ANTIBODY REPERTOIRE DYNAMICS IN THE CHANGING LANDSCAPE OF INFECTION

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

JOHN CHARLES SCHWARTZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

MICHAEL P. MURTAUGH, Ph.D. ADVISOR

MAY 2013

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Michael Murtaugh for the opportunity to work in his lab, for his enthusiasm and interest in this project, and for his tremendous guidance and support throughout my studies.

I would also like to thank my committee members, Dr. Reuben Harris, Dr. Thomas Molitor, and Dr. Kent Reed for their input and suggestions which helped to shape this project. Along with my advisor, their efforts have helped transform me into a competent scientist.

Numerous individuals contributed to the work contained herein. Dr. Marie-Paule Lefranc at the Université Montpellier 2, an expert in the field of immunogenetics, provided generous support in the annotation of the antibody light chain loci. She is co-author on two of the included manuscripts in this thesis (Chapters 2 &3). Dr. Chris Tuggle and Dr. James Reecy at Iowa State University and Dr. Jane Loveland at the Wellcome Trust Sanger Institute provided annotation support and helpful correspondence concerning the pig genome project. I would like to thank the rest of the swine genome sequencing consortium for including me as a member of that collaborative project which resulted in the co-authorship of two additional manuscripts. I thank Dr. Benjamin Hackel in the College of Engineering and Materials Science for generously providing support and reagents for the on-going development of an antibody yeast display library.

I thank my friends and colleagues at South Dakota State University, specifically Drs. Chris Chase, Ying Fang, Steve Lawson, Eric Nelson, Susan Gibson, Alan Young, and Scott Pedersen, as well as Nancy Thiex, Brian Steinlicht, and Travis and Liz Clement

for providing me with an educational foundation in the biological sciences, for allowing me to work in their labs, for introducing me to PRRSV research and molecular biology, and for their friendship.

I would especially like to thank Dr. Cheryl Dvorak, Dr. Juan Abrahante, and the late Dr. Craig Johnson for their efforts in the Murtaugh lab and to Dr. Abrahante specifically for his tremendous guidance, friendship, and knowledge in the realm of bioinformatics. I would also like to thank my wonderful colleagues in the Murtaugh lab: Erin Babineau and Jenny Zhang who provided excellent technical assistance, as well as Dr. Sally Robinson and Suzanne Stone who assisted with animal tissue collection and assayed serum for neutralizing antibody titers and for performing helpful random tasks around the lab.

Past members of the Murtaugh lab also require acknowledgement; namely, Dr. Sumathy Puvanendiran, Dr. Josephine Gnanandarajah, and Dr. Juan Li for providing me with guidance during my first few years in the lab. I would also like to thank additional members of the Murtaugh lab not mentioned hitherto: Dr. Xiong Wang, Dr. Kyra Martins, Xi Guo, Ashley Swenson, Heidi Schatz, and visiting scientist Dr. Jie Zhang for a stimulating and enjoyable work environment. All of these aforementioned names and their associated faces, I will never forget.

Throughout my studies, I was blessed to have received funding from multiple sources, without which, completion of this thesis would not have been possible: the Department of Veterinary Biosciences, National Pork Board grant 10-139, a pre-doctoral fellowship through the Institute for Molecular Virology (NIH T32 AI83196), and a

doctoral dissertation fellowship through the Graduate School. I am immensely grateful for their support.

I thank the Comparative and Molecular Biosciences program and the College of Veterinary Medicine for the opportunity to pursue my graduate studies, as well as Lisa Hubinger and Kate Berry for their tremendous efforts and logistical expertise in coordinating the graduate program. I would also like to thank the college faculty and students for their support and comradeship.

Lastly, and perhaps most importantly, I sincerely thank my family for their moral support and for assisting me in achieving my goals.

This dissertation is dedicated to my parents and to my sisters who taught me to experience the world with an open mind.

ABSTRACT

Antibody responses are fundamentally important to effector and memory mechanisms of disease resistance. In order to respond to a nearly infinite array of possible antigens, the antibody repertoire must be suitably diverse. To achieve this necessary high level of diversity, the antibody repertoire has evolved a unique recombinatorial system consisting of a large number of gene segments that can recombine in different combinations to yield an astronomical array of potential antigen-binding structures. Understanding the antibody repertoire of swine (*Sus scrofa*) can inform about host genetic differences that may affect disease susceptibility and resistance. Also, it may allow identification of antibody molecules that are important in the host immune response against specific pathogens. Such knowledge could potentially be used in the future to develop selective breeding programs for animals that possess desirable immunological traits, and to screen for specific antibody molecules that are of either therapeutic or diagnostic importance.

Knowledge of antibody repertoire diversity in swine has heretofore been lacking. While most previous studies have focused heavily on understanding the heavy chain repertoire by analyzing hundreds of cDNA clones, there have been few investigations of the porcine light chain repertoire. This study was designed to characterize the organization and complexity of both the kappa and lambda light chain loci in the pig genome. Findings revealed extensive allelic variation between both homologous pairs of chromosomes in a single sow and suggested non-crossover homologous recombination

(i.e. gene conversion) as a potential evolutionary mechanism to explain at least part of that variation.

Armed with this new information, and with that from the previously characterized heavy chain locus, antibody variable region amplicon libraries were generated from lymphoid tissues of pigs either infected (n=2) or mock-infected (heavy chain, n=2; light chain, n=3) with the major swine pathogen, porcine reproductive and respiratory syndrome virus (PRRSV). It is hypothesized that the major anti-PRRSV antibody responses would be detectible in infected animals compared to their control counterparts.

Approximately a half-million reads for each heavy and light chain library were generated. From this data, diversity of the expressed antibody repertoire was assessed, including gene segment usage and allelic variability, and anti-PRRSV responses. As predicted, due to biological necessity, the heavy and light chain repertoires possessed a rich array of putatively functional antibody transcripts (heavy chain richness estimate, $3x10^5$ molecules; kappa light chain, $1.5x10^5$; lambda light chain, 2.3×10^5), despite being restricted in their germline to a small number of functional D and J gene segments, a single heavy chain V gene segment family and four light chain V gene families. Interestingly, a power-law distribution of antibody abundances was detected similar to what has previously been reported in zebrafish (*Danio rerio*), whereby a small number of antibody sequences are exceptionally common and the vast majority are exceptionally rare. Substantial allelic variation was also detected, most notably in the lambda locus. Four out of 5 pigs possessed a functional copy of a previously undescribed V gene segment (IGLV3-1-1) which substantially contributed to the expressed repertoire of the

animals that possessed a copy. Importantly, a small number of antibody sequences were detected which were incredibly abundant (>1% of the entire repertoire) in PRRSV-infected pigs and rare in uninfected pigs. It is hypothesized that these highly abundant antibody molecules are PRRSV-specific.

Using the knowledge obtained from these studies, future investigations will examine the repertoire for specific heavy and light chain pairs from PRRSV-infected pigs that can neutralize PRRSV using an antibody yeast-display system. In addition, specific heavy and light chain pairs identified in our expression analysis and deemed putatively PRRSV-specific are to be tested for epitopic specificity against labeled PRRSV as well as individual PRRSV recombinant antigens. This last method represents a potential novel and non-lethal means of generating antigen-specific recombinant antibodies derived from lymphoid tissue of immunized animals.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER V

LIST OF FIGURES

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER V

LIST OF ABBREVIATIONS

AID	activation-induced cytidine deaminase
B2M	beta-2-microglobulin
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
bp	base pairs
BTA	*Bos taurus* chromosome
C	constant region
CDC	Centers for Disease Control and Prevention
cDNA	complementary DNA
CDR	complementarity determining region
CSR	class switch recombination
D	diversity gene segment
dA	(deoxy)adenine
dC	(deoxy)cytosine
dG	(deoxy)guanine
dT	(deoxy)thymine
dU	(deoxy)uracil
DNA	deoxyribonucleic acid
DNA-PK	DNA-dependent protein kinase
ELISA	enzyme-linked immunosorbent assay
EST	expressed sequence tag

FACS	fluorescence-assisted cell sorting
FAO	Food and Agriculture Organization of the United Nations
Fc	fragment, crystallizable (i.e. the antibody constant region)
FDA	[United States] Food and Drug Administration
FISH	fluorescence in situ hybridization
FR	framework region
HIV	human immunodeficiency virus
Ig	immunoglobulin
IGH	immunoglobulin heavy
IGK	immunoglobulin kappa
IGL	immunoglobulin lambda
IgSF	immunoglobulin superfamily
IMGT	ImMunoGeneTics (an IgSF numbering system and online database)
J	joining gene segment
kb	kilobase pairs
KDE	kappa deleting element
KIR	killer-cell immunoglobulin-like receptor
LILR	leukocyte immunoglobulin-like receptor
Mab	monoclonal antibody
MHC	major histocompatibility complex
mRNA	messenger RNA
NHEJ	non-homologous end joining

nsp	non-structural protein
ORF	open reading frame
PCR	polymerase chain reaction
PRRSV	porcine reproductive and respiratory syndrome virus
RAG	recombination activating gene
RNA	ribonucleic acid
RS	recombination signal
RSV	respiratory syncytial virus
scFv	single chain fragment, variable
SHM	somatic hypermutation
Sμ	switch region, mu
SNP	single nucleotide polymorphism
SSC	*Sus scrofa* chromosome
TCR	T-cell receptor
TdT	terminal deoxynucleotidyl transferase
TNF	tumor necrosis factor
UNG	uracil-DNA glycosylase
V	variable gene segment
VH	variable heavy
VL	variable light
XRCC4	X-ray repair cross-complementing factor 4

CHAPTER I

COMPARATIVE IMMUNOLOGY OF PORCINE ANTIBODIES:

A REVIEW

The present review explores current understanding of the antibody repertoire in agriculturally relevant species, largely focusing on the pig (*Sus scrofa*) as a model of antibody repertoire dynamics. Evolutionary mechanisms have molded the repertoire through selective pressure, presumably to avoid auto-immunity and to enhance the immunological outcomes against pathogens and prevent repeat infections. Additional diversification is achieved at the somatic level through gene segment recombination, through mutations arising from enzymatic activity at the gene segment junctions, and through somatic hypermutation in antigenically-stimulated B-cells as a means of improving antibody affinity. Pigs are typical in many ways to other artiodactyls, including having a relatively restricted germline antibody repertoire that relies heavily on somatically generated diversity and allelic variation. While perhaps beneficial at an evolutionary scale, these factors may lead to consequences for commercial animal production, where homogeneity and uniformity are desirable characteristics. Individual variation in immune responses to infection, variable disease outcomes, and variable cross-protective vaccine efficacy all aggravate efforts at controlling disease in these populations. Thus, a detailed understanding of how the antibody repertoire is generated and how that repertoire responds to infection is of importance to animal agriculture.

