Infection, Genetics and Evolution 12 (2012) 1111–1119

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Porcine reproductive and respiratory syndrome virus diversity of Eastern Canada swine herds in a large sequence dataset reveals two hypervariable regions under positive selection

Benjamin Delisle *, Carl A. Gagnon, Marie-Ève Lambert, Sylvie D'Allaire

Faculté de Médecine Vétérinaire, Université de Montréal, C.P. 5000, St-Hyacinthe, Québec, Canada J2S 7C6

ARTICLE INFO

Article history:
Received 6 October 2011
Received in revised form 13 March 2012
Accepted 18 March 2012
Available online 29 March 2012

Keywords:
Porcine reproductive and respiratory syndrome virus (PRRSV)
Evolution
ORF5
GP5
Positive selection
Hypervariable region

ABSTRACT

Porcine reproductive and respiratory syndrome virus (PRRSV) is known to be genetically highly variable, but knowledge of sequence diversity from Eastern Canada and its degree of genetic plasticity in or near the principal neutralizing epitope (PNE) in association with evolutionary selective pressure is limited. The purposes of our study were to investigate the extent of strain diversity, the existing glycotypes and the amino acid sites under selective evolutionary pressure in its encoded protein, GP5, for a dataset of 1301 sequences (1998–2009). This was addressed by partitioning and clustering into subgenotypes a large number of open reading frame 5 sequences from the province of Quebec and analyzing the content of these subgenotypes. The overall pairwise diversity was 12% and was comparable to what has been reported around the world. The mean diversity for sequences within subgenotypes was around 7%. No marked variations in subgenotype emergence could be observed through time. Thirty-eight GP5 glycotype patterns were observed which included a newly identified site at position N57 which was already present in 1998. These patterns possessed one to six N-glycosylation sites in total and could be located in eight different positions. No obvious grouping of glycotypes could be established in relation to subgenotypes. Positions N44 and N51 were confirmed to be fixed N-glycosylation positions, whereas other positions where found to be shifting and located in or near hypervariable regions (HVRs) 1 and 2. Both HVRs were under selective evolutionary pressure in half of all subgenotypes including vaccine-like groups. Conversely, the PNE flanked by both HVRs was well conserved among most subgenotypes demonstrating potential molecular constraint in a probable viral binding region. The analysis of this dataset increased knowledge of evolutionary change inferred from genetic data, more specifically regarding the implications of both HVRs in PRRSV diversity.

1. Introduction

Porcine reproductive and respiratory syndrome virus (PRRSV) is one of the major problems facing the swine industry at the present time. It is characterized by severe reproductive failure including late term abortions and early farrowing in sows, and respiratory disease and mortality in young pigs. PRRSV is also known to be genetically highly variable, but there is a lack of information on the extent of strain diversity present in Eastern Canada and its degree of genetic plasticity in viral epitopes recognized by neutralizing antibodies (NAb) (Kimman et al., 2009; Shi et al., 2010a,b). PRRSV belongs to the *Arteriviridae* family, genus *Arterivirus* and is a positive sense single strand RNA virus. The genome is composed of nine open reading frames (ORF): 1a–1b (replication protein), 2a–2b (structure protein), 3–5 (envelope protein), 6 (membrane protein) and 7 (nucleocapsid protein) (Music and Gagnon, 2010). Recently a new ORF was discovered, ORF5a, which is translated from an alternative reading frame mainly overlapping the beginning of ORF5 (Johnson et al., 2011). The possible roles for this new structural protein have not been completely clarified but involvement in virus production seems to be one of its attributes (Firth et al., 2011). Of these proteins, GP4 (ORF4) for the European (EU) type 1 genotype and GP5 (ORF5) for the North American (NA) type 2 genotype are outer component proteins of the virion considered to be the most important immunodominant epitope (Costers et al., 2010; Ostrowski et al., 2002). Knowing that in Canada, only NA genotypes are routinely observed in the field and combined with the key features (identification of strain relatedness, induction of neutralizing antibodies and viral attachment to host cell) of this genomic region, compelled us to use ORF5 nucleotide (nt) and GP5 amino acid (aa) sequences for describing strain diversity and evolution of PRRSV.

* Corresponding author. Tel.: +1 450 773 8521x8493; fax: +1 450 778 8120.
E-mail addresses: benjamin.delisle@umontreal.ca (B. Delisle), carl.a.gagnon@umontreal.ca (C.A. Gagnon), marie-eve.lambert@umontreal.ca (Marie-Ève Lambert), sylvie.dallaire@umontreal.ca (S. D'Allaire).

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2012.03.015

Among the useful information that can be obtained from viral aa sequences is the location of potential glycosylation sites. This is one of the strategies employed by some viruses to evade host immune system and has been reported in human immunodeficiency virus (HIV), lactate dehydrogenase virus and also PRRSV (Chen et al., 2000; Kimman et al., 2009; Zhang et al., 2004). The PRRSV GP5 protein ectodomain has some variable N-glycosylation (N-gly) sites, which play a critical role in both immune evasion by glycan shielding and infectivity (Ansari et al., 2006; Faaberg et al., 2006). Gain or loss of one or many glycosylation sites may have an impact on sterical masking of immunogenic epitopes and hence can change the outcome of infection by changing virus conformation and/or allowing evasion of host immune response (Zhang et al., 2004).

Specific adaptive genomic regions can be distinguished and characterized by pinpointing sites which preferentially undergo aa change. These sites under selective pressure are determined by evaluating positive selection at each aa site, which globally favors the survival of the fittest variants to infect and persist within the host. The analysis of these positive aa sites within a large dataset may generate useful information in perspective of known biologically meaningful regions and epitopes. Major regions of importance of GP5 have been identified in previous studies and are enumerated here in a sequential order: the peptide signal, the decoy or epitope A, hypervariable region 1 (HVR1), principal neutralizing epitope (PNE) or epitope B, HVR2 and transmembrane regions (TM1, TM2, TM3) (Music and Gagnon, 2010; Zhou et al., 2009b).

The objectives of this study were to investigate the extent of strain diversity, the existing glycotypes and the amino acid sites under selective evolutionary pressure in its encoded protein, GP5, for a dataset of 1301 sequences (1998–2009).

