RESEARCH REPORT
SWINE HEALTH

Title: PRRSV Modulation of the Porcine Antibody Repertoire - **NPB ID #: 10-139**

Investigator: Dr. Michael P. Murtaugh

Institution: University of Minnesota

Date Submitted: 19 Jan 2012

Industry Summary: Antibody responses play a crucial role in defense against many viral diseases, and knowledge of critical functions can be used to monitor and predict vaccine efficacy. We have identified the features of key antibodies produced in response to PRRSV, and have defined the genetic elements that can give rise to anti-PRRSV antibodies. This information can, and will be, used to identify antibodies originating in pigs responding to a PRRSV infection and characterize the neutralizing activity. This knowledge will help to focus vaccine strategies and provide therapeutic reagents to reduce disease severity.

Keywords: Swine, porcine reproductive and respiratory syndrome virus, immunology, disease resistance, antibody

Scientific Abstract: Extensive efforts have been directed to characterize the antibody and neutralizing antibody responses to PRRSV infection, in the hope of elucidating key insights into protective and cross-protective immunity. The main hurdle to overcome in obtaining these answers is knowledge of the antibody molecules that pigs express, and how PRRSV infection modulates this population. Here, we characterized the Ig heavy chain

(HC) repertoire of the pig and discovered major alterations in diversity due to PRRSV infection. We also sequenced the genomic Ig light chain (LC) loci to identify variable (V) and joining (J) gene segments that encode LC antigen binding. This information is essential to construct antigen binding libraries and identify PRRSV-specific antibodies. We expect the results of this research to open the door to development of therapeutic reagents to treat acute PRRS, genetic testing for PRRS resistance, and a mechanistic understanding of cross-protective immunity.

Introduction: During antibody development in animals, both heavy and light chain genes undergo a series of recombination events which result in a highly variable antigen-binding region (variable region) joined to a constant region. The Ig heavy chain variable region (VH) is generated from the recombination of a variable gene (V), a diversity gene (D), and a joining gene (J). Diversity in the Ig light chains is generated in a similar manner except that the light chains lack a D gene. Further diversity is generated in B cells from class switch recombination to a different constant region gene as well as from enzymatically-induced point mutations introduced into the variable region following activation (i.e. somatic hypermutation). The regions possessing the greatest germline diversity and which receive the highest degree of somatic hypermutation comprise the complementarity determining regions (CDRs) which are exposed to interact with antigen in the three-dimensional protein structure.

A highly conserved framework region separates each CDR and contributes to the structural integrity of the entire molecule. Lastly, a separate exon upstream of the V gene encodes for a leader peptide important for the trafficking of the nascent antibody molecule through the endoplasmic reticulum. This peptide is cleaved off prior to secretion. The potential antibody repertoire size is many orders of magnitude due to the number of theoretically possible V(D)J combinations and the junctional diversity between genes.

Previous investigations of antibody repertoire diversity in pigs, by John Butler's lab at the University of Iowa, bypassed this extreme complexity by relying on techniques such as CDR3 spectratyping and limited sequencing of cloned transcripts. Their work is the foundation of our current understanding of the pre-immune repertoire in fetal pigs, and the pathologic effect of PRRSV on antibody development (1-4). With the

development of high-throughput sequencing technologies, we now can examine the full range of antibody diversity through deep sequencing of Ig mRNA. The long reads enable one to analyze the diversity of the entire V(D)J regions. Recent studies investigated Ig heavy chain repertoire diversity in zebrafish and humans, but not the effect of disease state on the expression of the immune repertoire (5,6).

Here, we characterized the diversity of Ig heavy chain variable regions, determined the effect of PRRSV infection on diversity, identified specific antibody sequences produced in response to PRRSV infection, and characterized the organization and complexity of the porcine light chain genetic loci as a prelude to examination of light chain diversity and the impact of PRRSV infection.