1.1 - Introduction

Acquired humoral immune responses are a critical component of immunity to infectious disease. However, individual variation in disease outcomes and in vaccine efficacy complicate attempts at controlling disease through immunological means, especially for rapidly evolving pathogens, such as RNA viruses. Many advances in understanding human and mouse immunogenetics have been made, however, knowledge of the immune repertoires of important agricultural species is trailing. Knowledge of the humoral response and its repertoire is crucial to the maintenance and welfare of agricultural species and to the human food supply through vaccination and antibody engineering.

The ability of a vaccine to produce neutralizing antibodies is a major correlate of protection against viral pathogens (Burton 2002). It is therefore an ideal goal of future vaccines to incorporate components that stimulate the development of high neutralizing antibody titers that are effective against antigenically-diverse strains. Recently, significant strides have been made in understanding what the neutralizing epitopes of HIV are as well as what the antibodies are that recognize them. Problematically, these antibodies exert selective pressure on the virus, resulting in the development of escape mutants within a short time period (Wei et al. 2003). Intriguingly, however, a combination of several passively-administered broadly neutralizing antibodies were effective in controlling HIV infection in humanized mice (Klein et al. 2012).

Modern understanding of the human and mouse antibody repertoires have led to significant advances in medicine. Over 20 therapeutic monoclonal antibodies (Mabs) are

in clinical use, many of which target components of the immune system as therapeutics for auto-immune disorders and the prevention of graft rejection (Chames et al. 2009). However, at least one FDA-approved Mab, palivizumab, is used for the prevention of an infectious disease. Palivizumab is a humanized monoclonal IgG antibody derived from mice that is specific for the F (fusion) protein of respiratory syncytial virus (RSV), and is effective at preventing RSV-associated hospitalizations (Andabaka et al. 2013). In addition, a CD4-specific Mab, ibalizumab, is a promising candidate for the treatment and prevention of HIV infection, as it prevents viral entry into $CD4^+$ T-cells, yet does not interfere with the cells' functions (Pace et al. 2013).

In 1991, Barbas et al. established a means of reconstituting and assaying the entire antibody repertoire using phage surface display. The utility of this technology was immediately apparent as it allowed for deeper investigation into the humoral immune response, particularly against HIV. In particular, antibody surface display technology allows for high throughput characterization of antibody molecules generated entirely in vitro. Thus, such libraries have since become a powerful tool for developing new therapeutic and diagnostic antibodies. The most successful such antibody is, arguably, adalimumab (Humira), the first entirely-human FDA-approved Mab, which recognizes tumor necrosis factor (TNF)-α and is widely used for the treatment of various chronic inflammatory disorders including rheumatoid arthritis and ankylosing spondylitis (Bain and Brazil 2003).

As a prerequisite, the genetic structure of the antibody repertoire must be characterized in order to generate display libraries and to determine what, if any, genetic

factors play a role in the generation of effective antibody responses to pathogens. A lack of information regarding the antibody repertoires of species other than humans and mice is an impediment toward the discovery of protective antibody molecules in agricultural species and, perhaps, ultimately toward the development of efficacious vaccines with broad neutralizing ability.

Recently, however, two large consortiums have sequenced the genomes of the agriculturally important species *Sus scrofa* (swine) and *Bos taurus* (cattle), allowing for greater characterization of their immunogenetic loci (Groenen et al. 2012; Elsik et al. 2009; Dawson et al. 2013). Focusing on the former species, the present review summarizes the new information that has been gleaned from these studies and others and provides insight into what remains to be known and the implications for the health and welfare of agricultural species.

1.2 - Swine as a model of immunological dynamics

Pork is the most widely consumed meat in the world, followed by poultry and beef (FAO 2009). The United States is the world's second largest pork producing nation behind China and pork consumption in the U.S. lags slightly behind poultry. Thus, pigs are an immensely important species to humans, both economically and nutritionally. Swine possess characteristics that are somewhat intermediate between humans and other members of the order Cetartiodactyla, making them an ideal non-primate species for studying immune system development and evolution in domesticated species. Such characteristics include the production of large litters, making then amenable to study,

similar nutritional requirements and digestive system to humans, and antibody light chain and T-cell receptor repertoires similar to humans. Furthermore, pigs are a host for zoonotic pathogens, of which influenza virus is perhaps the most important (Butler et al. 2009a). Thus, structural and functional characterization of antibody production is important for swine health and as a comparative model for adaptive immune system dynamics.

Pigs belong to a large group of even-toed hoofed mammals (artiodactyls), which together with the Cetacea, form the order Cetartiodactyla comprising approximately 290 extant species (Price et al. 2005). The major clades within the order include the Suiformes (swine and peccaries), Camelidae (camels and llamas), Hippopotamidae (hippopotamuses), Cetacea (whales and dolphins), and Ruminantia, which contains the families of Bovidae (cattle, sheep, goats, and antelope), Cervidae (deer), Moshidae (musk deer), Antilocapridae (pronghorn antelope), Giraffidae (giraffes), and Tragulidae (chevrotains) (Price et al. 2005). Within the order, only the swine and cattle (*Bos taurus*) genomes have been sequenced and annotated (Groenen et al. 2012, Elsik et al. 2009). Apart from these two species, relatively little is known about antibody gene content and usage among the other cetartiodactyls. The most recent common ancestor shared by the cetartiodactyls lived between 65 and 71 million years ago, when the camelids diverged from the rest of the clade (Meredith et al. 2011). The suborder Suiformes represents the next earliest divergence, circa 65 million years ago, and shared a common ancestor with humans between 79 and 96 million years ago (Kumar and Hedges 1998; Groenen et al. 2012).

While there are many similarities in the immune systems of swine, humans, and cattle, there are numerous differences as well. The Suiformes, Hippopotamidae, and Dolphinidae of the cetartiodactyls, as well as the Rhinocerotidae and Elephantidae of the perissodactyls (odd-toed ungulates) possess an inverted lymph node structure and inverted lymphatic flow not found in other organisms (reviewed by Binns 1982). Also, antibody light chain isotype usage and light chain germline repertoire diversity in pigs is more similar to humans than cattle (Butler et al. 2006a, Schwartz et al. 2012a,b). However, antibody heavy chain germline diversity is severely restricted in both pigs and cattle compared to humans, with both artiodactyl species possessing only a single variable gene segment family (Butler et al. 2006a; Chapter 5). Compared to mice, however, and discussed later in more detail, pigs possess more similar antibody light chain isotype usage and gene structure to humans, perhaps making pigs a more appropriate immunological model for humans than mice.

1.3 - A brief history of antibodies

The origin of immunology as a modern field of scientific study is often credited to Edward Jenner. In 1796 he demonstrated that inoculating patients with material from cowpox lesions (i.e. *Variola vaccina*) conferred cross-protection against smallpox. This insight eventually led to one of the single greatest achievements in recent human history - the global eradication of smallpox by 1980 (Breman and Arita 1980). Most recently, in June 2011, as the result of an intense vaccination program, the Food and Agriculture Organization (FAO) of the United Nations formally declared that rinderpest had also

been globally eradicated, and was therefore the first agriculturally-relevant disease to become extinct (Horzinek 2011). Furthermore, the vaccination campaign against polio has resulted in its near-eradication with a mere 650 new cases reported worldwide in 2011 (CDC 2012); and, relevant to the swine industry, classical swine fever was eliminated from the United States in 1976 followed by pseudorabies in 2005 (Edwards et al. 2000; USDA-APHIS 2008). Thus, the sustained inquiry into the mechanisms of immunological protection and the practical use of that knowledge provides the potential to make a powerful impact on the well-being of people directly and indirectly through the protection of the food supply.

Investigation into the mechanisms of acquired immunity received its first leap forward in 1890, when Emil Adolf von Behring discovered that rabbits immunized with heat treated diptheria toxin were protected from challenge (Behring and Kitasato 1890). Importantly, this protection was conferred upon passive transfer of their sera to other animals - an observation which lead to the conclusion that immunized animals had formed *antitoxins* in their sera (Behring and Kitasato 1890).

Inspired by his work with Behring on developing serum therapy to combat outbreaks of diphtheria, Paul Ehrlich, in 1897 developed the side-chain theory (German: *seitenkettentheorie*) of antibody formation and antibody-antigen interaction (Ehrlich 1897). He postulated that antibodies are chemical substances produced by cells and expressed in either a receptor or a secreted form and are generated upon stimulation with antigen. Further, antigen-antibody interactions were imagined to be stereochemical in nature, similar to the "lock-and-key" (German: *schlüssel-schloss-prinzip*) model of

enzymatics proposed by Emil Fischer a few years earlier (Fischer 1894). Over time, Ehrlich's prophetic theory would largely prove to be true and laid the foundation for later characterizations of immunoglobulin structure and function.

In 1937, Swedish researcher Arne Tiselius was the first to electrophoretically separate serum proteins (Tiselius 1937). The resulting bands, in order of migration following serum albumin, were classified as globulins α, β, and γ. The relative concentrations of these fractions was determined in a single rabbit, both before and after immunization with ovalbumin. There was a substantial increase in the relative concentration of the γ-globulin fraction (from 17.2 percent to 33 percent of total) and a corresponding decrease in albumin and the β-globulin fractions. He also showed that antibodies generated in response to ovalbumin migrated with γ-globulin, as approximately 25 percent of the total nitrogen from this isolated fraction specifically precipitated upon addition of ovalbumin (Tiselius 1937).

