Published in final edited form as:

Infect Genet Evol. 2013 December ; 20: . doi:10.1016/j.meegid.2013.09.022.

Purifying selection in Porcine reproductive and respiratory syndrome virus ORF5a protein influences variation in envelope glycoprotein 5 glycosylation

Sally R. Robinson, Juan E. Abrahante, Craig R. Johnson, and Michael P. Murtaugh

Department of Veterinary and Biomedical Sciences, University of Minnesota, 1971, Commonwealth Avenue, St. Paul, MN 55108, USA

Sally R. Robinson: robin782@umn.edu; Juan E. Abrahante: abrah023@umn.edu; Michael P. Murtaugh: murta001@umn.edu

Abstract

Porcine Reproductive and Respiratory Syndrome Virus ORF5a protein is encoded in an alternate open reading frame upstream of the major envelope glycoprotein (GP5) in subgenomic mRNA5. Bioinformatic analysis of 3,466 Type 2 PRRSV sequences showed that the two proteins have co-evolved through a fine balance of purifying codon usage to maintain a conserved RQ-rich motif in ORF5a protein, while eliciting a variable N-linked glycosylation motif in the alternative GP5 reading frame. Conservation of the ORF5a protein RQ-motif also explains an anomalous uracil desert in GP5 hypervariable glycosylation region. The N-terminus of the mature GP5 protein was confirmed to start with amino acid 32, the hypervariable region of the ectodomain. Since GP5 glycosylation variability is assumed to result from immunological selection against neutralizing antibodies, these findings show that an alternative possibility unrelated to immunological selection not only exists, but provides a foundation for investigating previously unsuspected aspects of PRRSV biology. Understanding functional consequences of subtle nucleotide sequence modifications in the region responsible for critical function in ORF5a protein and GP5 glycosylation is essential for rational design of new vaccines against PRRS.

Keywords

natural selection; evolution; PRRSV

1. Introduction

Porcine Reproductive and Respiratory Syndrome (PRRS) virus is an Arterivirus in the Order Nidovirales. It causes PRRS in swine, a devastating disease with significant animal welfare and economic implications for which there is no specific treatment and incomplete protection by vaccination. Despite extensive research yielding substantial knowledge of PRRSV biology, epidemiology and genome structure since its emergence in the late 1980s, the precise mechanisms responsible for virulence, pathogenesis and protective immune

© 2013 Elsevier B.V. All rights reserved.

Corresponding author: Sally Robinson, robin782@umn.edu, Phone: +1-(612)-624-4926, Fax: +1-(612)-625-5203.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

response to PRRSV remain poorly understood. These factors limit progress toward development of effective measures for prevention and treatment of this disease.

The ~15kb PRRSV RNA genome is extremely compact, with 98% coding capacity. There are conserved untranslated regions (UTRs) at both the 5' and 3' end of 150 to 220 nt, and a single non-coding nucleotide between open reading frames (ORFs) 1b and 2, at the transition between coding sequences for the non-structural and structural proteins. Coding potential of PRRSV is maximized through a number of mechanisms, including ribosomal frameshifting and proteolytic cleavage of translated polypeptides to yield multiple non-structural proteins, and overlapping open reading frames and nested subgenomic (sg) mRNA in the structural protein coding sequences (Fang et al., 2012; Maclachlan and Edward, 2011).

The structural proteins of PRRSV are translated from a 3' co-terminally nested set of sgRNAs resulting in expression of each protein from separate sgRNA fragments. Two additional small envelope proteins, E and ORF5a protein, are expressed from overlapping bicistronic sgRNA 2 and sgRNA 5, respectively (Firth et al., 2011; Johnson et al., 2011; Snijder et al., 1999; Wu et al., 2001). In each case, overlapping GP2 and E proteins, and ORF5a protein and GP5 are translated from alternate reading frames. The novel PRRSV ORF5a protein is initiated 10 bp upstream of, and encoded in the +2 reading frame from, the major envelope glycoprotein GP5 in Type 2 PRRS viruses (Firth et al., 2011; Johnson et al., 2011). The presence of ORF5a in sgRNA5 is evolutionarily conserved throughout the Arteriviridae (Firth et al., 2011; Johnson et al., 2011).