Objectives: 1. Characterize the full range and relative abundances of Ig heavy chain antigen-binding domains in porcine spleen and tonsil. The expected outcomes are (a) the first comprehensive description of porcine Ig diversity and gene segment usage, (b) essential baseline data of Ig expression in healthy commercial swine for comparison to Ig expression in response to infectious disease challenges, and (c) baseline data necessary for determination of the effect of PRRSV infection on Ig expression.

2. Determine the effect of PRRSV infection on Ig heavy chain antigen-binding domain expression in spleen and tonsil. The expected outcomes are (a) identification of V, D, and J regions that play central roles in the antibody response to PRRSV, (b) determine systemic versus regional specialization in antibody repertoire expansion against PRRSV, (c) a solid foundation to investigate the effect of strain virulence on Ig repertoire diversity, and (d) obtain specific sequence information for future expression of recombinant molecules specific for PRRSV proteins.

Materials & Methods: Porcine lymphoid tissues (spleen, palatine tonsil, inguinal lymph node, and mesenteric lymph node) were leveraged from a previous study (7). Briefly, animals were either mock-infected or infected with PRRSV (JA142 strain) at three weeks of age and sacrificed 62 days later. For this study, all four tissues from two infected and two uninfected animals were homogenized, and total RNA was extracted and reverse transcribed. Primers for amplification of the entire variable region were developed using consensus framework

sequences obtained from approximately 140 GenBank porcine Ig heavy chain (IgH) cDNA entries (Table 1). The 454 Titanium FLX fusion adapter and a 10 base barcode for sample differentiation were synthesized at the 5' end of each primer. The resulting PCR amplicons were then pooled according to infection status before being pooled in equimolar amounts. The final pooled product was then agarose gel purified and purified again using the QIAquick PCR purification kit (Qiagen). The resulting sample was then sequenced at the University of Minnesota Biomedical Genomics Center using 454 GS FLX Titanium pyrosequencing.

Name	Region	454 Titanium Adapter	Barcode	Template Specific Sequence
1AF	Leader, 16-35F	5'- CGTATCGCCTCCCTCGCGCCATCAG	TGAACAATCG	AACTGGGTGGTCTTGTTTGC
2AF	Leader, 36-59F	5'- CGTATCGCCTCCCTCGCGCCATCAG	TGAACAATCG	TCTCTTACAAGGTRTCCAGGGTG
3AR	FR4, 491-513R	5'- CTATGCGCCTTGCCAGCCCGCTCAG	TGAACAATCG	TGAGGACACGACGACTTCAA
4AR	FR4, 520-539R	5'- CTATGCGCCTTGCCAGCCCGCTCAG	TGAACAATCG	AAGATTTTGGGGCTGGTTTC
1BF	Leader, 16-35F	5'- CGTATCGCCTCCCTCGCGCCATCAG	ATATCGCGAG	AACTGGGTGGTCTTGTTTGC
2BF	Leader, 36-59F	5'- CGTATCGCCTCCCTCGCGCCATCAG	ATATCGCGAG	TCTCTTACAAGGTRTCCAGGGTG
3BR	C, 491-513R	5'- CTATGCGCCTTGCCAGCCCGCTCAG	ATATCGCGAG	TGAGGACACGACGACTTCAA
4BR	C, 520-539R	5'- CTATGCGCCTTGCCAGCCCGCTCAG	ATATCGCGAG	AAGATTTTGGGGCTGGTTTC

Table 1. Primers used for the amplification of rearranged Ig heavy chain transcripts.

Obtained sequences were separated by barcode (infected vs. uninfected) and sequences containing multiple ambiguous bases were removed from analysis. BLAST (8) was used to identify variable (V), diversity (D), and joining (J) region gene segments based on their previously described germline sequences (4). Translated, full-length sequences were clustered based on similarity using cd-hit (9). Amino acid sequences for all three CDRs were extracted from the full-length sequences for diversity assessment using FileMaker Pro. Conserved framework sequences were used to identify the CDR boundaries. For diversity assessment, reads containing aberrant sequencing errors, ambiguous amino acids, stop codons, or short read length were trimmed. A threshold of 100 percent identity was used to obtain clusters for assessing total IgH diversity. The program EcoSim (10) was used to compute individual-based rarefaction curves. The Chao 1 non-parametric estimate of total diversity was calculated using the program EstimateS (11).