The antibody light chain was first discovered in 1848. In that year, Henry Bence Jones described an albumin-like substance in the urine of multiple myeloma patients. Much later, in 1922, Bayne-Jones and Wilson concluded that there were at least two groups of these so-called Bence Jones proteins based on their immunoprecipitative characteristics (Bayne-Jones and Wilson 1922). This finding was later confirmed by Korngold and Lipiri in 1956. Using antisera raised against Bence Jones proteins, passive double immunodiffusion assays were performed and with the conclusion that only two types of Bence Jones proteins exist. The two serotypes κ and λ were designated, after their last initials (Korngold and Lipari 1956). However, the importance of these proteins

to the field of immunology was unknown until 1962 when Edelman and Gally showed that Bence Jones proteins are equivalent to a light chain polypeptide component of immunoglobulin (Edelman and Gally 1962).

Finally, in 1970, Wu and Kabat used existing amino acid sequence data from Bence Jones proteins to characterize the primary structure of the light chain variable region (Wu and Kabat 1970). This characterization revealed three hypervariable regions, or complementarity-determining regions (CDRs), responsible for antibody-antigen interaction. In 1973, Poljac et al. used x-ray crystallography to conclusively determine the three dimensional structure of an antibody immunoglobulin domain (Poljac et al. 1973).

1.4 - Immunoglobulin structure

Antibodies belong to a large class of molecules collectively known as the immunoglobulin superfamily (IgSF). IgSF members not only include the genes of the B-cell receptor (i.e. antibodies), but also those of the T-cell receptor (TCR), the genes of the natural killer-cell immunoglobulin-like receptor (KIR), the genes of the major histocompatibility complex (MHC) and beta-2 microglobulin, leukocyte immunoglobulin-like receptors (LILR), certain cytokine receptors, immunoglobulin binding receptors (e.g. Fc receptors), antigen receptor and co-stimulatory molecules (e.g. CD3, CD28, CD80, and CD86), cellular adhesion molecules, growth factor receptors, and others (See Barclay 2003 for review).

Kabat et al. (1991) established early numbering schemes for various IgSF members

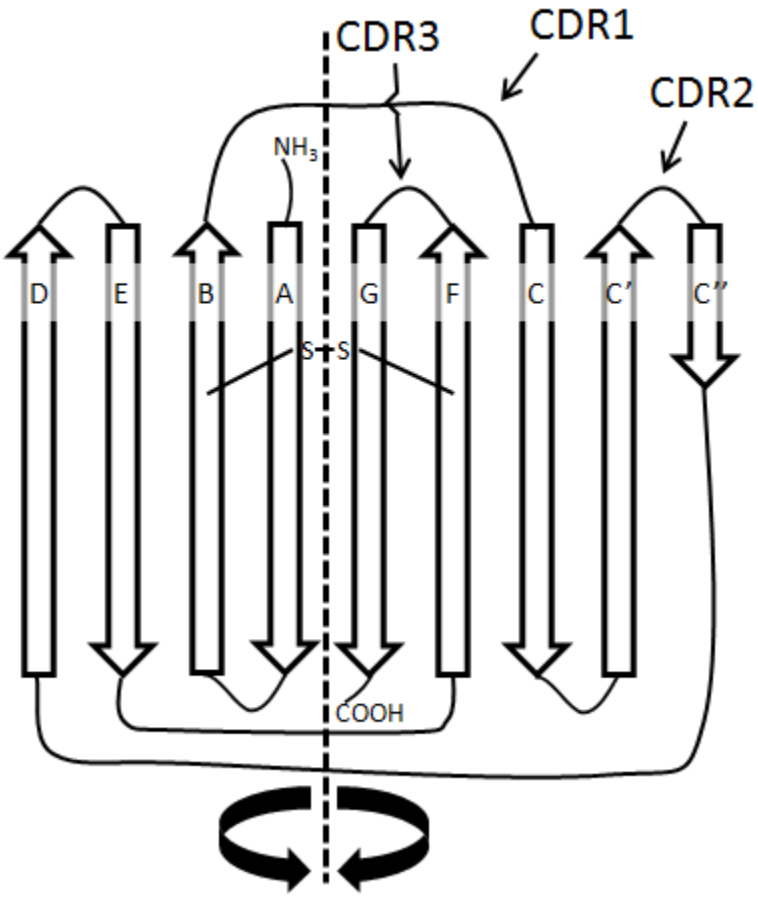
based on their initial analyses of aligned protein sequences (Kabat et al 1991). However, due to the limited availability of crystal structures these were largely based on sequence alone, rather than physical structure. More recently, ImMunoGeneTics (IMGT) residue numbering and nomenclature has standardized IgSF numbering based on an understanding of the physical structure of the Ig domain and is now widely used (Lefranc et al. 2003, 2005). An additional advantage of the IMGT numbering scheme is that conserved residues are also conserved in their numbered position, allowing for more efficient comparisons between various immunoglobulin (Ig) domains. For these reasons, residue numbers and immunoglobulin nomenclature in the present paper will henceforth utilize the IMGT scheme.

The Ig domain is comprised of approximately 100 amino acids arranged in anti-parallel β-strands (described as A through F) which form two β-sheets. Antibody variable domains include two additional β-strands not found in the constant domain, C' and C", which together form CDR2 in their intermediary loop. The two β-sheets are folded such that they face each other due to a disulfide bond between two conserved cysteine residues on strands B and F (Cys23 and Cys104, respectively). While the β-strands themselves comprise the highly conserved framework regions (FR), the three hypervariable CDRs, which confer antigenic binding and specificity, are found on loops BC (CDR1), C'C" (CDR2), and FG (CDR3) (Figure 1.1).

Antibody molecules consist of two identical heavy chain polypeptides and two identical light chain polypeptides with each chain containing a variable region comprised of a single Ig domain and a constant region. In the light chain, the constant region

contains a single Ig domain, whereas the heavy chain constant region contains multiple Ig domains, depending on the isotype. The variable regions of each heavy and light chain pair confer antigenic specificity, such that each antibody molecule contains two antigen binding sites. Additionally, depending on heavy chain constant region isotype, antibodies may increase the number of antigen-binding sites by forming multimers (i.e. dimeric IgA and pentameric IgM), thereby increasing the avidity of the molecule for antigen. Interestingly, camelids have evolved unusual, yet fully functional, heavy chain isotypes (IgG1 and IgG2) which are incapable of forming disulfide bonds with light chains (Hamers-Casterman et al. 1993). As a possible mechanism to compensate for the accompanying loss of diversity, camelids have evolved non-canonical CDR loop structures which are longer and more diverse than is typical of other species (Sircar et al. 2011). It has also been shown recently that a specific heavy chain CDR3 is capable of cross-neutralizing influenza A virus via the recognition of a small epitope on hemagglutinin, suggesting that this region is by itself sufficient for antigenic specificity (Ekiert et al. 2012).

FIGURE 1.1 Typical Ig domain structure of an antibody variable region. Wide bands indicate β-strands and are labeled A-G from N-terminus to C-terminus. The plane of folding between the two β-sheets is indicated with a dotted line. The conserved disulfide bond between strands B and F is indicated as are the three CDR loops at the N terminal pole (Adapted from Barclay 2003).



1.5 - Antibody gene structure

Among mammals, there are three distinct antibody loci in the genome: a single heavy chain locus, and two light chain loci (κ and λ). In pigs, these loci have been identified on *Sus scrofa* chromosome (SSC) 7q25-26 (heavy chain), SSC 3q12-14 (κ), and SSC 14q16-21 (λ) (Juneja et al. 1986; Frönicke et al. 1996, Retternberger et al. 1996; Yerle et al. 1997). Conceptually, each antibody locus can be thought of as a single gene that is extensively and irreversibly modified during B-cell development as a result of the recombination of various gene segments.

All three loci contain multiple variable (V) gene segments, each consisting of a single Ig domain of approximately 100 amino acids (Figure 1.2). Each V gene segment is preceded by promoter elements; that is an octamer with the canonical sequence ATTTGCAT and a TATA box approximately 85 bases downstream from the octamer. The ATG start site is approximately 15 bases further downstream followed by a 45-50 base pair coding region (IMGT: L-PART1). The coding region of this first exon, along with the first 7-11 bp of the second exon (IMGT: L-PART2) encode the post-translationally cleaved leader region which is crucial for the transport of the nascent molecule to the cell surface. An intron of approximately 150 to 330 base pairs in length separates the first exon from the second exon of the variable region. The first ~300 bp of this second exon (IMGT: V-EXON) encodes the bulk of the Ig domain of the variable region. Such is the structure of a single V gene segment.

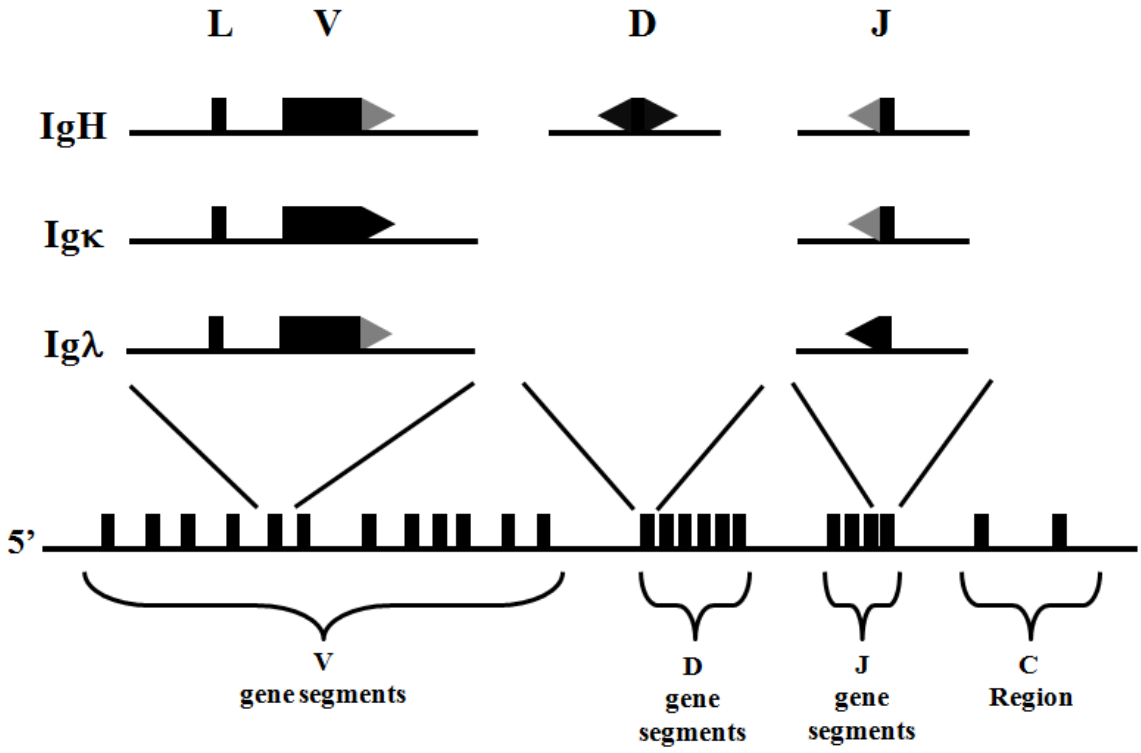
The V gene segments of the heavy chain locus are arranged upstream from multiple diversity (D) gene segments each typically encoding approximately 12 amino acids.