2. Materials and methods

2.1. Sample collection

Samples and information on sampling time were collected through passive (from January 1998 onto August 2009) and active (from May 2005 onto August 2008) surveillance of PRRSV in the province of Quebec. These samples originated either from herds (1) experiencing reproductive or respiratory problems, (2) monitoring regularly their PRRSV status, or (3) by participating in a research project on risk factors and active surveillance for PRRSV (Lambert et al., 2012). Samples came from 586 different pig production sites. Among these samples, 87 were from PRRSV vaccinated herds, 244 from non-vaccinated herds and for 970 samples the vaccine status of the herd was unknown. In Canada, live attenuated vaccines for type 2 PRRSV, available at the time of the study, included MLV (Ingelvac PRRS MLV; Boehringer Ingelheim, Canada) and ATP (Ingelvac PRRS ATP; Boehringer Ingelheim, Canada).

2.2. RT-PCR and sequencing

The following procedures were performed to obtain the ORF5 sequences. RNA was extracted from tissues or serum with QIAamp viral RNA mini kit according to the manufacturer's instructions (Qiagen Inc., Mississauga, Ontario, Canada). Subsequently, RT-PCR was performed using Qiagen OneStep RT-PCR Kit and primers 5FN and 5DN as previously described (Larochelle et al., 2003). Finally, PCR products were purified before sequencing with Qiaquick spin kit (Qiagen). ORF5 sequencing was done on both strands of PCR amplicons using the same RT-PCR primers with BigDye terminator on ABI PRISM 310 Genetic analyzer (Applied Biosystems Canada, Streetsville, Ontario, Canada). Described procedures were performed at the molecular diagnostic laboratory at the Faculty of Veterinary Medicine of the University of Montreal for sequences obtained through the University. Some sequences were also obtained from other laboratories using similar procedures.

2.3. K-means partitioning and phylogenetic analysis

A total of 1301 NA ORF5 PRRSV field sampled sequences were available for our initial investigation. These sequences ($n = 1301$) were aligned by Bionumerics software v6.5 (Applied Maths Inc., Austin, TX, USA) and screened for recombination by Recombination Detection Program v3.44 (RDP3) (Martin et al., 2010). Six potential recombinant sequences were identified by software and ascertained by visual comparison of ORF5 alignment. These natural recombinants and nine descendant strains with evidence of the same recombination were removed from the alignment along with two sequences from the province of Ontario leaving 1284 sequences. Lineage analysis was done on this whole dataset (1284 sequences) in a Neighbor Joining tree (Kimura 2 parameter model) with reference sequences described elsewhere (Shi et al., 2010b), and allowed us to distinguish wild-type sequences ($n = 1097$) from vaccine associated sequences ($n = 187$). Subsequently, partitioning of wild-type sequences into subgenotypes was determined using the K-means algorithm (maximum of 15 groups) with an iterative process in Bionumerics software. Through this process, sequences that showed partitioning instability as determined by jacknife statistical method and also groups having less than 10 sequences were removed after each iteration. Eleven iterations were required to obtain stable groupings from the dataset of which 45 sequences were unstable and another 41 sequences grouped in clusters of less than 10. These were removed from the dataset and left 1011 wild-type sequences to be analyzed by maximum likelihood for phylogenetic relatedness and determination of subgenotypes in PhyML web interface (Guindon et al., 2010). A general time-reversible nucleotide substitution model was chosen with four categories of gamma-distributed rate heterogeneity. The subtree pruning and regrafting (SPR) was used to optimize starting BIONJ tree and a Chi^2-based approximate likelihood ratio test (aLRT) was determined to provide branch support for maximum likelihood tree. Ultimately, final subgenotype clustering was determined by phylogenetic groupings with a good aLRT (e.g. $\geqslant 90\%$) at critical tree nodes and also those showing congruence with previously described K-means partitioning. The average pairwise distance within each subgenotype was also calculated including minimum and maximum comparison values within the matrix. A reference nt sequence for each subgenotype was submitted to GenBank database and were given the accession numbers JQ691584 to JQ691601 (see in Supplementary Table S1 for year of sampling and corresponding subgenotype).

2.4. Glycosylation sites

The distribution of N-gly sites were determined from the whole dataset ($n = 1284$) using NetNGlyc 1.0 Server web utility (http://www.cbs.dtu.dk/services/NetNGlyc/). A glycosylation "potential" score which is the result of the analysis of sequences by nine neural networks and the agreement of each network was determined. A minimum default threshold (0.50) was used to determine positive from negative sites, and additional thresholds (0.75 and 0.90) were also used to identify positive sites with a higher confidence level.

2.5. Amino acid substitution rates

The estimation of substitution rates of non-synonymous (dN) and synonymous (dS) for codon aligned nt sequences within wild-type and vaccine-like subgenotypes were calculated by Synonymous Non-synonymous Analysis Program (SNAP web

utility) based on the Nei and Gojobori method (Nei and Gojobori, 1986) (http://hcv.lanl.gov/content/sequence/SNAP/SNAP.html). Estimates of aa substitutions were represented by the method of dN–dS rates: dN–dS > 0, dN–dS < 0 and dN–dS = 0 representing tendencies for non-synonymous and synonymous variation, and neutral mutations, respectively (Kosakovsky Pond and Frost, 2005).

2.6. Selective pressure

The CODEML program in PAML 4.3 software was used to calculate the positive selection pressure at aa positions with the codon-substitution model M2a (positive selection) (Yang, 1997, 2007). Firstly, the existence of positive selection was tested by likelihood ratio tests (LRT) by comparison of the alternative model M2a with the null model M1a (nearly neutral). In LRT twice the log likelihood difference, $2\Delta\ell = 2(\ell_1 - \ell_0)$, was compared with a Chi^2 distribution to test whether the null model was to be rejected, where ℓ_1 and ℓ_0 were the log likelihood for the alternative model and the null model, respectively at Chi^2 critical value of 9.21 (df = 2) for 1% significance level (Yang, 2007). Afterwards, when positive selection was ascertained within a subgenotype, a dN/dS ratio (ω) for each codon was calculated by the CODEML program in which the Bayes Empirical Bayes analysis then calculated the posterior probability (P – 1) that a site belonged to the class of $\omega > 1$ so inference, at 95% and 99% significance, of positively selected sites could be made.

