de différentes manipulations qui sont réalisées chez les animaux, exemple le plus fréquent: liquide obtenu lors de la castration).

PLP : Protéinase de type papaine (papain-like protease)

PoT : Pool de tissu (pool of tissue)

PP : Polyprotéine

PRRS : Porcine reproductive and respiratory syndrome

PRRSV : Porcine reproductive and respiratory syndrome virus

RNA : Ribonucleic acid

RT-qPCR : Transcription inverse-réaction de polymérase en chaîne quantitative (reverse transcription- quantitative polymerase chain reaction)

SRCR : Récepteur éboueur riche en cystéine (Scavenger receptor cystein-rich)

SGE : Séquençage de génome entier

SMRTS : Séquençage en temps réel d'une seule molécule (single molecule real-time sequencing)

SRRP : Syndrome reproducteur et respiratoire porcin

TNF- $\alpha$  : Facteur de nécrose tumorale alpha (tumor-necrosis factor)

VSRRP : Virus du syndrome reproducteur et respiratoire porcin

WGS : Whole genome sequencing

OMZ : Ondes de mode zéro (zero mode wave)

*For the moon of my life, my eldoradita.*

## Remerciements

Il est important de prendre un moment pour remercier l'ensemble des gens qui ont contribué de près et de loin à la réalisation de ce projet. Que ce soit le Dr Gagnon pour l'opportunité, la Dre Provost pour son aide, sa patience, et le partage de ses connaissances, les Drs Abrahamyan, Barjesteh, Beaudry, Boyer, Chorfi, Harel et Thibodeau pour leurs conseils et appui ou mes collègues présents et passé, mesdames Burgher, Franco, Gallardin, Koszegi, Lafrance, Sanchez et Valle, vous avez tous permis, à votre façon, la réalisation de ce projet. J'aimerais aussi remercier l'équipe complète du service de diagnostic, je pense particulièrement à mesdames Allard, Bazin, Beaudry, Bournival, Bousquet, Boyer, Déry, Généreux, Grenier Saint-Sauveur, monsieur Fontaine et Dre Cornaglia.

Je ne peux non plus passer sous silence l'aide et le support colossale qui m'a été offert par l'ensemble du personnel de soutien et administratif de la Faculté de médecine vétérinaire, du GREMIP et du CRIPA, que ce soit mesdames Baribeau, Bilodeau, Bisailon, Blondin, Crost, Dumas, Guilmette et Flibotte qui ont toujours été dévouées à la tâche et ont tellement facilité mon parcours, ni le support moral offert par monsieur Allard et Dre Archambault lorsque les temps étaient moins faciles.

J'aimerais aussi remercier l'immense apport de la division FANI de la faculté, particulièrement monsieur Lacasse qui a toujours été là pour aider nos différents projets *in vivo*, ainsi que tous les professeurs qui ont participé à l'enseignement que j'ai reçu.

Finalement j'aimerais remercier l'Université de Montréal, de même que le MAPAQ, le FRQNT, le CRSNG, le CRIPA et le GREMIP pour leur soutien à notre recherche tout au long de mon passage à la faculté.

# **1 - Revue de littérature**

## **1.1 - Le syndrome reproducteur et respiratoire porcin**

Le syndrome reproducteur et respiratoire porcin est la maladie la plus importante économiquement pour l'industrie porcine mondialement. Elle coûte plus de 150 millions de dollars annuellement à l'industrie porcine canadienne[1] et plus de 663 millions de dollars annuellement à l'industrie porcine américaine[2].

### **1.1.1 - La découverte du syndrome reproducteur et respiratoire porcin**

Le SRRP a été rapporté pour la première fois en 1987 aux États-Unis[3, 4], bien qu'à l'époque la maladie était inconnue et a été nommée Maladie mystérieuse « Mystery Swine Disease » (MSD) dû au fait qu'aucun agent pathogène connu n'y était lié. En 1990, la maladie était rendue endémique partout là où l'élevage porcin était intensif aux États-Unis[5]. Puis, c'est en novembre 1990 que la maladie fut rapportée pour la première fois en Allemagne[6]. Très rapidement, la maladie fut décrite dans d'autres pays Européens. Avec la propagation de la maladie vint une panoplie de noms pour la décrire. Outre le MSD, il y a eu le « Porcine Epidemic abortion and respiratory syndrome », « Swine infertility and respiratory syndrome », « blue-eared pig disease » et « porcine reproductive and respiratory syndrome ». C'est en 1992 lors du symposium international sur la maladie que le nom de « porcine reproductive and respiratory syndrome » (d'où provient le nom francophone québécois SRRP) fut choisi comme nom officiel de la maladie[7].



Figure 1 Adapté de Karniychuk (2012)[8].

Figure 1. – Problèmes reproducteurs engendrés par le VSRRP

Exemple des différents problèmes reproducteurs associés aux VSRRP. Certains des porcelets sont morts au stade fœtal et ont été momifié *in-utero* alors que d'autres sont mort-nés. Les porcelets survivants peuvent être chétifs et faible à la naissance comme sur la photo.

### 1.1.2 - Les manifestations cliniques du SRRP

Comme son nom le suggère, le SRRP peut engendrer à la fois des symptômes reproducteurs et des symptômes respiratoires. Toutefois, il est important de mentionner que beaucoup de cas sont sous-cliniques, ce qui rend le contrôle de la maladie plus difficile et problématique[9]. Il a de plus été démontré qu'une infection sous-clinique par l'agent étiologique du SRRP peut tout de même générer un retard de croissance chez les porcs atteints[10].

Parmi les problèmes reproducteurs liés à la maladie, on recense des avortements tardifs, une augmentation de la fréquence de porcelets mort-nés, la momification des fœtus et des porcelets nouveau-nés plus faible et chétif[11-14] (voir Figure 1). Les problèmes respiratoires liés à la maladie peuvent se retrouver chez les porcs de tous âges et tendent à causer des retards de croissance qui sont principalement dû à une pneumonie interstitielle[13-18] et une susceptibilité accrue à d'autres pathogènes[19-21]. Aussi, comme l'un de ses nom l'indique, (blue-eared pig disease), il peut mener à une coloration bleu ou mauve des oreilles chez les porc atteints, cependant cette



manifestation clinique est beaucoup plus rare de nos jours [13] ou encore se manifeste lors d'une infection avec une souche hautement pathogène de son agent étiologique[22].

### 1.1.3 - L'agent étiologique du SRRP

C'est en juin 1991 au Pays-Bas que l'agent étiologique du SRRP fut découvert, alors qu'un nouveau virus fut isolé d'un grand nombre de porcelets et de truies atteints du SRRP[23]. Ce nouveau virus fut nommé le virus de Lelystad, (qu'on appelle maintenant la souche de Lelystad) d'après le nom de la ville où il a été identifié. Peu de temps après, le virus fut aussi isolé en Allemagne[24] et aux États-Unis[16]. Les souches apparentées au virus isolé premièrement en Europe furent étiquetées de « type Européen » puis de « type 1 » alors que les souches apparentées au virus isolé aux États-Unis furent étiquetées « type Nord-Américain » ou « type 2 ».

En 2015, le comité international sur la taxonomie des virus a finalement reconnu les deux types de virus du syndrome reproducteur et respiratoire porcin (VSRRP) comme des espèces distinctes. Ils ont formellement été renommé *Betaarterivirus suis 1* et *Betaarterivirus suis 2* en 2018[25], bien que le nom de VSRRP soit encore celui d'usage dans la littérature.

Les deux virus du syndrome reproducteur et respiratoire porcin (VSRRP) sont des virus de l'ordre viral *Nidovirales*, de la famille *Arteriviridae* et du genre *Betaarterivirus*. Ce sont des petits virus enveloppés d'un diamètre de 50 à 65 nm[26, 27] (Figure 2), avec des génomes d'ARN monocaténaire positif d'environ 15000 bases, coiffés à l'extrémité 5' et polyadénylés à l'extrémité 3'[28, 29]. Son génome contient au moins 12 cadres de lecture ouverts (ORF)[30] (Figure 4). Un ensemble de 7 protéines et glycoprotéines (GP) transmembranaires et une protéine accessoire ornent l'enveloppe qui recouvre la capsid monomérique et hélicoïdale du virion [31-33] (Tableau 1).

Parmi les pathogènes pouvant causer des problèmes lors de coinfections avec le VSRRP, on retrouve entre autres (mais pas exclusivement) le *circovirus* porcin de type 2, agent principal du syndrome de dépérissement post-sevrage[19], la bactérie *streptococcus suis* qui cause des septicémies et des méningites[20] et le virus de l'*influenza* porcin qui cause lui aussi des problèmes respiratoires exacerbés lors d'une co-infection avec le VSRRP[34].

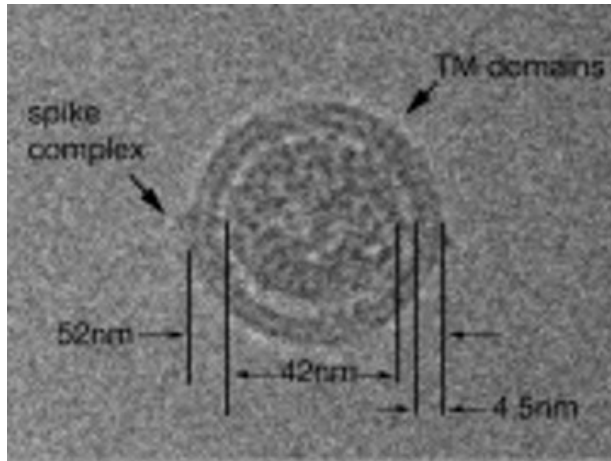


Figure 2 Tiré de Dokland et al., 2010.

Figure 2. – Structure d'un virion du VSRRP par cryo-microscopie électronique.

Particule virale typique de VSRRP. Les complexes d'épines protubérants sont des hétérotrimères des glycoprotéines GP2, GP3 et GP4 alors que les complexes transmembranaires sont des hétérodimères de la glycoprotéine GP5 et de la protéine M. Tiré de Dokland et al., 2010.

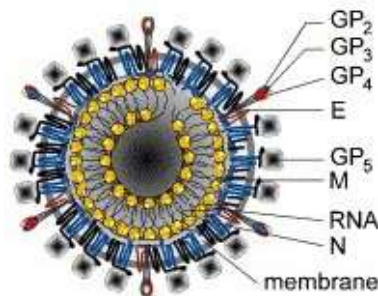


Figure 3 adapté de Gorbalenya et al., 2006.

Figure 3. – Schématisation de la structure d'un virion du VSRRP

Virion de VSRRP. Les protéines GP5 et M forment le complexe membranaire majeur et les glycoprotéines GP2, GP3 et GP4 forment le complexe membranaire secondaire. L'ARN génomique viral est recouvert de la protéine N et enroulé sur lui-même. Adapté de Gorbalenya et al., 2006.

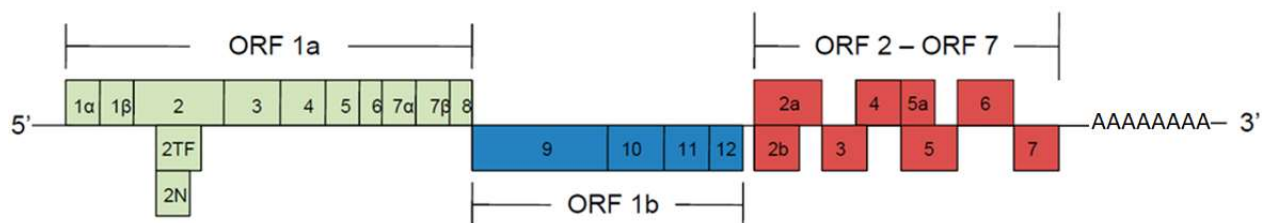


Figure 4 Adapté de Rascòn-Castelo et al., 2015[31].

#### Figure 4. – Organisation génomique du VSRRP

Le génome du VSRRP présente 8 cadres de lectures ouverts (ORFs) de protéines structurales (2a, 2b, 3, 4, 5a, 5, 6 et 7) ainsi que 4 ORFs de protéines non-structurales (1a, 1b, 2TF et 2N). L'ORF1b, de même que 2TF et 2N, sont traduits à la suite d'un changement de cadre de lecture ribosomale. Le génome est polyadénylé en 3' et coiffé en 5'. Adapté de Rascòn-Castelo et al., 2015[31].

Tableau 1. – Protéines structurales du VSRRP

Nom	Taille	Description
GP2	256 aa	Protéine transmembranaire glycosylée, forme un multimère avec GP3 et GP4
E	70 aa	Protéine transmembranaire non-glycosylée, fonction de canal ionique. Forme un homo-multimère.
GP3	254 aa	Protéine transmembranaire glycosylée, forme un multimère avec GP2 et GP4
GP4	178aa	Protéine transmembranaire glycosylée, forme un multimère avec GP2 et GP3
GP5	200 aa	Protéine transmembranaire glycosylée. Forme un dimère avec la protéine M
GP5a	51 aa	Protéine impliquée dans la glycosylation de GP5
M	174 aa	Protéine transmembranaire non-glycosylée. Forme un dimère avec la protéine GP5
N	128 aa	Protéine de la nucléocapside

**aa : acide aminé.**

Tableau 1

### **1.1.4 - Le génome du VSRRP**

Les deux premiers ORFs, les ORF1a et ORF1b, qui totalisent entre 11000 et 12000 bases (environ 75% du génome entier), codent les polyprotéines 1a et 1ab, des ensembles de protéines non-structurales qui sont scindées de façon post-traductionnelle[29]. La polyprotéine 1ab est le résultat d'un changement de cadre de lecture ribosomal[35]. Les ORF1a et 1b codent principalement pour 16 protéines non-structurales que l'on nomme NSP 1 à 12 (avec les protéines NSP 1 et NSP 7 chacune en deux partis; NSP1 $\alpha$ , NSP1 $\beta$ , NSP7 $\alpha$  et NSP7 $\beta$ ). NSP1 $\alpha$ , NSP1 $\beta$  et NSP2 sont des protéinases de cystéine papaïne-like (PLP) et NSP4 est la protéinase de sérine (SP) principale du virus[36]. NSP9 est l'ARN polymérase ARN-dépendante, NSP10 est une hélicase et NSP11 une endoribonucléase[37]. De plus, les protéines NSP1 $\alpha$ , NSP1 $\beta$ , NSP2 et NSP7 possède un potentiel antigénique et seraient impliquées dans l'immunité contre le VSRRP[38]. Aussi, la présence d'un autre site de changement de cadre de lecture ribosomal à l'intérieur de l'ORF1a a récemment été décrite, pouvant donner deux formes alternatives tronquées de la protéine non-structurale NSP2[39]

L'ORF5 du VSRRP mérite une courte présentation à cause de son importance dans l'épidémiologie et la classification actuelle des souches virales du VSRRP. Son importance vient en premier lieu du fait qu'elle code pour la protéine membranaire GP5, qui est la cible principale des anticorps neutralisant contre le VSRRP[40, 41]. De plus, l'ORF5 avait été identifié comme une région hypervariable du virus et était donc un candidat de choix pour le suivi épidémiologique et l'établissement de la phylogénie des souches virales du VSRRP[42, 43].

### **1.1.5 - Les protéines structurales du VSRRP**

Les ORFs 2, 3, 4 et 5 codent respectivement les glycoprotéines membranaires GP2, GP3, GP4, GP5, alors que les ORF2b et M codent les protéines membranaires non-glycosylées E et M (voir Figure 5) [29, 31, 44]. Les protéines GP2, GP3 et GP4, longues de 256, 254 et 178 acides aminés respectivement, forment un trimère qui est essentiel à l'incorporation de E dans la membrane[45]. Elles sont glycosylées en 2, 6 et 4 sites, et ces glycosylation sont essentiels à l'infectiosité du virus[46]. De plus leur absence n'empêche pas la production de particules virales mais celles-ci ne sont pas infectieuses[47]. La protéine E, longue de 70 acides aminés, forme des

homo-oligomères qui constituent un canal ionique membranaire qui aiderait au changement de pH nécessaire à la fusion de la membrane virale à la membrane cellulaire[48]. Les protéines GP5 et M, longues de 200 et 174 acides aminés respectivement, forment le complexe membranaire majeure. Elles forment un dimère par un lieu disulfure entre la cystéine en position 48 de GP5 et la cystéine en position 9 de M[33]. Le dimère GP5-M est essentiel au bourgeonnement des particules virales[49] et est probablement aussi impliqué dans la liaison initiale de la particule virale à la cellule hôte de même qu'à l'internalisation du virus[50]. Le mécanisme précis demeure encore toujours à élucider. La protéine GP5 possède aussi deux sites de glycosylation aux asparagines en positions 46 et 53 et celui à la position 46 est essentiel à la production de particules virales et à l'infectiosité. Un épitope neutralisant, appelé épitope B, est généralement retrouvé aux positions 37 à 45[51]. Il est toutefois suggéré que le virus peut échapper à la neutralisation par cette épitope grâce à la présence de N-glycan, d'anticorps-non neutralisant reconnaissant le même site ou par de léger changement de conformation entre les protéines GP5 et M de l'hétérodimère[52]. L'ORF7 code pour la protéine de la nucléocapside (N). Cette protéine de 128 acides aminés vient recouvrir le génome du VSRRP avant qu'il ne s'enroule sur lui-même. Puis finalement, l'ORF5a code pour une courte protéine de 51 acides aminés (GP5a) qui aiderait à la glycosylation de la protéine GP5[31, 53].

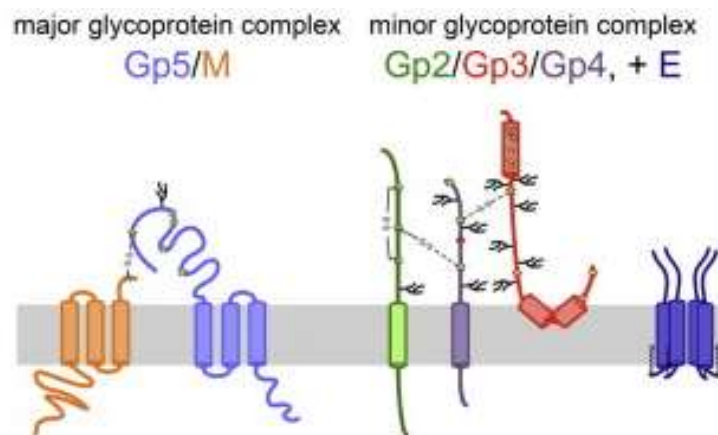


Figure 5 Tiré de Veit et al., 2014.

Figure 5. – Protéines transmembranaires de l'enveloppe du VSRRP

Les protéines GP5 et M forment un dimère par un lieu disulfure entre la cystéine en position 48 de GP5 et la cystéine en position 9 de M. Les glycoprotéines GP2, GP3 et GP4 forment des trimères essentiels à l'infectiosité mais pas à la production de particules virales. Elles possèdent respectivement 2, 6 et 4 sites de glycosylation. Tiré de Veit et al., 2014.

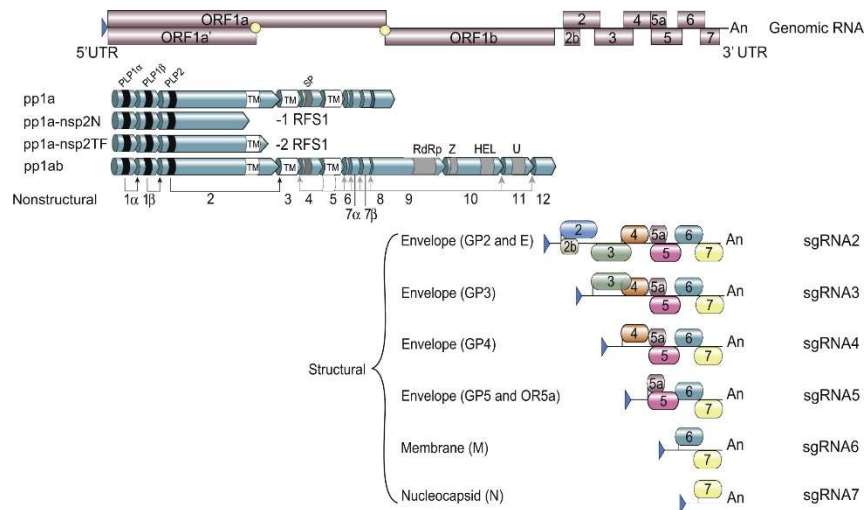


Figure 6 Tiré de Kappes et al, 2015[44].

## Figure 6. – Réseau d’ARN sous-génomique du VSRRP

Quatre protéines distinctes (pp1a, pp1a-nsp2N, pp1a-nsp2TF, pp1ab) sont traduites directement à partir du génome viral et du même codon de départ grâce à deux sites de changement de cadre de lecture (RFS). Les 8 protéines structurales sont traduites à l’aide d’un ensemble de 6 ARN sous-génomique (sgRNA 2 à 7), nommé selon l’ORF principal traduit à partir de cet ARN. Les ARNsg2 et ARNsg5 servent aussi à la traduction des ORFs 2b et 5a respectivement.

### 1.1.6 - La réplication du VSRRP

L’une des particularités des nidovirus est leur réplication à l’aide d’un réseau d’ARN sous-génomiques[54] (Figure 6). Après l’entrée du virus dans la cellule, l’ARN génomique sert d’abord à transcrire l’ORF1a et l’ORF1b en 4 polyprotéines qui sont ensuite clivées en 16 protéines non-structurales[55]. Par la suite, les protéines structurales sont exprimés à l’aide de l’ensemble des ARN sous-génomiques. Ceux-ci portent le numéro de l’ORF principal qu’il sert à traduire à leur extrémité 5’. Sur les ARN sous-génomiques, on retrouve en 5’ la coiffe suivi du gène (ou des deux gènes pour l’ARN sous-génomiques 2 et 5) allant être traduit, puis le reste des gènes en 3’ jusqu’à la queue poly(A) (voir Figure 6).

### 1.1.7 La pathogenèse du VSRRP

Le récepteur principal du VSRRP est encore inconnu à ce jour. Toutefois, le glycosaminoglycane sulfate d’héparane, présent à la surface de toutes les cellules, serait impliqué

dans l'adhésion initiale du virus à la membrane[50]. Puis, le récepteur éboueur riche en cystéine (SRCR) CD163 et spécifique aux macrophages permet l'entrée du virus dans la cellule[56]. L'internalisation du virus est aussi facilitée par la sialoadhésine porcine CD169[57]. La phase aiguë de l'infection dure jusqu'à un mois et correspond à la période avec la plus grande concentration virale dans les tissus avec une virémie importante[52]. Cette période post-infection est aussi caractérisée par une réplication abondante du virus dans les cellules cibles : les macrophages alvéolaires pulmonaires (PAM)[58]. Par la suite, on retrouve une phase d'infection chronique persistante[59]. Durant cette phase, le VSRRP persiste dans certains tissus, particulièrement les tissus lymphoïdes, et ce jusqu'à cinq mois post-infection[60, 61]. Les animaux restent infectieux pendant une période variant de 1 mois et demi à 3 mois post-infection[62]. Il est intéressant de noter que du matériel génétique viral reste parfois détectable dans certains tissus sur une période pouvant dépasser 200 jours puis donner jours à un deuxième épisode de virémie après 252 jours[63]. Cela suggère que certains animaux développent une infection persistante qui peut ressurgir plusieurs mois après l'infection initiale. La chronologie de l'infection peut être visualisée à la Figure 7.

Il est intéressant de noter que la pathogenèse varie énormément d'une souche à l'autre. Certaines souches, considérées hautement pathogéniques, induisent plus lentement la réponse en anticorps[64]. L'infection par ces souches est généralement associée à une fièvre plus élevée, un retard de croissance plus prononcé et des lésions pulmonaires plus volumineuses[64]. Par comparaison, certaines souches sont asymptomatiques et le virus se réplique à l'intérieur d'animaux sans l'apparition de signes cliniques quelconque[65]. Ces infections asymptomatiques entraînent toutefois une réponse en anticorps similaire à celle des infections avec les souches plus pathogéniques[65]. Les souches plus virulentes semblent toutefois réduire la production de certaines cytokines impliquées dans la réponse cellulaire[66].

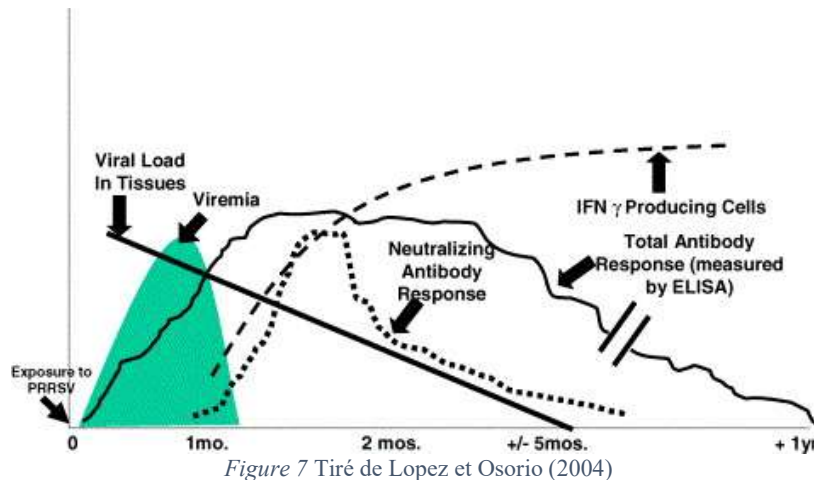


Figure 7. – Chronologie de l'infection au VSRRP

Suite à l'infection au VSRRP, une virémie est détectable pendant environ 5 semaines, avec une charge virale sanguine qui atteint son maximum environ 3 semaines après l'infection. Le virus peut toutefois persister plus de 5 mois dans les tissus. La réponse en anticorps apparaît rapidement mais n'est pas initialement neutralisante. La réponse en anticorps neutralisants et la réponse cellulaire apparaissent environ un mois (au plus tôt 21 jours) post-infection. Tiré de Lopez et Osorio (2004).

### 1.1.8 - L'immunité au VSRRP

L'immunité de l'hôte envers le VSRRP est médiée par plusieurs mécanismes. Une réponse cellulaire médiée par les lymphocytes T est présente de la 2<sup>e</sup> à la 14<sup>e</sup> semaines post-infection[67, 68]. Une réponse en anticorps contre le VSRRP est présente à partir de la 2<sup>e</sup> semaine. Une réponse en anticorps neutralisants apparaît entre la 3<sup>e</sup> et 5<sup>e</sup> semaine post-infection[52].

La réponse immunitaire innée semble grandement affectée par l'infection au VSRRP. Le VSRRP se réplique principalement à l'intérieur de macrophages et peut ainsi mener à une suppression de la réponse immunitaire[69]. Le VSRRP peut aussi induire une réduction de l'activité des lymphocytes cytotoxiques naturels pouvant durer jusqu'à 4 semaines[70]. La protéine NSP2 du virus permet aussi au virus d'évader la réponse dépendante de l'ubiquitine et diminue la production d'IFN $\gamma$ [37, 71, 72].