These D gene segments are arranged upstream from multiple joining (J) gene segments which encode approximately 17 amino acids each. Light chains are similar, except that they lack D gene segments. As the D gene segment of the heavy chain and the J gene segment each contribute sequence to the FG loop of CDR3, all three gene segments contribute to the antigenic specificity of the variable region. The intervening sequence between the J gene segment and the constant region exon(s) functions as an intron, complete with typical intronic structure, including 5' (GT) and 3' (AG) splice sites and a polypyrimidine tract. This region is therefore excised (including any intervening J gene segments) during mRNA processing.

Recombination signals (RS) are found immediately downstream from the V gene segments, flank both sides of the D gene segments, and are upstream from the J gene segments. These serve as recognition sequences for the recombination activating gene (RAG)1/RAG2 complex. This recombinase catalyzes a reaction which results in the ablation of all intervening DNA between two RS sites. These RS sites contain an obligatorily conserved heptamer (CACAGTG) immediately adjacent to the sequence of interest followed by a spacer sequence of either 12 or 23 bases and a highly conserved nonamer (ACAAAAACC). Recombination occurs between an RS with a spacer of 12 bases and one with 23 bases. In this manner, recombination is restricted so that only the correct final product is typically obtained. That is, only products containing V(D)J rearrangements are generally possible (Figure 1.2). Thus, a multitude of different gene segments can alternatively pair with one another, potentially resulting in a large number of combinatorial products.

The DNA cleavage generated by the RAG complex creates a hairpin loop at each end. These exposed ends result in the recruitment of the Ku70/Ku80 heterodimer which binds to the DNA ends and anchors DNA-dependent protein kinase (DNA-PK), which in turn recruits additional DNA repair enzymes, namely, the endonuclease Artemis, DNA-ligase IV, X-ray repair cross-complementing factor 4 (XRCC4), and Cernunnos (an XRCC4-like protein) (Reviewed by Gellert 2002). Upon recruitment, Artemis imprecisely digests the hairpin loop formed by RAG at the DNA ends allowing for further DNA end repair. DNA-ligase IV, XRCC4, and Cernunnos together form a complex that recruits additional exonucleases and terminal deoxynucleotidyl transferase (TdT) which act to further diversify the junction between gene segments by both the deletion of nucleotides and the addition of random nucleotides to the exposed ends. The DNA-ligase complex then mediates non-homologous end joining (NHEJ) by aligning and ligating the two ends together. The resulting junctional diversity greatly expands the repertoire of the heavy chain CDR3, making this region the longest and most variable of the six CDRs. Light chain V-J recombination follows a similar mechanism, except that while B-cells undergo V-J recombination they typically lack TdT and exonuclease activity, resulting in a light chain CDR3 that is restricted in both length and sequence variability. Due to their role in both B-cell and T-cell development, many of the enzymes involved with V(D)J recombination, and more generally, NHEJ, are associated with severe immunodeficiencies (Gennery 2006; Luo et al. 2004). Thus, a functionally rearranged V(D)J region forms a single complete exon (Figure 1.2).

FIGURE 1.2 Typical antibody gene structure. Shown is an example genomic backbone of a heavy chain locus. Actual number and spacing of individual gene segments varies greatly between species (and possibly individuals). The light chain loci (κ and λ) lack D gene segments and the lambda locus is organized with tandem J-C repeats in both primates and artiodactyls (not shown). Grey triangles represent recombination signal (RS) sequences containing a 23 bp spacer and black triangles represent RS sequences containing 12 bp spacers. This arrangement makes it highly unlikely that incorrect recombination events will occur. L, leader; V, variable; D, diversity; J, joining; C, constant.


L
V
D
J
IgH
Igκ
Igλ
5'
V
gene segments
D
gene
segments
J
gene
segments
C
Region

1.6 - The antibody germline repertoire

The genomic organization of the porcine heavy chain has been partially characterized (Eguchi-Ogawa et al. 2010, 2012). Swine possess five heavy chain J (IGHJ) gene segments, only one of which is functional (IGHJ5), and four heavy chain D (IGHD) gene segments, of which two are functional (IGHD3 and IGHD4). In the constant region from V-proximal to V-distal, swine possess IgM, IgD, six IgG subclasses: IgG3, IgG5-1, IgG5-2, IgG1, IgG6-1, and IgG6-2 followed by IgE and IgA (Butler et al. 2006b, 2009b; Eguchi-Ogawa et al. 2012). Unfortunately, the known IgG subclass effector functions from other species cannot be used to infer functions in the pig (or in other mammals) as the expansion of constant region subclasses occurred independently after speciation (Kehoe and Capra 1974). Thus, there is a lack of understanding regarding the effector functions for the various porcine IgG subclasses.

All porcine heavy chain variable (IGHV) gene segments derive from a single gene expansion indicated by their high level of framework region conservation. Consequently, all IGHV gene segments are closely related to and most similar to the IGHV3 gene family of the mammalian IGHV clan III (Figure 1.3; Sun et al. 1994). The first 15 IGHV gene segments have been mapped, of which 10 are putatively functional (Eguchi-Ogawa et al. 2010). Interestingly, there is a duplication of six IGHV gene segments which occurred approximately 1.6 million years ago. Four of these twelve gene segments (IGHV4/IGLV10 and IGHV6/IGHV12) are highly expressed, suggesting that this duplication had some immunological benefit to pigs (Eguchi-Ogawa et al. 2010). In total, swine are predicted to possess between 20 and 31 different IGHV gene segments (Butler

et al. 2006b).

All known expressed cattle IGHV gene segments are restricted to a single family (bovine subgroup IGHV1), most similar to the human IGHV4 family of clan II, as in sheep (*Ovis aries*) (Figure 1.3). A preliminary analysis of the bovine heavy chain repertoire mapped several apparently functional IGHV gene segments distributed between *Bos taurus* chromosomes (BTA) 7 and 21 (Pasman and Kaushik 2012). However, florescence in situ hybridization (FISH) has since established a single bovine heavy chain locus on BTA 21 and not elsewhere, suggestive of inaccuracies in the *Bos taurus* genome assembly (Niku et al. 2012). This most recent analysis revealed 36 IGHV gene segments in three distinct families spanning mammalian IGHV clans I and II. However, only 10 functional gene segments belonging to bovine IGHV1 of clan II were identified, while the remainder are pseudogenes (Niku et al. 2012).

The porcine light chain loci have also recently been characterized (Schwartz et al. 2012a,b). As in humans and cattle, pigs possess a single kappa constant region exon (IGKC) downstream from five J gene segments (IGKJ); however, IGKJ2 accounts for the vast majority of IGKJ usage, while IGKJ3 and IGKJ5 are unused (Butler et al. 2005; Chapter 4). Based on transcriptional analysis, there are two expressed kappa chain V (IGKV) gene segment families, IGKV1 and IGKV2, belonging to IGKV clan I and II, respectively (Schwartz et al. 2012b; Butler et al. 2005; Chapter 4). Of these, at least in pre-immune piglets, the latter appears to be the most abundantly expressed family, though this trend may flip in older animals (Butler et al. 2005; Chapter 4). Fourteen IGKV gene segments have been described in their genomic context (Schwartz et al.

2012b). Of these, nine are putatively functional (four IGKV1 and five IGKV2 family members). All five non-functional gene segments are found at the constant region-proximal end of the IGKV locus. The first four of these belong to the IGKV3, IGKV5, and IGKV7 families of the mammalian IGKV clan III and may represent ancestral gene segments undergoing deletion. Additional gene segments upstream of IGKV1-14 likely exist, however, as the flanking genes on SSC3 were unidentified and swine are estimated to possess upwards of 60 IGKV gene segments in total (Schwartz et al. 2012b; Butler and Wertz 2012). Two non-functional orphan gene segments (IGKV2/OR3-1 and IGKV3/OR3-2) were also identified on two separate BACs. As these orphans are also found on SSC3 of the swine genome, they may represent upstream kappa locus sequence, however, this seems unlikely as no additional IGKV gene segments were identified on their respective BACs (each approximately 150 kb in size) (Schwartz et al. 2012b).