3. Results

3.1. Nucleotide subgenotype diversity and temporal distribution

All sequences in this study were of the NA genotype and varied between 597 and 606 base pairs (bp) in length with 96.3% having 603 bp. Using the classification of Shi et al. (2010a,b), only lineages 1, 5 and 8 were identified in the whole dataset for the province of Quebec (data not shown). The largest pairwise distance found in the whole dataset matrix was 23.1% and the average pairwise diversity was 12.2%.

The analysis of the wild-type sequence dataset ($n = 1011$) revealed 18 subgenotypes (Fig. 1). This excluded vaccine-like sequences (ATP, $n = 29$ and MLV, $n = 158$) previously identified and removed from K-means and maximum likelihood analysis. All subgenotypes were cut-off at 7% or less diversity by K-means, with the exceptions of subgenotypes 2 and 17, for which a monophyletic natural break in the phylogenetic tree was done to obtain a meaningful clustering. All subgenotypes were supported by good aLRT values (e.g. $\geqslant$ 90%). Forty sequences did not regroup in any of the subgenotypes and were disregarded because of an insufficient number of sequences ($n \leqslant 9$). The final dataset with wild-type and vaccine associated subgenotypes consisted of 1158 sequences without the 40 sequences that did not regroup. The subgenotypes were of variable sizes, ranging from 10 sequences in the smallest to 158 for the largest (Table 1). Intra subgenotype pairwise diversity ranged from 9.0% in the most diverse subgenotype (# 2) as compared to 1.1% in the least divergent one (ATP). The mean pairwise diversity for wild-type subgenotypes was 5.8% and that of MLV subgenotype was of 2.4%. These subgenotypes were first identified between 1998 and 2005 and were still present up until the two last years of the project (Table 1), with one exception for subgenotype 5 which was only present from 2004 to 2006 and consisted of 14 sequences.

3.2. Glycosylation site variations

Potential N-gly sites were observed at eight different positions: N30, N32, N33, N34, N35, N44, N51 and N57 (Table 2). No obvious relation could be found between N-gly pattern and a subgenotype. The positions and total numbers of N-gly sites within a subgenotype exhibited variable patterns (data not shown). N-gly sites variations were located mainly between aa 30–35 near or within the HVR1 region (aa 32–34). The N44 and N51 N-gly sites were conserved in the PNE region except for a few sequences ($n = 30$ and $n = 6$, respectively), and inversely N35 was scarcely seen in sequences ($n = 43$). The N57 N-gly site in HVR2, aa 57–59, is a newly described region of interest and was found in 106 sequences. The N-gly site N57 was already present at the beginning of 1998. Globally, N57 was present in six subgenotypes (4, 6, 7, 8, 16 and 17), and accounted for 66% and 49% of all sequences in subgenotypes 6 and 7, respectively. These two subgenotypes both regrouped 69% of all N57 sites found. The proportion of sequences which harbored N57 was slightly greater from 1998 to 2003 (20%) than during the period 2004 to 2009 (6%) and these were associated with the glycotypes 9, 12 and 13 (Table 2). The most frequent positions with a higher confidence level ($\geqslant$ 0.75) in the glycosylation score were located at sites N34 and N44, while N51 was the most frequent site present in the dataset. The number of potential sites for each sequence varied from 1 to 6, but had generally either 3 or 4 for a high proportion of sequences (Table 2).

3.3. Amino acid variations and positive selective pressure

Positive selection was determined to be present in 17 out of 20 subgenotypes (Table 3) with LRT for positive selection model (M2a – M1a = $2\Delta\ell$) ranging from 14.8 to 244.4 with a Chi^2 critical value of 9.21 (df = 2) for 1% significance level. The overall analysis of aa substitution per site (non-synonymous and synonymous) showed that all the wild-type subgenotypes had more dS then dN variation. Nevertheless, aa changes occurred more frequently in some known or suspected to be important biological regions and in identified antigenic regions. As an example, the dN–dS rate tendencies per site are depicted for subgenotype 8 in Fig. 2 with significant positive selection (red lines), antigenic regions (upper rectangles) and important biological regions (lower rectangles) (Supplementary Fig. S2).

The signal peptide region, aa 1–26, was characterized by several non-synonymous changes across all sequences, particularly at aa positions 13 and 15 where significant positive selection was found in at least one of these sites for 12 subgenotypes (Table 3). Few synonymous variations were seen in peptide signal region in all subgenotypes. No positive selection was found in the decoy region, aa 27–30 in any of the subgenotypes. Both HVR1 and HVR2 had high dN values at different positions and at least one significant site under positive selection in 17 out of 20 subgenotypes. Significant positive selection was present in 11 subgenotypes for HVR1 (P-1 $\geqslant$ 95%) and 16 subgenotypes for HVR2 (P-1 $\geqslant$ 95%) (Table 3). HVR2 was the region that had the highest number of sites by subgenotype with positive selection. In HVR2, four subgenotypes had one aa position under selection, nine had two positions and three had all three positions under positive selection. The PNE region or epitope B, aa 37–52, had very few non-synonymous mutations. Even though this region is shown to be fairly well conserved, some synonymous mutations occurred for all subgenotypes in this region. Interestingly, positions 102 and 104 had a high dN rate in many subgenotypes and showed positive selection in 5 subgenotypes. Other amino sites under positive selection in one or two subgenotypes were 3, 4, 8, 10, 106, 151, 170, 173 and 196.