La réponse cellulaire est dirigée principalement contre les protéines M et GP5[73]. Les principaux médiateurs de l'immunité cellulaire lors de l'infection au VSRRP sont l'IL-1 $\beta$ , l'IL-10, l'IL-12, l'IFN $\alpha$ , l'IFN $\gamma$  et le TNF $\alpha$ [30], avec une importance particulière de l'IFN $\gamma$ [74]. Le VSRRP induit la sécrétion d'IL-10 et du facteur de croissance TGF- $\beta$ , ce qui mène à une



immunosuppression et un retardement de la réponse immunitaire via les lymphocytes Th1[70]. Des lymphocytes T cultivés avec des cellules présentatrices d'antigènes infectées avec une souche virulente de VSRRP présentent une expression significativement moins élevée de CD25, d'IFN- $\gamma$  et de TNF- $\alpha$ [66]. Beaucoup de mécanismes de l'immunité cellulaire au VSRRP de même que l'impact de cette réponse demeurent inconnus à ce jour[30]. De plus, différentes souches de VSRRP semblent induire l'activation de lymphocytes avec des profils différents[30].

La réponse initiale en anticorps ne semble pas protéger l'hôte contre le virus[75]. En fait, un phénomène de facilitation de l'infection par des anticorps (antibody-dependent enhancement) semble être présent lors de la phase initiale de l'infection au VSRRP[76]. Cet effet semble attribuable au fait que la présence d'anticorps non-neutralisant facilite l'internalisation du virus à l'intérieur des macrophages[76]. La réponse en anticorps neutralisant fini toutefois par apparaître. Ceux-ci sont dirigés en premier lieu contre la GP5 et ensuite contre la GP2, GP3 et GP4[40, 77, 78]. Le délai dans l'apparition des anticorps neutralisant serait dû, au moins en partie, à la présence de N-glycan sur la protéine GP5[79]. De plus, la présence d'anticorps neutralisants ne semble pas conférer d'immunité à long terme[52]. Aussi, plusieurs des protéines non-structurales comme NSP1, NSP2 et NSP7 sont aussi immunogènes, mais le rôle des anticorps contre ces protéines reste toujours à être déterminé[80].

La vaccination, bien qu'elle réduise la durée de la virémie chez les animaux, ne semble pas réduire de façon significative la transmission du virus[81].

### **1.1.9 - Les vaccins commerciaux contre le VSRRP**

Il existe présentement plusieurs vaccins commerciaux disponibles pour lutter contre le VSRRP de type 2 (voir Tableau 2). Au Canada, les trois principaux vaccins utilisés sont les vaccins Ingelvac PRRS® ATP (Boehringer Ingelheim Vetmedica), Ingelvac PRRS® MLV (Boehringer Ingelheim Vetmedica), et Fostera PRRS (Zoetis US)[82]. Ce sont des vaccins vivants atténués. Les vaccins ATP et MLV ont une durée rapportée d'immunité rapportée par le fabricant d'au minimum 4 mois, alors que le vaccin Fostera a une durée d'immunité rapportée par le fabricant d'au moins 26 semaines.

Bien que ces vaccins soient généralement considérés comme sécuritaires, la réversion à la pathogénicité est possible[83]. Un exemple est la situation du Danemark à la fin des années 1990,

où des souches virulentes apparentées à une souche vaccinale ont été introduites suite à la vaccination par un vaccin commercial vivant modifié[84]. De plus, bien que les vaccins contre le VSRRP soient très répandus, la protection qu'ils confèrent est limitée aux souches apparentées au vaccin utilisé[85].

Des vaccins autologues tués sont parfois aussi utilisés, mais la protection qu'ils confèrent est de durée limitée, n'offrant plus de protection 6 semaines après la vaccination[86].

Tableau 2. – Vaccins commerciaux contre le VSRRP de type 2

<b>Nom</b>	<b>Fabricant</b>	<b>Lignée génétique selon la classification ORF5.</b>
Ingelvac® PRRS ATP	Boehringer Ingelheim	8
Ingelvac® PRRS MLV	Boehringer Ingelheim	5
Fostera® PRRS	Zoetis	8
Prevacent® PRRS	Elanco US	1
Prime Pac® PRRS	Merck Animal Health USA	7

Tableau 2

### **1.1.10 - La recombinaison du génome virale**

Le VSRRP est un virus capable de faire de la recombinaison génomique[55]. La recombinaison du génome viral est un processus qui se produit lors de la réplication de l'ARN génomique en présence de deux souches distinctes[87]. Lors de la réplication de l'ARN génomique, il est possible pour la polymérase de sauter d'un génome d'une souche virale à l'autre en cours de réplication, donnant ainsi lieu à un génome chimérique composé des deux génomes parentaux (voir Figure 8). Lors d'une infection simultanée par deux souches distinctes de VSRRP, il a été démontré qu'environ 38% des virions produits sont potentiellement des recombinants[88, 89].

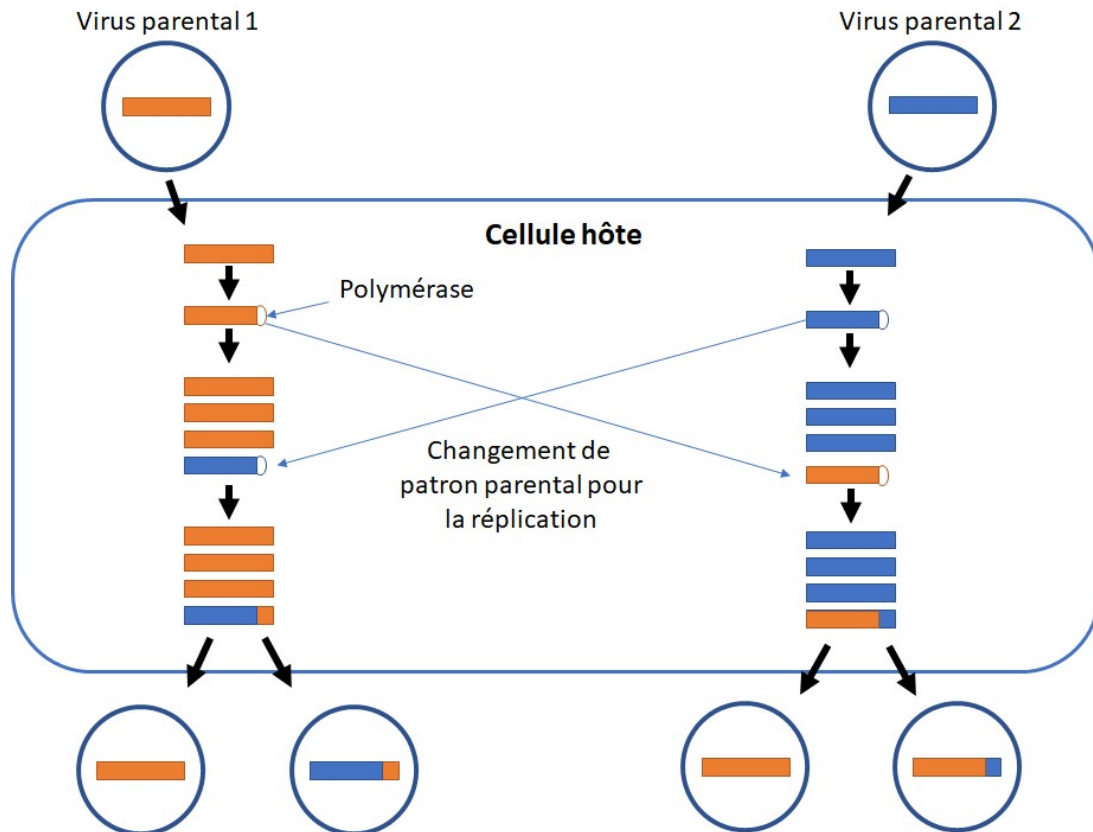


Figure 8

Figure 8. – Recombinaison du génome viral.

Lors de l'infection simultanée d'une cellule par deux souches distinctes, la polymérase peut parfois changer de souches comme patron lors de la répliation. Les virions ainsi produits ont des génomes chimériques des deux souches parentales.

### 1.1.11 - La variabilité génétique du VSRRP

Le VSRRP est l'un des virus possédant le taux de mutations le plus élevé rapporté[90]. Cela est dû tant à la fréquence élevée de recombinaison[88] qu'au manque de fiabilité de l'ARN polymérase ARN dépendante[87]. Cette grande variabilité est perçue comme un atout qui permet au virus d'échapper au système immunitaire de l'hôte[91-93].

L'homologie au niveau du nucléotide entre les deux espèces de VSRRP (*Betaarterivirus suid 1* et *Betaarterivirus suid 2*) est estimée être entre 55% et 79% selon les régions du génome[94, 95]. La figure 9 est une représentation schématique de la diversité des VSRRP.

Le VSRRP de type 1 est divisé en 3 sous-type, avec le sous-type 1 divisé en 12 clades[96]. Les clades sont principalement séparés géographiquement, avec le clade A retrouvé globalement, et étant aussi le clade de VSRRP de type 1 retrouvé aux États-Unis. La différence d'une souche à l'autre de VSRRP de type 1 peut monter jusqu'à 18%[97].

Le VSRRP de type 2 est divisé en 9 lignées monophylétiques distinctes[82, 98]. De ces neuf lignées, 5 sont retrouvées au Canada, les lignées 1, 2, 5, 8 et 9[82]. Aux États Unis, seules les lignées 3 et 4 sont absentes, qui sont retrouvées exclusivement en Asie[96]. Les lignées 1, 3, 4 et 9 sont celles retrouvées en Asie[98]. Les lignées 5 et 8 sont généralement considérées comme des souches apparentées aux vaccins commerciaux disponible (5 pour le vaccin MLV et 8 pour les vaccins ATP et Foster) [43]. Au Québec, les lignées 1, 2 et 5 sont les plus prévalentes[43]. La limite de différence génétique à l'intérieur d'une même lignée a été établie à 11%[82].

Malheureusement, les données sont présentement insuffisantes pour être capables de donner une évaluation juste de la variabilité génétique du génome entier du VSRRP. Toutefois, il est connu que pour l'ORF5 du VSRRP de type 2, il peut y avoir au moins jusqu'à 18% de différence au niveau nucléotidique d'une souche à l'autre[98].

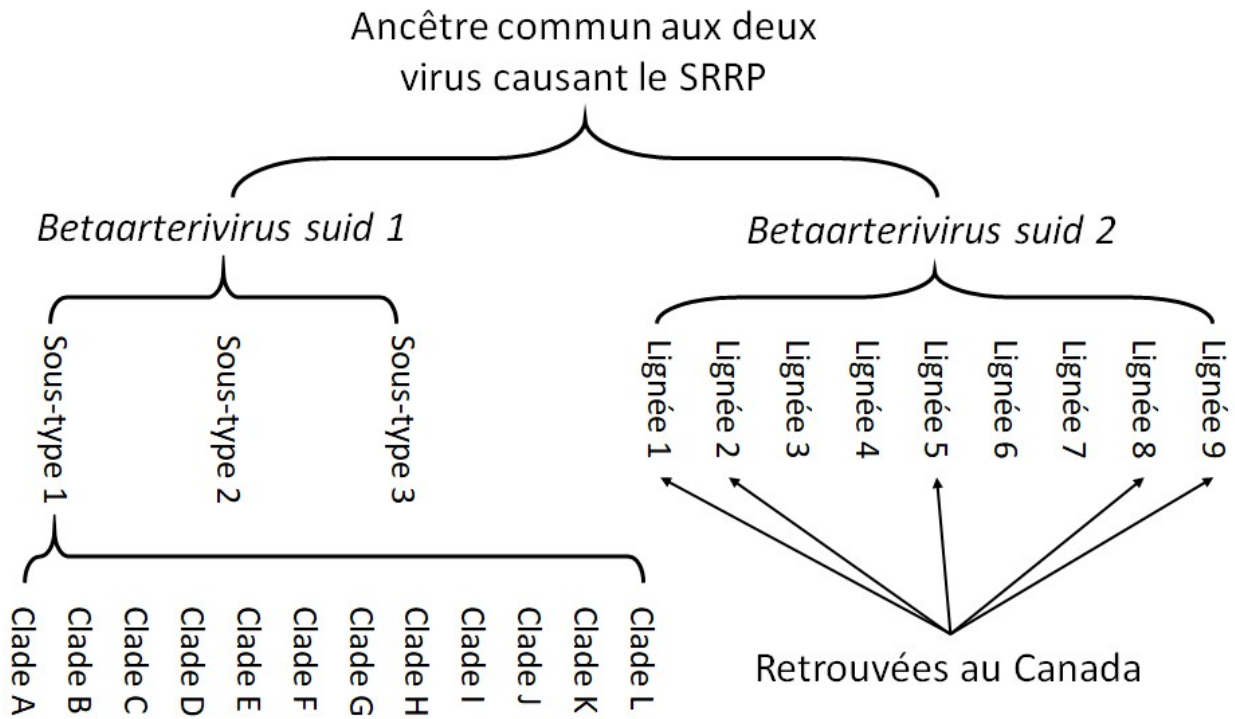


Figure 9

Figure 9. – Schématisation de la diversité génétique des virus causant le SRRP

*Betaarterivirus suid 1* est retrouvé principalement en Europe. Le sous-type 1 est le plus répandu des trois, la clade A étant aussi retrouvé aux États-Unis. *Betaarterivirus suid 2* était originalement limité à l'Amérique du Nord, mais on le retrouve maintenant partout dans le monde. Les lignées 1, 2, 5, 8 et 9 sont retrouvés au Canada.

## 1.2.0 - Les méthodes de séquençage de l'acides nucléiques

### 1.2.1 - L'histoire du séquençage

Le séquençage d'ADN moderne a été développé dans les années 70 par Frederik Sanger[99]. Ses avancées sur le séquençage lui ont valu le prix Nobel de 1980.

Le séquençage de type Sanger est en fait une forme de séquençage qu'on appelle séquençage par terminaison. Pour se faire, une amorce complémentaire à la région 3' du fragment à séquencer est utilisée pour débiter l'élongation à l'aide d'une ADN polymérase. En plus des nucléotides nécessaires à la synthèse du brin d'ADN (dNTPs), des di-désoxynucléotides (ddNTPs) sont aussi ajoutés en petite quantité dans la réaction. Ces ddNTPs n'ont pas l'oxygène nécessaire à l'élongation et ainsi, lorsque l'un de ces ddNTPs est ajouté, l'élongation va se terminer. Une multitude de fragments de différentes longueurs sont ainsi générés avec à leur extrémité 3' un ddNTP marqué soit par un fluorophore ou, anciennement, de la radioactivité. L'utilisation de quatre fluorophores différents (un par nucléotide) permet de réaliser l'ensemble du séquençage dans une seule réaction.

Les fragments sont alors séparés par leur taille et la couleur de fluorescence indique le dernier nucléotide de chaque fragment (voir Figure 10).

Le séquençage de type Sanger est la méthode de séquençage la plus abordable et elle est très efficace lorsque l'on connaît déjà ce que l'on veut séquencer et que l'on cherche à obtenir une séquence de moins de 1000 paires de bases. Ses principales limitations sont donc celles-ci : on ne peut séquencer des fragments dépassant 1000 paires de bases et on doit posséder une amorce correspondante à la région à séquencer. Malgré tout, le séquençage Sanger a été, entre autres choses, utilisé pour séquencer le génome complet de l'homme lors du Human Genome Project, nécessitant plus de 12 années de travail[100]. Par comparaison, les méthodes plus modernes de séquençage de nouvelle génération permettent le séquençage du génome humain complet en moins d'une semaine.

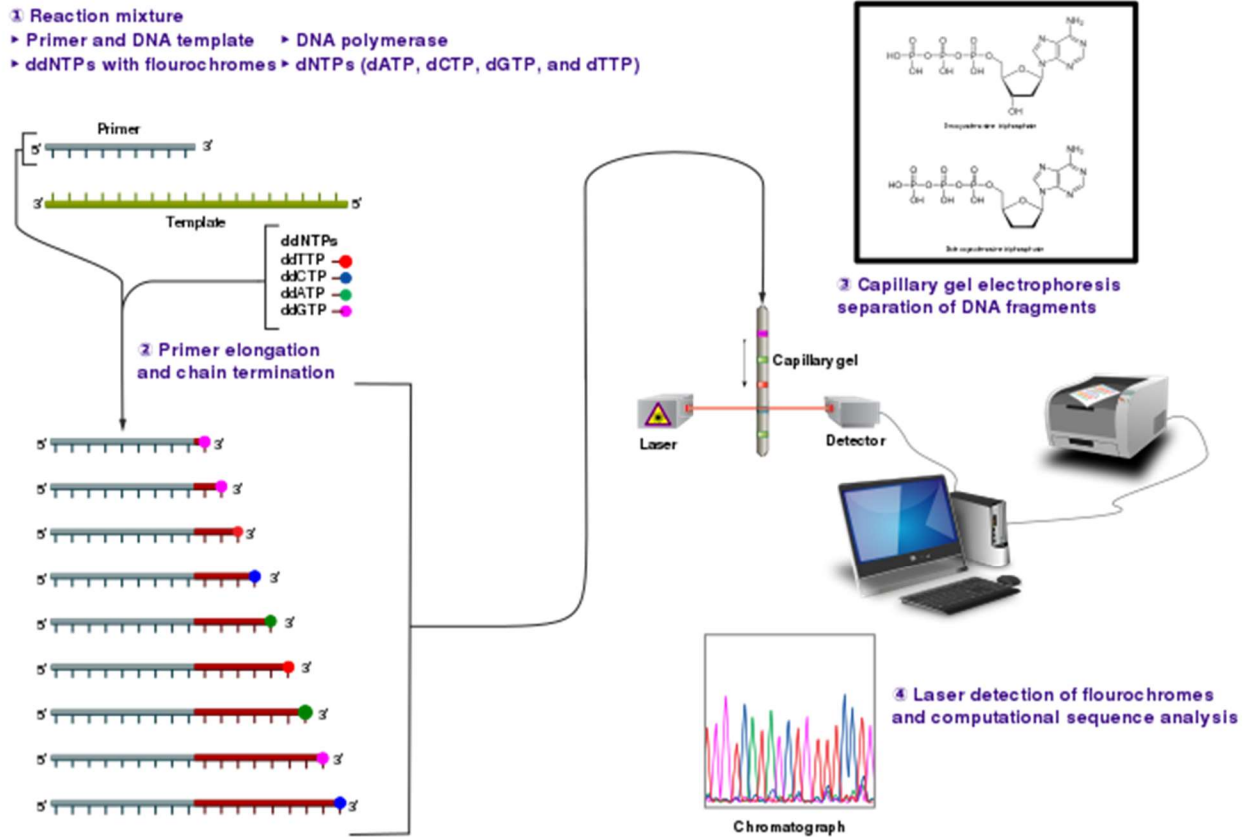


Figure 10 tiré de Wikipedia ([https://en.wikipedia.org/wiki/Sanger\\_sequencing](https://en.wikipedia.org/wiki/Sanger_sequencing)) dernière consultation le 7 mai 2020

Figure 10. – Schématisation du séquençage de type Sanger.

Un ensemble de nucléotides et de nucléotides liés à des fluorophores sont utilisés pour synthétiser le brin complémentaire en suivant le brin patron. Lorsqu'un nucléotide lié à un fluorophore est incorporé dans le brin d'ADN, l'élongation de ce brin s'arrête. La longueur du fragment est déterminée par électrophorèse et la longueur d'onde émise par le fluorophore incorporé indique la base qui a été ajoutée à la position correspondant à la longueur du fragment. Des milliers de fragments de différentes longueurs sont ainsi générés, permettant de reconstituer la séquence du patron utilisé. Wikipedia, dernière consultation le 7 mai 2020 ([https://en.wikipedia.org/wiki/Sanger\\_sequencing](https://en.wikipedia.org/wiki/Sanger_sequencing)).

### 1.2.2 - Le séquençage à haut débit

La grande avancée dans le monde du séquençage est apparue dans les années 90 mais n'a été commercialisé que dans les années 2000s[101]. Le séquençage à haut débit, aussi connu sous

le nom de séquençage de nouvelle génération, est un ensemble de technologies permettant de générer une quantité de données beaucoup plus grande que le séquençage Sanger. La méthode la plus courante, le séquençage par synthèse, regroupe en fait les trois méthodes les plus populaires : le pyroséquençage (maintenant désuet), le séquençage par la technologie Illumina et le séquençage par la technologie Ion-Torrent. Une autre dénomination pour ce type de séquençage est le « massively parallel high-throughput sequencing ».

La grande limitation de la technologie est le coût de l'équipement, un séquenceur coûtant des dizaines de milliers de dollars minimum, parfois des centaines de milliers de dollars. De plus, la quantité de données générées, de même que la complexité des analyses à effectuer, nécessitent souvent une analyse bio-informatique très poussée, nécessitant du même coup du personnel hautement qualifié et encore une fois de l'équipement spécialisé.

### **1.2.3 - Le séquençage par synthèse**

L'efficacité du séquençage par synthèse repose sur le séquençage en parallèle de plusieurs milliers (voir même millions ou milliards selon la technologie et l'équipement utilisé) de fragments d'ADN. À partir de l'échantillon à séquencer, un grand nombre de molécules d'ADN d'une grandeur prédéterminée (100-1000 bases environ, selon la technologie) sont générées, auxquelles des adaptateurs de séquences nucléotidiques connues sont ajoutés afin de permettre la liaison des amorces pour le séquençage. Ces séquences sont ensuite liées sur un substrat solide, comme une cellule d'écoulement, où chaque îlot représente une molécule d'ADN spécifique qui sera séquencée indépendamment des autres. Une polymérase catalysera ensuite l'élongation des brins complémentaires. Pour la technologie Illumina, chaque nucléotide est lié à un terminateur réversible et un fluorophore. Pour Ion torrent, les différents nucléotides sont ajoutés un à la fois, et lorsqu'ils se lient il y a libération d'un atome d'hydrogène (figure 14). Après chaque cycle, une caméra (Illumina) ou un transistor (Ion-Torrent) captent le signal libéré par le nucléotide ajouté. Le terminateur réversible est enlevé et un nouveau nucléotide est ajouté. Le cycle est répété jusqu'à 600 fois. Les séquences sont alors générées en mettant à la suite les différentes informations nucléotidiques pour chacun des îlots. Les données générées doivent alors être reconstituées afin de former une séquence continue (*de novo* sequencing) ou alors elles doivent être alignées sur un génome de référence (resequencing).



L'avantage de la technologie Illumina est qu'elle est relativement peu dispendieuse (coût par base), mais l'équipement est quant à lui très coûteux. La technologie Ion-Torrent est quant à elle plus rapide, ne nécessitant pas de détection optique, et l'équipement est moins dispendieux, mais le coût par base est plus élevé. La technologie de pyroséquençage n'est plus offerte car le coût par nucléotide était beaucoup trop élevé par comparaison à ses compétiteurs.

#### **1.2.4 - Séquençage *de novo***

Le séquençage *de novo* est l'une des applications concrètes du séquençage à haut débit. Son but est de reconstituer une séquence entière à partir des courts fragments obtenus (c.-à-d les lectures ou « reads ») sans connaître au préalable ce que l'on cherche dans l'échantillon. Les lectures sont alignées selon leurs extrémités concordantes afin d'essayer d'obtenir la plus longue séquence possible. Sa principale faiblesse est qu'elle nécessite énormément de ressources informatiques, à un tel point qu'il faut souvent avoir accès à des systèmes spécialisés pour réaliser les analyses.

#### **1.2.5 - Séquençage en temps réel d'une seule molécule (SMRTS)**

Il existe un autre type de séquençage à haut débit qu'on surnomme parfois séquençage de troisième génération. Il s'agit du SMRTS, et comprend les technologies de PacBio et d'Oxford Nanopore.

La technologie de PacBio est une approche un peu différente du séquençage à haut débit. Plutôt que de séquencer un très grand nombre de petite séquence simultanément, la SMRTS vise à séquencer une seule molécule beaucoup plus longue (jusqu'à plusieurs dizaines de kilobases) à la fois.

Cette approche utilise aussi le séquençage par synthèse. Une polymérase synthétise le brin complémentaire à l'aide de nucléotides liés à des fluorophores. Les fluorophores sont libérés au fur et à mesure de l'élongation, et la fluorescence se diffusant est capté en temps réel (Figure 11). La détection d'une quantité si petite de fluorescence est permise grâce à l'application de guides d'ondes de mode zéro (OMZ), qui guident l'énergie lumineuse à l'intérieur de trou plus petit que la longueur des ondes.

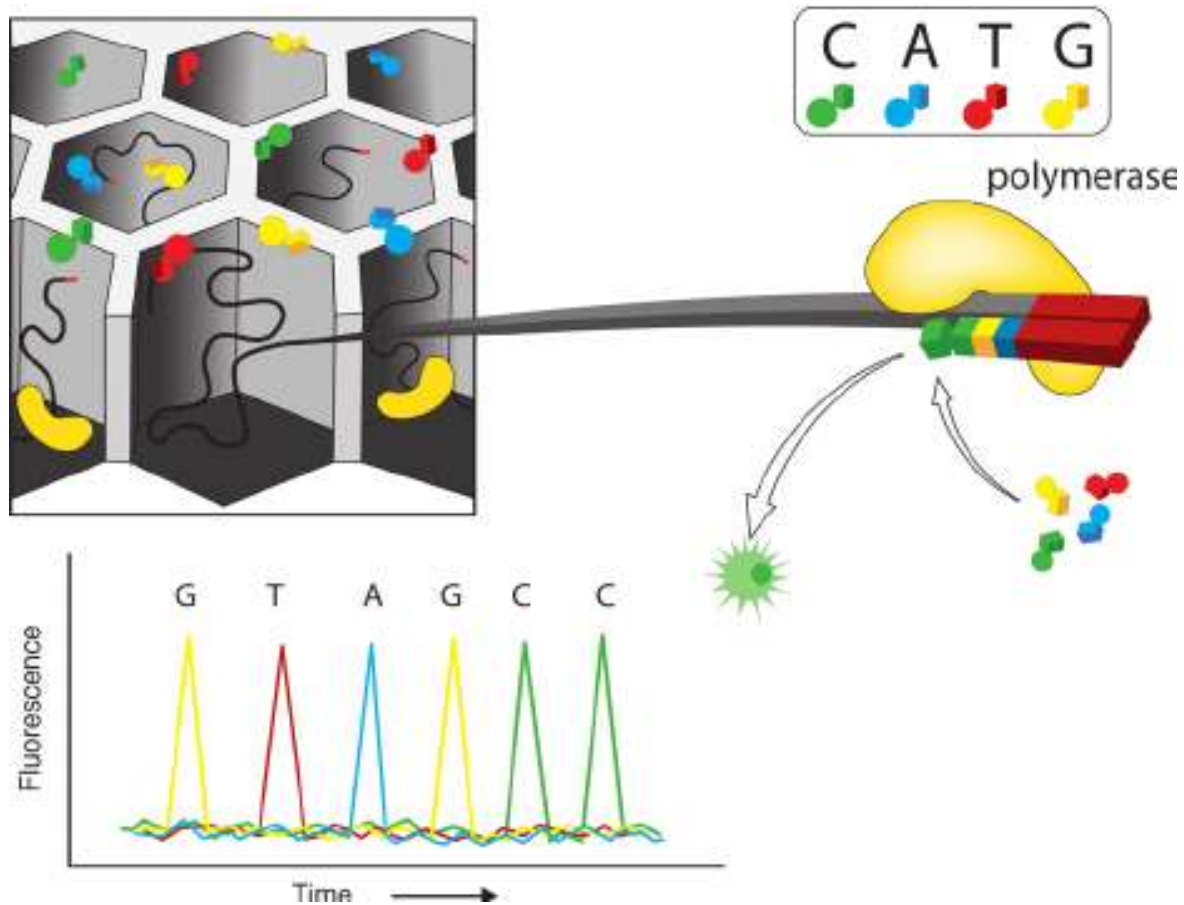


Figure 11 tiré de Escalante *et al*, 2014[102]

Figure 11. – Séquençage par la technologie PacBio (Escalante *et al*, 2014).