ORF5a protein is predicted to be a small (46 – 51 amino acid) type 1 membrane protein with a short ectodomain, a single transmembrane region and an endodomain tail containing a highly conserved 8 aa arginine-glutamine (RQ) motif (Johnson et al., 2011). Paradoxically, the conserved RQ-motif is encoded in the same nucleotide sequence that encodes a hypervariable glycosylation region of the GP5 ectodomain, first described in 1995 (Meng et al., 1995; Pirzadeh et al., 1998). GP5 is predicted by protein structure modeling software to contain an N-terminal signal peptide (Meulenberg et al., 1995). It directs GP5 to the secretory pathway of the endoplasmic reticulum/Golgi complex, where it can undergo maturation, including folding and N-linked glycosylation prior to virion assembly (Faaberg et al., 2011; Mardassi et al., 1996). The N-terminal portion of mature GP5 protein is proposed to contain a neutralizing epitope flanked by two hypervariable glycosylation regions that provide a glycan shield for immune evasion and that are under intense immunological selection (Ansari et al., 2006; Faaberg et al., 2006; Lopez and Osorio, 2004; Meng et al., 1995; Ostrowski et al., 2002; Vu et al., 2011). However, the significance of GP5 glycosylation and neutralization in PRRSV immunity remains controversial (Li and Murtaugh, 2012; Murtaugh and Genzow, 2011). Furthermore, the implications of differential selective forces on the overlapping reading frames of this potentially important region for PRRSV immunity have not been evaluated. To investigate the mechanism by which the sgRNA5 nucleotide sequence simultaneously accomplished conservation of the RQ-motif in ORF5a protein and hypervariability in GP5 glycosylation, we analyzed 3466 unique and diverse type 2 PRRSV ORF5 sequences.

2. Material and methods

PRRSV ORF5 nucleotide sequences of 603 bp in length (the coding sequence for GP5 from amino acid position 1 to 201) were obtained from the University of Minnesota Veterinary Diagnostic Laboratory PRRSV database, containing more than 8,000 ORF5 sequences, including field isolates and vaccine strains, and representing the full diversity of type 2 PRRSV (Shi et al., 2010). Sequences were manually edited in CLC sequence viewer (CLC

Bio, Cambridge MA), aligned using MUSCLE (Edgar, 2004), and translated in both ORF5a protein and GP5 reading frames. ORF5 sequences begin at the GP5 translation start codon, so are devoid of the first 10 nucleotides of ORF5a. Hence, translated ORF5a protein sequences start at amino acid 5. Sequences were clustered at 100% similarity using CD-HIT to remove identical sequences (Li and Godzik, 2006).

ORF5a and GP5 nucleotide and amino acid frequency logos were generated with WebLogo (Crooks et al., 2004). GP5 N-linked glycosylation motifs were identified with the N-GlycoSite tool (Zhang et al., 2004). The relative frequency of ORF5a and GP5 synonymous and non-synonymous nucleotide substitutions was determined using Molecular Evolutionary Genetics Analysis (MEGA) 5 (Tamura et al., 2011).

For N-terminal sequencing of GP5, PRRSV strain VR2332 (Genbank accession U87392.3) was grown in MARC145 cells and purified from cell culture supernatant as described, except that sucrose pellets were banded in 1.25 g/ml CsCl (Johnson et al., 2011). Purified virions were electrophoresed on a 10% SDS-polyacrylamide gel. Coomassie blue-stained bands in the molecular weight regions of 18, 20, 25 and 28 kD were excised and submitted to the Mayo Clinic Proteomics Core (Rochester, MN) for automated Edman degradation protein sequencing on an ABI Procise system (Applied Biosystems, Carlsbad, CA).

Codon usage was analyzed using MEGA (Tamura et al., 2011) for ORF5a R and Q residues and GP5 N and S residues in the overlapping region. In addition these were compared with codon usage in the complete PRRSV VR2332 genome (Genbank accession U87392.3) and expressed swine genes obtained from the codon usage database (<http://www.kazusa.or.jp/codon/>) (Nakamura et al., 2000). Statistical comparisons were made between observed codon usage frequencies in the 3466 ORF5 sequences and frequencies expected from random codon usage using the Chi-squared test for residues with 6 possible codons (R and S) and Fisher's exact test for those with 2 codons each (Q and N). A p value of 0.05 was considered significant.