Results: (1.) IgH gene segment usage. Heavy chain antibody variable regions are derived from V, D, and J segments. Approximately 49,000 sequences were used to analyze V segment usage. The sequences aligned to the 15 known V gene segments with high identity (>98%). Novel, low level expression of a previously described pseudogene, IGHV1, also was detected. When V gene segment usage was compared between infected and uninfected pigs significant differences were found for nearly all of the expressed V gene segments. Interestingly, expression of IGHV4 and IGHV11 were proportionally increased in PRRSV-infected animals, suggesting that antigen-binding structures produced from these V segments preferentially bind one or more PRRSV antigens (Figure 1).

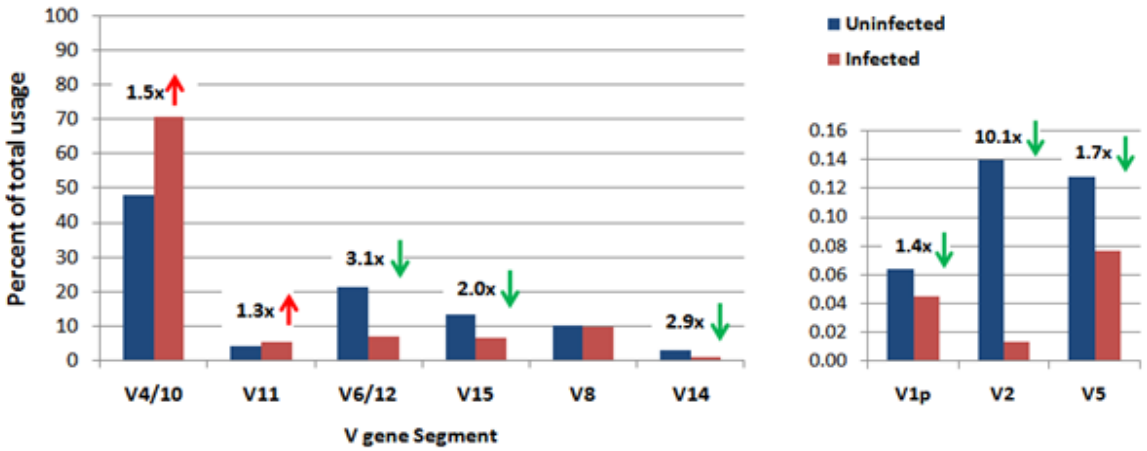

Figure 1. Percent usage of known IGHV gene segments showing fold differences between PRRSV-infected and uninfected groups. All infected/uninfected pairs have $P < 0.001$, except for IGHV8, using a two-tailed χ^2 test.

Approximately 100,000 sequences aligned with high similarity to one of the five known J segments. As expected, IGHJ5 expression was nearly ubiquitous; however, novel, very low level expression of both IGHJ1 and IGHJ3 was detected. Approximately 41,000 reads aligned with high similarity to one of the four known D gene segments. Of these, IGHD1 and IGHD2 were both highly utilized, which has been reported previously (4). In sharp contrast to V segment usage, we found no statistically significant differences in either D or J gene segment usage between PRRRSV-infected and uninfected pigs.

(2.) Total repertoire diversity. Antibody variable regions consist of regions of highly conserved amino acid sequence (framework) alternating with highly diverse complementarity determining regions (CDRs) that provide the contact points that bind to antigens. Heavy chain variable regions contain three CDRs, of which

CDR-H3 is the largest.. Rarefaction analysis, a method used to estimate total population diversity from a test sample population, revealed that CDR-H3 contributes the far larger portion of heavy chain diversity than either CDR-H1 or CDR-H2 (Figure 2).