L'avantage de ce type de séquençage est bien évidemment la longueur des fragments séquencés (moyenne de plusieurs milliers contre 100-300 pour Illumina et Ion Torrent). L'inconvénient est cependant que le taux d'erreur est beaucoup plus élevé, avec un taux d'erreur variant entre 3 et 13% (comparativement à 0.01 à 0.4% pour les autres technologies de séquençage à haut débit comme Illumina et Ion Torrent)[103]. Tiré de Escalante *et al*, 2014[89].

### 1.2.6 – Oxford Nanopore

La deuxième technologie de type SMRTS est la technologie de Nanopore. C'est une variante du SMRTS qui consiste à faire passer un brin d'ADN ou d'ARN à l'intérieur d'un nanopore de  $10^{-9}$  mètres de diamètre par électrophorèse. Lors du passage au travers du nanopore, la base en question va provoquer un changement dans la densité du courant électrique à la surface du pore. Ce changement, propre pour chacune des 4 bases, est capté par des électrodes. C'est donc la seule

technologie de séquençage qui ne requiert pas de polymérisation ou d'étiquetage chimique de l'échantillon. Il faut toutefois tout de même préparer une librairie de séquençage en laboratoire avant de pouvoir séquencer un échantillon. Cette technologie est aussi offerte en format portatif, étant de la grosseur d'une clé USB[104].

Son inconvénient majeur est, comme pour la technologie PacBio, un taux d'erreur élevé (3-7%)[105]. Toutefois, cette technologie offre l'avantage de ne nécessiter aucun investissement en équipement spécialisé et d'être offerte en format de poche. Elle est aussi la technologie offrant le séquençage des plus longs fragments, pouvant aller jusqu'à quelques centaines de milliers de bases.

### **1.3.0 - La surveillance épidémiologique moléculaire du VSRRP**

Le diagnostic du SRRP se fait souvent par une constatation des signes cliniques par le vétérinaire suivi d'un test de dépistage du virus par RT-qPCR. La région amplifiée lors de la réaction se trouve généralement dans l'ORF6 ou l'ORF7 du génome viral. De par son mode de répllication par réseau d'ARN sous-génomique, l'ORF7 et l'ORF6 sont sur-représentés dans les tissus infectés grâce à leur présence sur la totalité (ORF7) ou la quasi-totalité (ORF6) des ARN génomiques et sous-génomiques, conférant une grande sensibilité aux tests moléculaires basés sur ces séquences[106].

#### **1.3.1 - Le séquençage du VSRRP**

Le séquençage du VSRRP est très fortement limité au séquençage du gène ORF5. Le gène ORF5 a été sélectionné à l'époque car il était considéré comme hypervariable[42]. De plus, il est l'une des cibles principales des anticorps neutralisants contre le VSRRP[40, 41]. Il s'effectue par la méthode de Sanger et est donc relativement abordable. C'est la méthode la plus répandue mondialement, et sur laquelle la très grande majorité de la littérature sur l'épidémiologie et la classification des VSRRP est basée[43, 82, 96, 98].

#### **1.3.2 - La classification des souches de VSRRP**

La classification des VSRRP se fait généralement sur la base de la séquence du gène ORF5 codant pour la protéine GP5[98]. Des deux types de VSRRP reconnus, seul le type 2 est retrouvé au Canada et au Québec[43, 82, 98]. Le type 1 a été recensé une seule fois au Canada chez un

porcelet importé d'Europe se trouvant isolé en quarantaine[107], suite à une mesure de biosécurité due à son importation. Le type 1 est toutefois retrouvé depuis plusieurs années aux États-Unis[108].

Parmi les différentes souches de VSRRP de type 2, on retrouve 9 lignées monophylétiques distinctes[82, 98]. De ces neuf lignées, 5 sont retrouvées au Canada, les lignées 1, 2, 5, 8 et 9[82]. Les lignées 5 et 8 sont généralement considérées comme des souches apparentées aux vaccins commerciaux disponibles (5 pour le vaccin MLV et 8 pour les vaccins ATP et Foster)[43]. Au Québec, les lignées 1, 2 et 5 sont les plus prévalentes[43].

### **1.3.4 - La banque centralisée de séquence Québécoise**

La banque centralisée de séquence Québécoise (BCSQ), dont le laboratoire de diagnostic moléculaire du Service de diagnostic de la Faculté de médecine vétérinaire de l'Université de Montréal est dépositaire, est un registre qui contient l'ensemble de toutes les séquences des gènes ORF5 de VSRRP séquencés au Québec. À ce jour, elle contient plus de 4695 séquences de souches distinctes.

### **1.3.5 - Le diagnostic moléculaire du VSRRP au Québec**

Suite à un cas positif, le vétérinaire peut alors demander le séquençage du gène ORF5. Cette séquence sera ensuite automatiquement ajoutée à la BCSQ, puis comparée aux souches vaccinales disponibles commercialement ainsi qu'à l'ensemble des autres séquences disponibles dans la banque de donnée. Suite à un alignement de séquences, deux informations sont relayées au vétérinaire.

Premièrement, la souche sera classifiée soit, vaccinale, soit apparentée à un vaccin ou soit sauvage. Concrètement, si la séquence partage une identité de 98% ou plus au niveau du nucléotide avec l'une des souches vaccinales commerciales, la souche est considérée comme étant une souche vaccinale. Si elle partage entre 96% et 98% d'identité avec l'une des souches vaccinales, elle est considérée comme apparentée au vaccin. Finalement, si elle partage une identité inférieure à 96% avec les souches vaccinales, elle est considérée comme étant de type sauvage.

Ensuite, la souche est comparée à la totalité de la base de données afin de trouver à quelles autres souches elle s'apparente le plus. Cette information est relayée au vétérinaire qui pourra alors poser l'action appropriée selon la sévérité et la pathogénicité supposée de la souche en question.

- Problématique, hypothèse et objectifs

## 2.1 - Problématique

Considérant la très grande variabilité génétique[90], la présence de virus recombinants[88, 89], et le fait que l'ORF5 du VSRRP ne fait que 603 nucléotides de longueur (environ 4% du génome entier), une très grande quantité d'information est manquante lorsqu'uniquement l'ORF5 du VSRRP est séquencé. Aussi, il est connu que la pathogénicité du virus est multigénique[44], et que d'autres protéines que la GP5 sont fortement impliquées dans l'immunité envers le VSRRP[109].

La recombinaison peut particulièrement mener à une interprétation erronée des résultats car l'ORF5 et le reste du génome pourraient très bien ne pas provenir de la même souche parentale[55]. De plus, il a déjà été démontré pour d'autres pathogènes respiratoires porcins comme le virus influenza que les coinfections simultanées par différentes souches sont possibles[110]. Or, le séquençage de type Sanger ne permet pas la détection ou la différenciation de différentes souches présentes dans le même échantillon clinique.

Le séquençage à haut débit est très peu utilisé en diagnostic pour la production porcine en raison 1) du coût élevé de chaque réaction de séquençage (pour diverses raisons telles que : l'équipement dispendieux qui est nécessaire dans les laboratoires et les besoins d'avoir du personnel hautement qualifié); et 2) de la faible efficacité à obtenir le génome viral entier du VSRRP à partir des échantillons cliniques.

La classification des souches et des cas cliniques de VSRRP pourrait être améliorée de façon significative si le séquençage du génome entier était effectué plutôt que le séquençage Sanger de l'ORF5 uniquement. Cela permettrait de détecter les cas de coinfections, les virus recombinants et permettrait de donner une information plus juste aux vétérinaires et à leurs producteurs dans ces cas. Cela permettrait aussi d'acquérir de l'information sur l'impact pathogénique des autres gènes et régions codantes du VSRRP.

## **2.2 - Hypothèse**

Le séquençage de génome entier de VSRRP permettra une meilleure classification des différentes souches et une caractérisation plus représentative de la réalité pour les échantillons positifs au VSRRP.

## **2.3 - Objectifs**

### **2.3.1 - Objectif 1**

Mettre sur pied une méthodologie de séquençage du génome entier (SGE) efficace, plus abordable et plus rapide que les méthodes présentement employées.

### **2.3.2 - Objectif 2**

Faire le portrait de la diversité génétique des souches de VSRRP au Québec

### **2.3.3 - Objectif 3**

Mettre en évidence les divergences de classification entre la classification basée sur le génome entier et celle basée sur l'ORF5.

#### **2.3.3.1 - Objectif 3.1**

Mettre en évidence les cas de coinfections.

#### **2.3.3.2 - Objectif 3.2**

Mettre en évidence les cas de virus potentiellement recombinant.

### **3 - Premier article**

Cet article dont nous sommes trois auteurs principaux a pour but de démontrer l'efficacité de la méthode de préparation des échantillons cliniques porcins et de séquençage à haut débit que nous avons mise sur pied pour le VSRRP. Il répond directement à l'objectif 1 de mon mémoire et la méthodologie décrite est celle employée dans le second article.

Cet article a été soumis une première fois au Journal of Veterinary Diagnostic Investigation le 20 décembre 2019 puis resoumis avec corrections mineures le 30 mars 2020. Il a été accepté pour publication le 22 juillet 2020.

Contributions individuelles :

Dr Carl A. Gagnon : Conception du projet et de l'expérimentation, interprétation des résultats, écriture du premier jet du manuscrit et révision du manuscrit.

Christian Lalonde : Mise au point du protocole, acquisition, analyse et interprétations des résultats, révision du manuscrit.

Dr Chantale Provost : Conception du projet et mise au point du protocole, acquisition, analyse et interprétation des résultats, révision du manuscrit.

**Porcine reproductive and respiratory syndrome virus whole genome sequencing efficacy with field clinical samples using a poly(A)-tail viral genome purification method**

Journal of veterinary diagnostic investigation

Carl A. Gagnon<sup>1</sup>, Christian Lalonde, Chantale Provost

Swine and Poultry Infectious Diseases Research Center (Gagnon, Lalonde, Provost), and Molecular Diagnostic Laboratory (Gagnon, Provost) of the Service de diagnostic; Faculté de médecine vétérinaire, Université de Montréal, St-Hyacinthe, Québec Canada.

<sup>1</sup>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](mailto:carl.a.gagnon@umontreal.ca) .  
Phone: 450-773-8521 (ext: 8681).

Running head: Whole genome sequencing of PRRS virus.



### 3.1 - Abstract.

The genomic surveillance of porcine reproductive and respiratory syndrome virus (PRRSV) is based on sequencing of the ORF5 gene of the virus, which covers only 4% of the entire viral genome. It is expected that PRRSV whole genome sequencing (WGS) will improve the PRRSV genomic data and allow a better understanding of clinical discrepancies observed in the field when using ORF5 sequencing. Our main objective was to put in place an efficient method for WGS of PRRSV from clinical samples. The viral genome was purified using a poly(A)-tail viral genome purification method and sequenced using Illumina technology. We tested 149 PRRSV positive samples (80 sera, 33 lungs, 33 pools of tissues, 2 oral fluids, and 1 processing fluid (i.e., castration liquid.) Overall, WGS of 67.1% PRRSV positive cases was successful. The viral load, in particular for tissues, had a major impact on the PRRSV WGS success rate. Serum was the most efficient type of sample to conduct PRRSV WGS poly(A)-tail assays, with a success rate of 76.3%, and this result can be explained by improved sequencing reads dispersion matching throughout the entire viral genome. WGS was unsuccessful for all pool of tissues and lungs samples with Cq values > 26.5, whereas it could still be successful with sera at Cq  $\leq$  34.1. Evaluation of results of highly qualified personnel confirmed that laboratory skills could affect PRRSV WGS efficiency. Oral fluid samples seem very promising and need to be further investigated because with only 2 samples of low viral load (Cq = 28.8, 32.8), PRRSV WGS was successful.

**Key words:** animal viral disease, MiSeq, poly(A)-tails, PRRS, swine virus, whole genome sequencing, WGS

### 3.2 - Introduction

Porcine reproductive and respiratory syndrome (PRRS) is an important viral disease causing severe economic losses in North America, with \$663 million and \$150 million per year in the United States and Canada, respectively[111, 112]. Therefore, a lot of effort and resources are allocated to attempt to control and contain the virus. Among those efforts, biosecurity measures, vaccination, and epidemiologic surveillance are the most prominent[113, 114]. The disease has many clinical manifestations, but the 2 most common are reproductive disorders in sows and gilts (characterized by late abortions with an increased number of stillbirths and mummified foetuses, including weak-born piglets)[115-119] and respiratory problems in pigs of all ages characterized by interstitial pneumonia[115, 119-122]. PRRS virus (PRRSV; Nidovirales, Arteriviridae, Betaarterivirus suis) the etiological agent of PRRS[123, 124], is an enveloped virus of 50-65 nm diameter[125]. The PRRSV genome is composed of a positive single-stranded RNA of ~15 kb nucleotides (nt) and it encodes at least 11 open reading frames (ORFs)[126]. Around three-quarters of the genome is composed of the ORF1 at the 5' end of the genome, which encodes for non-structural proteins; the ORFs located at the 3' end (ORF2-7) encode mostly for virion structural proteins[124, 127, 128]. Moreover, its viral genome extremities are capped at the 5' end, and with a polyadenylated tail (poly(A)-tail) at the 3' end[127, 128]. The pathogenicity of PRRSV is multigenic whereas several lineages of PRRSV exist within both PRRSV species (i.e. *Betaarterivirus suis 1* and 2, commonly named PRRSV type 1 or PRRSV-1, and PRRSV type 2 or PRRSV-2, respectively) and consequently, the virulence of the strains is variable (from low to high)[124]. Molecular epidemiologic tools have been developed to help in the control of PRRSV, and to understand the links between a specific PRRSV strain with: 1) the origin of an outbreak, and 2) its genetic proximity with vaccine strains to estimate their potential efficacy[129-131].

The ORF5 gene of PRRSV has been selected for the epidemiologic surveillance of PRRSV strains because: 1) it is hypervariable[129, 132] and 2) it encodes a protein (GP5) that induces the synthesis of neutralizing antibodies in pigs[133, 134]. This GP5 antigenicity property has contributed to the selection of ORF5 for PRRSV surveillance despite the fact that other viral proteins are involved in the neutralizing antibody recognition of the virion[135]. Most reported studies have used ORF5 for the classification of PRRSV[129]. Moreover, ORF5 sequencing is part of the strategy that has been put in place to control PRRSV[130, 131]. As an example, with the agreement of all the swine veterinarians in the province of Québec (Canada), the Molecular diagnostic laboratory (MDL) of the Service de diagnostic; Faculté de médecine vétérinaire (FMV), Université de Montréal (UdeM), is the depository of a databank containing almost all PRRSV ORF5 nt sequences that have been sequenced over the past 10 years. This agreement has set up the rules for data sharing between veterinarians. To date, at least 4,695 PRRSV ORF5 nt sequences are included in this databank. Noteworthy, some researchers have recently reported that 11% of the PRRSV strains were misclassified by the ORF5 sequencing method (i.e. PCR amplification of the ORF5 gene followed by the sequencing of ORF5 PCR amplicons by the Sanger-based sequencing method) compared to PRRSV whole genome sequencing, because of viral genome recombination events and simultaneous coinfection of 2 significantly different PRRSV strains in some swine clinical samples[136]. Thus, the ORF5 genomic methodology that is currently used can cause classification errors and therefore mislead veterinary interventions.

Over the past few years, next-generation sequencing (NGS) has been a very powerful tool for the detection and discovery of viruses in a large diversity of environments, type of samples and hosts[137-140]. In fact, NGS has allowed the discovery and sequencing of the entire viral genome of new swine viruses, such as the atypical porcine pestivirus and porcine circovirus 3[141, 142]. Therefore, there is an urge to improve the present ORF5 molecular epidemiologic tool for the

surveillance of PRRSV through the use of NGS tools. During the last North American PRRS Symposium and Conference of research workers in animal diseases (CRWAD) joint meeting (November 2019, Chicago, IL <https://www.vet.k-state.edu/na-prrs/> and <https://crwad.org/crwad2019/>; last access date: December 17<sup>th</sup> 2019), several speakers reported the whole genome sequencing (WGS) of PRRSV using different technologies, such as MiSeq (Illumina) and MinION (Oxford Nanopore Technologies). Unfortunately, a lack of data was obvious regarding the efficacy of those WGS methods with various swine clinical samples.

Therefore, the goal of our present work was to put in place an efficient method for WGS of PRRSV from different types of swine clinical samples that are routinely submitted for diagnostic and surveillance purposes, allowing its use in a real veterinary diagnostic laboratory environment.

### **3.3 - Materials and methods**

#### **3.3.1 - Swine clinical samples**

Convenience swine clinical samples that tested positive for PRRSV by a reverse transcription quantitative PCR (RT-qPCR) assay were selected for WGS with the goal of obtaining the entire PRRSV viral genome. Those clinical samples were submitted between December 2015 and July 2019 (most of them were collected in 2017 [n=43] and 2018 [n=46]). They were mainly tested by the MDL, FMV, UdeM using a PRRSV RT-qPCR assay 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 into herds. Four clinical samples were confirmed to be PRRSV positive by another veterinary diagnostic laboratory (Biovet). The swine clinical samples originated from different swine herds and types of production systems throughout the province of Quebec, Canada. We included 149 PRRSV RT-qPCR positive clinical samples: 80 sera (from which 48 samples are pooled sera), 33 lung tissues, 2 oral fluids (OF), and 1 processing fluid (PF; i.e., in the present case it was collected

from testes after castration) (Tableau 3). An additional category of clinical samples that were obtained from pathologists conducting carcass macroscopic examinations were submitted as a pool of tissues (PoT) including mainly lungs with several other type of tissues such as lymph nodes, spleen, liver and intestine (Tableau 3). We included 33 PRRSV RT-qPCR positive PoT clinical samples. The PRRSV RT-qPCR assay cycle quantification (Cq) values of all clinical samples were 11.5-34.3 (mean:  $22.6 \pm 4.5$ ). More specifically, the mean Cq values of all specific types of clinical samples were:  $22.4 \pm 3.9$ ,  $22.3 \pm 4.5$ ,  $22.5 \pm 4.6$ ,  $30.8 \pm 2.8$  and 28.9 (only 1 sample) for lung, PoT, sera, OF, and PF, respectively.

### **3.3.2 - PRRSV RT-qPCR assay**

The PRRSV RT-qPCR assay conducted by MDL (American Association of Veterinary Laboratory Diagnosticians [AAVLD] accredited) was an in-house assay (protocol PON-MOL-029). Briefly, PRRSV RNA was isolated from clinical samples (QIAamp cadof Pathogen Mini Kit, Qiagen; KingFisher Flex automated nucleic acid extraction apparatus, ThermoFisher Scientific) as recommended by the manufacturers. Thereafter, 5  $\mu$ L of RNA was used to conduct RT-qPCR assays (Rotor-Gene real-time PCR; Qiagen). The primers and probe sequences of the PRRSV RT-qPCR assay cannot be disclosed because they are the property of MDL. Noteworthy, a log-linear equation to quantify the amount of PRRSV genome copies per mL or gram of clinical samples was established using a standard curve determined with commercially available standards (Tetracore). The equation was established as follow:  $y = 10^{(-0.307x + 10.390)} \times 100$ ;  $y$  = number of viral genome copies,  $x$  = RT-qPCR Cq value, and 100 = RNA elution volume in  $\mu$ L.

### **3.3.3 - Genome extraction and purification**

We ground 100 mg of lung or tissues (Beadbeater; BioSpec Products) in PBS, then centrifuged at full speed for 1 min and used the supernatant for viral extraction. Two hundred  $\mu$ L of sera, OF, and

PF were centrifuged for 5 min at  $10,000 \times g$  and the supernatant was used for viral extraction. Viral RNA was extracted (Quick-RNA Viral Kit; Zymo Research) as described in the company's protocol. Thereafter, RNA was eluted using 50  $\mu\text{L}$  of nuclease-free water (Corning). Total elution volume was used to isolate RNA with poly(A)-tails (NEBNext Poly(A) mRNA Magnetic Isolation Module; New England BioLabs), as described in the company's protocol, and poly(A)-tails RNA was resuspended in 15  $\mu\text{L}$  Tris Buffer (New England BioLabs). Then, first strand cDNA was synthesized (Non-directional Reaction Step Up protocol, NEBNext RNA First Strand Synthesis Module; New England BioLabs), starting with 10  $\mu\text{L}$  of isolated poly(A)-tails RNA. Immediately after first strand cDNA synthesis, the second DNA strand was synthesized (NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module; New England BioLabs), as described by the manufacturer's protocol with a minor modification at the incubation step in the thermocycler for 2 h (instead of 1.5 h) at  $16^\circ\text{C}$ . dsDNA was then purified (AxyPrep Mag PCR Clean-up Kits; Axygen Corning) using 1.8X of beads and 70% ethanol. The purified dsDNA was diluted in 30  $\mu\text{L}$  of 10mM Tris-HCl at pH 8.0 and was stored at  $-20^\circ\text{C}$ .

### **3.3.4 - PRRSV whole genome sequencing**

dsDNA was quantified (Qubit dsDNA HS Assay Kit, Qubit fluorometer; ThermoFisher Scientific). Sequencing libraries (Nextera XT DNA Library Preparation Kit; Illumina) were still performed even when the dsDNA quantification results were lower than the Qubit dsDNA HS Assay Kit threshold of detection (0.2 ng). Briefly, 0.2 to 0.3  $\text{ng}/\mu\text{L}$  of dsDNA was used to construct the sequencing libraries (5  $\mu\text{L}$  total volume was used, even when dsDNA quantification results were below the threshold of detection). Fragmentation and tagmentation was performed as suggested by the manufacturer's protocol. Amplification and indexation were also performed as described by the manufacturer's protocol, except that 14 amplification cycles were run. Sequencing libraries

were then purified (AxyPrep Mag PCR Clean-up Kits; Axygen Corning) as described in the Nextera XT protocol. Library quality was assessed (Agilent High Sensitivity DNA Kit, Bioanalyzer; Agilent). Sequencing libraries were normalized (LNB1 beads; Nextera XT protocol), or with the manual normalization protocol if the concentrations of the libraries were at the lower part of the Bioanalyzer curves. Sequencing libraries were sequenced in a v3 600-cycle cartridge using a MiSeq instrument, and PhiX was included at ~ 1% of the total sequencing libraries as a control to establish the sequencing run efficacy (Illumina, CA, USA). A variable amount of PRRSV sequencing libraries were indexed per MiSeq run to take into account the type of samples being sequenced. In fact, several types of samples obtained from different animal species were processed simultaneously by the high-throughput sequencing laboratory and were subsequently combined, to allow optimal use of the v3 600-cycle cartridges. Those samples may have included: 1) bacteria and 2) viruses contained within various clinical samples (or may have been previously isolated), and 3) PCR amplicons (for different purposes such as 16S microbiome characterization and swine influenza A virus sequencing). Large amounts of MiSeq high-throughput sequencing (HTS) reads data were obtained from each successful PRRSV WGS case (Suppl. Table 1).

### **3.3.5 - Highly qualified personnel for PRRSV WGS efficiency**

During our study, 2 highly qualified personnel (HQP) constructed sequencing libraries from the PRRSV RT-qPCR positive swine clinical samples. A total of 83 and 66 clinical samples were processed by HQP-1 and HQP-2, respectively. HQP-1 processed 19 lungs, 17 PoT, and 44 sera; HQP-2 processed 14 lungs, 16 PoT, and 36 sera. HQP-1 also processed the 2 OF and the only PF clinical sample. All clinical samples that were processed by both HQPs are distinct, meaning that a specific clinical sample was processed by only one HQP, with rare exceptions. Both HQPs followed the same PRRSV WGS protocol described above.

### **3.3.6 - Bioinformatic analyses**

At first, reads were trimmed for adaptors and quality by the MiSeq software during FastQ generation. Using CLC Genomic Workbench software (v.12.0.3, Qiagen), reads from each sample were mapped using Map Reads to Reference application with default settings against a list of PRRSV full-length genomes obtained from GenBank and from the PRRSV nt sequences obtained by our laboratory. Thereafter, all reads from each sample were mapped against the closest PRRSV full-length genome in the list. Consensus sequences were extracted from alignments that had a full-length coverage. Moreover, reads were trimmed again for quality and adaptors in CLC Genomic Workbench software, usually only a few reads, but some needed more trimming. This second step of trimming is highly recommended by Qiagen before doing de novo analysis. Then, de novo analysis was performed using the application “De novo Assemble Metagenome” with 2000 minimum contigs length with the scaffolding settings option. Contigs were used to confirm the PRRSV full-length genome obtained by re-sequencing.

### **3.3.7 - Reads dispersion**

A typical MiSeq Illumina sequencing experiment involves fragmentation of the genome to be sequenced from a clinical sample into millions of molecules. The set of fragments, after different modifications, is referred to as a sequencing library, which is sequenced to produce a set of reads (a read length depends on the protocol and may vary between 100 to 300 bp) (BREDA Genetics srl. Sequencing library: what is it? 2016, <https://bredagenetics.com/sequencing-library-cosa-e/>; Last access date: December 17, 2019). To calculate the reads dispersion throughout the entire viral genome, a critical step needed to ensure the efficiency of PRRSV WGS, the viral genome was divided into 2 sections: the 5' region of the viral genome containing the ORF1 gene and that represents ~ 75% of the entire viral genome; and the 3' region of the viral genome containing all other genes (ORF2 to 7). The number of reads targeting each genomic region was determined,



including the overlapping reads that were included in both genomic regions. The ratio of reads was calculated as follow: 5' viral genome-specific reads / 3' viral genome-specific reads. The higher the ratio of reads value is, the higher the reads dispersion throughout the viral genome is expected.

### **3.3.8 - Amount of PRRSV-specific reads**

The % of PRRSV-specific reads per clinical sample was calculated as follow: (number of PRRSV-specific reads / total reads) X 100%.

### **3.3.9 - Statistical analyses**

All statistical analyses were performed using GraphPad Prism software (v.8.3.0, GraphPad Prism Software). Different types of analyses were performed including: parametric 2-tailed unpaired t tests; ordinary 1-way ANOVA with Tukey post-tests; and ordinary 2-way ANOVA.