4. Discussion

The analysis of nt and aa diversity of a large sequence dataset of the ORF5 gene increased our knowledge on evolutionary change

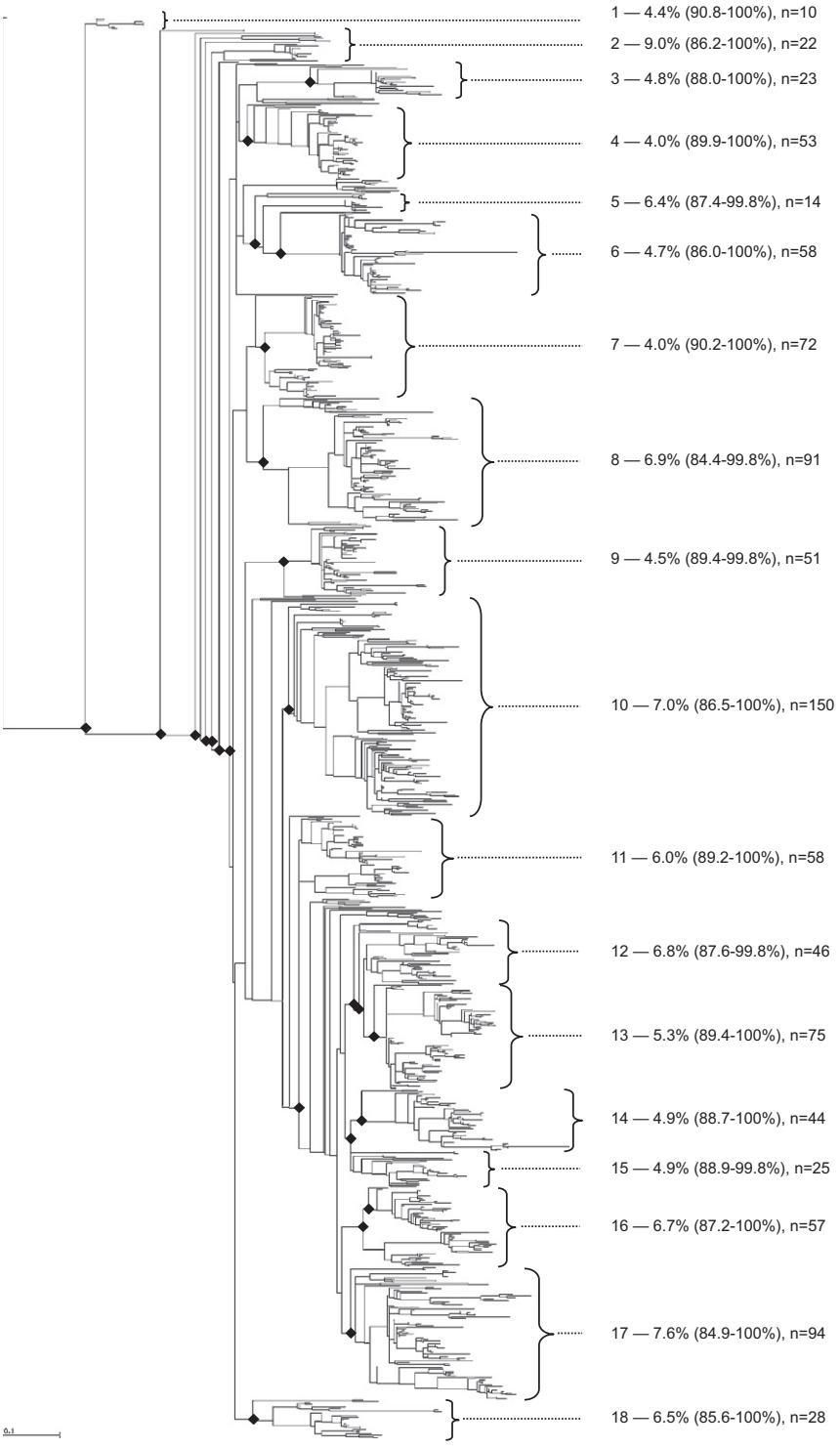

Fig. 1. Phylogenetic tree obtained by maximum likelihood forming 18 subgenotypes from 1011 wild-type PRRSV ORF5 sequences. Subgenotype numbering, average pairwise diversity, minimum and maximum similarity and the number of sequences per cluster are shown. Black diamond at nodes between the root and the subgenotypes show the approximate likelihood-ratio test values $\geqslant$90%.

inferred from genetic data. The average pairwise diversity of 12.2% found between all NA sequences in the province of Quebec was surprisingly close to that of 12.5% determined for all available NA sequences around the world (Shi et al., 2010b) and this, even though vaccine strains were also included in a proportion of 14%. This should only have lessened the average diversity, since these subgenotypes are highly homologous. This diversity is also represented by the largest pairwise distance found in Quebec of 23.1%, versus that found in the whole world 27.8% (Shi et al., 2010b). In an evolutionary perspective, these results are interesting considering that the latter study did not have many sequences from Quebec. Even though the number of sequences analyzed in our study was considerably larger than that of a previous Quebec study (Larochelle et al., 2003) and spanned over a 11 ½ year period for a comparable geographic area, a similar heterogeneity was observed. The average pairwise diversity within each subgenotype

Table 1
Yearly distributions of ORF5 PRRSV sequences for 20 subgenotypes in the province of Quebec from 1998 to mid-2009 (n = 1158[a]).

Subgenotype id	Year												Sequences/subgenotype[a]
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	
ATP-like							3	20	3	1	2		29
MLV-like		4	6	4	5	7	31	35	28	24	12	2	158
1							1	1	1	1	4	2	10
2		1				1	3	8	4	3	1	1	22
3		1	1	1	1	3	3	5	2		3	3	23
4	1	1	1	2	2	5	9	21	9		1	1	53
5							5	8	1				14
6		7	7	1	5	17	6	4	3	6	2		58
7			1	6		4	6	14	8	10	15	8	72
8		4	8	3	2	6	13	15	22	10	4	4	91
9					3	8	8	12	11	4	1	4	51
10	2	7	7	10	11	7	12	19	24	27	20	4	150
11	1	1	2	1	2	1	7	23	9	2	6	3	58
12			1	1	2	1	7	7	8	6	9	4	46
13								2	10	16	36	11	75
14					2	4	2	6	14	2	8	6	44
15						1	2		2	6	6	8	25
16					1		6	5	17	8	18	2	57
17	1	1		4	1	15	11	27	13	12	6	3	94
18		3	3	1	3	4	3	8	1	2			28
Sequences/year	5	30	37	34	40	84	138	240	190	140	154	66	1158
% Sequences/year	0	3	3	3	3	7	12	21	16	12	13	6	100

[a] Forty sequences did not regroup in any subgenotype.

Table 2
Glycotype patterns observed in the province of Quebec from 1998 to 2009 (n = 1284).