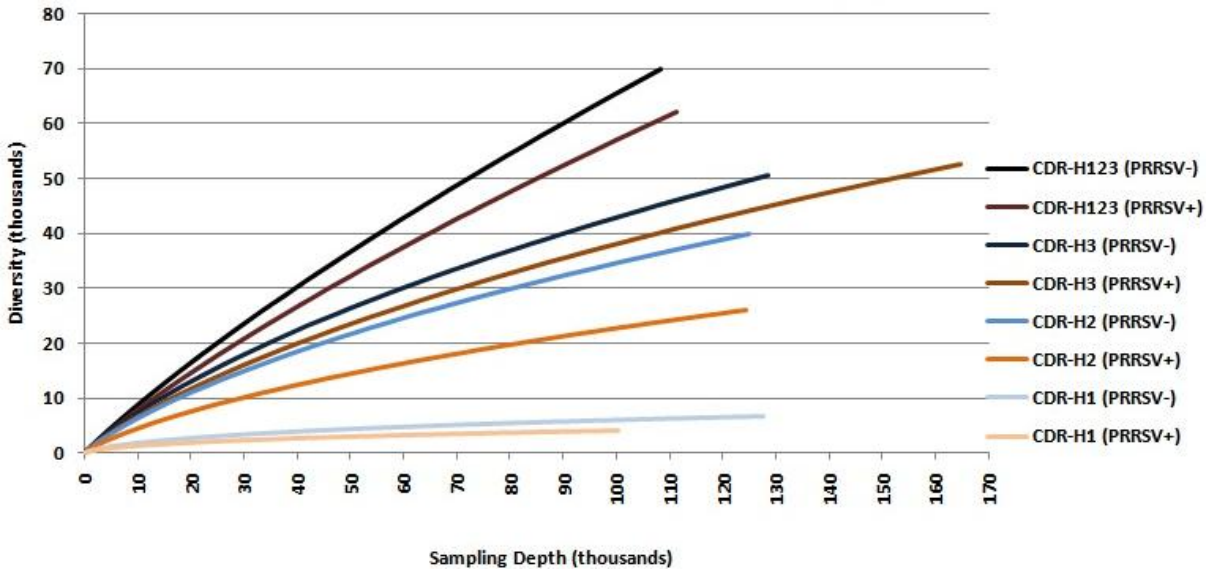

Figure 2. Individual-based rarefaction curves of all three IgH CDRs. Clusters were generated using translated amino acid sequences and 100 percent identity cutoff. Calculated 95 percent confidence intervals (not shown) are not significant.

Interestingly, there was a decrease in antibody diversity as a result of PRRSV infection, likely as a result of clonal selection and proliferation of a relatively small number of B cells with affinity for PRRSV epitopes. Despite the extensive diversity in CDRs, there is a significant constraint on amino acid usage in each of them, particularly CDR-H3. If any, for example, any amino acid were theoretically possible within CDR-H3, we would predict as many as $8x10^{16}$ for a CDR-H3 region of 13 amino acids in length. The actual estimate of total CDR-H3 diversity is approximately 10^5, suggesting that only relatively few amino acid combinations are tolerated (Table 2).

			Chao 1, 95% C.I.				
Status	**CDR**	**Chao 1**	**Lower Bound**	**Upper Bound**	**ACE**	**Dominance**	**Hurlburt's PIE**
PRRSV+	CDR-H1	6,555.83	6,252.96	6,900.05	5,388.04	0.13235	0.96004
	CDR-H2	57,861.06	56,478.00	59,307.53	58,650.13	0.05171	0.99380
	CDR-H3	116,574.53	114,639.69	118,570.41	115,486.05	0.02405	0.99888
	CDR-H123	246,597.04	241,223.00	252,133.17	295,934.44	0.01710	0.99940
PRRSV-	CDR-H1	10,906.47	10,528.27	11,323,89	8,478.47	0.12282	0.97203
	CDR-H2	90,644.95	88,868.74	92,486.07	90,517.40	0.02645	0.99796
	CDR-H3	117,143.53	114,199.97	118,282.14	117,143.53	0.00571	0.99969
	CDR-H123	297,429.28	291,128.64	303,910.51	364,137.32	0.00420	0.99991