3. Theory

Coincident nucleotide sequence coding for a highly conserved motif in one protein and a highly variable motif in another is paradoxical. We hypothesized that selective codon usage driven by purifying selection to maintain the conserved ORF5a protein RQ-motif simultaneously facilitates nonsynonymous mutations increasing glycosylation site variability in GP5. Glycosylation site diversity as an indirect consequence of RQ-motif conservation provides an alternative to immunological selection, for which little direct evidence has been marshaled.

4. Results

Genetic selection pressure on 3466 Type 2 PRRSV ORF5 sequences is purifying across the majority of the 201 aa GP5 sequence, with the exception of the first 35 N-terminal aa. Selection pressure here is neutral or diversifying, and a hypervariable glycosylation region is present in aa 32 to 35 (Fig 1). The difference in selection pressure was previously utilized to predict the dual-coding nature of ORF5 (Firth et al., 2011). Of the 3466 ORF5a protein sequences, seventy six percent were 46 aa long; the remaining 24% were 51 aa in length, consistent with previous findings (Firth et al., 2011).

Ninety-six percent of 3466 ORF5a proteins exclusively contained R and Q residues in the RQ-motif (amino acid position 32–39) (Fig. 2a). In overlapping amino acid positions 29–36 of the GP5 protein, N-linked glycosylation motifs (defined by N-X-S/T, where N = asparagine, X = any amino acid, S/T = serine or threonine) were identified; the majority of

full-length GP5 sequences (94%) were found to contain 2 to 4 N-linked glycosylation motifs (Fig. 3a). Along with the essentially fixed motifs at GP5 positions 44 and 51, additional N-linked glycosylation motif variation occurs in the ORF5a overlapping region at positions 30, 32, 33, 34 and 35 (Fig. 3b), and extensive use of N and S residues is evident (Fig. 2a).

The discordance between R/Q conservation in ORF5a protein and glycosylation diversity in GP5 is associated with differences in the reading frame of specific nucleotide changes that are shared between the two ORFs. The nucleotide variation in this region of ORF5a is largely restricted to the third codon position, with mean frequency of sequence variation 21%, compared to codon position 1 (1%) or 2 (6%). Third codon position variation frequently results in synonymous substitutions due to the degenerate nature of the amino acid code, preserving the R/Q motif (Fig. 2a). By contrast, the same mutation occurs in the second codon position of ORF5, causing a non-synonymous substitution that disrupts glycosylation motifs with high frequency (Fig. 2a).

The overall selective pressure in this region, assessed by the relative frequency of synonymous and non-synonymous nucleotide substitutions, was neutral ($dN=dS$) or diversifying ($dN>dS$) for GP5, whereas it was strongly purifying ($dN<dS$) or neutral in ORF5a, with some diversification evident at residues 37 and 38 (Figure 2b). ORF5a codons in which C was followed by A/G variation in the second position (positions 33, 35, 36, 37) resulted in a change from R to Q or Q to R. Thus, simple analysis of dN and dS to infer selective forces underestimates conservation of the RQ-motif as the R/Q substitution is a non-synonymous change.

Strongly purifying selection is observed at ORF5a protein residues 32, 33, and 34, which overlap the GP5 signal peptide. The signal peptide is cleaved between amino acids 31 and 32 to form mature GP5, as predicted by SignalP (www.cbs.dtu.dk/services/SignalP/). N-terminal amino acid sequencing of GP5 confirmed the prediction. Electrophoresis of purified PRRSV VR2332 virions revealed two visible bands were present in the molecular weight regions of 25 and 28 kD, consistent with molecular weight heterogeneity observed for GP5 due to a combination of N-linked high-mannose and complex oligosaccharide types (Fig. 4a) (Mardassi et al., 1996). The first 10 residues of each band yielded the sequence S-DSSSHLQL, which corresponds to VR2332 GP5 residues 32–41 (Fig. 4b). The second cycle did not yield a product, consistent with the presence of a glycosylated asparagine. Thus, amino acid 32 defines the amino terminus of mature GP5. Consequently, the predicted N-linked glycosylation motif at amino acid position 30 in 32% of 3466 ORF5 sequences (Fig. 3b) is not present in PRRSV virions.