## **3.4 - Results**

Complete PRRSV genome sequences (Tableau 3) were obtained from 100 samples (67.1%) (Figure 12). More interestingly, the WGS sequencing success rate was more efficient with sera (76.3%) compared to lungs (60.6%) and PoT (51.5%) clinical samples (Figure 12). Noteworthy, OF and PF samples were not included in the statistical analyses because of the low number of samples, i.e., 2 OF and 1 PF. Nonetheless, WGS sequences of PRRSV were obtained from both OF samples, whereas the PRRSV entire viral genome was not obtained with PF.

One hypothesis was that WGS success rate is highly dependent on the viral load. To test this hypothesis and to establish the efficiency of the poly(A)-tail method for PRRSV WGS, it was essential to compare the WGS sequencing success rate with PRRSV viral load in each clinical sample. The success of PRRSV WGS from clinical samples was significantly higher with a higher viral load (i.e. lower PRRSV RT-qPCR Cq values; Figure 12) for all types of samples, i.e., lungs, PoT and sera ( $p < 0.0001$ ; 2-way ANOVA). Therefore, as expected, 100% PRRSV WGS positive

results were obtained for all types of clinical samples at the higher viral load, i.e.,  $\leq 15$  Cq, which is the lower PRRSV RT-qPCR Cq value. Moreover, the WGS success rate was also dependent on the type of samples being analyzed ( $p = 0.004$ ; 2-way ANOVA). In fact, PRRSV WGS was more successful from sera, with 75.4% to 72.4% WGS positive results, for Cq values of  $\leq 25$  and  $\leq 35$ , respectively (Figure 13). The type of clinical samples from which WGS was less successful was PoT with 66.7% to 54.8% WGS positive results for Cq values of  $\leq 25$  and  $\leq 35$ , respectively (Figure 13). The overall WGS success % (i.e., at Cq  $\leq 35$ ) for all clinical samples with known Cq values was 65.2% (Figure 13). The viral load of each clinical samples was plotted and separated into 2 groups, i.e., the samples from which WGS positive and negative results were obtained (Figure 13). WGS positive results were obtained with tissues (i.e., lungs and PoT) for up to the lower viral load of 26.5 Cq, whereas for sera, the lower viral load with a WGS positive result was 34.1 Cq (Figure 14). Very interesting results were obtained with OF regarding PRRSV WGS sensitivity. Although only 2 OF samples were tested, with 28.8 and 32.8 Cq values, WGS positive results were obtained for both samples (data not shown), whereas the higher Cq values obtained with lungs and PoT were 26.5 and 26.0, respectively. Unfortunately, the only PF clinical sample being tested (with 28.9 Cq value) was WGS negative (data not shown). The viral loads of WGS positive results were significantly different compared to WGS negative results for 2 types of samples in addition to all the samples (Figure 14). More precisely, the WGS positive and negative cases viral loads in all samples were  $21.6 \pm 4.2$  and  $24.4 \pm 4.5$  ( $p < 0.001$ ); in lungs were  $20.9 \pm 2.9$  and  $24.6 \pm 4.1$  ( $p < 0.01$ ); and in PoT were  $20.7 \pm 3.4$  and  $24.4 \pm 4.6$  ( $p < 0.05$ ), respectively (Figure 14). Interestingly, the viral loads in sera were not significantly different when WGS positive and negative results groups were compared (Figure 14).

The efficiency and sensitivity of WGS of a virus from a clinical sample is highly dependent on the amount of target-specific reads being obtained during a NGS run and on the viral load

contained in the clinical sample. Therefore, we evaluated and compared the % of PRRSV-specific reads obtained from clinical samples to PRRSV viral loads contained within WGS positive and negative samples (Figure 15). Overall, the % of PRRSV-specific reads was higher in sera at all viral loads compared to lungs and PoT clinical samples (Figure 15b, c and d). Noteworthy, the amount of PRRSV-specific reads did not correlate with the PoT sample viral loads (Fig 4C). In addition to the fact the PoT PRRSV-specific reads were very low compared to sera and lung PRRSV-specific reads, PoT sample composition may differ because they contains various amounts of different swine tissues and subsequently, may also have contained various amount of high throughput sequencing inhibitors. This could explain why no correlation was found between the viral loads and PRRSV-specific reads for PoT samples.

The synthesis of sequencing libraries is a critical phase that may fail at different steps along their construction. Therefore, sequencing libraries need HQP to ensure the validity of NGS results. Consequently, we evaluated the HQP efficiency (Figure 16). Before comparing the WGS results between HQPs, it was essential to determine if the clinical sample viral loads were similar between the samples tested by each HQP. Noteworthy, the viral load of the clinical samples tested by HQP-2 was significantly higher than the viral load of clinical samples tested by HQP-1 ( $21.6 \pm 3.9$  versus  $23.4 \pm 4.8$ , respectively) (Figure 15a,  $p = 0.022$ ). Surprisingly, even if HQP-2 had tested samples with lower Cq values, their PRRSV WGS success rate was overall lower compared to HQP-1 ( $p = 0.021$ ; Figure 15b). In fact, the % of PRRSV WGS positive results for all samples, lungs, PoT, and sera were: 78.3% and 53.0%; 79.0 % and 35.7%; 58.8% and 43.8%; 86.4% and 63.9%; for HQP-1 and -2, respectively (Figure 15b). As expected, clinical sample PRRSV Cq values were significantly lower in PRRSV WGS positive cases compared to WGS negative cases for both HQPs (Figure 15c).

In order to obtain the complete viral genome sequence of PRRSV directly from a clinical sample, it is necessary to ensure good dispersion of PRRSV-specific reads throughout the entire viral genome. Therefore, we calculated ORF1/ORF2-7 reads ratios to estimate the reads dispersion throughout the entire viral genome. Surprisingly, almost opposite results were obtained by both HQPs (Figure 16). The reads dispersion was significantly higher in sera compared to lungs clinical samples as expected for HQP-1 ( $p = 0.020$ ;  $1.61 \pm 0.56$  versus  $1.21 \pm 0.54$ , respectively), whereas the reads dispersion seems to be higher in lungs compared to serum clinical samples for HQP-2 ( $1.20 \pm 0.50$  versus  $0.82 \pm 0.87$ , respectively) but both sample groups were not significantly different ( $p = 0.359$ ; Figure 16). When reads dispersion results of both HQPs were combined, no statistical difference was found between serum and lung clinical samples ( $p = 0.459$ ). An example of HQP-1 reads dispersion obtained with a serum and a lung sample is illustrated in Figure 16b.

### **3.5 - Discussion**

Based on the previous reports, it was obvious that NGS of PRRSV genome from clinical samples needed to be improved significantly[143-145]. In fact, in one research report, the metagenomic approach that was proposed to sequence PRRSV genome, which was based on Illumina technology, was unsuccessful with nasal swabs (PRRSV RT-qPCR Cq values of 25-35) and OF (PRRSV RT-qPCR Cq values of 20-29)[143]. The authors suggested that these types of samples were not appropriate because of the low viral load, viral genome degradation and competition for sequencing reads given that several other viral and bacterial genome sequences were found using their metagenomics approach, in particularly in OF. Unfortunately, in the present study, no nasal swab was included in the analyses, therefore no comparison can be made. Noteworthy, WGS of PRRSV was successful with the 2 tested OF samples even if the PRRSV viral load was lower in one sample (PRRSV RT-qPCR Cq values of 28.8 and 32.8) indicating that

the poly(A)-tail approach, to increase the specific selection of PRRSV genome, has a great impact on the sensitivity of the method. The metagenomic method that had been proposed[143] was successful for PRRSV WGS with sera in 38.6% of the tested samples (PRRSV RT-qPCR Cq values of 14-25)[143] ; whereas we achieved a success rate of 76.2%, illustrating the significant improvement that has been brought about by the poly(A)-tail method. Using a similar Illumina metagenomic method[143], with the addition of cDNA synthesis using random primers (hexamers), others have reported the ability to generate PRRSV sequences by WGS from different types of clinical samples, but they were successful only with clinical samples containing higher viral loads compared to our study. [144]. They reported PRRSV WGS to be only successful from sera with PRRSV RT-qPCR  $Cq \leq 23.6$  (whereas in the present study:  $Cq \leq 34.1$ ); only successful from lung tissues with  $Cq \leq 21$  (whereas in the present study:  $Cq \leq 26.5$ ); and only successful from OF with  $Cq \leq 20.6$  (whereas in the present study:  $Cq \leq 32.8$ ), clearly illustrating that the poly(A)-tail method improves the sensitivity of PRRSV WGS. Beside MiSeq Illumina technology, others have reported the use of the Oxford Nanopore MinION direct RNA sequencing to generate PRRSV whole genome sequences[145]. Unfortunately, in that report, only 6 swine clinical samples (all sera) were tested, reducing the statistical significance and impact of their results for field application. Moreover, the sensitivity of the Oxford Nanopore MinION PRRSV WGS with these clinical serum samples was very low[145] compared to the present study. The reported sensitivity with sera of the Oxford Nanopore Technology was  $4.65 \times 10^4$  times lower compared to the MiSeq poly(A)-tail method (Oxford Nanopore Technology sensitivity was  $3.8 \times 10^6$  viral genome copies per sequencing reaction, and PRRSV WGS was unsuccessful, compared to  $8.17 \times 10^1$  viral genome copies per sequencing reaction at 34.1 Cq value with the MiSeq poly(A)-tail method and PRRSV WGS was successful). Other investigators have also reported the low sensitivity of the direct RNA sequencing method for WGS of several RNA viruses (with or without viral poly(A) tail

genomes)[146]. Moreover, some concerns were raised regarding possible high sequence error rates of the direct RNA sequencing method of Oxford Nanopore Technology that can limit the virus strain identification[146]. The low sensitivity of the Oxford Nanopore MinION direct RNA sequencing with swine clinical samples illustrates that further investigations need to be done to improve this method.

Our results demonstrate that PRRSV WGS is dependent on the viral load, a conclusion supported by other reports[143, 144]. Noteworthy, the viral load seems to have a negligible impact on PRRSV WGS success rate with sera (Figure 14). Is it possible that others factors may influence the PRRSV WGS success rate with sera? Our study seems to support this hypothesis. In molecular testing, several PCR inhibiting substances and factors may exist (such as proteases, calcium, iron, gel media, anticoagulants, degradation of the genomic content) in a clinical sample and need to be considered to ensure the quality of a result[147]. The same inhibiting factors may have an impact on WGS success from a specific clinical sample. In fact, PRRSV WGS poly(A)-tail method was unsuccessful with some clinical samples that possess a high viral load (e.g., serum at 17.2 Cq and lung at 15.8 Cq) but overall, the amount of PRRSV high-throughput sequencing reads was also higher in sera compared to lung samples at the same viral load..

It would have been interesting to determine from one individual infected pig, at different times post-infection, which is the most efficient type of sample to conduct PRRSV WGS. Unfortunately, our databank does not contain this type of samples to allow WGS efficiency comparison among samples of one infected individual. Nonetheless, we are confident that the obtained results indicate that serum is a very good sample choice to conduct PRRSV WGS.

Diverse genomes have been found by metagenomics in clinical samples (including host, bacterial, viral and other high-throughput sequencing reads), and they are present in different ratios depending on the type of clinical sample being analyzed[148]. The authors reported that serum is

the sample type with the highest ratio of reads related to viral genome (12%, whereas 50% for host genome) compared to tissue (0.2% whereas > 95% for host genome)[148]. Therefore, it may explain why we found sera to be the most sensitive and appropriate type of sample to conduct PRRSV WGS assays, a finding also supported by 2 previous PRRSV WGS studies[143, 144]. It may also explain why the PRRSV WGS poly(A)-tail method, which allows the specific selection of the viral genome, may be more sensitive compared to MiSeq random primer approaches that will produce, in theory, a significantly much larger quantity of indexed host genome.

The extracted RNA quality may have a major impact on PRRSV WGS efficiency. Unfortunately, the quantity of extracted and purified poly(A)-tail RNA from serum was too low to allow RNA quality assessment. In fact, it is possible to assess the RNA quality with a Bioanalyzer (Agilent) using a specific application (RNA Integrity Number, RIN). Unfortunately, RIN values are not accurate for samples that contain < 25 ng RNA / $\mu$ L[149]. It is well known that the amount of extracted RNA from serum may be very low, i.e. 2-10 ng/ $\mu$ L[150, 151] and therefore, we could not assess accurately the RIN value or RNA quality from serum in our study.

In theory, although PRRSV virions should be free in sera and therefore their viral genomes should not be in a replicative state, the PRRSV viral genome in tissues may replicate in infected cells. During the viral genome replication of arteriviruses, a subset of viral subgenomic mRNA (i.e. poly(A)-tails mRNA) possessing the same 3' end nucleotide sequences of the viral genome are synthesized[152]. Consequently, the presence of viral subgenomic mRNA in infected tissues should increase the ORF2-ORF7 viral gene copies (these ORFs are located at the 3' end of the viral genome) over the ORF1 viral gene copies and therefore reduce the ORF1/ORF2-7 read ratio. This phenomenon should subsequently decrease the high throughput sequencing reads dispersion throughout the PRRSV viral genome from infected tissue in clinical samples. The results obtained

by HQP1 are in accordance with this theory, meaning that a better reads dispersion throughout the PRRSV genome was found with sera compared to lungs samples. Unfortunately, opposite results regarding reads dispersion were obtained by HQP2. This discrepancy may have arisen because the HQP2 PRRSV WGS success rate was lower compared to HQP1 even if HQP2 has tested clinical samples possessing a higher viral load. Beside the fact that PCR inhibiting factors may be present in a clinical sample, HQP skills to conduct high-throughput sequencing laboratory handling are critical to ensure WGS assay success.

Several reports have described PRRSV WGS that were generated from viruses previously isolated in cell culture or from several PCR amplicons covering the entire viral genome[153-156]. Our goal was to put in place a method that can efficiently work directly from clinical samples without virus isolation or PCR amplicon prior steps because some PCR amplicons could be missing, and because mutations and isolation of more specific PRRSV cell culture adapted strains can be promoted[156-158]. The poly(A)-tail viral genome purification method improved significantly the efficiency of PRRSV WGS from clinical sample compared to the methods that have been reported previously[143-145]. Furthermore, serum was the most appropriate type of sample for sequencing the entire viral genome of PRRSV.

### **3.6 - Acknowledgments**

We thank all the veterinarians who have submitted clinical samples to their university diagnostic laboratory; their trust in the FMV diagnostic service is tremendously appreciated.

### **3.7 - Declaration of conflicting interests**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.



### **3.8 - Funding**

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).

Tableau 3. – PRRSV RT-qPCR positive sample description.

	Whole genome sequencing results	
	Success	Failure
PRRSV positive clinical samples tested (n)	100	49
% of PRRSV genome coverage	98-100	0-85
Mean coverage	9.85-49,314.88	0-4.84
% of PRRSV-specific reads	0.06-88.11	0.01-3.59
<u>Type of samples tested:</u>		
Lungs	20	13
Pool of tissues (PoT) <sup>a</sup>	17	16
Processing fluid (PF) <sup>b</sup>	0	1
Oral fluid (OF)	2	0
Sera	61	19

<sup>a</sup>Including mainly lungs with several other type of tissues such as lymph nodes, spleen, liver, intestine.

<sup>b</sup>Liquid collected from testes after castration.

*Tableau 3*

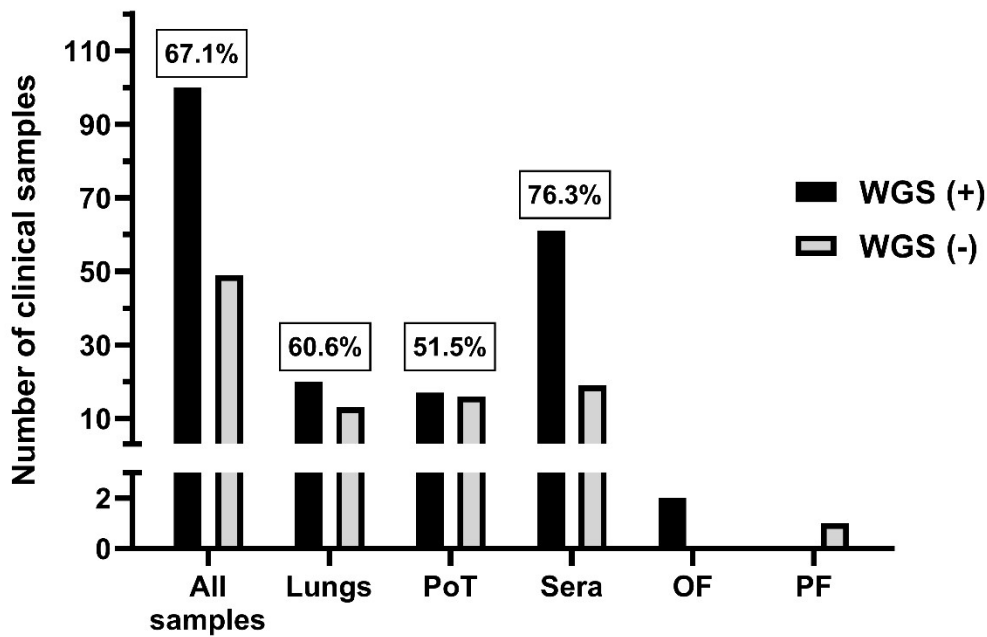


Figure 12

Figure 12. – Number of PRRSV whole genome sequencing (WGS) positive cases per type of clinical sample.

PRRSV WGS was attempted only on PRRSV RT-qPCR positive cases. The “all samples” group includes all types of samples that we tested (i.e., lungs, sera, OF, PF, and PoT). The number in boxes is the % of PRRSV WGS successful cases. OF = oral fluid; PF = processing fluid (i.e. castration liquid); PoT = pool of tissues; WGS (+) = PRRSV WGS success; WGS (-) = PRRSV WGS failure.

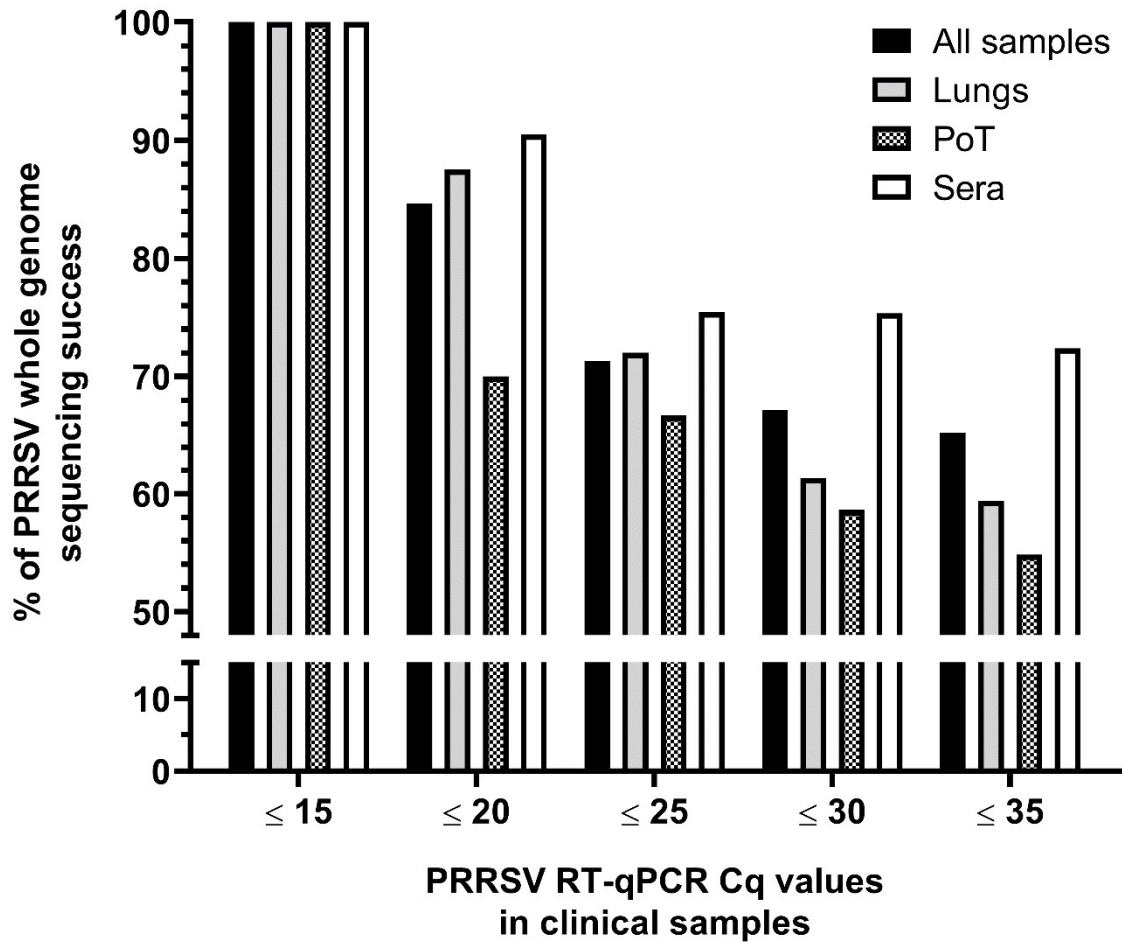


Figure 13

Figure 13. – Percentage of PRRSV whole genome sequencing success (WGS) compared to samples viral load.

The success of PRRSV WGS from clinical samples was significantly higher with a higher viral load. The “all samples” group includes all types of samples that we tested (i.e., lungs, sera, OF, PF, and PoT). See Fig. 1 legend for abbreviations.

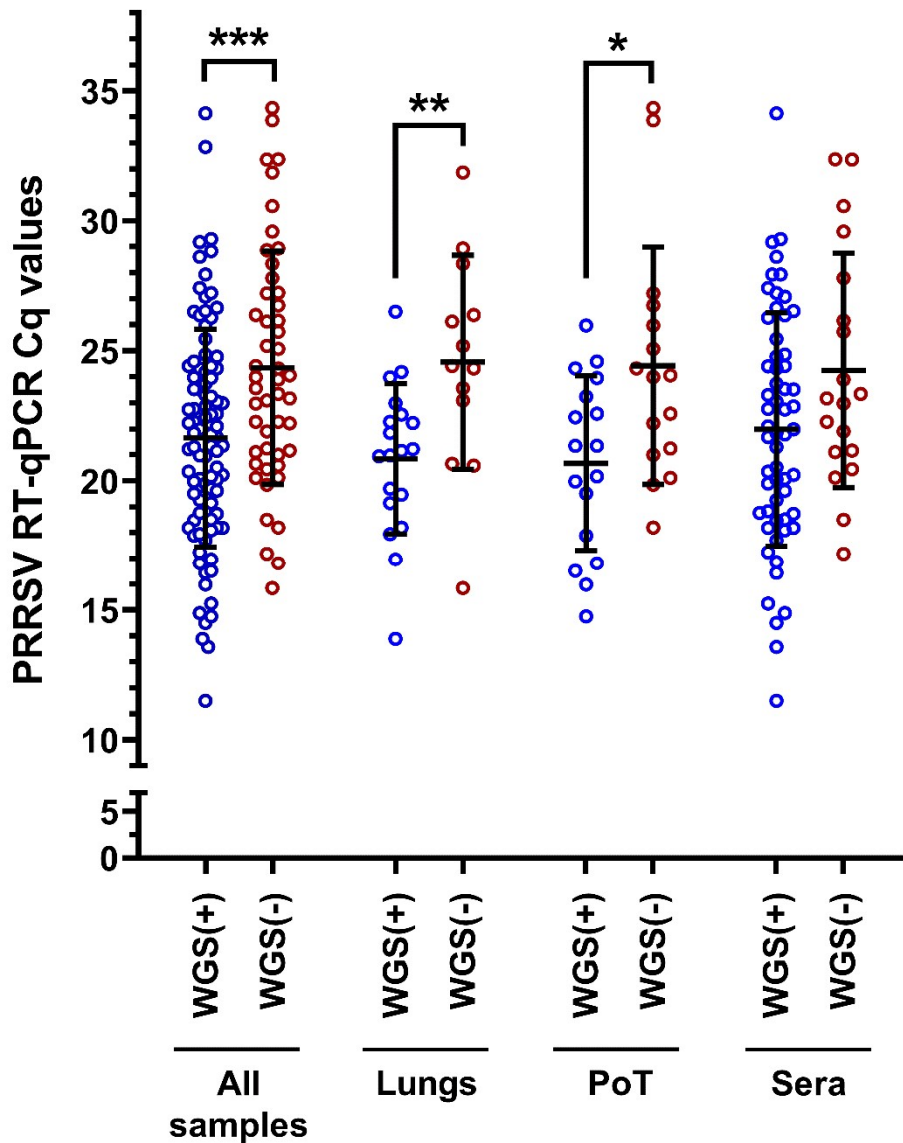


Figure 14

Figure 14. – PRRSV whole genome sequencing success compared to samples viral load.

The “all samples” group includes all types of samples we tested (i.e., lungs, sera, OF, PF, and PoT). The bars and whiskers represent the PRRSV RT-qPCR Cq mean  $\pm$  standard deviation; the empty dots represent individual Cq sample values. WGS (+) = PRRSV WGS successful cases; WGS (-) = PRRSV WGS failure cases. WGS (+) and WGS (-) groups were compared by parametric 2-tailed unpaired t tests. \* =  $p \leq 0.05$ ; \*\* =  $p \leq 0.01$ ; \*\*\* =  $p \leq 0.001$ .