Table 2. Richness estimates for the antibody repertoires of pigs. The non-parametric estimators (Chao 1 and ACE) represent a lower bound of total repertoire richness. ACE: abundance-based coverage estimator. Hurlbert's PIE: probability of an interspecific encounter; that is, the probability that any two randomly chosen sequences are non-identical.

(4.) CDR-H3 diversity and the effect of PRRSV infection. Even though the length and sequence of a CDR is fixed in the genome, both can vary in an mRNA due to hypersomatic mutation. CDR-H3, in particular, shows extensive diversity in length and sequence. Analysis of CDR-H3 length distribution reveals that the average length in swine is 13 amino acids, two amino acids longer than recently reported in humans using a similar method (6). Additionally, swine appear to possess an increased preference for short CDR-H3s compared to humans. In pigs that have been infected with PRRSV, the distribution appears to trend the same; however, there is disproportionately greater usage of CDR-H3 lengths 3, 9, 11, 13, 15, 18, 19, and 20 in infected animals. Interestingly, sorting the CDR-H3 clusters by length reveals that many of the CDR-H3s of these lengths possess comparatively large clusters dominated by a single sequence in the PRRSV-infected group. Hence, a small number of clusters account for a large fraction of the total antibody repertoire, especially in PRRSV-infected animals. The distributions for all three CDRs follow a similar power law distribution as previously described for zebrafish antibodies (5, Figure 3).

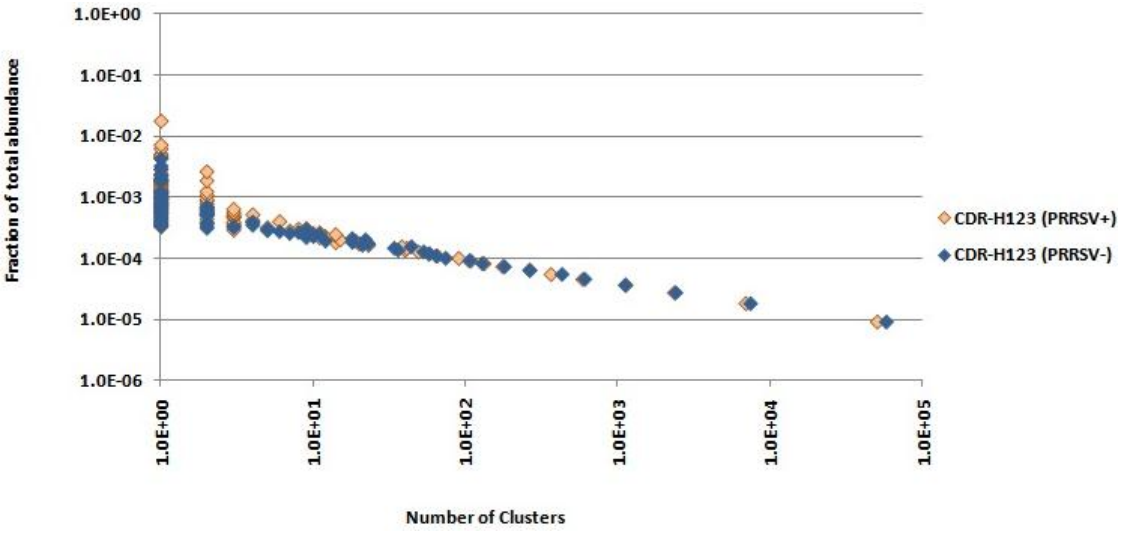

Figure 3. Abundance distributions of the concatenated IgH CDRs (D). All individual CDRs (not shown) follow a similar power law distribution, with very few clusters representing a substantial fraction of the total antibody pool. This phenomenon is accentuated in the PRRSV-infected population, likely owing to expanded PRRSV-specific B cell populations.