A notable feature of the hypervariable glycosylation region is the absence of U nucleotides in GP5 amino acid residues 30 to 36 (Fig. 5a). To analyze the mechanism for U residue depletion, the theoretical usage of U nucleotides in the possible codons for N, S, R and Q were compared to the actual observed codon use in this region. U nucleotides are present in 6 of the 8 codons for N and S, accounting for 7 of the 24 possible codon nucleotide positions, or 29%, in these key glycosylation residues. By contrast, U occurs only once in the 8 possible codons for R and Q, at only 1 of 24, or 4%, of potential coding nucleotides, in the overlapping ORF5a coding sequence. However, the observed frequency of U nucleotides in 3466 sequences was substantially lower, at 0.1%. The low observed frequency of U in this region is due to codon usage bias in N and S residues imposed by ORF5a protein (Fig. 5b), since U is absent in Q and is only in 1 of 18 possible R codon positions. The observed codon usage bias differs from random utilization of N and S codons significantly (χ^2 test, $p \sim 0$) and also differs from the frequency of use in other regions of the PRRSV genome and from codon usage in expressed swine genes (Fig. 5b). Strikingly, of the 8 possible S and N codons, AGC (coding for serine) and AAC (coding for asparagine) are utilized almost

exclusively in this region (Fig. 5b). These are the only codons for S and N that do not contain U and are the only codons that allow dual coding of S/N in the GP5 reading frame and R/Q in the ORF5a protein +2 reading frame. A striking codon bias is also observed for R residues than would be expected from random codon usage (χ^2 test, $p \sim 0$) (Fig. 5c).

5. Discussion

The paradox of a highly conserved RQ-motif coexisting with a hypervariable glycosylation region existing in a uracil desert is the mechanistic result of ORF5a codon wobble whose concomitant position two change in the overlapping GP5 codon alters the pattern of N-linked glycosylation motifs. Previously, immunological selection to escape neutralizing antibodies was proposed to account for variable glycosylation pattern evolution in GP5. However, direct functional evidence for immunological selection is lacking, and the significance of neutralizing antibody responses to PRRSV immunity is not firmly established (reviewed in (Murtaugh and Genzow, 2011)). Disruption of N-linked glycosylation motifs is reported to result in increased susceptibility to virus neutralization *in vitro* (Ansari et al., 2006; Faaberg et al., 2006; Vu et al., 2011). Reversion of mutant virus to the glycosylation phenotype has been observed *in vivo*, with corresponding increased resistance to *in vitro* neutralization (Vu et al., 2011). However, neutralizing monoclonal antibodies have been shown to bind both glycosylated and non-glycosylated forms of GP5 (Pirzadeh and Dea, 1997), and immunization with non-glycosylated GP5-M ectodomains did not elicit neutralizing antibodies even though partial protection was observed (Li and Murtaugh, 2012). Our analysis revealed that GP5 S and N codon usage is highly restricted to accommodate ORF5a protein R and Q residues in the alternate reading frame. Selective codon usage in ORF5a R, in particular, serves to maintain hypervariable glycosylation sites in GP5. Thus, in addition to purifying selection to maintain the conserved ORF5a protein motif, selective pressure also exists to maintain glycosylation sites in GP5. Our results, combined with lack of compelling evidence for an immunological benefit to GP5 glycosylation or for phylogenetic clustering of GP5 glycotypes indicative of selection (Delisle et al., 2012), indicate that glycosylation motif position variation in GP5 is a mechanism to maintain a functional requirement for glycan addition in the GP5 ectodomain, such as for correct folding (Wei et al., 2012), whilst preserving an undetermined function of the RQ-motif in ORF5a protein.

The strong, purifying selection to maintain Type 2 PRRSV ORF5a protein RQ-motif, combined with genetic ablation studies (Sun et al., 2013) indicate that ORF5a is necessary for production of viable progeny virions. While an RQ-motif is present in Type 1 PRRSV and other arteriviruses, its conservation appears to be less pronounced (Johnson et al., 2011; Firth et al., 2011). A relative paucity of ORF5 and ORF5a sequences for other arteriviruses at this time precludes a detailed comparative analysis of selection pressures. However, genetic ablation of ORF5a in equine arteritis virus (EAV) only reduced, but did not eliminate, virion production in infected cells (Firth et al., 2011; Sun et al., 2013), suggesting that ORF5a is not essential in EAV. Also, translation of ORF5a of Type 1 PRRSV, lactate dehydrogenase-elevating virus (LDV) and simian hemorrhagic fever virus (SHFV) is initiated downstream of ORF5, rather than upstream as occurs in Type 2 PRRSV and EAV (Firth et al., 2011). The significance of this difference is not known. The specific ORF5a amino acid residue patterns in the RQ-rich region vary among the arteriviruses, as does the overlapping GP5 sequences (Firth et al., 2011; Johnson et al., 2011). The exact position within this region of N-linked glycosylation motifs, and proposed neutralizing epitopes also differs between Type 1 and 2 PRRSV (Ostrowski et al., 2002; Plagemann et al., 2002; Wissink et al., 2003). It is possible that the high degree of RQ-motif conservation and overlap with the hypervariable glycosylation motif of GP5 contribute to the higher virulence of Type 2 PRRSV.