Glycotype pattern	N-gly sites								Number of sequences	% of total
	N30	N32	N33	N34	N35	N44	N51	N57		
1				X		X	X		363	28
2			X			X	X		181	14
3	X			X		X	X		146	11
4	X		X			X	X		145	11
5						X	X		62	5
6		X	X			X	X		58	5
7	X		X	X		X	X		53	4
8		X				X	X		46	4
9	X			X		X	X	X	30	2
10			X	X		X	X		28	2
11	X					X	X		25	2
12						X	X	X	23	2
13				X		X	X	X	22	2
14	X				X	X	X		18	1
15				X			X		10	1
16					X	X	X		9	1
17	X			X	X	X	X		9	1
18			X				X		7	1
19	X		X	X		X	X	X	7	1
20	X		X			X	X	X	5	0
21			X			X	X	X	4	0
22	X					X	X	X	4	0
23		X				X	X	X	4	0
24	X			X			X		4	0
25							X		3	0
26	X		X	X		X			3	0
27					X	X	X	X	2	0
28		X			X	X	X		2	0
29				X			X	X	2	0
30		X		X		X	X		1	0
31		X	X			X	X	X	1	0
32		X				X			1	0
33					X	X			1	0
34	X				X		X		1	0
35		X	X				X		1	0
36	X		X	X			X		1	0
37	X				X	X		X	1	0
38	X			X			X	X	1	0

Table 3
Significant amino acid positions under positive selective pressure in GP5 for nt subgenotype in the province of Quebec from 1998 to 2009 (n = 1158).

Subgenotype id	GP5 region and amino acid position																			Number of sequences/ subgenotype
	Peptide signal						HVR1[a]			HVR2[a]			TM2[b]			T cell 2			B cell	
	3	4	8	10	13	15	32	33	34	57	58	59	102	104	106	151	170	173	196	
ATP-like		**					**				**									29
MLV-like	**				**		**	**	**		**					**				158
1																				10
2							**									*				22
3							*	**		**	**	**						**		23
4						*			**	**	**	**								53
5																				14
6											**	**							**	58
7						**		**			**	**								72
8		**	*			**	**	**	**	**	**	**								91
9					**	**					**	**	**	**	*					51
10				**		**				**	**		**	**						150
11						**	**	*		**	**									58
12						*	**		*	*	**		*	*						46
13					**		**		**	**	**			**						75
14						**				**	*						**			44
15											*									25
16						**	**	**			**		*	**						57
17						**				**	**		**	**					**	94
18																				28

* P-1 ⩾ 95%.
** P-1 ⩾ 99%.
[a] HVR: hypervariable region.
[b] TM: transmembrane.

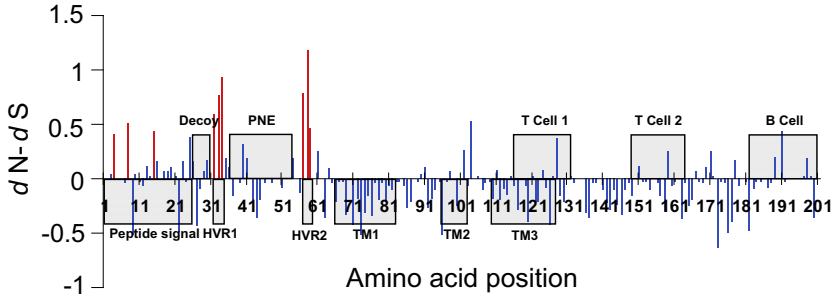

Fig. 2. GP5 amino acid variations of the subgenotype (# 8) having the most diversity and the greater number of positively selected sites. dN–dS represent substitution rates dN (non-synonymous) and dS (synonymous) of codon aligned nucleotide sequences. Blue lines show the dN–dS estimates by site position and red lines indicate sites under significant positive selection (P-1 ⩾ 95%). Upper rectangles indicate antigenic regions: decoy epitope, PNE principal neutralizing epitope, T cell and B cell epitopes. Lower rectangles indicate biologically significant regions: peptide signal region, HVR (hypervariable regions) and TM (transmembrane regions).

did not seem to be influenced by the number of sequences in it, as our largest wild-type subgenotype contained 150 sequences and had 7% diversity and our most diverse subgenotype having 9%, consisted of only 22 sequences. These values give an indication of the extent of plasticity exhibited by the NA wild-type ORF5 gene in the province of Quebec.

The results of sequence distribution in time represent case submissions over the years. Although no major trends in temporal evolution of strains could be observed, these results showed that subgenotype 13 has probably emerged in more recent years (2005) (Table 1) and that subgenotype 5 was transient through time (2004–2006). On the opposite, subgenotype 10 was observed every year and accounted for a high proportion of all sequences. The fact that 14 out of 18 wild-type subgenotypes had their oldest strain originating in the years 1998 to 2003, suggests that some PRRSV strains in the province of Quebec have potentially evolved by radial spread (Murtaugh et al., 2003) leading to the multiple clades that are observed today.

N-gly sites on the GP5 of PRRSV have been described in some studies, but our dataset revealed a greater diversity in potential N-gly sites, either by aa location (eight different positions) or total number of sites (one to six sites) than previously reported, including a newly identified site at position N57 (Ansari et al., 2006; Kimman et al., 2009; Pirzadeh et al., 1998; Zhou et al., 2009a,b). These sites are found at particular locations: within or near the first HVR region (N30 to N35), followed by N44, N51 and in some strains in the second HVR region at position N57. Positive selection in HVR2 for involved subgenotypes could not be related to subgenotypes having N57 glycosylation sites. The frequency of positions N34, N44 and N51 with a higher confidence level (⩾ 0.75) in the glycosylation score supports the important role of these three N-gly sites in virus life cycle already demonstrated by experimental procedures (Ansari et al., 2006). These rather well defined positions tend to confirm a shielding of the conserved PNE region by these sugar moieties from immune detection and/or a possible biological role in receptor binding (Vu et al., 2011).