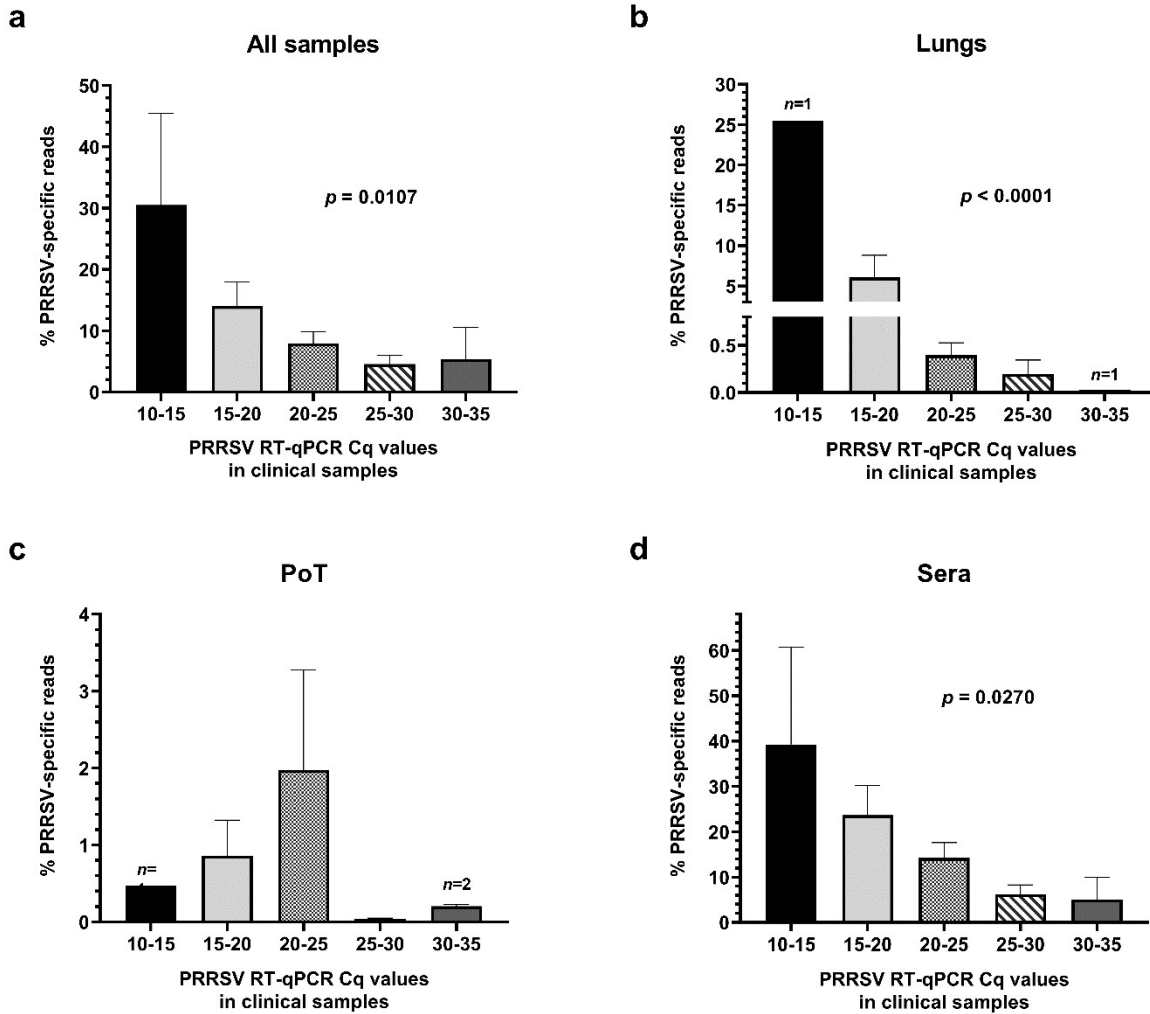


Figure 15

Figure 15. – Amount of PRRSV high-throughput sequencing specific reads in clinical samples. Graphs were built with all clinical samples, i.e. from both WGS successful and unsuccessful samples. The bars and whiskers represent the % of PRRSV-specific reads mean  $\pm$  standard error of the mean. Parametric ordinary one-way ANOVA tests were done to compare the % of PRRSV-specific reads to the clinical sample viral load. **a.** all samples (including PF and OF); **b.** lungs samples; **c.** PoT samples; and **d.** sera. The significant  $p$  values in graphs indicate that the % of PRRSV-specific reads result is dependent on the viral load (i.e., PRRSV RT-qPCR Cq value).

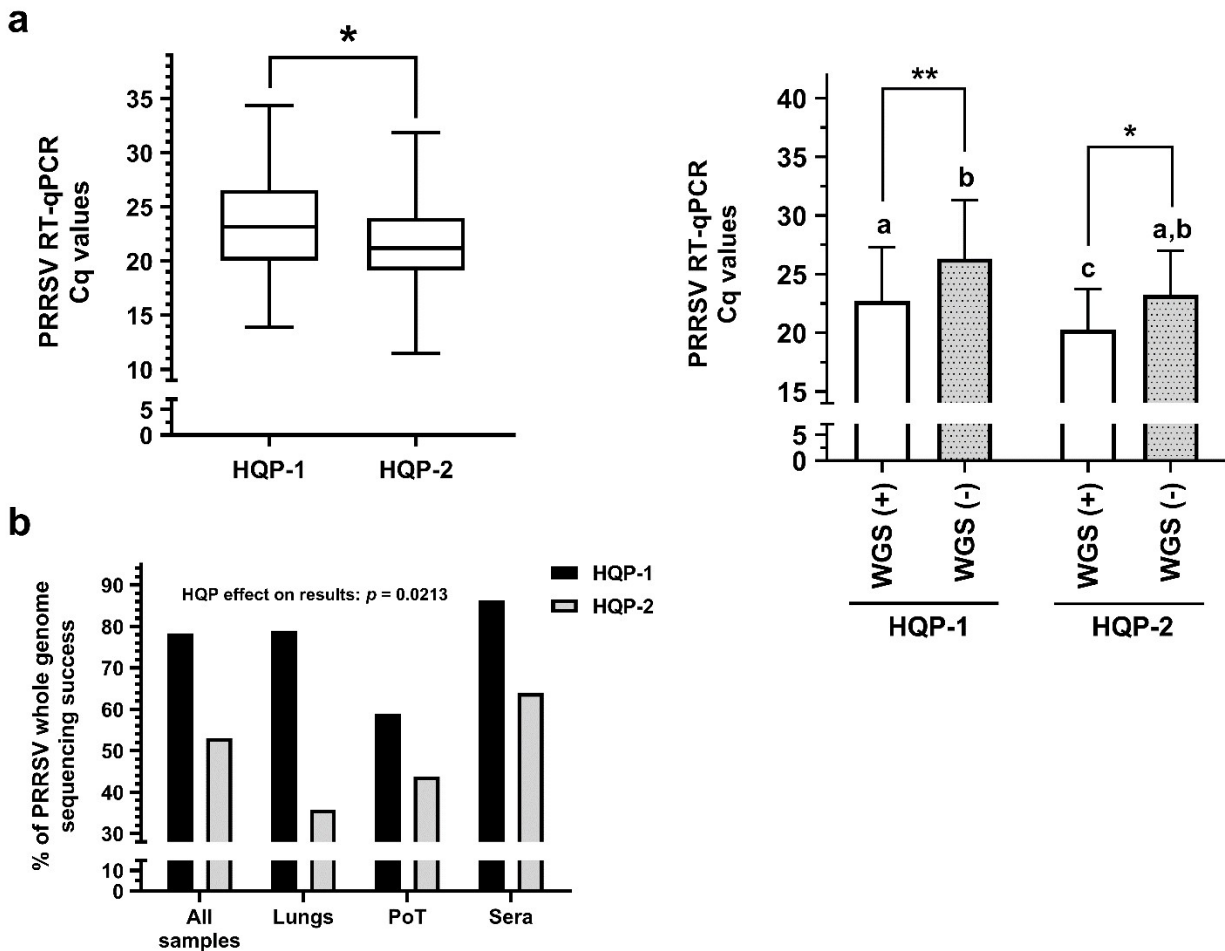


Figure 16

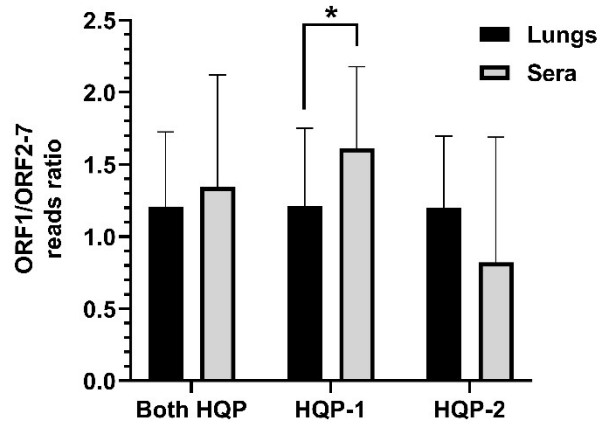
Figure 16. – Highly qualified personnel (HQP) whole genome sequencing efficiency (WGS).

**a.** Viral load of PRRSV positive samples processed by both HQP. The whiskers go down to the smallest PRRSV RT-qPCR Cq values and up to the highest Cq values; the box extends from the 25th to 75th percentiles. The bar within the box represent the median Cq values. Parametric 2-tailed unpaired t test was done to compare HQP-1 and HQP-2. \* =  $p \leq 0.05$ . **b.** Percentage of PRRSV WGS success for different types of samples for each HQP. The HQP effect on the results was determined using an ordinary 2-way ANOVA. **c.** The viral load (Cq values) of PRRSV WGS successful and unsuccessful samples per each HQP was determined. The bars represent the Cq mean  $\pm$  standard deviation. Parametric ordinary one-way ANOVA with Tukey post-tests were done to compare each group. When 2 sets of data are labeled with superscripts of different letters or when only one set is labeled with a superscript, it indicates that these 2 sets of data are statistically

different ( $p \leq 0.05$ ). The  $C_q$  values of PRRSV WGS successful samples were significantly different from WGS failure samples for both HQP (\* =  $p \leq 0.05$ ; \*\* =  $p \leq 0.01$ ).



**a**



**b**

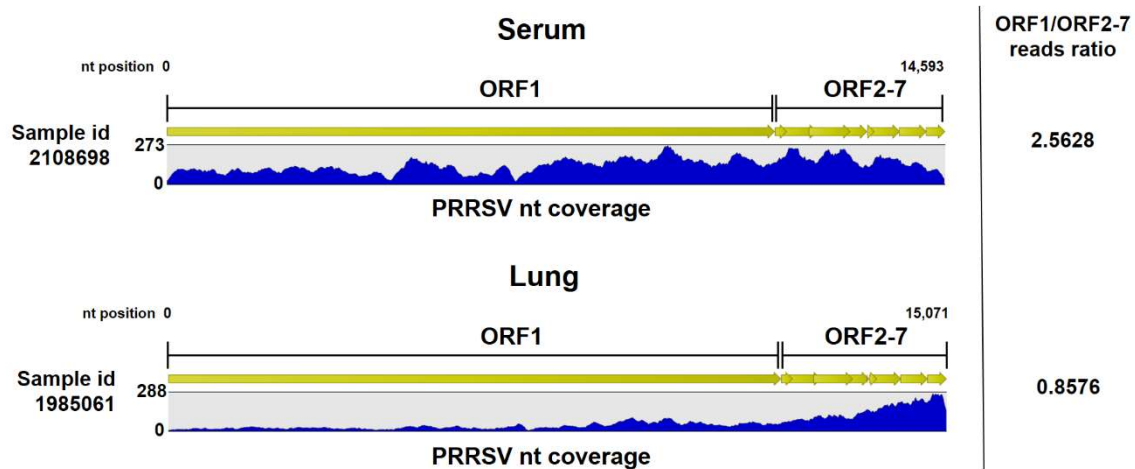


Figure 17

Figure 17. – PRRSV-specific reads dispersion throughout the viral genome.

**a.** ORF1/ORF2-7 reads ratio obtained from WGS (+) successful clinical samples. The bars and whiskers represent the ORF1/ORF2-7 reads ratio mean  $\pm$  standard deviation. Parametric 2-tailed unpaired t tests were done to compare lungs and serum sample groups. \* =  $p \leq 0.05$ . **b.** Graphic representation of PRRSV genome nucleotide (nt) coverage, obtained after WGS of a lung and a serum sample by HQP-1. The nt coverage represent how many times a specific nt was sequenced following reads alignment to the viral genome. Two samples with a similar range of coverage (up to 273 and 288 for the serum and lung samples, respectively) were selected. Green arrows represent the open reading frames (ORF 1 to 7) contained in the viral genome. HQP = highly qualified personnel.

### 3.9 - Supplemental data

Tableau 4. – PRRSV samples information and sequencing coverage

Case identification number	PRRSV RT-qPCR Cq value in the sample	Type of sample	Number of animals pooled in the sample	Total reads	Number of PRRSV mapped reads	% of PRRSV mapped reads	Mean coverage
1844370	n.a.	Serum	1	148210	104619	70.588	1493
1890826	21.98	Serum	1	2073558	1386641	66.873	19809
1926452	18.2	Lung	1	4437900	12489	0.281	143
1927781	23.74	Serum	1	935500	28537	3.050	364
1943176	21.3	Sera	6	283896	7766	2.736	88
1943177	14.5	Sera	6	1738040	1080360	62.160	15185
1943798	14.88	Sera	6	4098788	3611571	88.113	49315
1951167	22.75	Sera	7	3257468	1268970	38.956	17111
1952821	23.95	PoT <sup>b</sup>	1	2529498	1427	0.056	19
1953024	21.68	Sera	7	467026	75971	16.267	839
1953769	22.45	PoT	1	2504292	2381	0.095	36
1956308	20.17	PoT	1	1664089	14931	0.897	204
1957367	23.24	PoT	1	107602	25219	23.437	372
1961450	24.42	Sera	5	1452828	173038	11.910	2561
1961538	19.98	PoT	1	3672398	49274	1.342	679
1966315	20.08	Sera	4	344968	49276	14.284	478
1966316	18.48	Sera	4	1099770	447562	40.696	4632
1968868	34.13	Serum	1	1543552	2504	0.162	35
1968869	27.08	Serum	1	1702286	361199	21.218	4519
1978525	21.79	Sera	8	171332	48987	28.592	482
1980019	22.27	Lung	1	3380328	5649	0.167	83
1982018	21.15	Lung	1	1925924	27669	1.437	381
1982260	20.36	Sera	10	998032	168453	16.879	1866
1982353	18.19	Sera	6	1437798	1147565	79.814	13678
1982354	16.86	Sera	6	1610730	1129865	70.146	10864
1982412	13.89	Lung	1	1888322	481256	25.486	6805
1985061	24.19	Lung	1	2064428	4239	0.205	59
1985928	15.25	Serum	1	35366	10304	29.135	126
1994368	19.7	Lung	1	854464	33184	3.884	300
2012512	17.94	Lung	1	1427924	247213	17.313	3210
2013955	22.99	Lungs	2	1730426	12038	0.696	164
2016194	16.97	Lung	1	641910	98392	15.328	1172

2017769	22.23	Lungs	2	1560084	1615	0.104	21
2018110	21.85	Sera	5	3772466	3280540	86.960	37953
2018114	19.62	Sera	5	95326	42080	44.143	435
2020590	24.59	PoT	1	1464302	3552	0.243	46
2021685	21.22	Lung	1	1985182	14994	0.755	193
2023030	18.73	Sera	3	684202	487383	71.234	5096
2023323	22.55	Lung	1	2860316	8313	0.291	99
2029734	21.85	Lung	1	1536754	2059	0.134	25
2035290	22.86	Sera	5	1515352	113970	7.521	1034
2048380	20.23	Sera	10	1156950	52039	4.498	495
2050445	20.06	Serum	1	2901278	587223	20.240	8348
2050887	23.52	Sera	10	893244	22467	2.515	230
2057423	22.77	Sera	10	1222100	36670	3.001	334
2057984	24.4	Sera	5	175942	40968	23.285	53
2058132	19.51	PoT	1	237606	10389	4.372	137
2059368	18.09	Sera	10	129220	60072	46.488	758
2059830	27.22	Sera	5	92458	1571	1.699	10
2061291	22.58	PoT	2	1271382	5050	0.397	70
2063747	24.77	Sera	2	78076	935	1.198	14
2065929	26.37	Sera	10	529728	8033	1.516	90
2067407	23.06	Sera	5	179552	76.297	0.042	1002
2067408	24.33	Sera	5	1135916	66336	5.840	634
2067423	26.28	Serum	1	129260	3019	2.336	41
2069767	24.33	PoT	2	1124518	22100	1.965	287
2070378	25.98	PoT	2	520846	30243	5.807	432
2072533	26.53	Serum	1	152588	10736	7.036	134
2073509	26.5	Lung	1	1798804	3096	0.172	41
2073979	18.83	Sera	10	2090940	92881	4.442	749
2075305	20.95	Lung	1	723612	13719	1.896	203
2078279	18.19	Sera	5	819850	59698	7.282	552
2078481	n.a.	Serum	1	30710	5132	16.711	58
2082089	19.47	Lung	1	439064	19620	4.469	252
2084371	23.98	Lung	1	2399650	7414	0.309	88
2087409	20.98	Lung	1	1172768	4641	0.396	61
2087619	21.35	PoT	1	2613914	26422	1.011	374
2089141	17.89	PoT	1	6251778	12888	0.206	134
2089260	16.81	PoT	2	4467442	2608	0.058	27
2092399	19.26	Sera	2	1170672	2852	0.244	29
2101209	23.53	Sera	3	1550790	8291	0.535	85
2101687	16.44	Sera	5	693544	78940	11.382	834
2104407	13.57	Serum	1	1933694	9873	0.511	106

2104469	n.a.	Lung	1	3057156	88502	2.895	1071
2108625	19.89	Sera	6	1550226	70296	4.535	655
2108698	29.18	Serum	1	33476	7650	22.852	111
2109640	17.7	Serum	1	794786	9411	1.184	71
2125706	28.83	Oral fluid	Pen <sup>c</sup>	3372740	3943	0.117	55
2132295	27.42	Serum	1	1749770	270942	15.484	4030
2137004	29.3	Serum	1	47964	14201	29.608	204
2137961	28.62	Serum	1	1632382	19284	1.181	271
2146871	24.85	Serum	1	1613492	191318	11.857	2808
2147100	32.84	Oral fluid	Pen	1036475	487981	47.081	4762
2153073	27.93	Serum	1	485839	46014	9.471	443
2154774	11.5	Serum	1	1221852	77485	6.342	556
2158115	18.77	Serum	1	395810	1258	0.318	13
2158116	17.24	Serum	1	1128925	84829	7.514	817
2126306	21.35	PoT	2	2423832	2376	0.098	33
2153250	23.3	Sera	9	1184256	105787	8.933	1492
2172065	22.1	Sera	n.a.	748088	79475	10.624	1138
2164092	19.14	Lung	2	6222062	10670	0.171	160
2190451	26.66	Sera	5	716850	4584	0.639	43
2190452	25.45	Sera	5	659272	9371	1.421	86
2191288	n.a.	Sera	10	353150	136394	38.622	1517
2206200	14.75	PoT	1	3651528	17109	0.469	198
2206201	16.51	PoT	1	2384358	15394	0.646	292
2208392	18.51	Sera	10	614140	44917	7.314	506
2210458	15.99	PoT	1	1788710	16412	0.918	175
2236039	n.a.	Serum	1	217310	60139	27.674	794
2236850	20.52	Serum	1	282080	38646	13.700	403

<sup>a</sup>n.a.: means the data is not available.

<sup>b</sup>PoT or pool of tissues include mainly lungs with several other type of tissues such as lymph nodes, spleen, liver, intestine, etc.

<sup>c</sup>The number of animals per pen during oral fluid sampling is unknown.

*Tableau 4*

## 4 - Deuxième article

Cet article dont je suis le premier auteur a pour but de démontrer l'avantage du séquençage de génome entier de VSRRP par rapport au séquençage Sanger de l'ORF 5. Il répond directement aux objectifs 2 et 3 de mon mémoire.

Il a d'abord été soumis au Journal of Clinical Microbiology le 16 janvier 2020 et suite à une demande de corrections majeures, il a été resoumis le 1<sup>er</sup> juillet 2020 avec les améliorations demandées.

Contributions individuelles :

Christian Lalonde : Mise au point du protocole, acquisition, analyse et interprétations des résultats, écriture du premier jet et révision du manuscrit

Dr Chantale Provost : Conception du projet et mise au point du protocole, acquisition, analyse et interprétation des résultats, révision du manuscrit

Dr Carl A. Gagnon : Conception du projet et révision du manuscrit.

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

Journal of Clinical Microbiology

Christian Lalonde<sup>1</sup>, Chantale Provost<sup>1,2</sup>, Carl A. Gagnon<sup>1,2</sup>

<sup>1</sup>Swine and Poultry Infectious Diseases Research Center (CRIPA) and <sup>2</sup>Molecular 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](mailto: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

## 4.1 - 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 C<sub>q</sub> 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 genome 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 strains 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.

## 4.2 - Introduction

Porcine reproductive and respiratory syndrome (PRRS) is a major economic concern worldwide, costing 663 million dollars yearly to the American swine industry[2] and over 150 million dollars yearly to the Canadian swine industry[1]. 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)[11-14] and retarded growth in piglets, mainly caused by respiratory problems (characterized by interstitial pneumonia)[13-18] and increased animal susceptibility to other pathogens[19-21]. 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 [159, 160].

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[26]. They contain a single stranded positive-sense RNA genome of around 15 kb in length, is capped at the 5' end and polyadenylated at the 3' end and contains at least 11 known open reading frames (ORF)[28, 31]. 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) [28, 161, 162]. 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%[94, 95].



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 were 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 lineages under 89%. Of these 9 lineages, lineages 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 in 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. 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 act 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[109]. Given the high prevalence of PRRSV infections, concomitant infection 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, a 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.

## **4.3 - Material and Methods**

### **4.3.1 - Swine samples.**

Convenience 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 pool of tissues (PoT; mainly lungs with several other type of tissues such as lymph nodes, spleen, liver, intestine, etc.), with PRRSV RT-qPCR diagnostic assay C<sub>q</sub> values between 11.5 and 34.34. The PRRSV viral genome was considered complete if at least 98% of the coding sequence was obtained.

### **4.3.2 - 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 were 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 were 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 were 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 16°C (#E6111; New England BioLabs, ON, Canada). dsDNA were then purified using AxyPrep Mag™ PCR Clean-up Kits (Axygen – Corning, NY, USA) using 1.8X of beads and 70% ethanol. The purified dsDNA were diluted in 30 ul of 10mM Tris-HCl at pH 8.0 and stored at -20°C for later use.

#### **4.3.3 - PRRSV whole genome sequencing.**

dsDNA were 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 library's 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).

#### **4.3.4 - 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 than 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 than 1% of the positions for each complete sequence. Thus, the worse 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  $\geq 96\%$ . 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).

#### **4.3.5 - Accession Numbers**

All entire viral genomes of successfully sequenced PRRSV strains were submitted to GenBank databank, with accession numbers MN865482 to MN865573 (Supplemental data, Tableau 7).

## 4.4 - 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 (Supplemental data, Table 7). 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 identities below 96% were identified in each case (listed in Tableau 5). 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. 18, Table 6). 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. 18). 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 18b). 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. 19). Moreover, PRRSV strain 2072533 was identified to be a recombinant virus which explain the discrepancy between the ORF5 and WGS classification methods (Table 5). No difference was found for the ATP vaccine-like and Fostera vaccine-like classification status (Fig 18a compared to 18b). Consequently, an overall 3.23% PRRSV vaccine-like strains ORF5 misclassification has been revealed by WGS.

As illustrated in Figure 19, 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. 19) 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 5. 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 20 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 coinfecting samples. This is however to be expected given that coinfection is a prerequisite for recombination event to occur, taking into account that more than one third of produced virions during coinfections can be PRRSV recombinant viruses[88].

In four of the 88 successfully sequenced PRRSV positive samples, two different PRRSV strains were found (Table 5), 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 sample 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 21 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 strain (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 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) (Supplemental data Table 7 and Table 5).

## 4.5 - 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. 18). 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-ISC15 system of the innate immune response. There are also many reports of highly pathogenic or epidemics strains recognizable by nsp2 deletions. The importance of nsp2 in the pathogenicity of PRRSV is what make 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[163]. 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 6). Interestingly, the most conserved nt encoded protein gene was found to be the serine protease nsp4 (Table 6), which is the main PRRSV protease responsible to cleave most of the replicase polyprotein into its smaller functional parts[164].

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[165]. 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 [166-168]. 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 than 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[110].

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 change 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[169]. 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 does 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. This improvement

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.

#### **4.6 - 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.

Tableau 5. – List of clinical samples containing more than one PRRSV strain (coinfection) or from which PRRSV strains were genetically misclassified by the ORF5 method.

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



2087409	Lung	Recombinant	Strain clustered differently in the same lineage compared to ORF5 method
---------	------	-------------	--

<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.

*Tableau 5*

Tableau 6. – Nucleotide conservation status of the main PRRSV protein encoded genes.

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

<sup>a</sup>The highest nucleotide identities obtained between strains were  $\geq 99.8\%$  for all protein encoded genes.

*Tableau 6*

The lowest inter-strains nucleotide identities between major PRRSV proteins are indicated.

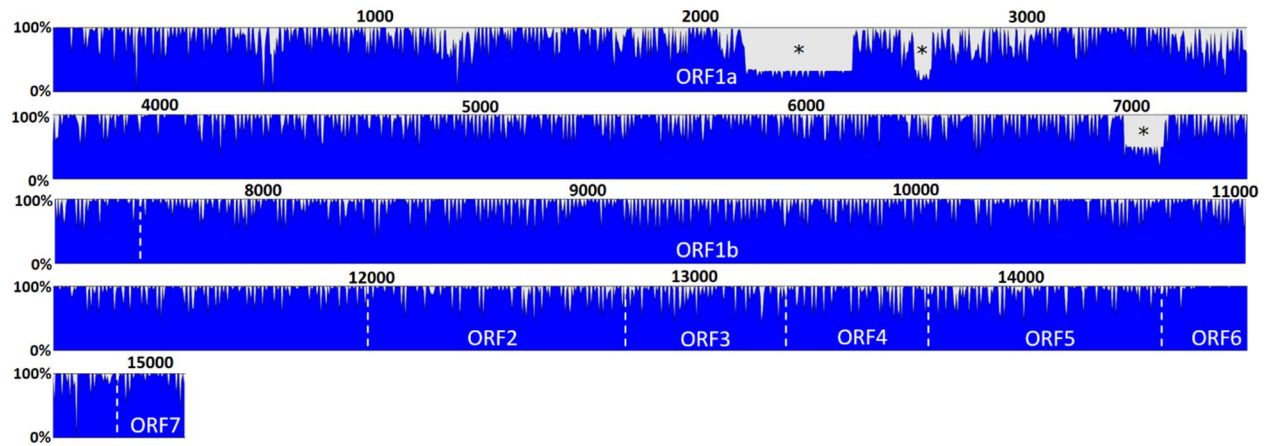


Figure 18

Figure 18. – 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 indicate 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.