6. Conclusions

PRRSV has evolved a mechanism of highly selective codon usage, utilizing the ORF5a wobble position to maintain ORF5a protein RQ-motif conservation in the face of positive selective pressure for GP5 N-linked glycosylation motifs. The extreme bias for codon usage resulting in a uracil desert indicates evolution of a fine-tuned mechanism to accommodate the dual coding nature of the nucleotide sequence region and suggests co-evolution of the two proteins to accommodate the functional needs of both a GP5 N-linked glycosylation motif and an ORF5a protein RQ-motif. Since GP5 glycosylation variability is assumed to result from immunological selection against neutralizing antibodies, these findings show that an alternative possibility unrelated to immunological selection not only exists, but provides a foundation for investigating previously unsuspected aspects of PRRSV biology.

ORF5a protein appears to play a key role in the Arteriviridae since its genetic ablation in equine arteritis virus severely impacts growth in permissive cells (Firth et al., 2011), and abolishes growth in PRRSV (Sun et al., 2013). The high level of ORF5a protein RQ-motif conservation throughout Type 2 PRRS viruses indicates that this region, in particular, is critical. Hence, subtle modifications of nucleotide sequence in this region based on rare, viable field isolates may provide an avenue to rational development of new vaccines.

Acknowledgments

The study was supported by an undesignated gift from Boehringer Ingelheim Vetmedica. SRR was the recipient of an NIH T32RR018719 fellowship.

References

- Ansari IH, Kwon B, Osorio FA, Pattnaik AK. 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; 80:3994–4004. [PubMed: 16571816]
- Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A Sequence Logo Generator. *Genome Research*. 2004; 14:1188–1190. [PubMed: 15173120]
- Delisle B, Gagnon CA, Lambert ME, D’Allaire S. Porcine reproductive and respiratory syndrome virus diversity of Eastern Canada swine herds in a large sequence dataset reveals two hypervariable regions under positive selection. *Infection, Genetics and Evolution*. 2012; 12:1111–1119.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 2004; 32:1792–1797. [PubMed: 15034147]
- Faaberg, KS.; Balasuriya, UB.; Brinton, MA.; Gorbalenya, AE.; Leung, FC.; Nauwynck, HJ.; Snijder, EJ.; Stadejek, T.; Yang, H.; Yoo, D. Family *Arteriviridae*. In: King, AMQ.; Adams, MJ.; Carstens, EB.; Lefkowitz, EJ., editors. *Virus Taxonomy, Classification and Nomenclature of Viruses*. Ninth report of the International Committee on Taxonomy of Viruses. Elsevier; Oxford: 2011. p. 796-805.
- Faaberg KS, Hocker JD, Erdman MM, Harris DL, Nelson EA, Torremorell M, Plagemann PG. Neutralizing antibody responses of pigs infected with natural GP5 N-glycan mutants of porcine reproductive and respiratory syndrome virus. *Viral Immunol*. 2006; 19:294–304. [PubMed: 16817772]
- Fang Y, Treffers EE, Li Y, Tas A, Sun Z, van der Meer Y, de Ru AH, van Veelen PA, Atkins JF, Snijder EJ, Firth AE. Efficient -2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein. *Proceedings of the National Academy of Sciences*. 2012; 109:E2920–E2928.
- Firth AE, Zevenhoven-Dobbe JC, Wills NM, Go YY, Balasuriya UB, Atkins JF, Snijder EJ, Posthuma CC. Discovery of a small arterivirus gene that overlaps the GP5 coding sequence and is important for virus production. *J Gen Virol*. 2011; 92:1097–1106. [PubMed: 21307223]