The linear amino acid distance (6–7 aa) between sites N44, N51 and N57 also seems to be defined or constrained, compared to position shifting of glycosylated sites, located in or near HVR1, this of course without taking into account the three dimensional

conformation of the protein. This distinction between shifting and fixed N-gly sites has also been observed for HIV (Zhang et al., 2004). Overall, these results for N-gly location in aa sequence might indicate a different role for sites upstream of PNE region versus inside and downstream of PNE. The variations in number of N-gly sites upstream might be more involved in viral persistence whereas other sites downstream, especially in a conserved region, would be more implicated in functional role as described for PRRSV and HIV (Ansari et al., 2006; Vigerust and Shepherd, 2007). In addition, the duality between immune escape through hyperglycosylation and cell receptor function of this region, that includes HVR1, PNE and HVR2 must take into account a potential fitness cost of increased N-gly sites over diminished receptor avidity as seen for the influenza virus (Das et al., 2011).

Analysis of biologically significant regions in the ORF5 gene for all subgenotypes by *d*N and *d*S substitution rates showed that a majority of non-synonymous substitutions were occurring in the N terminal part and ectodomain region of the encoded GP5 protein, mainly in the putative peptide signal and the two hypervariable regions. These results are consistent with previously reported information (Li et al., 2009; Thanawongnuwech et al., 2004; Wang et al., 2008), although there is no general agreement on the size of the hypervariable regions in the literature (Hu et al., 2009; Zhou et al., 2009a). The high rate of non-synonymous variations in the peptide signal for many subgenotypes and the positive selective pressure at positions 13 and 15 in some subgenotypes partly corroborate results of Hanada et al. (2005) who found significant positive selection for aminos 13 and 14. Although these authors have discussed a role for this region in host cell attachment the importance of these findings in immune selection remains to be clarified.

The PNE region was fairly well conserved among all subgenotypes. The aa 38, 39, 40, 42, 43 and 44 which were described as being important residues either contributing to binding of antibodies or being part of the main recognition core of epitope B (Ostrowski et al., 2002) were particularly well conserved among almost all subgenotypes in our study. So, the high level of aa conservation in the PNE of a large number of strains and in an important functional domain suggests that mutations in this region would represent a disadvantage to the virus life cycle (Lopez and Osorio, 2004), and would corroborate the existence of a molecular constraint in this region.

Conversely, the presence of several sites under selective pressure in both HVR1 and 2, which flank the PNE region, might be linked with maintaining a conserved PNE region while evading the host immune system and involved in mechanism of diversification as recombination hotspots. Recently, a hypervariable region in GP4 (aa 57–68) of PRRSV EU strains was demonstrated to be targeted by antibody mediated immune response and involved in selection of variants resistant to these antibodies (Costers et al., 2010). A role for hypervariable regions in evasion of host immunity has also been extensively documented for hepatitis C virus (HCV), for which evolutionary rates and quasispecies complexity are associated with outcome of acute infections (clearance or persistence) (Curran et al., 2002). In this case, it was demonstrated that a mechanism behind the persistence of HCV, involved virus-specific neutralizing antibodies applying selective pressure on hypervariable regions, thereby driving the evolution of viral variants (Dowd et al., 2009). In this sense, some comparisons can be made between hypervariable regions of HCV and PRRSV. Firstly, both viruses have a HVR in an envelope viral glycoprotein associated with potential binding to host cell and seemingly involved in immune evasion systems (Cuevas et al., 2009). Also, both of these HVRs or adjacent regions are recognized as major determinants for neutralizing antibodies, with cross protection from antibodies of this epitope between homologous strains and some cross-neutralization between heterologous strains *in vivo* for PRRSV and *in vitro* for HCV (Curran et al., 2002; Lin et al., 2008; Lopez and Osorio, 2004; Plagemann et al., 2002). Finally, both viruses have an acute and persistent phase where epitope specific neutralizing antibodies could contribute to viral clearance in sera (Batista et al., 2004; Curran et al., 2002; Lopez and Osorio, 2004). These points of comparison and considerations regarding HCV show the importance of comparing the appropriate specific genomic regions to characterize immune selection and in the case of PRRSV, at a level where a certain amount of variability has accumulated between phylogenetically related strains. These HVRs are also potentially involved in recombination, as aa 59 (or nt 177) at the end of HVR2 and nt 349 have been described as an ORF5 recombination hotspots in an *in vivo* coinfection study on PRRSV (Liu et al., 2011). A preliminary break-point analysis from natural field recombinants of this study suggests that HVR1 region and a region near nt 349 are possibly involved in the observed intragenic cross-over (unpublished results).

The decoy region or epitope A underwent some non-synonymous variations in many subgenotypes, but not at any specific position. Even though this region is known to induce non-neutralizing antibodies (Lopez and Osorio, 2004), renewal of this epitope must happen to continue the decoy effect. Therefore, it would be inherent to suppose that epitope A is under some selective pressure by host immune cells at least in the earlier weeks post-infection. Other studies have found some positive selection in the epitope A region (Hanada et al., 2005; Hu et al., 2009). Two other positive sites for evolutionary selection were 102 and 104 at the end of the TM2 region. The TM regions are known to be involved in host cell membrane recognition and attachment (Hanada et al., 2005). The aa sites 32, 33, 34, 35, 58, 59 and 102 were also shown to be under positive selection (Xu et al., 2010).

Even though PRRSV ORF5 gene constantly renews itself (genetic plasticity) as seen by the large number of different sequences and genomic variations of our database, a certain variability threshold seems to exist and limit the range of heterogeneity seen in the field through time, at least in our study region. In this perspective, the implications of having positive selection in the HVR1 of ORF5 superposed over a conserved region in a different frameshift of ORF5a will bring new challenges in comparative genomics in regards to this exceptional feature of viral evolution (Drew, 2011; Johnson et al., 2011).

Finally, the hypothesis that a viral binding site located in the PNE of GP5 (Kimman et al., 2009; Murtaugh et al., 2010) which has a conformational limitation linked to its involvement in viral attachment would be plausible as seen for other RNA viruses, like measles and influenza (Ruigrok and Gerlier, 2007). This hypothesis would also be comforted by the presence of fixed N-gly sites, as discussed above, potentially involved in a functional role, e.g. receptor recognition. The N-gly sites are another factor for viral binding as demonstrated by a research group in Belgium, who showed a reduced infection of porcine macrophages by PRRSV when all the N-gly sites were removed by enzymatic treatment. Knowing that sialic acids are most likely present on glycan chains (Delputte and Nauwynck, 2004), and that glycosylation at position N44 was determined to be necessary for a sialic acid dependant binding of PRRSV to macrophage cells (Van Breedam et al., 2010) contribute at indicating that these regions are playing a major functional role in the virus life cycle and evolution of PRRSV diversity.