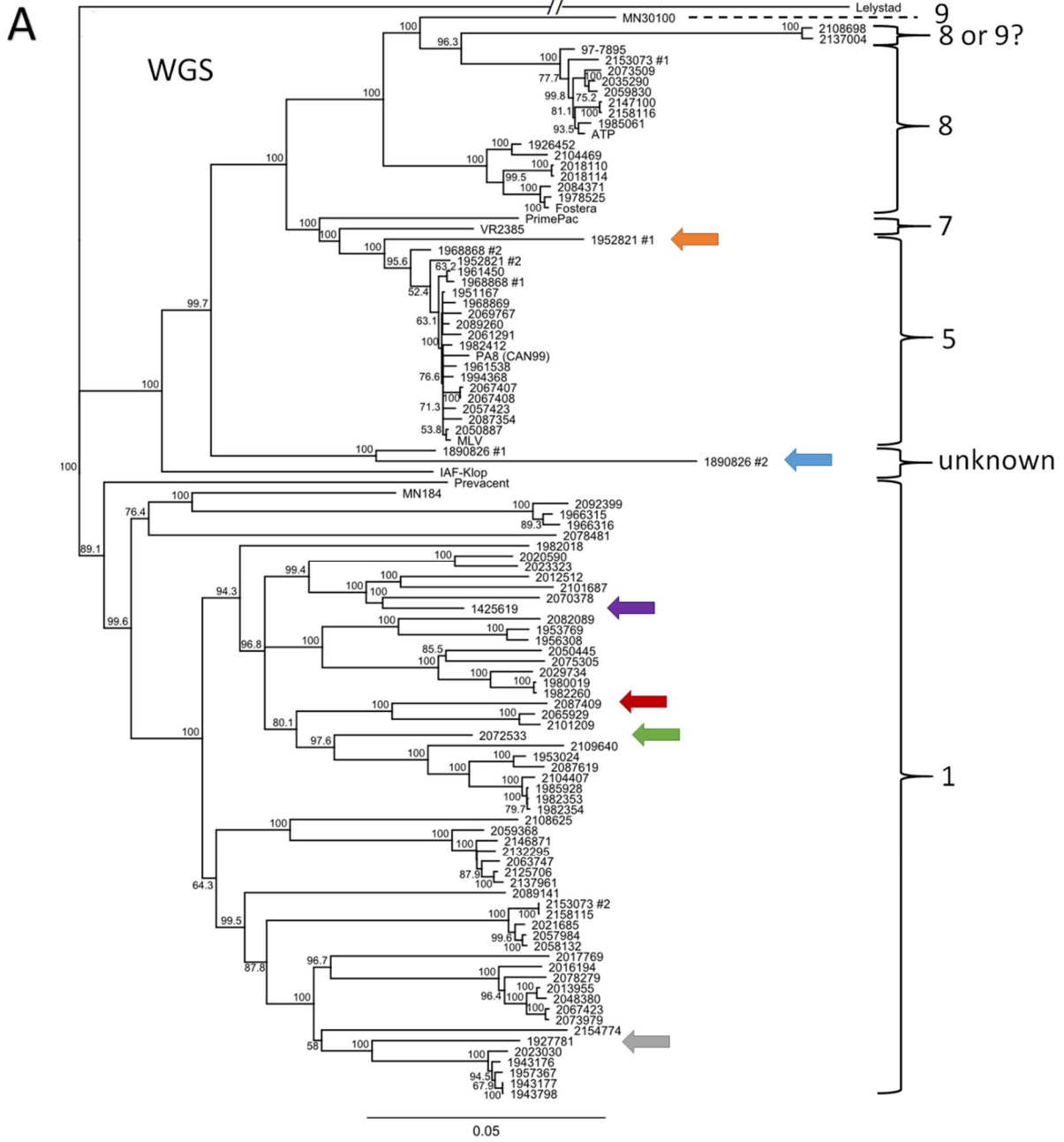


Figure 19

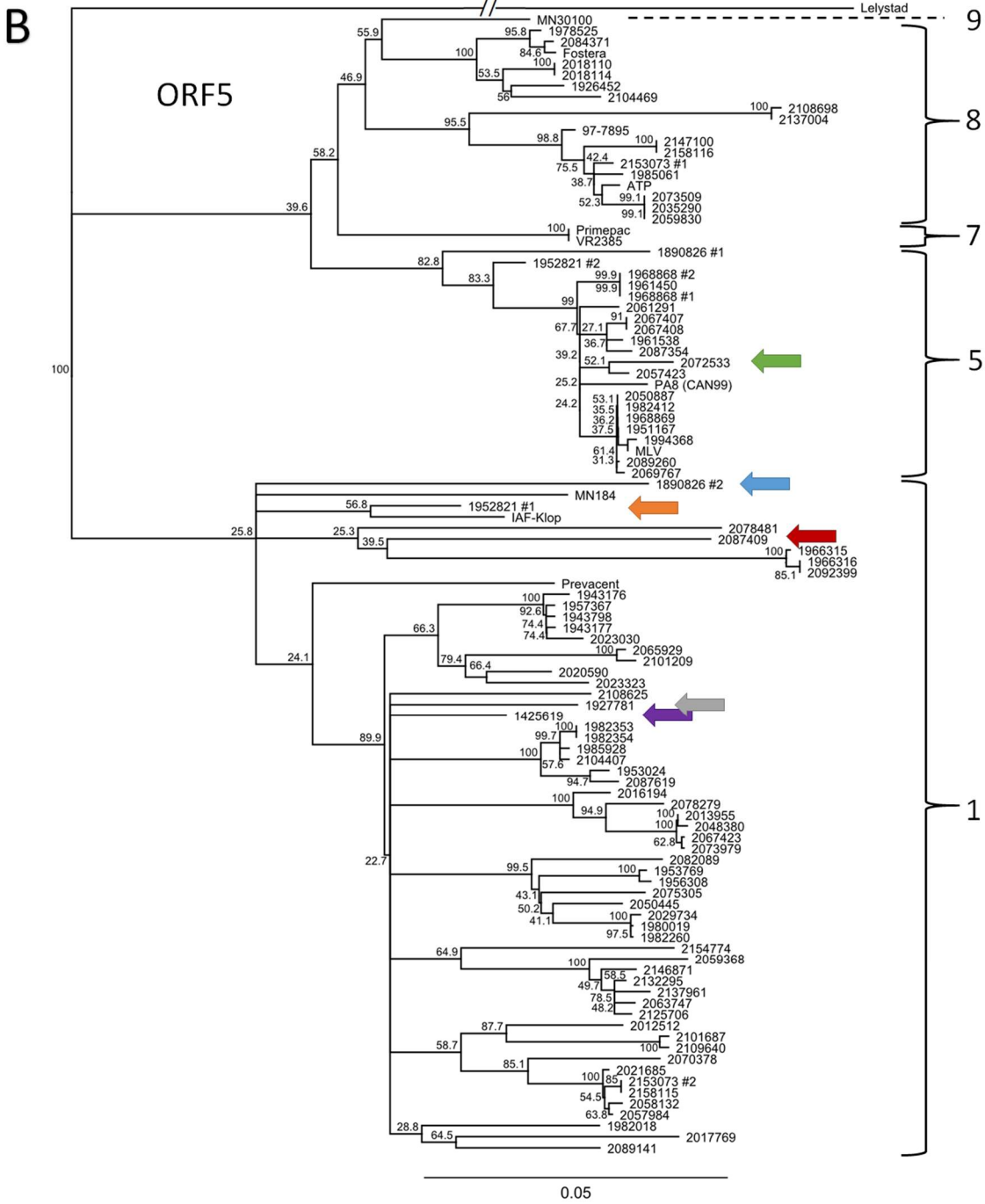


Figure 19. – PRRSV whole genome and ORF5 gene nucleotide phylogenetic trees comparison.

A) PRRSV WGS nucleotide phylogenetic tree and B) PRRSV ORF5 nucleotide phylogenetic tree. The brackets indicate the PRRSV type 2 lineages. MLV, ATP, Foster, 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.

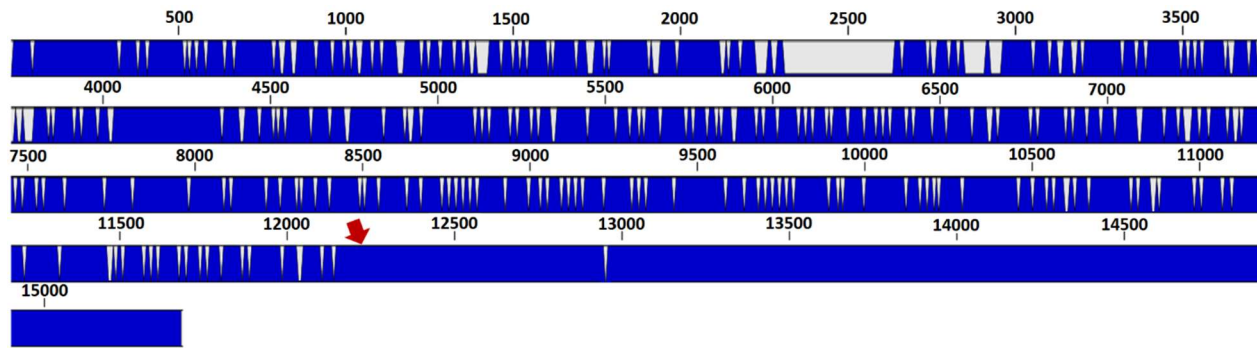
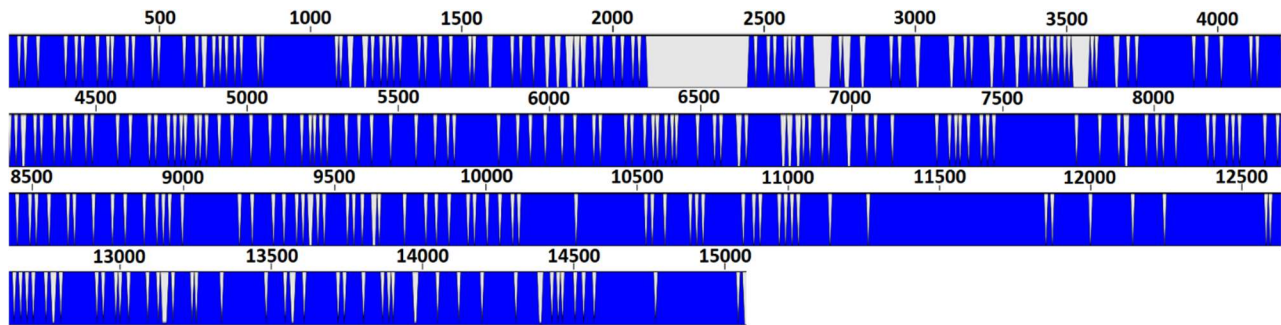


Figure 20

Figure 20. – Comparison of the PRRSV recombinant strain 2072533 with one of its possible parental strain, the MLV vaccine strain.

The red arrow point 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 21*

Figure 21. – 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.



## 4.7 - Supplemental data

Tableau 7. – PRRSV whole genome sequenced strains description

Strain identification number	PRRSV RT-qPCR Cq value in the sample	Type of sample	Number of animals pooled in the sample	GenBank accession number
1425619	n.a. <sup>a</sup>	Lung	1	MN865482
1890826 #1 <sup>b</sup>	21,98	Serum	1	MN865566
1890826 #2	21,98	Serum	1	MN865567
1926452	18,2	Lung	1	MN865483
1927781	23,74	Serum	1	MN865484
1943176	21,3	Sera	6	MN865485
1943177	14,5	Sera	6	MN865486
1943798	14,88	Sera	6	MN865487
1951167	22,75	Sera	7	MN865488
1952821 #1	23,95	PoT <sup>c</sup>	1	MN865568
1952821 #2	23,95	PoT	1	MN865569
1953024	21,68	Sera	7	MN865489
1953769	22,45	PoT	1	MN865490
1956308	20,17	PoT	1	MN865491
1957367	23,24	PoT	1	MN865492
1961450	24,42	Sera	5	MN865493
1961538	19,98	PoT	1	MN865494
1966315	20,08	Sera	4	MN865495
1966316	18,48	Sera	4	MN865496
1968868 #1	34,13	Serum	1	MN865570
1968868 #2	34,13	Serum	1	MN865571
1968869	27,08	Serum	1	MN865497
1978525	21,79	Sera	8	MN865498
1980019	22,27	Lung	1	MN865499
1982018	21,15	Lung	1	MN865500
1982260	20,36	Sera	10	MN865501
1982353	18,19	Sera	6	MN865502
1982354	16,86	Sera	6	MN865503
1982412	13,89	Lung	1	MN865504
1985061	24,19	Lung	1	MN865505
1985928	15,25	Serum	1	MN865506
1994368	19,7	Lung	1	MN865507
2012512	17,94	Lung	1	MN865508
2013955	22,99	Lungs	2	MN865509

2016194	16,97	Lung	1	MN865510
2017769	22,23	Lungs	2	MN865511
2018110	21,85	Sera	5	MN865512
2018114	19,62	Sera	5	MN865513
2020590	24,59	PoT	1	MN865514
2021685	21,22	Lung	1	MN865515
2023030	18,73	Sera	3	MN865516
2023323	22,55	Lung	1	MN865517
2029734	21,85	Lung	1	MN865518
2035290	22,86	Sera	5	MN865519
2048380	20,23	Sera	10	MN865520
2050445	20,06	Serum	1	MN865521
2050887	23,52	Sera	10	MN865522
2057423	22,77	Sera	10	MN865523
2057984	24,4	Sera	5	MN865524
2058132	19,51	PoT	1	MN865525
2059368	18,09	Sera	10	MN865526
2059830	27,22	Sera	5	MN865527
2061291	22,58	PoT	2	MN865528
2063747	24,77	Sera	2	MN865529
2065929	26,37	Sera	10	MN865530
2067407	23,06	Sera	5	MN865531
2067408	24,33	Sera	5	MN865532
2067423	26,28	Serum	1	MN865533
2069767	24,33	PoT	2	MN865534
2070378	25,98	PoT	2	MN865535
2072533	26,53	Serum	1	MN865536
2073509	26,5	Lung	1	MN865537
2073979	18,83	Sera	10	MN865538
2075305	20,95	Lung	1	MN865539
2078279	18,19	Sera	5	MN865540
2078481	n.a.	Serum	1	MN865541
2082089	19,47	Lung	1	MN865542
2084371	23,98	Lung	1	MN865543
2087354	n.a.	Lung	n.a.	MN865544
2087409	20,98	Lung	1	MN865545
2087619	21,35	PoT	1	MN865546
2089141	17,89	PoT	1	MN865547
2089260	16,81	PoT	2	MN865548
2092399	19,26	Sera	2	MN865549
2101209	23,53	Sera	3	MN865550
2101687	16,44	Sera	5	MN865551
2104407	13,57	Serum	1	MN865552

2104469	n.a.	Lung	1	MN865553
2108625	19,89	Sera	6	MN865554
2108698	29,18	Serum	1	MN865555
2109640	17,7	Serum	1	MN865556
2125706	28,83	Oral fluid	Pen <sup>d</sup>	MN865557
2132295	27,42	Serum	1	MN865558
2137004	29,3	Serum	1	MN865559
2137961	28,62	Serum	1	MN865560
2146871	24,85	Serum	1	MN865561
2147100	32,84	Oral fluid	Pen	MN865562
2153073 #1	27,93	Serum	1	MN865572
2153073 #2	27,93	Serum	1	MN865573
2154774	11,5	Serum	1	MN865563
2158115	18,77	Serum	1	MN865564
2158116	17,24	Serum	1	MN865565

<sup>a</sup>n.a.: means the data is not available.

<sup>b</sup>The # after the sample identification number represent one of the two PRRSV strains that was identified during coinfection within the same clinical sample.

<sup>c</sup>PoT or pool of tissues include mainly lungs with several other type of tissues such as lymph nodes, spleen, liver, intestine, etc.

<sup>d</sup>The number of animals per pen during oral fluid sampling is unknown.

*Tableau 7*

## 5 - Discussion

### 5.1 - Mise au point du protocole

L'une de nos principales réalisations est le séquençage de génome entier du VSRRP directement à partir d'échantillons cliniques. D'autres méthodes ont précédemment été décrites mais nécessitent des pré-amplifications des génomes viraux par PCR[155, 156] ou un isolement préalable du virus en culture cellulaire[157]. De plus, si l'on compare la sensibilité de notre méthodologie à d'autres équivalentes (*i.e.* d'autre séquençage de génome entier directement à partir d'échantillons cliniques), nous avons obtenu des génomes entiers à partir d'échantillons possédant des charges virales beaucoup moins élevées (Figure 13) que ce qui est rapporté dans la littérature[143, 144], avec des valeurs de Cq d'échantillons plus élevé de 10.53 pour les sérums, de 5.5 pour les poumons et de à 12.24 pour les fluides oraux.

Le succès de la méthodologie employée semble s'expliquer principalement par l'utilisation des billes magnétiques couplées à des poly(dT) pour la concentration des génomes ARN viraux. Puisque les ARN polyadénylés ne correspondent qu'à environ 3% des ARN totaux dans les cellules eucaryotes[170] et que les ARN génomiques et sous-génomique du VSRRP possèdent tous une queue poly(A), l'isolement des ARN avec queue poly(A) laisse présumer que les ARN viraux devraient être 33 fois plus concentrés que dans les ARN totaux (puisque'on conserve tous les ARN viraux mais seulement 3/100 des ARN totaux). Ceci peut aisément expliquer le succès de notre méthodologie avec des échantillons aux valeurs de Cq plus élevées de 5.

Un autre facteur pouvant avoir joué sur le succès de notre méthodologie, particulièrement pour les échantillons de sérum, est l'utilisation de l'ensemble Nextera XT (Illumina) pour la création de bibliothèques et l'indexation. Plusieurs équipes ont démontré que cet ensemble était très efficace dans la génération de longs fragments de grande qualité pour le séquençage sur une plateforme Illumina MiSeq[171, 172]. Toutefois, comparé à d'autres méthodes, l'ensemble Nextera XT a tendance à générer des biais lorsque les contenus en GC sont élevés[173]. Cela n'est toutefois pas un problème pour le VSRRP qui possède un contenu en GC d'environ 52%[174].

Un autre inconvénient avec la technologie Nextera XT est une tendance à générer moins de d'îlots de séquençage sur les cartouches et donc une profondeur de séquençage moins élevée[171].

Lors du séquençage de nos premières cartouches, nous avons fait face à un problème similaire avec une utilisation incomplète de l'espace sur les cartouches et donc d'une diminution significative du nombre de lectures obtenues. Cet inconvénient a été surmonté grâce à quelques changements au niveau du protocole de préparation des librairies. Voici les changements qui ont été apportés : en premier lieu, lorsque possible une concentration d'ADNdb allant jusqu'à 0.3 ng/μl (au lieu de 0.2ng/μl dans le protocole du fabricant) a été utilisée; ensuite, on a procédé à 14 cycles d'amplifications à l'étape d'indexation des librairies au lieu de 12 comme suggéré par le fabricant; et finalement, une quantité jusqu'à 50% plus élevée du pool de librairies a été chargée sur les cartouches.

Malgré le succès de la méthodologie décrite, plusieurs obstacles restent à surmonter pour une implémentation à grande échelle. Le premier est le besoin en personnel hautement qualifié. De plus, il a été démontré que le taux d'efficacité de séquençage (succès) peut significativement différer entre deux individus ayant reçu la même formation méthodologique (Figure 15b). Le grand nombre d'étapes à réaliser lors de notre protocole et la précision requise pour leurs exécutions suggèrent un haut potentiel d'erreurs humaines, et nécessitent donc du personnel hautement qualifié mais aussi une rigueur interne dans l'exécution des manipulations.

Le second obstacle est le coût associé, autant celui des réactifs, que celui du temps du personnel, de la formation requise et de l'achat initial d'équipement. La rentabilisation de tels investissements nécessite un volume d'échantillons difficilement atteignable et un amortissement des coûts initiaux sur une longue période.

Au niveau de la reconstitution bio-informatique des génomes, la plus grande limite de la méthode décrite plus haut est l'utilisation de logiciels à licence commerciale. L'utilisation de logiciels validés est nécessaire pour les laboratoires accrédités et à des fins diagnostiques, mais leur coût parfois très élevé limite grandement l'utilisation par d'autres laboratoires ou chercheurs. Il existe des alternatives en libre accès sur le web mais les différents logiciels utilisent des algorithmes variés lors des analyses et peuvent donc avoir une incidence sur les résultats finaux[175, 176].

Une autre limitation au niveau bio-informatique est la base de données de génomes de références utilisée. Dû à la grande variabilité génétique du VSRPP tant au niveau régional (Figure 188) qu'international[98], il est possible que les génomes des souches génétiquement éloignées (par exemple de lignées différentes), soient difficiles à reconstituer en se servant des souches qui sont

présentement disponibles comme référence. Cet obstacle devrait toutefois disparaître graduellement au fur et à mesure que plus de séquences différentes deviendront disponibles, tant sur les bases de données en libre accès que dans les bases de données des différents laboratoires.

Une autre approche présentement employée est le séquençage de génome entier de VSRRP par la plateforme MinIon d'Oxford Nanopore[145]. Cette approche offre plusieurs avantages évidents par rapport à notre méthodologie. Le MinIon ne nécessite aucun besoin en équipement dispendieux, alors qu'un séquenceur de table coûte plusieurs dizaines de milliers de dollars. De plus, le temps de génération des lectures à partir d'un échantillon clinique est inférieur[177]. Cette différence est en partie attribuable au temps de séquençage du MinIon de 36 à 48 heures, comparativement au 65h requise pour les cartouches MiSeq v3 600. Le nombre d'étapes nécessaires est aussi inférieur, réduisant le risque d'erreur de manipulations. Cela ne tient toutefois pas compte du fait que notre protocole comprend une étape de concentrations des ARN viraux alors que les protocoles de MinIon n'en comprennent pas. Finalement, le MinIon génère des lectures beaucoup plus grandes[177].

Le MinIon n'est pas sans inconvénient toutefois. Le taux d'erreur est très élevé, nécessitant une plus grande profondeur de séquençage afin d'obtenir une séquence consensus[178]. Il nécessite aussi une très grande quantité de matériel génétique pour la création des bibliothèques (500ng contre 1ng pour notre protocole)[145].

En comparant nos résultats (sensibilité de  $8.17 \times 10^1$  copies de génome viral par réaction) avec ceux rapportés dans la littérature qui utilise la technologie MinIon (sensibilité de  $3.8 \times 10^6$  copies de génome viral par réaction) [145], nous avons obtenus une sensibilité  $4.65 \times 10^4$  fois plus élevée avec notre protocole de séquençage pour les échantillons de sérums. Cela semble indiquer que la technologie d'Oxford Nanopore doit encore être peaufinée afin de compétitionner avec les méthodes de séquençage par synthèse.

## **5.2 - Le type d'échantillons**

La plus grande sensibilité et le meilleur taux de succès de séquençage du génome entier du VSRRP ont été obtenus avec les échantillons de sérum (Figure 11 et Figure 12) Ce résultat était attendu car Zhang et collègues ont obtenus un résultat similaire[144]. Ils ont réussi le séquençage de génome entier à partir d'échantillons de sérum de valeur de Cq de 2.6 supérieurs aux échantillons

de poumons. Le succès du séquençage du génome entier avec des échantillons de sérum comparativement à des échantillons de poumons semble attribuable en partie à la plus faible présence d'acides nucléiques de l'hôte dans le sérum que dans les tissus[148]. Cet effet est observable par la proportion significativement plus élevée de lectures provenant du virus dans ces échantillons (Figure 14). Un autre phénomène semblant expliquer le succès du séquençage à partir d'échantillons de sérum est la dispersion des lectures. Dû à la nature du mode de réplication des nidovirus par réseau d'ARN sous-génomique[55], il est attendu que l'ARN viral provenant du site de réplication virale soit constitué principalement d'ARN sous-génomique. Cela peut mener à une couverture non-homogène du génome viral, telle qu'illustrée à la figure 16. Ce phénomène est aussi présent dans les échantillons de sérum mais il est moins prononcé (Figure 16). Dû à son tropisme principalement dirigé vers les macrophages alvéolaires pulmonaires, la réplication du VSRRP se produit principalement dans les poumons[58]. Une proportion plus grande de virions intact et donc d'ARN génomique viral, par rapport aux ARN sous-génomiques, se retrouverait donc dans le sérum. Les fluides oraux semblent offrir une sensibilité et un succès comparable aux sérums (Figure 11) mais un nombre trop faible (n=2) d'échantillons de ce type ont été analysés pour en tirer des conclusions.

Nous pouvons donc recommander l'utilisation d'échantillons de sérum étant donné que notre méthodologie de séquençage est la plus sensible et efficace avec ce genre d'échantillon.

### **5.3 - Diversité génétique**

La grande diversité génétique des souches Québécoises de VSRRP (Figures 18 et 19) était déjà prévisible selon les travaux de Lambert et collaborateurs[129]. Il est déjà connu que le virus mute rapidement[90] et donc une grande variabilité génétique est attendue au fil du temps. Ce qui est particulièrement intéressant est d'observer la diversité de nos souches virales sur une période de temps relativement courte de 24 mois (Figure 188).

Il est raisonnable de se demander si la grande variabilité génétique du VSRRP pouvait être liée à la présence de quasi-espèces. La théorie des quasi-espèces suggère que pour les virus à haut taux de mutation, la présence simultanée d'un éventail varié de mutants est plus avantageuse pour le virus que la sélection d'un seul des mutants [179-181] favorisant ainsi une plus grande diversité virale.

## 5.4 - Délétions

Les trois délétions décrites dans l'article 2 (Figure 188) sont particulièrement intéressantes. Les deux premières, aux positions 2159 à 2491 et 2717 à 2773 par rapport à la souche vaccinale MLV, sont intéressantes à cause de leur localisation dans l'ORF1a, plus précisément dans la protéine NSP2. Il a été démontré précédemment que NSP2 et les délétions s'y retrouvant étaient importantes dans la pathogenèse virale et dans l'évasion du système immunitaire[80, 166-168, 182, 183].

NSP2 est une protéine relativement grosse d'environ 1196 acides aminés pour le VSRRP de type 2 et 1077 acides aminés pour le VSRRP de type 1[37]. La taille de NSP2 est toutefois très variable chez les *arterivirus*, allant de 571 à 1196 acides aminés de long [184, 185]. Tant la présence de nombreuses délétions dans différentes souches que cette variabilité de taille entre les espèces semblent indiquer que plusieurs régions de la protéine ne sont pas essentielles à la réplication ou à l'infection virale. Parmi les fonctions connues de la protéine NSP2, on retrouve la désubiquitination, l'antagonisme de la production d'interféron, une activité protéinase de cystéine accessoire et une interaction potentielle avec les membranes intracellulaires[37, 71, 72]. L'étude de ces deux délétions pourrait permettre de mieux comprendre la pathogenèse du virus et les mécanismes par lesquelles le virus évite et/ou contrôle la réponse immunitaire innée de l'hôte infecté.

La troisième délétion, de la position 7106 à 7228 par rapport à la souche vaccinale MLV, est intéressante pour une autre raison. Il y a très peu d'études sur la protéine NSP7 et sa fonction principale n'est pas encore très claire. On sait par les travaux de Zhang et collègues qu'elle serait impliquée dans la réplication virale sans toutefois préciser sa fonction [163]. L'étude de cette délétion pourrait aider à élucider la fonction précise de la protéine NSP7.

## 5.5 - Coinfections

Le séquençage de deux souches de VSRRP distinctes à partir du même échantillon est, au meilleur de notre connaissance, une nouveauté. Tan et collègues ont précédemment séquencé sur MinIon des échantillons contenant deux souches mais il était question d'échantillons artificiellement additionnés pour créer le phénomène de coinfection[145]. Les 4 cas de coinfections décrits au Tableau 4 représentent donc une avancée majeure tant au niveau de la recherche qu'au niveau clinique. Le séquençage de nouvelle génération offre un avantage indéniable sur le



séquençage de type Sanger standard en permettant la détection et le séquençage simultanés de plusieurs souches. Toutefois, il est intéressant de mentionner qu'il existe des protocoles de séquençage de type Sanger permettant de détecter des variants de virus moins fréquents au sein d'un même échantillon[186]. Ces protocoles ne semblent jamais avoir été testés pour le VSRRP. De plus, il a été démontré pour d'autres virus que le séquençage Sanger n'était pas aussi efficace pour la détection des variants génomiques[187]. Cela suggère donc que pour le moment, notre protocole est le plus efficace pour la détection et le séquençage de souches virales coinfectantes au sein d'un échantillon clinique, du moins pour le PRRSV.