- Johnson CR, Griggs TF, Gnanandarajah J, Murtaugh MP. Novel structural protein in porcine reproductive and respiratory syndrome virus encoded by an alternative ORF5 present in all arteriviruses. *J Gen Virol*. 2011; 92:1107–1116. [PubMed: 21307222]
- Li J, Murtaugh MP. Dissociation of porcine reproductive and respiratory syndrome virus neutralization from antibodies specific to major envelope protein surface epitopes. *Virology*. 2012; 433:367–376. [PubMed: 22981434]
- Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006; 22:1658–1659. [PubMed: 16731699]
- Lopez OJ, Osorio FA. Role of neutralizing antibodies in PRRSV protective immunity. *Veterinary Immunology and Immunopathology*. 2004; 102:155–163. [PubMed: 15507302]
- MacLachlan, NJ.; Edward, JD. Chapter 25 - Arteriviridae and Roniviridae. In: MacLachlan, NJ.; Edward, JD., editors. *Fenner's Veterinary Virology*. 4. Academic Press; San Diego: 2011. p. 415–424.
- Mardassi H, Massie B, Dea S. Intracellular synthesis, processing, and transport of proteins encoded by ORFs 5 to 7 of porcine reproductive and respiratory syndrome virus. *Virology*. 1996; 221:98–112. [PubMed: 8661418]
- Meng XJ, Paul PS, Halbur PG, Morozov I. Sequence comparison of open reading frames 2 to 5 of low and high virulence United States isolates of porcine reproductive and respiratory syndrome virus. *Journal of General Virology*. 1995; 76:3181–3188. [PubMed: 8847527]
- Meulenbergh JJ, Petersen-den Besten A, De Kluyver EP, Moormann RJ, Schaaper WM, Wensvoort G. Characterization of proteins encoded by ORFs 2 to 7 of Lelystad virus. *Virology*. 1995; 206:155–163. [PubMed: 7831770]
- Murtaugh MP, Genzow M. Immunological solutions for treatment and prevention of porcine reproductive and respiratory syndrome (PRRS). *Vaccine*. 2011; 29:8192–8204. [PubMed: 21925560]
- Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research*. 2000; 28:292. [PubMed: 10592250]
- Ostrowski M, Galeota JA, Jar AM, Platt KB, Osorio FA, Lopez OJ. Identification of Neutralizing and Nonneutralizing Epitopes in the Porcine Reproductive and Respiratory Syndrome Virus GP5 Ectodomain. *Journal of Virology*. 2002; 76:4241–4250. [PubMed: 11932389]
- Pirzadeh B, Dea S. Monoclonal antibodies to the ORF5 product of porcine reproductive and respiratory syndrome virus define linear neutralizing determinants. *Journal of General Virology*. 1997; 78:1867–1873. [PubMed: 9266981]
- Pirzadeh B, Gagnon CA, SD. Genomic and antigenic variations of porcine reproductive and respiratory syndrome virus major envelope GP5 glycoprotein. *Canadian Journal of Veterinary Research*. 1998; 62:170–177. [PubMed: 9684045]
- Plagemann PGW, Rowland RRR, Faaberg KS. The primary neutralization epitope of porcine respiratory and reproductive syndrome virus strain VR-2332 is located in the middle of the GP5 ectodomain. *Archives of Virology*. 2002; 147:2327–2347. [PubMed: 12491101]
- Shi M, Lam TTY, Hon CC, Murtaugh MP, Davies PR, Hui RKH, Li J, Wong LTW, Yip CW, Jiang JW, Leung FCC. Phylogeny-Based Evolutionary, Demographical, and Geographical Dissection of North American Type 2 Porcine Reproductive and Respiratory Syndrome Viruses. *Journal of Virology*. 2010; 84:8700–8711. [PubMed: 20554771]
- Snijder EJ, van Tol H, Pedersen KW, Raamsman MJB, de Vries AAF. Identification of a novel structural protein of Arteriviruses. *Journal of Virology*. 1999; 73:6335–6345. [PubMed: 10400725]
- Sun L, Li Y, Liu R, Wang X, Gao F, Lin T, Huang T, Yao H, Tong G, Fan H, Wei Z, Yuan S. Porcine reproductive and respiratory syndrome virus ORF5a protein is essential for virus viability. *Virus Research*. 2013; 171:178–185. [PubMed: 23219923]
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*. 2011; 28:2731–2739. [PubMed: 21546353]