Swine possess three lambda constant region exons (IGLC) and four J (IGLJ) gene segments arranged in tandem repeats (i.e. J-C-J-C-J-C-J), similar to other artiodactyls as well as humans and different from the restricted V-J-C repeats found in the murine lambda locus (Butler et al. 2006b; Schwartz et al. 2012a; Sanchez et al. 1991). IGLJ4 is unique in that it lacks a corresponding downstream IGLC exon, although is otherwise apparently functional. IGLJ1 possesses a mutated heptamer and mutated W/F-G-X-G motif, rendering it non-functional. Thus, pigs possess only two functional IGLJ-IGLC cassettes (IGLJ2-IGLC2 and IGLJ3-IGLC3) which are nearly identical in sequence. Unlike the heavy chain locus and kappa locus, the porcine lambda locus is completely sequenced, based on the identification of flanking genes syntenic with cattle, yet distinct

from humans. The lambda locus contains 23 IGLV gene segments, of which 10 to 11 are functional, depending on haplotype (Schwartz et al. 2012a; Chapter 4). Of these, four to five belong to the IGLV3 family, five belong to the IGLV8 family, and one belongs to the IGLV5 family. An additional gene segment belonging to the IGLV2 family is putatively functional based on IMGT nomenclature; however, this gene segment possesses both a non-canonical octamer and heptamer and does not appear to be expressed as deep sequencing failed to identify it in mRNA form in any of five pigs (Schwartz et al. 2012a; Chapter 4). Additionally, expression of IGLV3 and IGLV8 dominate the expressed repertoire, such that IGLV5 expression is difficult to detect without deep sequencing (Chapter 4). There also appear to be age-specific differences in IGLV gene segment usage. One previous analysis found a higher usage of the constant region-proximal IGLV3 gene segments early in life (Butler et al. 2006b); while in another study, IGLV8 gene segments were more heavily utilized in older pigs (Schwartz et al. 2012a; Chapter 4). A more recent investigation, however, has found that in yolk sac and fetal liver, only IGLV8 gene segments are used, one of which, IGLV8-10, accounts for >50 percent of all lambda rearrangements at that early time point (Wertz et al. 2013).

Cattle, unlike pigs and humans, utilize the lambda locus in >90% of expressed antibodies (Butler et al. 2006a). As one may predict from this, the lambda locus in cattle is less compact, albeit only slightly, and is very similar in organization to swine (Pasman et al. 2010). Cattle possess four IGLJ-IGLC cassettes, and only use either IGLJ2-IGLC2 or IGLJ3-IGLC3, though the latter dominates expression. As in pigs, cattle IGLV gene segments are arranged in distinct clusters. The first two in both species are separated by

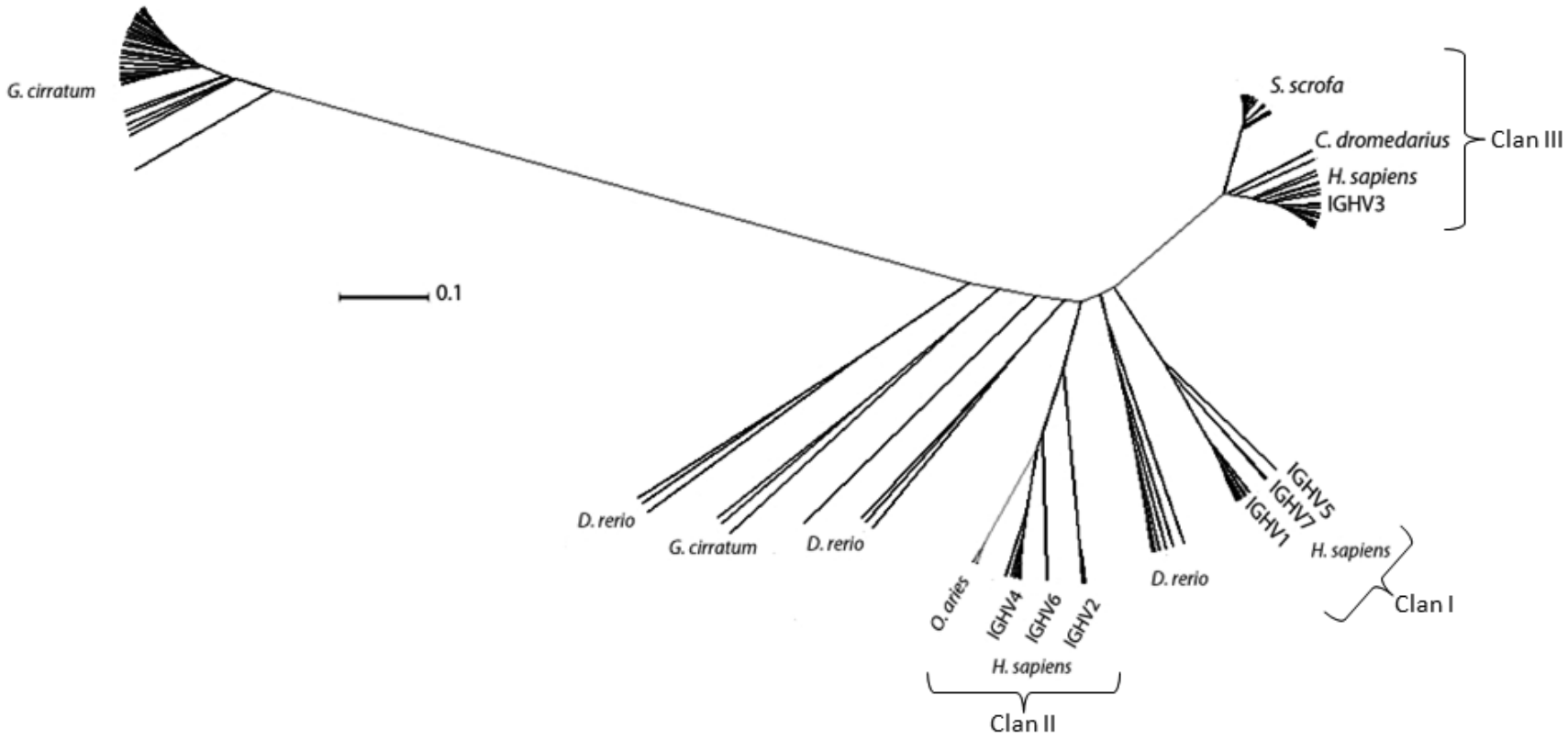
the non-Ig genes *ZNF280b* and *prame*. The first cluster, as in pigs contains IGLV3 and IGLV2 family members, which are rarely expressed in cattle. The cattle locus contains at least two additional clusters comprised exclusively of IGLV1 family members, which dominate the expressed repertoire (Pasman et al. 2010). Although, this more recent characterization did not identify other IGLV families, a previous analysis of the bovine genome revealed the existence of IGLV5 (i.e. bovine subgroup 7), IGLV8, and IGLV9 family members, though their locations in the genome were not conclusively determined (Ekman et al. 2009). In contrast, the bovine kappa locus is less well characterized, though it is known to contain at least 24 IGKV gene segments, of which eight are putatively functional (Ekman et al. 2009). Additionally, there are only five known unique CDR1/CDR2 pairs, making the bovine kappa locus highly restricted.

In contrast to the artiodactyls, humans possess a diverse germline repertoire of 123 IGHV gene segments from seven families, and of these, 40 gene segments are functional (Matsuda et al. 1998). Humans also possess 25 functional IGHD and 6 IGHJ gene segments, which together allow for greater combinatorial diversity than can be found in the comparatively limited germline repertoire of swine. Matsuda et al. (1998) estimate that the human IGHV locus has undergone DNA duplication events at least nine times in the last 133 million years, and seven of these have occurred since humans and mice diverged approximately 75 million years ago.

The human light chain loci are likewise more extensive. The human lambda locus contains seven J-C repeats and approximately 70 IGLV gene segments, about half of which are functional (Kawasaki et al. 1995). A comparative look at the organization of

the various gene segment families between pigs and humans suggest that significant gene segment expansions and deletions have occurred during the divergent evolution of these two species (Schwartz et al. 2012b; Kawasaki et al. 1995). Regarding the human kappa locus: 132 IGKV gene segments have been identified, of which 46 are apparently functional (Kawasaki et al. 2001). Included in the human kappa locus is an inverted repeat spanning 360 kb and containing 118 IGKV gene segments that appears to have originated around 5 million years ago.

FIGURE 1.3 Phylogenetic tree (unrooted) of concatenated heavy chain framework regions (FR1, FR2, and FR3) from select species. All known IGHV gene segments from five species, obtained from www.imgt.org, minus the CDRs, were used in the analysis. All 15 of the characterized porcine IGHV gene segments (Eguchi-Ogawa et al. 2010) and the 40 most abundantly expressed porcine heavy chain sequences (generated with data from Chapter 5) were used to represent the porcine clade. The major gene families for *Homo sapiens* and the mammalian clans I-III are also labeled. Two additional artiodactyls (*Ovis aries* and *Camelus dromedarius* (sheep and camel, respectively)) and two distantly related species (*Danio rerio* and *Ginglymostoma cirratum* (zebrafish and nurse shark, respectively)) are included for comparison. The apparent lack of heavy chain framework diversity among the artiodactyls is clearly demonstrated. Cattle (not included) are also restricted to IGHV1-like family members. The diversity of CDR1 and CDR2 (not shown) is typically reduced among family members. Thus, artiodactyl heavy chain diversity is highly dependent upon diversity generated during VDJ recombination (i.e. CDR3). Analysis was conducted using UPGMA with 1000 bootstrap iterations. The scale bar represents the number of substitutions per site.


G. cirratum
0.1
S. scrofa
C. dromedarius
H. sapiens
IGHV3
Clan III
D. rerio
G. cirratum
D. rerio
O. aries
IGHV4
IGHV6
IGHV2
H. sapiens
Clan II
D. rerio
IGHV5
IGHV7
IGHV1
H. sapiens
Clan I

1.7 - Mechanisms of germline evolution and somatic diversification

In order to mount a successful adaptive immune response against a nearly infinite number of potential antigenic structures, vertebrates have evolved a multitude of mechanisms to ensure that their antibody repertoires are suitably diverse. Antibody gene segments are believed to evolve primarily through segmental duplication and subsequent divergence. This is evidenced by the identification of evolutionarily recent V gene segment duplications. Notable characterizations include the previously mentioned 33 kb duplication in the pig heavy chain locus containing 6 IGHV gene segments (Eguchi-Ogawa et al. 2010), duplicated tandem repeats of IGHV genes in the human (Sasso et al. 1992), and a 360 kb inverted repeat containing 118 IGKV gene segments also in the human (Kawasaki et al. 2001).