The amino acid profile for several CDR-H3 lengths also reveal that single amino acids are preferentially selected for at each position among infected pigs. Comparatively, lengths for which usage reciprocally decreases as a result of infection show no such substantial difference in amino acid profile. Indeed, the CDR-H3

amino acid sequence ADCYSDGGICYFFDHGVMDL accounts for approximately 36 percent of all length 20 CDR-H3s in PRRSV-infected animals. Likewise, GYVYICGWACMDL, the most abundant CDR-H3 in infected animals, accounts for approximately 31 percent of all length 13 CDR-H3s. In uninfected animals, however, these sequences account for only approximately 0.1 percent of CDR-H3s of the same respective lengths. We identified six such highly expressed CDR-H3 sequences in the PRRSV infected group (Table 3). These sequences appear to represent clonal populations.

FR1	CDR-H1	FR2	CDR-H2	FR3	CDR-H3	FR4
EEKLVESGGGLVQPGGSLRLSCVGSGFTFS	STYIN	WVRQAPGKGLEWLA	AISTGGGSTYYADSVKG	RFTISRDNSQNTAYLQMNSLRTEDTARYYCAR	ADCYSDGGICYFFDHGVMDL	WGPGVEVVVSS
EEKLVESGGGLVQPGGSLRLSCVGSGFTFS	SYAVS	WVRQAPGKGLEWLA	CIYSSGSATYYADSVKG	RFTISRDNSQNTAYLQMNSLTTEDTARYYCAK	SVAVGIAFMSFAMDL	WGPGVEVVVSS
EEKLVESGGGLGQPGGSLRLSCVGSGFAFS	SSYIN	WVRQAPGKGLEWLA	AISTSGIGTYYADSVKG	RFTISSDNSQNTAYLQMNSLRTEDTARYYCSR	GYVYICGWACMDL	WGPGVEVVVSS
EEKLVESGGGLVQPGGSLRLSCVGSGFTFS	STYIN	WVRQAPGKGLEWLA	VISTDGVDTYYADSVKG	RFTISRGNSQNTAFLQMNSLRTEDTARYYCVR	GYIYGASYLDL	WGPGVEVVVSS
EEKLVESGGGLVQPGGSLRLSCVGSGFTFS	STYIN	WVRQAPGKGLEWLA	TIYRSDGNTDYEDSVKG	RFTISRDNSQNTAYLQMNSLRTEDTARYYCVR	DVYPSTMDL	WGPGVEVVVSS
EVKLVESGGGLVQPGGSLRLSCVGSGFTFS	VYNMV	WVRQRPGKGLEWLA	CITSRGSSTYYADSVKG	RFTISRDNSQNTAYLQMNSLRTEDTARYYCAR	DSDMDL	WGPGVEVVLSS

Table 3. PRRSV-specific IgH aa sequences. Shown are the most abundant sequences for each PRRSV-specific CDR-H3.

Discussion: As a result of these analyses, several PRRSV-specific CDR-H3 sequences can be derived and associated with full length amino acid sequences. Cluster analyses of all full length sequences bearing these CDR-H3s reveal that they are nearly clonal. The remaining smaller clusters within each CDR-H3 group are nearly identical to each other, suggesting that the most representative sequences are the end result of clonal selection and affinity maturation.