L'impact des coinfections sur la présentation clinique du VSRRP n'est cependant pas bien documenté. Une étude *in-vivo* sur l'infection simultanée par deux souches de VSRRP a déjà été réalisée mais elle ne s'intéressait pas à la présentation clinique de l'infection virale[88]. Dû à la variabilité de pathogénicité entre les souches de VSRRP et de son effet immunomodulateur[188], il est vraiment difficile de prédire l'impact d'une coinfection. Il est connu que différentes souches de VSRRP interagissent de manière différente avec l'hôte [189-191]. Cela peut laisser présager qu'une coinfection avec deux souches de VSRRP n'aurait pas nécessairement le même impact si l'une des souches impliquées dans la coinfection était changée pour une souche différente.

## **5.6 - Classification**

La classification déjà établie selon l'ORF5 s'est avérée applicable pour la majorité des échantillons suite aux résultats obtenus avec les analyses de génome entier du virus (9). Cette classification de l'ORF5 est basée sur 9 lignées monophylétiques distinctes[98]. Toutefois, certaines souches ne semblent pas correspondre aux lignées précédemment décrites (9A). C'est entre autres le cas pour les souches 2108698 et 2137004, identifiées par le crochet « 8 or 9? » à la Figure 199A. Ces deux souches semblent être à mi-chemin entre les lignées 8 ou 9, mais il est impossible de déterminer avec certitude à quelle lignée elles appartiennent réellement en se basant sur l'ensemble de génomes entiers que nous avons obtenus. Un autre cas particulier sont les souches 1890826 #1 et #2, identifiées par le crochet « unknown » à la Figure 199A. Elles sont trop éloignées des autres souches pour être incluses dans les mêmes lignées qui ont été créées à partir des données de l'ORF5.

L'autre cas intéressant est la classification de la souche IAF-Klop, une souche de référence québécoise de la lignée 1[98, 192]. Elle fait aussi partie des souches considérées « unknown »

lorsque son génome entier est utilisé pour la classification. L'explication la plus simple et plausible pour ce résultat est que la souche, vieille de plus de 25 ans, est trop éloignée temporellement pour être aisément classifiée avec des souches contemporaines. Cela illustrerait à quel point le VSRRP mute et change rapidement, comme cela a déjà été démontré[90].

Il est intéressant de noter que 6 souches n'appartiennent pas à la même grappe lorsque l'on compare l'arbre phylogénétique des génomes entiers à celui des ORF5 (Figure 199). De ces 6 souches, une seule pas du statut de souche vaccinale par ORF5 à souche sauvage par génome entier (flèche verte à la Figure 199). Deux autres n'appartiennent plus à la même lignée mais ne sont pas considérées comme des souches vaccinales (flèches orange et bleu à la Figure 199). Les trois autres ne font que changer de grappe au sein de la même lignée. L'explication la plus probable est que ces souches ont fait l'objet d'une recombinaison de génome viral entre des souches génétiquement différentes. La détection des recombinants et des sites de recombinaison est beaucoup plus simple lorsque les séquences des souches parentales sont disponibles. Cela devrait donc être plus facile à établir au fur et à mesure qu'un nombre plus élevé de séquences de génome entier seront disponibles.

Ces exemples suggèrent que la classification devrait être revue ou modifiée pour mieux représenter les génomes entiers de VSRRP. Malgré tout, la classification des lignées basées sur les données de génome entier semble être similaire à celle basée sur les données de l'ORF5 pour la plupart des souches. Ceci semble indiquer que des modifications au système actuel pourraient suffire pour inclure la description et la classification des souches basées sur les génomes entiers.

L'une des façons envisageables pour améliorer la classification serait de garder la classification actuelle de l'ORF5 tout en ajoutant une seconde classification réalisée selon une deuxième région du génome. La nomenclature des souches pourrait alors s'apparenter à celles des souches du virus *Influenza A*[165].

La candidate la plus prometteuse serait la région codante pour la protéine NSP2 à l'intérieur de l'ORF1a. Différentes raisons me poussent à suggérer cette région. Premièrement, 3 protéines distinctes sont traduites à partir de ce gène, la protéine NSP2 et les protéines NSP2N et NSP2TF grâce à un site de changement de cadre de lecture ribosomale (Figure 6). Deuxièmement, c'est l'une des régions les plus variables du génome (Tableau 5). Troisièmement, c'est une région génomique qui semble très propice aux délétions, tel que rapporté non seulement dans notre étude

(Figure 188) mais aussi par d'autres équipes de recherche[167, 168, 183]. De plus, c'est une région qui semble être impliquée dans la pathogenèse virale, et certaines délétions s'y retrouvant ont été décrites comme des marqueurs de souches hautement pathogènes[166]. Elle se trouve aussi assez près de l'extrémité 5' du génome viral, et donc l'utilisation de 2 régions aux deux extrémités du génome viral pourrait améliorer les chances de détecter les virus recombinants. Toutefois, comme la région codante pour NSP2 peut contenir beaucoup de délétions, qu'elle est très variable et qu'elle se trouve dans l'ORF1a (et donc qu'on la retrouve uniquement sur les ARN génomiques et pas sur les ARN sous-génomiques), l'amplification par Sanger de la région est très difficile. Il est donc plus efficace de séquencer le génome entier pour obtenir la séquence codante pour NSP2 que d'essayer l'amplification par PCR nécessaire au séquençage de type Sanger.

Une autre région potentiellement intéressante est la région codant pour la protéine NSP7. Elle est aussi très variable (Tableau 5), en plus d'être très immunogène[193]. La protéine NSP7 a cependant été beaucoup moins étudiée que NSP2 et sa fonction demeure peu connue[37, 163]. Elle est aussi très variable et se retrouve elle aussi dans l'ORF1a.

## **5.7 - Implication pour l'industrie**

Pour la filière porcine, l'avantage principal offert par notre méthode est une classification plus représentative de la réalité dans environ 9,09% des cas d'infection au VSRRP (Tableau 5). Cette meilleure interprétation peut mener à des interventions plus adéquates sur le terrain et pourrait donc avoir un impact positif sur la santé des troupeaux. Le taux de succès à des valeurs de Cq relativement élevé, donc avec une charge virale faible (Figure 12 et Figure 13), est aussi un avantage évident, facilitant le suivi et la surveillance. Parmi les méthodes employées pour le contrôle du VSRRP, la plus efficace et la moins dispendieuse demeure la biosécurité dans les élevages; c'est-à-dire de ne pas introduire le virus dans des troupeau négatifs ou de ne pas introduire de nouvelles souches dans les troupeaux positifs[194]. Avoir un portrait plus juste de la ou des souches circulantes dans un troupeau pourrait donc permettre d'améliorer les mesures de prévention et éviter l'introduction de nouvelles souches dans les troupeaux positifs. Cet avantage ne s'applique toutefois pas aux troupeaux négatifs pour lesquels les fermes emploient déjà des mesures pour empêcher l'introduction de toutes les souches de VSRRP. De plus, l'absence de traitement efficace contre la maladie signifie que la prévention demeure la meilleure mesure.

Il existe toutefois plusieurs facteurs limitants l'utilisation à grande échelle de notre méthode. La principale limitation est au niveau du coût. Le coût d'une réaction de séquençage peut être considéré en trois parties : le coût fixe par échantillon pour l'ensemble des réactifs nécessaires jusqu'à la création de la librairie, le coût en personnel pour le temps requis du personnel qui fait les manipulations et le coût par cartouche. Cependant, le prix de la cartouche peut être amorti sur l'ensemble des échantillons lorsqu'ils sont séquençés lors de la même réaction. Bien que nous ayons mis au point un protocole permettant de maximiser le nombre d'échantillons par cartouche et donc de minimiser ce coût par échantillon, le coût en réactifs et en temps demeure le même par échantillon, ne permettant pas au séquençage à haut débit du génome entier de compétitionner avec le séquençage de type Sanger de cibles génétiques précises au niveau de sa rentabilité.

Concrètement, la facturation du séquençage du génome entier à la clientèle vétérinaire est de 2 à 3 fois plus élevée que celle du séquençage de l'ORF5 seul. Il y a aussi l'investissement initial nécessaire par les laboratoires qui est très élevé, devant à la fois se procurer de l'équipement et des licences de logiciels souvent dispendieuses mais aussi se doter de et/ou former du personnel hautement qualifié tant en biologie moléculaire qu'en bio-informatique. Dans le contexte de la production animale, ces investissements en séquençage à haut débit doivent engendrer des économies au moins aussi importantes pour en justifier leur utilisation. Bien que nous ne puissions recommander le séquençage de routine de génomes entier de VSRRP, nous croyons qu'il est important de sélectionner les cas cliniques qui pourraient bénéficier du séquençage à haut débit comme par exemple : lors d'échec de la vaccination ou encore lorsque les résultats de séquençage de l'ORF5 ne corrélerent pas avec la sévérité des signes cliniques. Néanmoins, nos travaux ont fait un grand pas en améliorant significativement la sensibilité et l'efficacité du séquençage du génome entier du VSRRP directement à partir d'échantillons cliniques.

## **5.8 - Implication pour la recherche**

La méthode que nous avons mise au point pour le séquençage de génome entier de VSRRP a de grandes implications. Avec une efficacité et une sensibilité beaucoup plus grande que ce qui a été fait auparavant (Figure 11 et Figure 12), elle devrait aider d'autres chercheurs ou institutions à séquençer des génomes entiers de souches virales de VSRRP. Plus la communauté scientifique sera en mesure d'accumuler des données sur ce pathogène et plus nous en comprendrons la

pathogénèse et l'évolution, ce qui pourrait éventuellement mener à des avancées significatives dans la lutte au VSRRP. De façon plus concrète, cela pourrait permettre l'étude à grande échelle des coinfections par différentes souches en situation de ferme, et aider à mieux comprendre les mécanismes qui favorisent le haut taux de recombinaison entre les différentes souches. Cela pourrait aussi mener à la découverte de meilleur marqueur de pathogénicité outre l'ORF5 car il arrive parfois que des souches apparentées à des vaccins soit hautement pathogéniques[195]. Cela pourrait aussi aider à mieux comprendre les fonctions de certaines des protéines non-structurales qui sont toujours inconnues[37].

Les 92 séquences de génome entier de VSRRP de type 2 que nous avons obtenues et qui ont été envoyées à GenBank (Tableau 6) sont aussi une avancée majeure. Elles représentent environ 10% des séquences de génome entier de VSRRP de type 2 présentement disponibles en ligne.

Parmi les avenues à explorer, lorsqu'un plus grand nombre de génomes entiers seront disponibles, il sera possible d'essayer de trouver une corrélation entre certains motifs génétiques et la présentation des signes cliniques. Le recensement de toutes les délétions importantes retrouvées dans différentes souches comme celles observées lors de notre étude (Figure 188) de même que leur implication sur la réplication, l'immunogénicité et la pathogénèse pourraient permettre de mieux comprendre les étroites interactions entre le virus et l'hôte.

## Références bibliographiques

1. Mussel, A., *A Risk, Benefit, Strength, Weakness, Opportunity and Threat Analysis for the Control and Possible Eradication of Porcine Reproductive and Respiratory Syndrome (PRRS) Virus Within the Canadian Swine Herd*. Canadian Swine Health Board, 2011: p. 1-122.
2. Holtkamp, D.J., et al., *Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers*. Journal of Swine Health and Production, 2013. **21**(2): p. 72.
3. KEFFABER, K., *Reproductive failure of unknown aetiology*, AASP Newsletter. Sep. Oct, 1989.
4. Dial, G., et al., *Mystery swine disease: Implications and needs of the North American swine industry*. Prof Mystery Swine Dis Committee, 1990: p. 3-6.
5. Zimmerman, J. *A survey of the American Association of Swine Practitioners for the presence of mystery swine disease herds*. in Proceedings of Livestock Conservation Institute. 1991.
6. Lindhaus, W. et B. Lindhaus, *Mystery swine disease*. Praktische Tierarzt (Germany, FR), 1991.
7. Collins, J., *International symposium on SIRS/RRS*. American Association of Swine Practitioner Newsletter, 1992. **4**(4): p. 1.
8. Karniychuk, U., *Congenital porcine reproductive and respiratory syndrome virus infection in naïve and vaccinated sows*. 2012, Ghent University.
9. Zimmerman, J., et al., *General overview of PRRSV: a perspective from the United States*. Veterinary Microbiology, 1997. **55**(1-4): p. 187-196.
10. Regula, G., et al., *Comparison of serologic testing and slaughter evaluation for assessing the effects of subclinical infection on growth in pigs*. Journal of the American Veterinary Medical Association, 2000. **217**(6): p. 888-895.
11. Christianson, W.T., *Stillbirths, mummies, abortions, and early embryonic death*. Veterinary Clinics of North America: Food Animal Practice, 1992. **8**(3): p. 623-639.

12. Pol, J., J. van Dijk, et G. Wensvoort, *Pathological, ultrastructural, and changes caused by Lelystad virus in experimentally induced infections of mystery swine disease (synonym porcine epidemic abortion and respiratory syndrome (PEARS))*. *Veterinary Quarterly*, 1991. **13**(3): p. 137-143.
13. Loula, T., *Mystery pig disease*. *Agri-practice*, 1991. **12**(1): p. 23-34.
14. Bilodeau, R., et al., *'Porcine reproductive and respiratory syndrome' in Quebec*. *The Veterinary Record*, 1991. **129**(5): p. 102.
15. Albina, E., *[Porcine reproductive and respiratory syndrome: ten years of experience (1986-1996) with this undesirable viral infection]*. *Veterinary Research*, 1997. **28**(4): p. 305-52.
16. Collins, J.E., et al., *Isolation of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North America and experimental reproduction of the disease in gnotobiotic pigs*. *Journal of Veterinary Diagnostic Investigation*, 1992. **4**(2): p. 117-126.
17. Halbur, P., et al., *Comparison of the antigen distribution of two US porcine reproductive and respiratory syndrome virus isolates with that of the Lelystad virus*. *Veterinary Pathology*, 1996. **33**(2): p. 159-170.
18. Rossow, K.D., et al., *Lymph node lesions in neonatal pigs congenitally exposed to porcine reproductive and respiratory syndrome virus*. *Journal of Veterinary Diagnostic Investigation*, 1994. **6**(3): p. 368-371.
19. Harms, P., et al., *Experimental reproduction of severe disease in CD/CD pigs coinfecting with PRRSV and type 2 porcine circovirus*. *Veterinary Pathology*, 2001. **38**(5)
20. Galina, L., et al., *Interaction between Streptococcus suis serotype 2 and porcine reproductive and respiratory syndrome virus in specific pathogen-free piglets*. *The Veterinary Record*, 1994. **134**(3): p. 60-64.
21. Cho, J.G., et al., *The impact of animal age, bacterial coinfection, and isolate pathogenicity on the shedding of porcine reproductive and respiratory syndrome virus in aerosols from experimentally infected pigs*. *Canadian Journal of Veterinary Research*, 2006. **70**(4): p. 297.
22. Tian, K., et al., *Emergence of fatal PRRSV variants: unparalleled outbreaks of atypical PRRS in China and molecular dissection of the unique hallmark*. *PloS one*, 2007. **2**(6): p. e526.

23. Wensvoort, G., et al., *Mystery swine disease in The Netherlands: the isolation of Lelystad virus*. *Veterinary Quarterly*, 1991. **13**(3): p. 121-130.
24. Ohlinger, V., et al., *The German experience with the swine infertility and respiratory syndrome (SIRS)*. *Proc. MN Confor Vet*, 1991.
25. King, A.M., et al., *Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018)*. *Archives of Virology*, 2018. **163**(9): p. 2601-2631.
26. Spilman, M.S., et al., *Cryo-electron tomography of porcine reproductive and respiratory syndrome virus: organization of the nucleocapsid*. *Journal of General Virology*, 2009. **90**(3): p. 527-535.
27. Dokland, T., *The structural biology of PRRSV*. *Virus Research*, 2010. **154**(1-2): p. 86-97.
28. Meulenberg, J.J., *PRRSV, the virus*. *Veterinary Research*, 2000. **31**(1): p. 11-21.
29. Dea, S., et al., *Current knowledge on the structural proteins of porcine reproductive and respiratory syndrome (PRRS) virus: comparison of the North American and European isolates*. *Archives of Virology*, 2000. **145**(4): p. 659-688.
30. Lunney, J.K., et al., *Porcine reproductive and respiratory syndrome virus (PRRSV): pathogenesis and interaction with the immune system*. *Annual Review of Animal Biosciences*, 2016. **4**: p. 129-154.
31. Johnson, C.R., et al., *Novel structural protein in porcine reproductive and respiratory syndrome virus encoded by an alternative ORF5 present in all arteriviruses*. *Journal of General Virology*, 2011. **92**(5): p. 1107-1116.
32. Gorbalenya, A.E., et al., *Nidovirales: evolving the largest RNA virus genome*. *Virus Research*, 2006. **117**(1): p. 17-37.
33. Veit, M., et al., *Membrane proteins of arterivirus particles: structure, topology, processing and function*. *Virus Research*, 2014. **194**: p. 16-36.
34. Van Reeth, K., H. Nauwynck, et M. Pensaert, *Dual infections of feeder pigs with porcine reproductive and respiratory syndrome virus followed by porcine respiratory coronavirus or swine influenza virus: a clinical and virological study*. *Veterinary Microbiology*, 1996. **48**(3-4): p. 325-335.



35. Snijder, E.J. et J. Meulenbergh, *The molecular biology of arteriviruses*. Journal of General Virology, 1998. **79**(5): p. 961-979.
36. Snijder, E.J., M. Kikkert, et Y. Fang, *Arterivirus molecular biology and pathogenesis*. Journal of General Virology, 2013. **94**(10): p. 2141-2163.
37. Fang, Y. et E.J. Snijder, *The PRRSV replicase: exploring the multifunctionality of an intriguing set of nonstructural proteins*. Virus Research, 2010. **154**(1-2): p. 61-76.
38. Brown, E., et al., *Antibody response of nonstructural proteins: implications for diagnostic detection and differentiation of type I and type II porcine reproductive and respiratory syndrome viruses*. Clinical and Vaccine Immunology, 2009.
39. Li, Y., et al., *Nonstructural proteins nsp2TF and nsp2N of porcine reproductive and respiratory syndrome virus (PRRSV) play important roles in suppressing host innate immune responses*. Virology, 2018. **517**: p. 164-176.
40. Gonin, P., et al., *Seroneutralization of porcine reproductive and respiratory syndrome virus correlates with antibody response to the GP5 major envelope glycoprotein*. Journal of Veterinary Diagnostic Investigation, 1999. **11**(1): p. 20-26.
41. Pirzadeh, B. et S. Dea, *Immune response in pigs vaccinated with plasmid DNA encoding ORF5 of porcine reproductive and respiratory syndrome virus*. Journal of General Virology, 1998. **79**(5): p. 989-999.
42. Pirzadeh, B., C.A. Gagnon, et S. Dea, *Genomic and antigenic variations of porcine reproductive and respiratory syndrome virus major envelope GP5 glycoprotein*. Canadian Journal of Veterinary Research, 1998. **62**(3): p. 170.
43. Lambert, M.-È., et al., *Positioning Quebec ORF5 sequences of porcine reproductive and respiratory syndrome virus (PRRSV) within Canada and worldwide diversity*. Infection, Genetics and Evolution, 2019. **74**: p. 103999.
44. Music, N. et C.A. Gagnon, *The role of porcine reproductive and respiratory syndrome (PRRS) virus structural and non-structural proteins in virus pathogenesis*. Animal Health Research Reviews, 2010. **11**(2): p. 135-163.

45. Wissink, E., et al., *Envelope protein requirements for the assembly of infectious virions of porcine reproductive and respiratory syndrome virus*. Journal of Virology, 2005. **79**(19): p. 12495-12506.
46. Das, P.B., et al., *Glycosylation of minor envelope glycoproteins of porcine reproductive and respiratory syndrome virus in infectious virus recovery, receptor interaction, and immune response*. Virology, 2011. **410**(2): p. 385-394.
47. Wieringa, R., et al., *Structural protein requirements in equine arteritis virus assembly*. Journal of Virology, 2004. **78**(23): p. 13019-13027.
48. Lee, C. et D. Yoo, *The small envelope protein of porcine reproductive and respiratory syndrome virus possesses ion channel protein-like properties*. Virology, 2006. **355**(1): p. 30-43.
49. Nam, H.-M., et al., *Immune responses in mice vaccinated with virus-like particles composed of the GP5 and M proteins of porcine reproductive and respiratory syndrome virus*. Archives of Virology, 2013. **158**(6): p. 1275-1285.
50. Van Breedam, W., et al., *Porcine reproductive and respiratory syndrome virus entry into the porcine macrophage*. Journal of General Virology, 2010. **91**(7): p. 1659-1667.
51. Ostrowski, M., et al., *Identification of neutralizing and nonneutralizing epitopes in the porcine reproductive and respiratory syndrome virus GP5 ectodomain*. Journal of Virology, 2002. **76**(9): p. 4241-4250.
52. Lopez, O. et F. Osorio, *Role of neutralizing antibodies in PRRSV protective immunity*. Veterinary Immunology and Immunopathology, 2004. **102**(3): p. 155-163.
53. Robinson, S., et al., *Novel PRRSV ORF5a protein is not immunoprotective but drives GP5 glycosylation*. University of Minnesota, 2011.
54. Miller, W.A. et G. Koev, *Synthesis of subgenomic RNAs by positive-strand RNA viruses*. Virology, 2000. **273**(1): p. 1-8.
55. Kappes, M.A. et K.S. Faaberg, *PRRSV structure, replication and recombination: origin of phenotype and genotype diversity*. Virology, 2015. **479**: p. 475-486.
56. Welch, S.-K.W. et J.G. Calvert, *A brief review of CD163 and its role in PRRSV infection*. Virus Research, 2010. **154**(1-2): p. 98-103.

57. Vanderheijden, N., et al., *Involvement of sialoadhesin in entry of porcine reproductive and respiratory syndrome virus into porcine alveolar macrophages*. Journal of Virology, 2003. **77**(15): p. 8207-8215.
58. Pol, J., et al., *Pathological, ultrastructural, and immunohistochemical changes caused by Lelystad virus in experimentally induced infections of mystery swine disease (synonym: porcine epidemic abortion and respiratory syndrome (PEARS))*. Veterinary Quarterly, 1991. **13**(3): p. 137-143.
59. Zimmerman, J., et al., *Transmission of SIRS virus in convalescent animals to commingled penmates under experimental conditions*. Am. Assoc. Swine Pract. Newsl, 1992. **4**: p. 25.
60. Benfield, D., et al., *Pathogenesis and persistence of PRRS*. University of Minnesota, 1998.
61. Allende, R., et al., *Porcine reproductive and respiratory syndrome virus: description of persistence in individual pigs upon experimental infection*. Journal of Virology, 2000. **74**(22): p. 10834-10837.
62. Bierk, M., et al., *Diagnostic investigation of chronic porcine reproductive and respiratory syndrome virus in a breeding herd of pigs*. Veterinary Record, 2001. **148**(22): p. 687-690.
63. Wills, R.W., et al., *Duration of infection and proportion of pigs persistently infected with porcine reproductive and respiratory syndrome virus*. Journal of Clinical Microbiology, 2003. **41**(1): p. 58-62.
64. Sun, Z., et al., *Pathogenicity comparison between highly pathogenic and NADC30-like porcine reproductive and respiratory syndrome virus*. Archives of Virology, 2016. **161**(8): p. 2257-2261.
65. Cai, J.-P., et al., *Detection of asymptomatic antigenemia in pigs infected by porcine reproductive and respiratory syndrome virus (PRRSV) by a novel capture immunoassay with monoclonal antibodies against the nucleocapsid protein of PRRSV*. Clinical Vaccine Immunology, 2009. **16**(12): p. 1822-1828.
66. Charentantanakul, W., R. Platt, et J.A. Roth, *Effects of porcine reproductive and respiratory syndrome virus-infected antigen-presenting cells on T cell activation and antiviral cytokine production*. Viral Immunology, 2006. **19**(4): p. 646-661.

67. Bautista, E.M. et T.W. Molitor, *Cell-mediated immunity to porcine reproductive and respiratory syndrome virus in swine*. *Viral Immunology*, 1997. **10**(2): p. 83-94.
68. Fuertes, L.L., et al., *Analysis of cellular immune response in pigs recovered from porcine respiratory and reproductive syndrome infection*. *Virus Research*, 1999. **64**(1): p. 33-42.
69. Molitor, T.W., E. Bautista, et C. Choi, *Immunity to PRRSV: double-edged sword*. *Veterinary Microbiology*, 1997. **55**(1-4): p. 265-276.
70. Renukaradhya, G.J., et al., *Porcine reproductive and respiratory syndrome virus-induced immunosuppression exacerbates the inflammatory response to porcine respiratory coronavirus in pigs*. *Viral Immunology*, 2010. **23**(5): p. 457-466.
71. Sun, Z., et al., *The cysteine protease domain of porcine reproductive and respiratory syndrome virus nonstructural protein 2 possesses deubiquitinating and interferon antagonism functions*. *Journal of Virology*, 2010. **84**(15): p. 7832-7846.
72. Van Der Meer, Y., et al., *ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex*. *Journal of Virology*, 1998. **72**(8): p. 6689-6698.
73. Bautista, E., P. Suarez, et T.W. Molitor, *T cell responses to the structural polypeptides of porcine reproductive and respiratory syndrome virus*. *Archives of Virology*, 1999. **144**(1): p. 117-134.
74. Piras, F., et al., *Porcine Reproductive and Respiratory Syndrome (PRRS) virus-specific interferon- $\gamma$  T-cell responses after PRRS virus infection or vaccination with an inactivated PRRS vaccine*. *Viral Immunology*, 2005. **18**(2): p. 381-389.
75. Labarque, G.G., et al., *Effect of cellular changes and onset of humoral immunity on the replication of porcine reproductive and respiratory syndrome virus in the lungs of pigs*. *Journal of General Virology*, 2000. **81**(5): p. 1327-1334.
76. YOON, K.-J., et al., *Antibody-dependent enhancement (ADE) of porcine reproductive and respiratory syndrome virus (PRRSV) infection in pigs*. *Viral Immunology*, 1996. **9**(1): p. 51-63.
77. Costers, S., et al., *GP4-specific neutralizing antibodies might be a driving force in PRRSV evolution*. *Virus Research*, 2010. **154**(1-2): p. 104-113.