1.7.1 - Gene conversion

One consequence of having a large number of highly similar segmentally-duplicated elements in the antibody loci is that gene segments are capable of templating off of each other, thus allowing for germline (meiotic) non-crossover homologous recombination (i.e. germline gene conversion). Evidence for evolution through germline gene conversion has been observed among porcine IGKV (Schwartz et al. 2012b), human IGKV (Bentley and Rabbitts 1983), murine IGHV (Cohen and Givol 1983), and human IGHC (Flanagan et al. 1984; Huck et al. 1989; Lefranc et al. 1986). Investigation of pig heavy chain transcripts has also revealed that many IGHV gene segments share either CDR1 or CDR2 with the non-shared CDR being different in sequence (Butler et al. 2006c), suggesting that swine may also rely on this mechanism for generating heavy

chain diversity.

Somatic gene conversion is believed to be the main mechanism for generating antibody diversity in developing B-cells in the rabbit (Becker and Knight 1990) and chicken (Reynaud et al. 1985,1989; Arakawa et al. 1996). In these species, a limited functional V gene segment repertoire is compensated for via a diverse array of pseudogenic V gene segments as templates for somatic gene conversion. In swine, the apparent lack of diversity among heavy chain V, D, and J gene segments also initially suggested that this species may similarly diversify their heavy chain repertoire (Butler 1997). However, scant evidence for this exists in pigs, and the multitude of seemingly chimeric PCR products from pig heavy chain cDNA are explained as phenomena resulting from gene duplication and recombination at the germline level (Butler et al. 2006b,c; Eguchi-Ogawa et al. 2010).

In B-cells, somatic gene conversion, somatic hypermutation (SHM), and heavy chain class switch recombination (CSR) all rely on the activation-induced cytidine deaminase (AID) enzyme to diversify the antibody repertoire (Arakawa et al. 2002; Muramatsu et al. 2000). AID is specifically expressed in activated B-cells, is active on single-stranded (ss)DNA (such as within a transcription bubble), and preferentially deaminates the dC to dU in W-R-C-Y motifs giving the enzyme a certain degree of target sequence specificity (Bransteitter et al. 2004; Chadhuri et al. 2003; Pham et al. 2003; Muramatsu et al. 1999). Of the three mechanisms mentioned above, gene conversion and SHM are in competition, the outcome of which is dependent upon additional factors (Arakawa et al. 2004).

Gene conversion is favored in the presence of recombinases such as RAD51 and RAD54 as well as a suitably diverse array of DNA templates (e.g. pseudogenes) (Arakawa et al. 2004); whereas, in the absence of these factors, SHM is favored. In chickens, gene conversion is utilized to diversify the otherwise limited immune repertoire during primary B-cell lymphogenesis in the bursa and then later diversify their repertoires via SHM in the spleen (Reynaud et al. 1987; Arakawa et al. 1996). In gene conversion, the lesion generated as a result of AID is repaired via homologous recombination mediated by RAD51 and strand invasion onto a suitable homologous template, such as another V gene segment or pseudogene. The invading strand is then replicated and subsequently repaired in a RAD54-dependent manner. Following replication, a conversion tract is generated containing sequence identical to the template and distinct from the original sequence (Arakawa et al. 2004).

1.7.2 - Somatic hypermutation

During a primary immune response, activated B-cells undergo both affinity maturation and class switch recombination. In the former, antigen-stimulated B-cells clonally expand and further diversify so that they may gain increased affinity for antigen. This process of affinity maturation specifically involves the somatic hypermutation (SHM) of the antibody variable regions in an AID-dependent manner and greatly expands the repertoire of the adaptive immune system in response to infection.

Given the absence of the aforementioned recombinases and target templates necessary for gene conversion, under SHM, there are three potential mechanisms which can occur following dC to dU deamination (Di Noia and Neuberger 2007). First,

replication may proceed past the mutation, recognizing the uracil as thymine and resulting in dC to dT or dG to dA transition mutations. Second, recognition of dU may recruit uracil-DNA glycosylase (UNG) and generate an abasic site which may then either proceed through error prone replication resulting in either transition or transversion mutations (Arakawa et al. 2004). Third, the dU:dG mismatch may be recognized by the heterodimer Msh2/Msh6, resulting in the error-prone replication of several nucleotides around the lesion and allowing for mutations at nearby dA:dT pairs (Rada et al. 1998).

1.7.3 - Class switch recombination

Naïve B-cells and those participating early in a primary immune response express IgM and IgD. Upon activation by an antigen-stimulated $CD4^+$ T-cell via CD40/CD40L signaling, the expression of AID is specifically induced in the interacting B-cell, initiating SHM and CSR (Muramatsu et al. 2000). However, these two processes need not necessarily occur simultaneously, as hypermutated IgM^+ B-cells are observable, as are non-mutated IgG^+ B-cells (Durandy 2003). With the exception of IgD, switch regions exist upstream from each of the constant region exons. The switch region upstream from IgM (Sμ), for example, in swine shares high similarity with both humans and sheep, is 3.2 kb in length and contains a large number of pentameric and decameric repeats (e.g. GAGCT, GGGCT, GGGTT, GAGCTGAGCT, and GAGCTGGGCT) making the switch region particularly rich in AID-specific motifs (WRCY) (Sun and Butler 1997; Eguchi-Ogawa et al. 2012; Muramatsu et al. 1999). As AID activity is transcription-dependent, the transcription of individual switch regions require certain cytokines to induce isotype-specific class switching; for example, interleukin-4 signaling is a prerequisite for class

switching to IgG1 in mice (Goodman et al 1993). Thus, class switching is tightly regulated and dependent upon the local cytokine milieu.

Once the switch regions are exposed as ssDNA, AID is able to deaminate dC to dU as in previous mechanisms. This dU is then recognized by UNG which generates an abasic site that is nicked by apurinic/apyrimidinic (AP) endonuclease. Closely spaced nicks on opposite strands in both 5' and 3' switch regions (for example, Sμ and the IgA switch region) can generate double-stranded breaks in the DNA (Rada et al. 2002). The intervening DNA sequence between both switch regions is then ligated to form a switch circle and deleted from the genome. The NHEJ repair pathway then mediates the re-assembly of the genomic DNA in a Ku80-dependent manner, as in V(D)J recombination (Casellas et al. 1998).

1.7.4 - Light chain isotype usage

Along with humans and unlike many species, pigs use both light chain isotypes in approximately equal ratios (Arun et al. 1996; Butler et al. 2005; Hood et al. 1967; Sinkora et al. 2001). However, the mechanisms which determine this stoichiometry have yet to be fully elucidated. During B-cell lymphogenesis, the light chain is typically rearranged in a kappa-before-lambda fashion. This is at least true for both humans and mice (Hieter et al. 1981; Moore et al. 1985). However, a recent report suggests that the opposite occurs in swine as lambda rearrangements were detectable by PCR at least 30 days prior to kappa rearrangements in the fetal pig (Sun et al. 2012). Direct sequencing of the lambda locus in kappa producing B-cells should be carried out in order to determine if lambda rearrangement does in fact occur prior to kappa rearrangement in swine and

other species. If true, swine would be the first species to be recognized to rearrange lambda before kappa. This would be predicted to allow for a shift in favor of lambda usage in species that rearrange lambda first, as there would potentially be more opportunity for them to develop functional lambda rearrangements. Interestingly, in this regard, both cattle and sheep almost exclusively prefer lambda over kappa, while mice almost exclusively prefer kappa (Arun et al. 1996; Hood et al. 1967).

The molecular mechanisms for controlling locus rearrangement order (i.e. kappa before lambda or vice versa) are poorly understood. In humans and mice, a recombining element (RE) consisting of a conserved heptamer in the IGKJ-IGKC intron and a downstream kappa deleting element containing a typical RS sequence with a 23-base spacer are believed to be involved in the ablation of the kappa locus prior to lambda rearrangements, as these sites undergo recombination in a RAG-dependent manner (Siminovitch et al. 1985). However, these sites are conserved in cattle and pigs, suggesting they are functionally constrained (Das et al. 2009; Schwartz et al. 2012b). A similar control mechanism has not been identified which would mediate the ablation of the lambda locus, although that does not preclude the existence of one.

Locus complexity, such as the number of gene segments and their diversity, almost certainly plays a role in light chain isotype usage as well, however. For instance, mice have a severely restricted lambda repertoire compared to humans leaving them with little diversity with which to generate functional immune responses compared to kappa. Likewise, as previously described, the cattle kappa locus is severely restricted compared to the lambda locus, while the porcine light chain loci are comparable to each other in

their germline diversity. Rabbits, as previously described are different in this regard, relying instead on a vast array of diverse pseudogenes with which they can use as templates for somatic gene conversion.

1.8 - Diversity of the antibody repertoire

A combination of the previously described evolutionary mechanisms has resulted in extensive individual variation in the immunogenetic loci. Recent insights into this diversity has been gleaned from the development and use of high-throughput second generation sequencing platforms such as Roche 454 pyrosequencing and Illumina HighSeq and MiSeq which can cost-effectively generate hundreds of thousands to tens of millions of reads per sequencing run (Fischer 2011). For instance, a recent cDNA amplicon-based 454 characterization of 10 aboriginal humans from Papua New Guinea revealed a new IGHV gene segment and 16 new IGHV alleles, including a functional allele of a previously described IGHV pseudogene (Wang et al. 2011). In another investigation of heavy chain variation among 12 clinical human specimens, extensive heterozygosity, and instances of V gene segment copy number variation and D gene segment deletions were inferred from high-throughput sequencing data (Boyd et al. 2010). Additionally, the latter resulted in a significant perturbation of the expressed repertoire in terms of individual gene segment usage. Other recent high-throughput investigations of the human antibody repertoire have revealed insights into allelic variation, gene segment usage, junctional diversity, CDR3 length variability, and overall repertoire diversity and richness (Prabakaran et al. 2011, Glanville et al. 2009).

Characterization of the antibody repertoire among zebrafish (*Danio rerio*) also revealed extensive differences between individuals in the expressed heavy chain repertoire and revealed that a small number of antibody sequences comprise a relatively large fraction of the expressed VDJ repertoire (Weinstein et al. 2009).