Role of funding sources

The funding sources did not have any participation in study design; in the collection, analysis and interpretation of data; in the writing or submission of manuscript.

Acknowledgements

The authors thank our financial contributors and participants: Fédération des Producteurs de Porcs du Québec (FPPQ), Centre d'Insémination Porcine du Québec Inc. (CIPQ), Conseil pour le Développement de l'Agriculture du Québec (CDAQ), veterinarians and swine producers. We also thank the molecular diagnostic laboratory of Dr. Josée Harel and Mr. Donald Tremblay, at the Faculty of Veterinary Medicine of the University of Montreal for their participation in the laboratory work of this project.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2012.03.015.

References

Ansari, I.H., Kwon, B., Osorio, F.A., Pattnaik, A.K., 2006. Influence of N-linked glycosylation of porcine reproductive and respiratory syndrome virus GP5 on virus infectivity, antigenicity, and ability to induce neutralizing antibodies. J. Virol. 80, 3994–4004.

Batista, L., Pijoan, C., Dee, S., Olin, M., Molitor, T., Joo, H.S., Xiao, Z., Murtaugh, M., 2004. Virological and immunological responses to porcine reproductive and respiratory syndrome virus in a large population of gilts. Can. J. Vet. Res. 68, 267–273.

Chen, Z., Li, K., Plagemann, P.G., 2000. Neuropathogenicity and sensitivity to antibody neutralization of lactate dehydrogenase-elevating virus are determined by polylactosaminoglycan chains on the primary envelope glycoprotein. Virology 266, 88–98.

Costers, S., Lefebvre, D.J., Van Doorsselaere, J., Vanhee, M., Delputte, P.L., Nauwynck, H.J., 2010. GP4 of porcine reproductive and respiratory syndrome virus contains a neutralizing epitope that is susceptible to immunoselection in vitro. Arch. Virol. 155, 371–378.

Cuevas, J.M., Gonzalez, M., Torres-Puente, M., Jimenez-Hernandez, N., Bracho, M.A., Garcia-Robles, I., Gonzalez-Candelas, F., Moya, A., 2009. The role of positive selection in hepatitis C virus. Infect. Gen. Evol. 9, 860–866.

Curran, R., Jameson, C.L., Craggs, J.K., Grabowska, A.M., Thomson, B.J., Robins, A., Irving, W.L., Ball, J.K., 2002. Evolutionary trends of the first hypervariable region of the hepatitis C virus E2 protein in individuals with differing liver disease severity. J. Gen. Virol. 83, 11–23.

Das, S.R., Hensley, S.E., David, A., Schmidt, L., Gibbs, J.S., Puigbo, P., Ince, W.L., Bennink, J.R., Yewdell, J.W., 2011. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. Proc. Natl. Acad. Sci. USA 108, E1417–E1422.

Delputte, P.L., Nauwynck, H.J., 2004. Porcine arterivirus infection of alveolar macrophages is mediated by sialic acid on the virus. J. Virol. 78, 8094–8101.

Dowd, K.A., Netski, D.M., Wang, X.H., Cox, A.L., Ray, S.C., 2009. Selection pressure from neutralizing antibodies drives sequence evolution during acute infection with hepatitis C virus. Gastroenterology 136, 2377–2386.

Drew, T.W., 2011. The emergence and evolution of swine viral diseases: to what extent have husbandry systems and global trade contributed to their distribution and diversity? Rev. Sci. Technol. 30, 95–106.

Faaberg, K.S., Hocker, J.D., Erdman, M.M., Harris, D.L., Nelson, E.A., Torremorell, M., Plagemann, P.G., 2006. Neutralizing antibody responses of pigs infected with natural GP5 N-glycan mutants of porcine reproductive and respiratory syndrome virus. Viral Immunol. 19, 294–304.

Firth, A.E., Zevenhoven-Dobbe, J.C., Wills, N.M., Go, Y.Y., Balasuriya, U.B., Atkins, J.F., Snijder, E.J., Posthuma, C.C., 2011. Discovery of a small arterivirus gene that overlaps the GP5 coding sequence and is important for virus production. J. Gen. Virol. 92, 1097–1106.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Hanada, K., Suzuki, Y., Nakane, T., Hirose, O., Gojobori, T., 2005. The origin and evolution of porcine reproductive and respiratory syndrome viruses. Mol. Biol. Evol. 22, 1024–1031.

Hu, H., Li, X., Zhang, Z., Shuai, J., Chen, N., Liu, G., Fang, W., 2009. Porcine reproductive and respiratory syndrome viruses predominant in southeastern China from 2004 to 2007 were from a common source and underwent further divergence. Arch. Virol. 154, 391–398.

Johnson, C.R., Griggs, T.F., Gnanandarajah, J., Murtaugh, M.P., 2011. Novel structural protein in porcine reproductive and respiratory syndrome virus encoded by an alternative ORF5 present in all arteriviruses. J. Gen. Virol. 92, 1107–1116.

Kimman, T.G., Cornelissen, L.A., Moormann, R.J., Rebel, J.M., Stockhofe-Zurwieden, N., 2009. Challenges for porcine reproductive and respiratory syndrome virus (PRRSV) vaccinology. Vaccine 27, 3704–3718.

Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208–1222.

Lambert, M.E., Arsenault, J., Poljak, Z., D'Allaire, S., 2012. Epidemiological investigations in regard to porcine reproductive and respiratory syndrome (PRRS) in Quebec, Canada. Part 2: Prevalence and risk factors in breeding sites. Prev. Vet. Med. 104, 84–93.

Larochelle, R., D'Allaire, S., Magar, R., 2003. Molecular epidemiology of porcine reproductive and respiratory syndrome virus (PRRSV) in Quebec. Virus Res. 96, 3–14.

Li, Y., Wang, X., Jiang, P., Chen, W., Wang, K., 2009. Genetic variation analysis of porcine reproductive and respiratory syndrome virus isolated in China from 2002 to 2007 based on ORF5. Vet. Microbiol. 138, 150–155.