78. Popescu, L.N., et al., *GP5 of porcine reproductive and respiratory syndrome virus (PRRSV) as a target for homologous and broadly neutralizing antibodies*. *Veterinary Microbiology*, 2017. **209**: p. 90-96.
79. Ansari, I.H., et al., *Influence of N-linked glycosylation of porcine reproductive and respiratory syndrome virus GP5 on virus infectivity, antigenicity, and ability to induce neutralizing antibodies*. *Journal of Virology*, 2006. **80**(8): p. 3994-4004.
80. Brown, E., et al., *Antibody response to porcine reproductive and respiratory syndrome virus (PRRSV) nonstructural proteins and implications for diagnostic detection and differentiation of PRRSV types I and II*. *Clinical Vaccine Immunology*, 2009. **16**(5): p. 628-635.
81. Nodelijk, G., et al., *A quantitative assessment of the effectiveness of PRRSV vaccination in pigs under experimental conditions*. *Vaccine*, 2001. **19**(27): p. 3636-3644.
82. Shi, M., et al., *The spread of type 2 porcine reproductive and respiratory syndrome virus (PRRSV) in North America: a phylogeographic approach*. *Virology*, 2013. **447**(1-2): p. 146-154.
83. Nielsen, H.S., et al., *Reversion of a live porcine reproductive and respiratory syndrome virus vaccine investigated by parallel mutations*. *Journal of General Virology*, 2001. **82**(6): p. 1263-1272.
84. Storgaard, T., M. Oleksiewicz, et A. Bøtner, *Examination of the selective pressures on a live PRRS vaccine virus*. *Archives of Virology*, 1999. **144**(12): p. 2389-2401.
85. Mengeling, W.L., et al., *Strain specificity of the immune response of pigs following vaccination with various strains of porcine reproductive and respiratory syndrome virus*. *Veterinary Microbiology*, 2003. **93**(1): p. 13-24.
86. Zuckermann, F.A., et al., *Assessment of the efficacy of commercial porcine reproductive and respiratory syndrome virus (PRRSV) vaccines based on measurement of serologic response, frequency of gamma-IFN-producing cells and virological parameters of protection upon challenge*. *Veterinary Microbiology*, 2007. **123**(1-3): p. 69-85.
87. Worobey, M. et E.C. Holmes, *Evolutionary aspects of recombination in RNA viruses*. *Journal of General Virology*, 1999. **80**(10): p. 2535-2543.

88. Liu, D., et al., *Recombination analyses between two strains of porcine reproductive and respiratory syndrome virus in vivo*. Virus Research, 2011. **155**(2): p. 473-486.
89. Yuan, S., et al., *Recombination between North American strains of porcine reproductive and respiratory syndrome virus*. Virus Research, 1999. **61**(1): p. 87-98.
90. Murtaugh, M.P., et al., *The ever-expanding diversity of porcine reproductive and respiratory syndrome virus*. Virus Research, 2010. **154**(1-2): p. 18-30.
91. Petry, D., et al., *Differential immunity in pigs with high and low responses to porcine reproductive and respiratory syndrome virus infection*. Journal of Animal Science, 2007. **85**(9): p. 2075-2092.
92. Petry, D., et al., *Biological responses to porcine respiratory and reproductive syndrome virus in pigs of two genetic populations*. Journal of Animal Science, 2005. **83**(7): p. 1494-1502.
93. Kim, W.-I., et al., *Effect of genotypic and biotypic differences among PRRS viruses on the serologic assessment of pigs for virus infection*. Veterinary Microbiology, 2007. **123**(1-3): p. 1-14.
94. Nelsen, C.J., M.P. Murtaugh, et K.S. Faaberg, *Porcine reproductive and respiratory syndrome virus comparison: divergent evolution on two continents*. Journal of Virology, 1999. **73**(1): p. 270-280.
95. Murtaugh, M.P., M. Elam, et L. Kakach, *Comparison of the structural protein coding sequences of the VR-2332 and Lelystad virus strains of the PRRS virus*. Archives of Virology, 1995. **140**(8): p. 1451-1460.
96. Shi, M., et al., *Molecular epidemiology of PRRSV: a phylogenetic perspective*. Virus Research, 2010. **154**(1-2): p. 7-17.
97. Stadejek, T., et al., *Porcine reproductive and respiratory syndrome virus strains of exceptional diversity in eastern Europe support the definition of new genetic subtypes*. Journal of General Virology, 2006. **87**(7): p. 1835-1841.
98. Shi, M., et al., *Phylogeny-based evolutionary, demographical, and geographical dissection of North American type 2 porcine reproductive and respiratory syndrome viruses*. Journal of Virology, 2010. **84**(17): p. 8700-8711.

99. Sanger, F., S. Nicklen, et A.R. Coulson, *DNA sequencing with chain-terminating inhibitors*. Proceedings of the National Academy of Sciences, 1977. **74**(12): p. 5463-5467.
100. Collins, F.S., M. Morgan, et A. Patrinos, *The Human Genome Project: lessons from large-scale biology*. Science, 2003. **300**(5617): p. 286-290.
101. Kawashima, E., L. Farinelli, et P. Mayer, *Method of nucleic acid sequencing*. 2002, Google Patents.
102. Escalante, A.E., et al., *El estudio de la biodiversidad en la era de la secuenciación masiva*. Revista Mexicana de Biodiversidad, 2014. **85**(4): p. 1249-1264.
103. Besser, J., et al., *Next-generation sequencing technologies and their application to the study and control of bacterial infections*. Clinical Microbiology and Infection, 2018. **24**(4): p. 335-341.
104. Jain, M., et al., *The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community*. Genome Biology, 2016. **17**(1): p. 239.
105. Schneider, G.F. et C. Dekker, *DNA sequencing with nanopores*. Nature Biotechnology, 2012. **30**(4): p. 326-328.
106. Kelly, L., W.L. Gerlach, et P.M. Waterhouse, *Characterisation of the subgenomic RNAs of an Australian isolate of barley yellow dwarf luteovirus*. Virology, 1994. **202**(2): p. 565-573.
107. Dewey, C., et al., *Lelystad-like strain of porcine reproductive and respiratory syndrome virus (PRRSV) identified in Canadian swine*. The Canadian Veterinary Journal, 2000. **41**(6): p. 493.
108. Fang, Y., et al., *Diversity and evolution of a newly emerged North American Type 1 porcine arterivirus: analysis of isolates collected between 1999 and 2004*. Archives of Virology, 2007. **152**(5): p. 1009.
109. Chen, Z., et al., *Genetic diversity in envelope genes of contemporary US porcine reproductive and respiratory syndrome virus strains influences viral antigenicity*. Research in Veterinary Science, 2017. **115**: p. 432-441.
110. Choi, Y., et al., *Detection and subtyping of swine influenza H1N1, H1N2 and H3N2 viruses in clinical samples using two multiplex RT-PCR assays*. Journal of Virological Methods, 2002. **102**(1-2): p. 53-59.

111. Holtkamp, D.J., et al., *Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers*. Journal of Swine Health and Production, 2013. **21**(2): p. 72-84.
112. Mussell, A., *A Risk, Benefit, Strength, Weakness, Opportunity and Threat Analysis for the Control and Possible Eradication of Porcine Reproductive and Respiratory Syndrome (PRRS) Virus Within the Canadian Swine Herd*. Canadian Swine Health Board (CSHB) report, 2011: p. 1-122.
113. Linhares, D.C., C. Johnson, et R.B. Morrison, *Economic Analysis of Immunization Strategies for PRRS Control [corrected]*. PLoS One, 2015. **10**(12): p. e0144265.
114. Nathues, H., et al., *Cost of porcine reproductive and respiratory syndrome virus at individual farm level - An economic disease model*. Preventive Veterinary Medicine, 2017. **142**: p. 16-29.
115. Bilodeau, R., et al., *'Porcine reproductive and respiratory syndrome' in Quebec [letter] [see comments]*. Veterinary Records, 1991. **129**(5): p. 102-3.
116. Christianson, W.T., *Stillbirths, mummies, abortions, and early embryonic death*. Veterinary Clinics of North America: Food Animal Practice, 1992. **8**(3): p. 623-39.
117. Keffaber, K.K., *Reproductive failure of unknown etiology*. A.A.S.P. Newsletter, 1989. **1**: p. 1-10.
118. Loula, T., *Mystery pig Disease*. Agri-Practice, 1991. **12**: p. 23-34.
119. Pol, J.M., et al., *Pathological, ultrastructural, and immunohistochemical changes caused by Lelystad virus in experimentally induced infections of mystery swine disease (synonym: porcine epidemic abortion and respiratory syndrome (PEARS))*. Veterinary Quarterly, 1991. **13**(3): p. 137-43.
120. Albina, E., *Porcine reproductive and respiratory syndrome: ten years of experience (1986-1996) with this undesirable viral infection*. Veterinary Research, 1997. **28**(4): p. 305-52.
121. Collins, J.E., et al., *Isolation of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North America and experimental reproduction of the disease in gnotobiotic pigs*. Journal of Veterinary Diagnostic Investigation, 1992. **4**(2): p. 117-26.



122. Halbur, P.G., et al., *Comparison of the antigen distribution of two US porcine reproductive and respiratory syndrome virus isolates with that of the Lelystad virus*. *Veterinary Pathology*, 1996. **33**(2): p. 159-70.
123. King, A.M.Q., et al., *Virus Taxonomy, Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses*. 2012: Elsevier Academic Press.
124. Music, N. et C.A. Gagnon, *The role of porcine reproductive and respiratory syndrome (PRRS) virus structural and non-structural proteins in virus pathogenesis*. *Animal Health Research Review*, 2010: p. 135-163.
125. Spilman, M.S., et al., *Cryo-electron tomography of porcine reproductive and respiratory syndrome virus: organization of the nucleocapsid*. *Journal of General Virology*, 2009. **90**(Pt 3): p. 527-35.
126. Lunney, J.K., et al., *Porcine Reproductive and Respiratory Syndrome Virus (PRRSV): Pathogenesis and Interaction with the Immune System*. *Annual Review in Animal Bioscience*, 2016. **4**: p. 129-54.
127. Johnson, C.R., et al., *Novel structural protein in porcine reproductive and respiratory syndrome virus encoded by an alternative ORF5 present in all arteriviruses*. *Journal of General Virology*, 2011. **92**(Pt 5): p. 1107-16.
128. Dea, S., et al., *Current knowledge on the structural proteins of porcine reproductive and respiratory syndrome (PRRS) virus: comparison of the North American and European isolates*. *Archives of Virology*, 2000. **145**(4): p. 659-88.
129. Lambert, M.E., et al., *Positioning Quebec ORF5 sequences of porcine reproductive and respiratory syndrome virus (PRRSV) within Canada and worldwide diversity*. *Infection, Genetics and Evolution*, 2019. **74**: p. 103999.
130. Lambert, M.E., et al., *Porcine reproductive and respiratory syndrome virus: web-based interactive tools to support surveillance and control initiatives*. *Porcine Health Management*, 2019. **5**: p. 10.

131. Lambert, M.E., et al., *Evaluating an automated clustering approach in a perspective of ongoing surveillance of porcine reproductive and respiratory syndrome virus (PRRSV) field strains*. Infection, Genetics and Evolution, 2019. **73**: p. 295-305.
132. Pirzadeh, B., C.A. Gagnon, et S. Dea, *Genomic and antigenic variations of porcine reproductive and respiratory syndrome virus major envelope GP5 glycoprotein*. Canadian Journal of Veterinary Research, 1998. **62**(3): p. 170-7.
133. Gonin, P., et al., *Seroneutralization of porcine reproductive and respiratory syndrome virus correlates with antibody response to the GP5 major envelope glycoprotein*. Journal of Veterinary Diagnostic Investigation, 1999. **11**(1): p. 20-6.
134. Pirzadeh, B. et S. Dea, *Immune response in pigs vaccinated with plasmid DNA encoding ORF5 of porcine reproductive and respiratory syndrome virus*. Journal of General Virology, 1998. **79**(Pt 5): p. 989-99.
135. Chen, Z., et al., *Genetic diversity in envelope genes of contemporary U.S. porcine reproductive and respiratory syndrome virus strains influences viral antigenicity*. Research in Veterinary Science, 2017. **115**: p. 432-441.
136. Lalonde, C., C. Provost, et C.A. Gagnon. *Whole genome sequencing of PRRSV lead to a better classification of clinical cases*. in Conference of Research Workers in Animal Diseases (CRWAD) proceeding. 2019. Chicago, IL, United-States.
137. Datta, S., et al., *Next-generation sequencing in clinical virology: Discovery of new viruses*. World Journal of Virology, 2015. **4**(3): p. 265-76.
138. Fellers, J.P., et al., *Wheat Virus Identification Within Infected Tissue Using Nanopore Sequencing Technology*. Plant Disease, 2019. **103**(9): p. 2199-2203.
139. Gregory, A.C., et al., *Marine DNA Viral Macro- and Microdiversity from Pole to Pole*. Cell, 2019. **177**(5): p. 1109-1123 e14.
140. Hjelmso, M.H., et al., *Metagenomic analysis of viruses in toilet waste from long distance flights-A new procedure for global infectious disease surveillance*. PLoS One, 2019. **14**(1): p. e0210368.

141. Palinski, R., et al., *A Novel Porcine Circovirus Distantly Related to Known Circoviruses Is Associated with Porcine Dermatitis and Nephropathy Syndrome and Reproductive Failure*. Journal of Virology, 2017. **91**(1).
142. Hause, B.M., et al., *Discovery of a novel putative atypical porcine pestivirus in pigs in the USA*. Journal of General Virology, 2015. **96**(10): p. 2994-8.
143. Hause, B.M., R.A. Hesse, et J. Bai, *Development of next generation sequencing methodology for full genome characterization of porcine reproductive and respiratory syndrome virus (PRRSV) from oral fluids and nasal swabs*. 2015, Pork checkoff.
144. Zhang, J., et al., *High-throughput whole genome sequencing of Porcine reproductive and respiratory syndrome virus from cell culture materials and clinical specimens using next-generation sequencing technology*. Journal of Veterinary Diagnostic Investigation, 2017. **29**(1): p. 41-50.
145. Tan, S., C.M. Dvorak, et M.P. Murtaugh, *Rapid, Unbiased PRRSV Strain Detection Using MinION Direct RNA Sequencing and Bioinformatics Tools*. Viruses, 2019. **11**(12): p. 1132.
146. Wongsurawat, T., et al., *Rapid Sequencing of Multiple RNA Viruses in Their Native Form*. Frontiers in Microbiology, 2019. **10**: p. 260.
147. Yan, L., et al., *Inhibition monitoring in veterinary molecular testing*. Journal of Veterinary Diagnostic Investigation, 2019: p. 1040638719889315.
148. Naccache, S.N., et al., *A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples*. Genome Research, 2014. **24**(7): p. 1180-92.
149. Mueller, O., S. Lightfoot, et A. Schroeder, *RNA Integrity Number (RIN) – Standardization of RNA Quality Control*. 2016, Agilent Technologies. p. 1-8.
150. Li, Y. et K.V. Kowdley, *Method for microRNA isolation from clinical serum samples*. Anal of Biochemistry, 2012. **431**(1): p. 69-75.
151. Qi, P., et al., *Serum microRNAs as biomarkers for hepatocellular carcinoma in Chinese patients with chronic hepatitis B virus infection*. PLoS One, 2011. **6**(12): p. e28486.
152. Snijder, E.J. et J.J. Meulenber, *The molecular biology of arteriviruses*. Journal of General Virology, 1998. **79**(Pt 5): p. 961-79.

153. Balka, G., et al., *Full genome sequence analysis of a wild, non-MLV-related type 2 Hungarian PRRSV variant isolated in Europe*. *Virus Research*, 2015. **200**: p. 1-8.
154. Balint, A., et al., *Full-length genome sequence analysis of a Hungarian porcine reproductive and respiratory syndrome virus isolated from a pig with severe respiratory disease*. *Archives of Virology*, 2015. **160**(2): p. 417-22.
155. Brar, M.S., et al., *Genomic evolution of porcine reproductive and respiratory syndrome virus (PRRSV) isolates revealed by deep sequencing*. *PLoS One*, 2014. **9**(4): p. e88807.
156. Kvisgaard, L.K., et al., *A fast and robust method for full genome sequencing of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Type 1 and Type 2*. *Journal of Virological Methods*, 2013. **193**(2): p. 697-705.
157. Lu, Z.H., et al., *Genomic variation in macrophage-cultured European porcine reproductive and respiratory syndrome virus Olot/91 revealed using ultra-deep next generation sequencing*. *Virology Journal*, 2014. **11**: p. 42.
158. Dong, J., et al., *Quantitative estimation of the replication kinetics of genotype 2 PRRSV strains with different levels of virulence in vitro*. *Journal of Virological Methods*, 2016. **234**: p. 87-9.
159. Linhares, D.C., C. Johnson, et R.B. Morrison, *Economic analysis of vaccination strategies for PRRS control*. *PloS one*, 2015. **10**(12): p. e0144265.
160. Nathues, H., et al., *Cost of porcine reproductive and respiratory syndrome virus at individual farm level—an economic disease model*. *Preventive Veterinary Medicine*, 2017. **142**: p. 16-29.
161. Cao, Q.M., et al., *The non-structural protein Nsp2TF of porcine reproductive and respiratory syndrome virus down-regulates the expression of Swine Leukocyte Antigen class I*. *Virology*, 2016. **491**: p. 115-124.
162. Li, Y., et al., *Transactivation of programmed ribosomal frameshifting by a viral protein*. *Proceedings of the National Academy of Sciences*, 2014. **111**(21): p. E2172-E2181.
163. Zhang, M., et al., *Mutagenesis analysis of porcine reproductive and respiratory syndrome virus nonstructural protein 7*. *Virus Genes*, 2013. **47**(3): p. 467-477.

164. van Dinten, L.C., et al., *Proteolytic processing of the open reading frame 1b-encoded part of arterivirus replicase is mediated by nsp4 serine protease and is essential for virus replication*. Journal of Virology, 1999. **73**(3): p. 2027-2037.
165. Memorandums, M.-L., *A revision of the system of nomenclature for influenza viruses: a WHO memorandum*. Bulletin of the World Health Organization, 1980. **58**(4): p. 585-591.
166. Yu, L.x., et al., *The emergence of a highly pathogenic porcine reproductive and respiratory syndrome virus with additional 120aa deletion in Nsp2 region in Jiangxi, China*. Transboundary and Emerging Diseases, 2018. **65**(6): p. 1740-1748.
167. Choi, H.-W., et al., *Genomic analysis and pathogenic characteristics of type 2 porcine reproductive and respiratory syndrome virus nsp2 deletion strains isolated in Korea*. Veterinary Microbiology, 2014. **170**(3-4): p. 232-245.
168. Kim, D.-Y., et al., *Insertion and deletion in a non-essential region of the nonstructural protein 2 (nsp2) of porcine reproductive and respiratory syndrome (PRRS) virus: effects on virulence and immunogenicity*. Virus Genes, 2009. **38**(1): p. 118-128.
169. Wang, A., et al., *Genetic diversity of porcine reproductive and respiratory syndrome virus 1 in the United States of America from 2010 to 2018*. Veterinary Microbiology, 2019. **239**: p. 108486.
170. Watson, J.D., *Molecular biology of the gene*. 2004: Pearson Education India.
171. Ring, J.D., et al., *A performance evaluation of Nextera XT and KAPA HyperPlus for rapid Illumina library preparation of long-range mitogenome amplicons*. Forensic Science International: Genetics, 2017. **29**: p. 174-180.
172. King, J.L., et al., *High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq*. Forensic Science International: Genetics, 2014. **12**: p. 128-135.
173. Lan, J.H., et al., *Impact of three Illumina library construction methods on GC bias and HLA genotype calling*. Human Immunology, 2015. **76**(2-3): p. 166-175.
174. Amonsin, A., et al., *Comparative analysis of complete nucleotide sequence of porcine reproductive and respiratory syndrome virus (PRRSV) isolates in Thailand (US and EU genotypes)*. Virology Journal, 2009. **6**(1): p. 143.

175. Zhang, W., et al., *A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies*. PloS one, 2011. **6**(3).
176. Kumar, S. et M.L. Blaxter, *Comparing de novo assemblers for 454 transcriptome data*. BMC Genomics, 2010. **11**(1): p. 571.
177. Mikheyev, A.S. et M.M. Tin, *A first look at the Oxford Nanopore MinION sequencer*. Molecular Ecology Resources, 2014. **14**(6): p. 1097-1102.
178. Laver, T., et al., *Assessing the performance of the oxford nanopore technologies minion*. Biomolecular Detection and Quantification, 2015. **3**: p. 1-8.
179. Domingo, E., J. Sheldon, et C. Perales, *Viral quasispecies evolution*. Microbiology and Molecular Biology Review, 2012. **76**(2): p. 159-216.
180. Wagner, N., Y. Atsmon-Raz, et G. Ashkenasy, *Theoretical models of generalized quasispecies*, in *Quasispecies: From Theory to Experimental Systems*. 2015, Springer. p. 141-159.
181. Takahashi, K., et al., *Deep-sequence identification and role in virus replication of a JC virus quasispecies in patients with progressive multifocal leukoencephalopathy*. Journal of Virology, 2017. **91**(1): p. e01335-16.
182. Frias-Staheli, N., et al., *Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses*. Cell Host & Microbe, 2007. **2**(6): p. 404-416.
183. Zhou, L., et al., *The 30-amino-acid deletion in the Nsp2 of highly pathogenic porcine reproductive and respiratory syndrome virus emerging in China is not related to its virulence*. Journal of Virology, 2009. **83**(10): p. 5156-5167.
184. Meulenbergh, J.J., et al., *Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV*. Virology, 1993. **192**(1): p. 62-72.
185. Ziebuhr, J., E.J. Snijder, et A.E. Gorbalenya, *Virus-encoded proteinases and proteolytic processing in the Nidovirales*. Journal of General Virology, 2000. **81**(4): p. 853-879.
186. Davidson, C.J., et al., *Improving the limit of detection for Sanger sequencing: a comparison of methodologies for KRAS variant detection*. Biotechniques, 2012. **53**(3): p. 182-188.

187. Mohamed, S., et al., *Comparison of ultra-deep versus Sanger sequencing detection of minority mutations on the HIV-1 drug resistance interpretations after virological failure.* Aids, 2014. **28**(9): p. 1315-1324.
188. Nedumpun, T., et al., *Negative immunomodulatory effects of type 2 porcine reproductive and respiratory syndrome virus (PRRSV)-induced interleukin-1 receptor antagonist (IL-1Ra) on porcine innate and adaptive immune functions.* Frontiers in Immunology, 2019. **10**: p. 579.
189. Thanawongnuwech, R., et al., *Differential production of proinflammatory cytokines: in vitro PRRSV and Mycoplasma hyopneumoniae co-infection model.* Veterinary Immunology and immunopathology, 2001. **79**(1-2): p. 115-127.
190. Cecere, T., et al., *Co-infection of porcine dendritic cells with porcine circovirus type 2a (PCV2a) and genotype II porcine reproductive and respiratory syndrome virus (PRRSV) induces CD4+ CD25+ FoxP3+ T cells in vitro.* Veterinary Microbiology, 2012. **160**(1-2): p. 233-239.
191. Grau-Roma, L. et J. Segalés, *Detection of porcine reproductive and respiratory syndrome virus, porcine circovirus type 2, swine influenza virus and Aujeszky's disease virus in cases of porcine proliferative and necrotizing pneumonia (PNP) in Spain.* Veterinary Microbiology, 2007. **119**(2-4): p. 144-151.
192. Mardassi, H., S. Mounir, et S. Dea, *Molecular analysis of the ORFs 3 to 7 of porcine reproductive and respiratory syndrome virus, Quebec reference strain.* Archives of Virology, 1995. **140**(8): p. 1405-1418.
193. Janková, J. et V. Celer, *Expression and serological reactivity of Nsp7 protein of PRRS genotype I virus.* Research in Veterinary Science, 2012. **93**(3): p. 1537-1542.
194. Papatsiros, V., *Porcine herd health management practices for the control of PRRSV infection.* A Bird's-Eye View of Veterinary Medicine. Rijeka, Croatia: InTech, 2012: p. 281-300.
195. Zhao, H., et al., *Emergence of mosaic recombinant strains potentially associated with vaccine JXA1-R and predominant circulating strains of porcine reproductive and*

*respiratory syndrome virus in different provinces of China*. Virology Journal, 2017. **14**(1): p. 67.



## **Annexes**

### **Liste des réalisations en cours de Maîtrise.**

#### **Bourses et prix:**

- Septembre 2017      Bourse de dépannage du CRIPA.
- Juin 2018            Bourse de Congrès du CRIPA.
- Mai 2019            Best graduate student presentation, 18<sup>th</sup> symposium of the Canadian Animal Health Laboratorians Network.
- Novembre 2019      2019 NA PRRSV Symposium travel fellowship.

#### **Implications sociales et académiques**

- Juin 2017 – Mai 2018      Président du comité étudiant, CRIPA.
- Septembre 2018 - 2019      Coordonnateur aux cycles supérieurs, Association étudiante. (AEMVQ).
- Octobre 2018 & 2019      Conférencier invité sur les vitamines hydrosolubles, cours de 2<sup>e</sup> année au doctorat en médecine vétérinaire de l'université de Montréal.
- Octobre 2019 – Mars 2020      Auxiliaire d'enseignement, Faculté de médecine vétérinaire.

#### **Conférences et présentations:**

- 12 Mars 2018            Formation continue aux médecins vétérinaires, conférence de 45 minutes. Génome complet du virus SRRP: Surprises lors des premières analyses.
- 13-15 Juin 2018          Genomic characterization of porcine reproductive and respiratory syndrome virus (PRRSV) Quebec strains within clinical samples using whole genome

sequencing, présentation par affiche, 2<sup>nd</sup> Symposium of the Canadian Society for Virology.

- 8 Novembre 2018 Mise au point du protocole de préparation d'échantillons cliniques pour le séquençage à haut débit (SHD) du génome entier du virus du syndrome reproducteur et respiratoire porcin (VSRRP), présentation par affiche, Journée de la recherche de la FMV.
- 8 Novembre 2018 Characterization and quantification of transcobalamin cell receptor (TCII-R, CD320) in dog small intestine, présentation par affiche, Journée de la recherche de la FMV.
- 26-29 Mai 2019 Whole genome sequencing of PRRSV from clinical samples, présentation orale de 15 minutes, 18<sup>th</sup> annual meeting of the Canadian Animal Health Laboratorians Network.
- 2-5 novembre 2019 Whole genome sequencing of PRRSV leads to a better classification of clinical cases, présentation orale de 15 minutes, 100<sup>th</sup> Conference of research workers on animal diseases.

## **Publications**

Le Net, Rozenn, Chantale Provost, Christian Lalonde, Lyette Régimbald, Francois Vézina, Carl A. Gagnon, and Stéphane Lair. "Whole Genome Sequencing of an avipoxvirus associated with infections in a group of aviary-housed snow buntings (*Plectrophenax nivalis*).” *Journal of Zoo and Wildlife Medicine* 50, no. 4 (2020): 803-812.

### **En révision:**

Lalonde, Christian, Provost, Chantale, Gagnon, Carl A. “Whole genome sequencing of porcine reproductive and respiratory syndrome virus (PRRSV) from field clinical samples improves the genomic surveillance of the virus”, *Journal of Clinical Microbiology* X....

Gagnon, Carl A., Lalonde, Christian, Provost, Chantale. “Porcine reproductive and respiratory syndrome virus (PRRSV) whole genome sequencing (WGS) efficacy with field clinical samples using a poly(A)-tail viral genome purification method”, *Journal of Veterinary Diagnostic Investigation* X...