The high degree of individual variation that is known to exist among humans led to the proposed use of immunogenetic variation as a means of studying human evolution (Sanchez-Mazas et al. 2011). Quantifying immunogenetic variation can lead to insights into disease resistance and susceptibility differences between individuals. An immediate practical application of such high-throughput characterizations was demonstrated in a recent study where the antibody repertoires of mice immunized with three different antigens were compared to mice treated with adjuvant only (Reddy et al. 2010). Several highly abundant heavy chain CDR3 sequences were identified in the immunized mice that accounted for 1-10% of all heavy chain CDR3 sequences. By pairing and synthesizing the most abundant heavy and light chains in the immunized animals, the researchers obtained 21/26 antigen-specific clones from six mice immunized with three different protein antigens (Reddy et al. 2010). Thus, such an approach provides a more efficient method of selecting antigen-specific monoclonal antibodies than current hybridoma and ELISA-based screening methods.

Unfortunately, identifying alleles is complicated by the fact that true individual germline variation may be masked as a result of SHM. Indeed, Wang et al. (2008) suggest that nearly half of the described 226 human IGHV alleles are misclassified as a result of SHM. This underscores the need to sequence at the germline level in order to

understand true genetic variation. However, due to the existence of highly repetitive regions in the antibody loci, assembly is complicated and laborious and the antibody loci for any species have essentially never been sequenced more than once. Therefore, much of our understanding of allelic variation is derived from the analysis of cDNA sequence data.

Porcine heavy chain CDR3 is estimated to account for >95 percent of IGH diversity (Butler et al. 2000). Indeed, as swine possess only two functional IGHD and one functional IGHJ gene segment (Eguchi-Ogawa et al. 2010), somatically generated junctional diversity generated during V(D)J recombination must play a major role in determining the total diversity of the antibody repertoire. Indeed, the CDR3 region of porcine heavy chain transcripts exhibits a tremendous amount of sequence variation compared to CDR1 and CDR2 and especially the framework regions (Figure 1.4A). In contrast, light chain CDR3 is largely invariant in length and lacks diversity (Figure 1.4B-E; Chapter 4; Butler et al. 2013; Wertz et al. 2013). Light chain CDR1 and CDR2 also lack diversity, likely as a consequence of within-family gene segment expansion, particularly among the more recently expanded families, such as IGLV8 (Figure1.4D). Somatically-generated diversity as well as allelic variation between five pigs among approximately 100 transcripts containing IGLV3-3 is shown in Figure 1.4E. Almost all of the diversity among these transcripts is found in CDR3, suggesting the importance of either CDR3-targeted SHM, allelic variation, and/or junctional diversity in this region (Figure 1.4E).

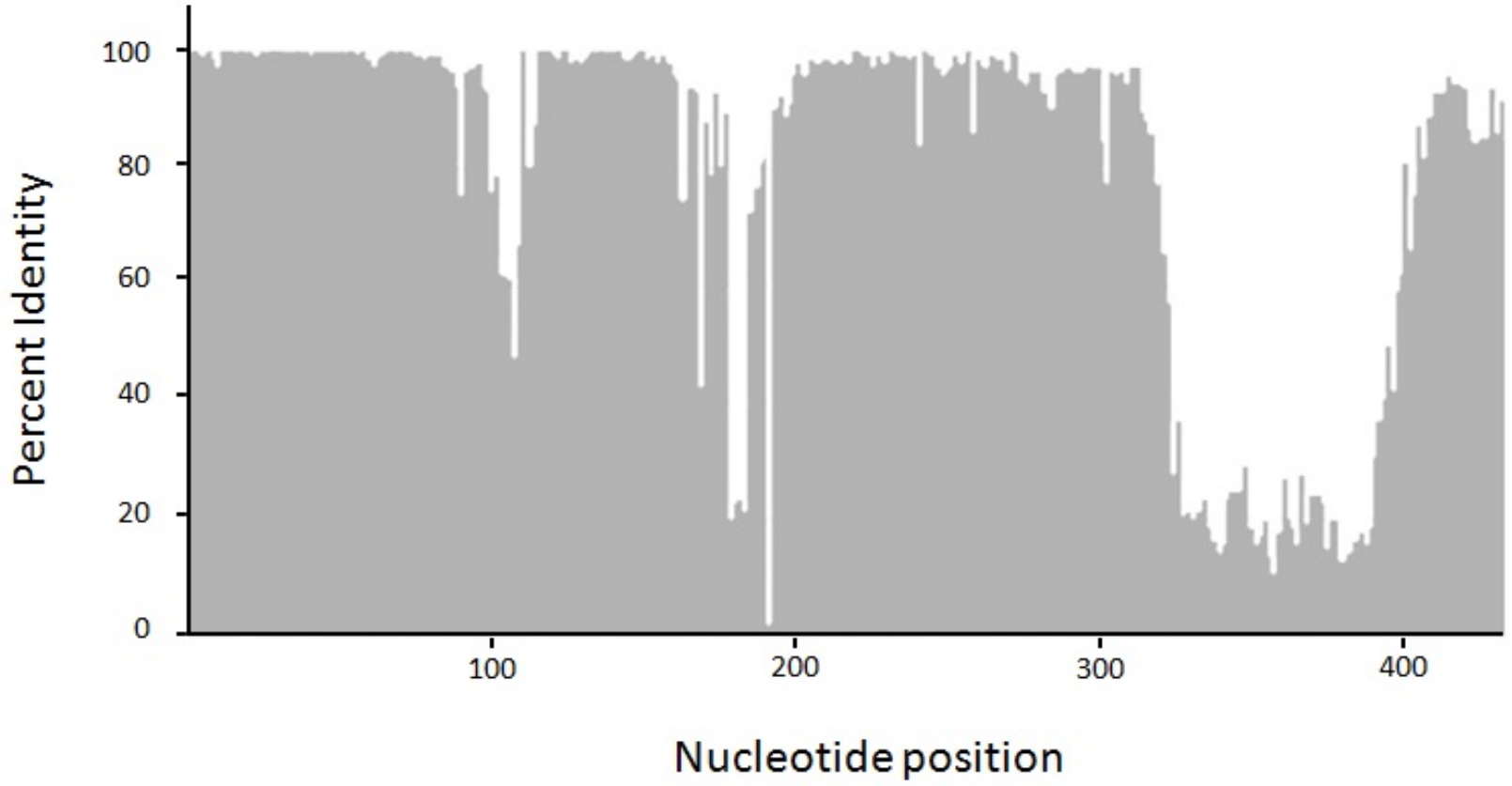
The paradigm of CDR3 diversity (i.e. junctional diversity) generation via V(D)J

recombination in pigs is the same as in other species. As such, their heavy chain CDR3 is variable in length with a minimum size of four amino acids, a maximum of around 30 amino acids, and a mode of 15 amino acids; making the porcine CDR3 repertoire slightly longer than humans on average (Glanville et al. 2009; Prabakaran et al. 2012; Chapter 5). Swine also lack the unusually long (~61 amino acid) CDR3s sometimes found in cattle as these are thought to result from the usage of a single bovine IGHD that is 148 bp in length (Shojaei et al. 2003).

Figure 1.4 Nucleotide conservation across the variable regions of porcine heavy (**A**) and light (**B**-**E**) chains. Approximately 200 heavy chain cDNA sequences were obtained from GenBank and approximately 100 random light chain cDNA sequences derived from five pigs were trimmed to include only the V(D)J region (light chain data obtained from Chapter 4). Sequences were then aligned using ClustalW (Thompson et al. 1994) and the percent conservation at each nucleotide position was determined. Light chain families were separated to highlight their similarity in the framework regions. IGLV3 was separated based on its two subgroups that possess different frameworks. The three regions of low similarity seen in all three figures correspond to the three CDRs, and the high similarity framework regions are found between them. The low similarity within CDR1 of (**D**) is the result of a three nucleotide insertion/deletion between IGLV3-1-1 and the other IGLV3 members. IGLV3-3 (**E**) is the only member of its expressed sub-family. Thus, all variation shown for IGLV3-3 is either allelic in origin or somatically-generated.