Our results suggest that relatively few B cell clones are highly responsive to PRRSV-infection. Of the six PRRSV-specific antibody sequences described here, two (bearing the CDR-H3s described in Table 3) match the translated germline sequence of IGHV4 to within a single amino acid. A third sequence matches the previously described V_HZ gene segment, though its germline position is uncharacterized and described entirely from cDNA data (1). The remaining three do not closely match any previously described IGHV gene segment; however, they may be derived as a result of extensive somatic hypermutation or they may represent allelic variants of a previously characterized gene segment. We are currently using the high-throughput 454 sequencing facilities at the University of Illinois at Urbana-Champagne to characterize the expression of light

chain transcripts under the same conditions as presented here. This was made possible by our characterization of the light chain loci (12,13) as there was a lack of information on the organization and diversity of the porcine light chain loci. The work that resulted in these two manuscripts was partially funded from this grant. A third manuscript is being developed based on the data presented in this report. The analysis of the light chain expressed repertoire is expected to produce an additional manuscript. Both of these papers will be targeted at the *Journal of Immunology*.

References

1. Butler, J.E., P. Weber and N. Wertz. 2006. Antibody repertoire development in fetal and neonatal piglets. XIII. Hybrid VH genes and the preimmune repertoire revisited. J Immunol 177:5459-5470.
2. Lemke, C.D., J.S. Hayes, R. Spaete, D. Adolphson, A. Vorwald, K. Lager, and J.E. Butler. 2004. Lymphoid hyperplasia resulting in immune dysregulation is caused by porcine reproductive and respiratory syndrome virus infection in neonatal pigs. J. Immunol. 172:1916.
3. Sinkora, M., J. Sinkorova, and J.E. Butler. 2002. B cell development and VDJ rearrangement in the fetal pig. Vet. Immunol. Immunop. 87:341.
4. Eguchi-Ogawa, T., N. Wertz, X.Z. Sun, F. Puimi, H. Uenishi , K. Wells, P. Chardon, G.J. Tobin, and J.E. Butler. 2010. Antibody repertoire development in fetal and neonatal piglets. XI. The relationship of variable heavy chain gene usage and the genomic organization of the variable heavy chain locus. J. Immunol. 184:3734.
5. Weinstein, J.A., N. Jiang, R.A. White III, D.S. Fisher, and S.R. Quake. 2009. High-throughput sequencing of the zebrafish antibody repertoire. Science. 324:307-10.
6. Glanville, J., W. Zhai, J. Berka, D. Telman, G. Huerta, G.R. Mehta, I. Ni, L. Mei, P.D. Sundar, G.M.R. Day, D. Cox, A. Rajpal, and J. Pons. 2009. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. P Natl Acad Sci USA. 106:20216-21.
7. Klinge, K.L., E.M. Vaughn, M.B. Roof, E.M. Bautista, and M.P. Murtaugh. 2009. Age-dependent resistance to Porcine reproductive and respiratory syndrome virus replication in swine. Virol. J. 6:177.

8. Altschul S.F, W. Gish, W. Miller, E.W. Meyers, and D.J. Lipman. 1990. Basic local alignment search tool. J Mol Biol 215:403-410.

9. Li, W. and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22(13):1658-9.

10. Gotelli, N.J. and G.L. Entsminger. 2011. EcoSim: Null models software for ecology. Version 7. Acquired Intelligence Inc. & Kesey-Bear. Jericho, VT 05465. http://garyentsminger.com/ecosim.htm.

11. Colwell, R. K. 2009. EstimateS: Statistical estimation of species richness and shared species from samples. Version 8.2. User's Guide and application published at: http://purl.oclc.org/estimates.

12. Schwartz, J. C., M-P. Lefranc, and M. P. Murtaugh. 2011. Evolution of the porcine (*Sus scrofa domesticus*) immunoglobulin kappa locus through germline gene conversion. Immunogenetics. doi:10.1007/s00251-011-0589-6:epub ahead of print

13. Schwartz, J. C., M-P. Lefranc, and M. P. Murtaugh. 2011. The organization, complexity and allelic diversity of the porcine (*Sus scrofa domesticus*) immunoglobulin lambda locus. Immunogenetics. doi:10.1007/s00251-011-0594-9:epub ahead of print