- Vu HLX, Kwon B, Yoon KJ, Laegreid WW, Pattnaik AK, Osorio FA. Immune evasion of porcine reproductive and respiratory syndrome virus through glycan shielding involves both glycoprotein 5 as well as glycoprotein 3. *Journal of Virology*. 2011; 85:5555–5564. [PubMed: 21411530]
- Wei Z, Lin T, Sun L, Li Y, Wang X, Gao F, Liu R, Chen C, Tong G, Yuan S. N-linked glycosylation of GP5 of porcine reproductive and respiratory syndrome virus is critically important for virus replication in vivo. *Journal of Virology*. 2012; 86:9941–9951. [PubMed: 22761373]
- Wissink EH, van Wijk HA, Kroese MV, Weiland E, Meulenbergh JJ, Rottier PJ, van Rijn PA. The major envelope protein, GP5, of a European porcine reproductive and respiratory syndrome virus contains a neutralization epitope in its N-terminal ectodomain. *J Gen Virol*. 2003; 84:1535–1543. [PubMed: 12771423]
- Wu W, Fang Y, Farwell R, Steffen-Bien M, Rowland RRR, Christopher-Hennings J, Nelson EA. A 10-kDa structural protein of Porcine Reproductive and Respiratory Syndrome Virus encoded by ORF2b. *Virology*. 2001; 287:183–191. [PubMed: 11504553]
- Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, Kuiken C, Korber B. Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology*. 2004; 14:1229–1246. [PubMed: 15175256]

Highlights

- PRRSV RNA sequence dually codes for ORF5a and GP5 in overlapping reading frames (82)
- Paradoxically, a conserved ORF5a motif overlaps a variable GP5 glycosylation motif (85)
- Select codon use provides fine balance for functional requirements of both proteins (85)
- GP5 glycosylation motif pattern is influenced by ORF5a RQ- coding sequence (77)

Figure 1.

Analysis of 3466 PRRSV glycoprotein 5 (GP5) coding sequences. Comparison of rate of nonsynonymous (dN) and synonymous (dS) changes per site as a measure of selective pressure across the GP5 coding sequence (top figure, and shown in blue), and in the region of ORF5a protein overlap (bottom figure, and shown in red). Shaded area represents the region of interest containing the GP5 hypervariable glycosylation region and the ORF5a protein RQ-motif.

Figure 2.

PRRSV sgmRNA5 dual coding sequence at the GP5 hypervariable region and ORF5a protein RQ-motif overlap. a) Relative frequency of nucleotide residues in the 3466 sgmRNA5 nucleotide sequences (from nucleotides 84–104 of the GP5 reading frame), and amino acid sequences translated in the GP5 and ORF5a protein reading frames. Amino acid positions are shown for GP5 and ORF5a protein. GP5 N-linked glycosylation motif positions are marked by asterisks and signal peptide cleavage site is shown by a dashed line. b) Comparison of synonymous and nonsynonymous changes in GP5 and ORF5a protein coding sequences, calculated in MEGA5 (Tamura et al., 2011).

Figure 3. Predicted GP5 N-linked glycosylation patterns. a) Variation in number of predicted N-linked glycosylation motifs across 3466 GP5 sequences. b) Frequency of predicted N-linked glycosylation motifs at each position in GP5.

Figure 4.

N-terminus of mature GP5. a) Purified PRRSV VR2332 virions electrophoresed on a 10% SDS-polyacrylamide gel, and stained with Coomassie blue. Molecular weight marker is shown in lane 1. Bands identified as GP5, M, and N were excised from the gel and subjected to N-terminal amino acid sequencing by Edman degradation. b) PRRSV VR2332 GP5 amino acid sequence. N-terminal amino acid residues identified by protein sequencing are shown in red, and the signal peptide cleavage site is marked by an arrow between residues 31 and 32. Asparagine (N) residues in bold denote N-linked glycosylation motifs at positions 33, 34 and 51.

Figure 5.

Uracil desert in the GP5/ORF5a protein overlap. a) Frequency of uracil residues at ORF5 nucleotide positions 66–126 for 3466 sequences. b) Codon usage for GP5 N and S residues in the hypervariable glycosylation region lacking uracil. Relative codon usage for the entire coding sequence of PRRSV VR2332 and of 3000 pig coding sequences is shown for comparison. c) Codon usage for ORF5a protein R and Q residues in the uracil desert.