Lin, Y., Kwon, T., Polo, J., Zhu, Y.F., Coates, S., Crawford, K., Dong, C., Wininger, M., Hall, J., Selby, M., Coit, D., Medina-Selby, A., McCoin, C., Ng, P., Drane, D., Chien, D., Han, J., Vajdy, M., Houghton, M., 2008. Induction of broad CD4+ and CD8+ T-cell responses and cross-neutralizing antibodies against hepatitis C virus by vaccination with Th1-adjuvanted polypeptides followed by defective alphaviral particles expressing envelope glycoproteins gpE1 and gpE2 and nonstructural proteins 3, 4, and 5. J. Virol. 82, 7492–7503.

Liu, D., Zhou, R., Zhang, J., Zhou, L., Jiang, Q., Guo, X., Ge, X., Yang, H., 2011. Recombination analyses between two strains of porcine reproductive and respiratory syndrome virus in vivo. Virus Res. 155, 473–486.

Lopez, O.J., Osorio, F.A., 2004. Role of neutralizing antibodies in PRRSV protective immunity. Vet. Immunol. Immunopathol. 102, 155–163.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.

Murtaugh, M.P., Wees, C.E., Johnson, J.E., Dwan, C., Faaberg, K.S., 2003. Molecular evolution of porcine reproductive and respiratory syndrome virus (PRRSV) envelope glycoprotein 5. In proceedings of the 4th International Symposium on emerging and re-emerging pig diseases, Rome, Italy 60–61.

Murtaugh, M.P., Stadejek, T., Abrahante, J.E., Lam, T.T., Leung, F.C., 2010. The ever-expanding diversity of porcine reproductive and respiratory syndrome virus. Virus Res. 154, 18–30.

Music, N., Gagnon, C.A., 2010. The role of porcine reproductive and respiratory syndrome (PRRS) virus structural and non-structural proteins in virus pathogenesis. Anim. Health Res. Rev. 11, 135–163.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Ostrowski, M., Galeota, J.A., Jar, A.M., Platt, K.B., Osorio, F.A., Lopez, O.J., 2002. Identification of neutralizing and nonneutralizing epitopes in the porcine reproductive and respiratory syndrome virus GP5 ectodomain. J. Virol. 76, 4241–4250.

Pirzadeh, B., Gagnon, C.A., Dea, S., 1998. Genomic and antigenic variations of porcine reproductive and respiratory syndrome virus major envelope GP5 glycoprotein. Can. J. Vet. Res. 62, 170–177.

Plagemann, P.G., Rowland, R.R., Faaberg, K.S., 2002. The primary neutralization epitope of porcine respiratory and reproductive syndrome virus strain VR-2332 is located in the middle of the GP5 ectodomain. Arch. Virol. 147, 2327–2347.

Ruigrok, R.W., Gerlier, D., 2007. Structure of the measles virus H glycoprotein sheds light on an efficient vaccine. Proc. Natl. Acad. Sci. USA 104, 20639–20640.

Shi, M., Lam, T.T., Hon, C.C., Hui, R.K., Faaberg, K.S., Wennblom, T., Murtaugh, M.P., Stadejek, T., Leung, F.C., 2010a. Molecular epidemiology of PRRSV: A phylogenetic perspective. Virus Res. 154, 7–17.

Shi, M., Lam, T.T., Hon, C.C., Murtaugh, M.P., Davies, P.R., Hui, R.K., Li, J., Wong, L.T., Yip, C.W., Jiang, J.W., Leung, F.C., 2010b. Phylogeny-based evolutionary, demographical, and geographical dissection of North American type 2 porcine reproductive and respiratory syndrome viruses. J. Virol. 84, 8700–8711.

Thanawongnuwech, R., Amonsin, A., Tatsanakit, A., Damrongwatanapokin, S., 2004. Genetics and geographical variation of porcine reproductive and respiratory syndrome virus (PRRSV) in Thailand. Vet. Microbiol. 101, 9–21.

Van Breedam, W., Van Gorp, H., Zhang, J.Q., Crocker, P.R., Delputte, P.L., Nauwynck, H.J., 2010. The M/GP(5) glycoprotein complex of porcine reproductive and respiratory syndrome virus binds the sialoadhesin receptor in a sialic acid-dependent manner. PLoS Pathog. 6, e1000730.

Vigerust, D.J., Shepherd, V.L., 2007. Virus glycosylation: role in virulence and immune interactions. Trends Microbiol. 15, 211–218.

Vu, H.L., Kwon, B., Yoon, K.J., Laegreid, W.W., Pattnaik, A.K., Osorio, F.A., 2011. Immune evasion of porcine reproductive and respiratory syndrome virus through glycan shielding involves both glycoprotein 5 as well as glycoprotein 3. J. Virol. 85, 5555–5564.

Wang, C., Lee, F., Huang, T.S., Pan, C.H., Jong, M.H., Chao, P.H., 2008. Genetic variation in open reading frame 5 gene of porcine reproductive and respiratory syndrome virus in Taiwan. Vet. Microbiol. 131, 339–347.

Xu, Z., Chang, X., Xiao, S., Chen, H., Zhou, R., 2010. Evidence for the adaptive evolution of ORF5 gene of porcine reproductive and respiratory syndrome virus isolated in China. Acta Virol. 54, 281–285.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Zhang, M., Gaschen, B., Blay, W., Foley, B., Haigwood, N., Kuiken, C., Korber, B., 2004. Tracking global patterns of N-linked glycosylation site variation in highly

variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. Glycobiology 14, 1229–1246.

Zhou, L., Chen, S., Zhang, J., Zeng, J., Guo, X., Ge, X., Zhang, D., Yang, H., 2009a. Molecular variation analysis of porcine reproductive and respiratory syndrome virus in China. Virus Res. 145, 97–105.

Zhou, Y.J., Yu, H., Tian, Z.J., Li, G.X., Hao, X.F., Yan, L.P., Peng, J.M., An, T.Q., Xu, A.T., Wang, Y.X., Wei, T.C., Zhang, S.R., Cai, X.H., Feng, L., Li, X., Zhang, G.H., Zhou, L.J., Tong, G.Z., 2009b. Genetic diversity of the ORF5 gene of porcine reproductive and respiratory syndrome virus isolates in China from 2006 to 2008. Virus Res. 144, 136–144.