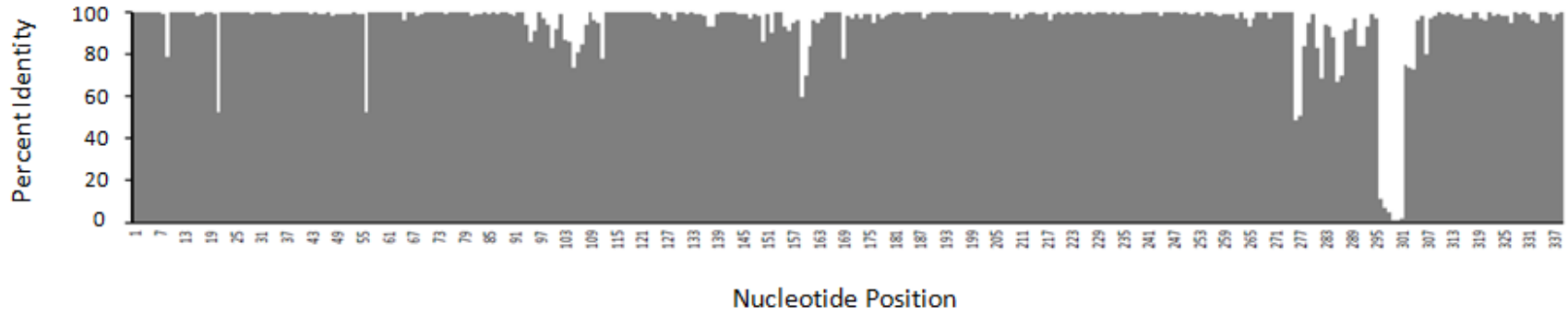
A

Heavy Chain



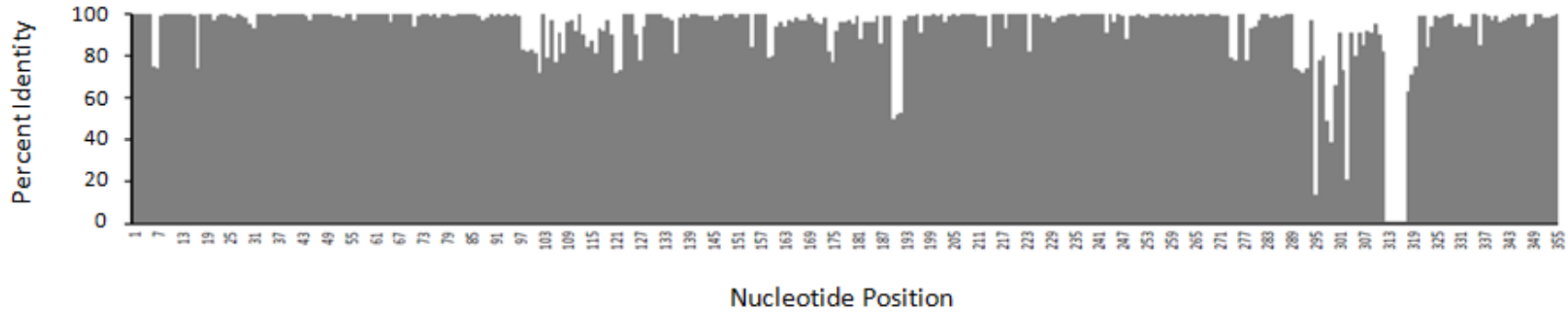
B

IGKV1


Percent Identity
100
80
60
40
20
0
1 7 13 19 25 31 37 43 49 55 61 67 73 79 85 91 97 103 109 115 121 127 133 139 145 151 157 163 169 175 181 187 193 199 205 211 217 223 229 235 241 247 253 259 265 271 277 283 289 295 301 307 313 319 325 331 337
Nucleotide Position

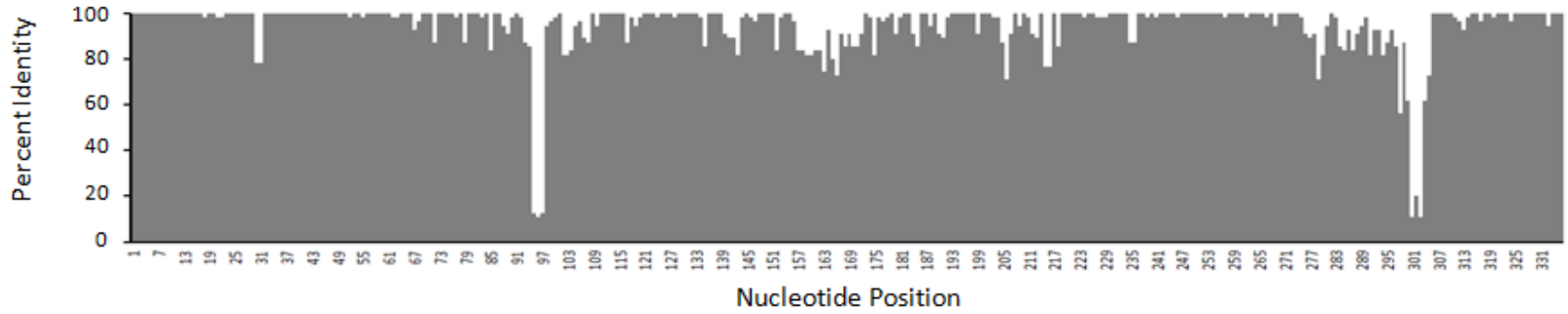
C

IGKV2


Percent Identity
100
80
60
40
20
0
1 7 13 19 25 31 37 43 49 55 61 67 73 79 85 91 97 103 109 115 121 127 133 139 145 151 157 163 169 175 181 187 193 199 205 211 217 223 229 235 241 247 253 259 265 271 277 283 289 295 301 307 313 319 325 331 337 343 349 355
Nucleotide Position

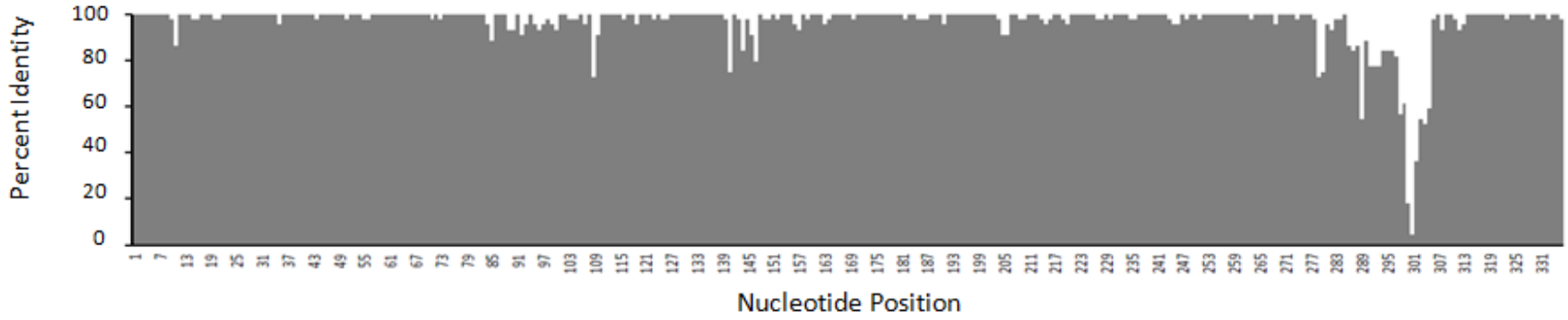
D

IGLV3-1-1, IGLV3-4, and IGLV3-5



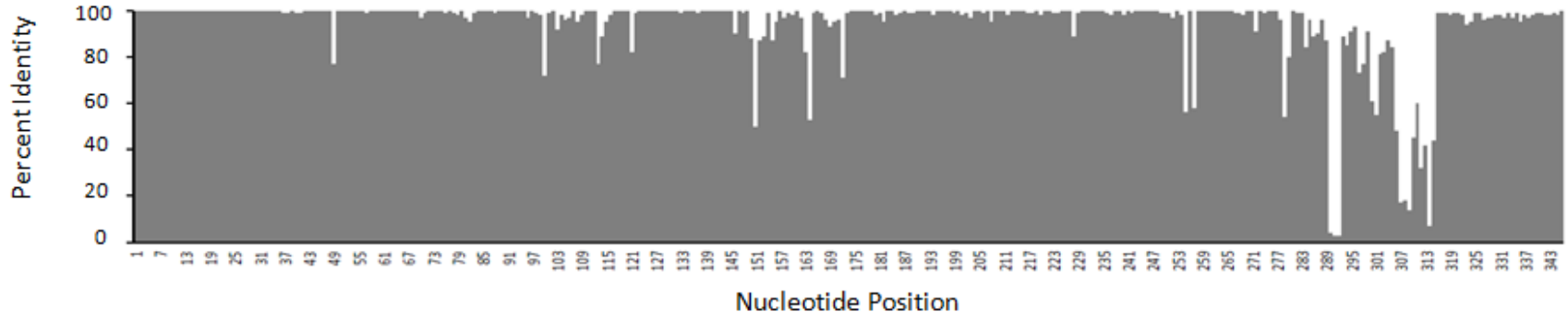
E

IGLV3-3


Percent Identity
100
80
60
40
20
0
1
7
13
19
25
31
37
43
49
55
61
67
73
79
85
91
97
103
109
115
121
127
133
139
145
151
157
163
169
175
181
187
193
199
205
211
217
223
229
235
241
247
253
259
265
271
277
283
289
295
301
307
313
319
325
331
Nucleotide Position

F

IGLV8


Percent Identity
100
80
60
40
20
0
1 7 13 19 25 31 37 43 49 55 61 67 73 79 85 91 97 103 109 115 121 127 133 139 145 151 157 163 169 175 181 187 193 199 205 211 217 223 229 235 241 247 253 259 265 271 277 283 289 295 301 307 313 319 325 331 337 343
Nucleotide Position

1.9 - Antibody responses to infection: PRRSV

Porcine reproductive and respiratory syndrome (PRRS) is a severe disease in pigs that concurrently emerged in the early 1990's in both the United States and Europe. The etiologic agent, PRRS virus (PRRSV), has since disseminated across the globe and is now the most important disease agent of swine worldwide. Epidemic and endemic disease cause widespread animal losses and reduced production that imperil food security and increase costs. In the United States alone, PRRS costs the pork industry an estimated $664 million per year (Holtkamp et al. 2011).

The immunological mechanisms that control PRRSV are unknown. In infected animals, sterilizing immunity is eventually achieved, as the virus is completely cleared after several months post infection (Molina et al. 2008). Unfortunately, current vaccines are only partially effective in the control of the disease, especially against genetically diverse strains, and new PRRSV variants continue to arise and threaten the health of swine herds (Kimman et al. 2009). Previously infected animals, however, are at least partially resistant to re-infection (Lager et al. 1997; Murtaugh et al. 2002). Thus, immunity to PRRSV must ultimately be driven by an adaptive immune response.

Cell-mediated immunity, as measured by the frequency of PRRSV-specific interferon gamma secreting cells following attenuated vaccine administration, is reportedly delayed and weak (Meier et al. 2000). Furthermore, an extensive analysis failed to reveal an association between PRRSV-specific T-cell responses and control of infection (Xiao et al. 2004), and experimental ablation of cytotoxic T-cells with anti-CD8 did not exacerbate PRRSV infection or disease (Lohse et al. 2004).

In contrast, a broad and robust B-cell response is generated to PRRSV (Molina et al. 2008; Mulupuri et al. 2008). PRRSV-specific IgM first appears 5-7 days post-inoculation and disappears after 3-4 weeks. IgG appears after 7-10 days, peaks after 2-3 weeks, then slowly declines to low levels after 300 days (Murtaugh et al. 2002). Neutralizing antibodies appear approximately 3 weeks post infection, which correlates with the reduction of viremia and eventual disappearance of PRRSV from circulation around 35-42 days post infection. However, neutralizing antibody responses vary widely between individual animals and studies report that some pigs fail to produce neutralizing titers despite resolution of viremia (Murtaugh et al. 2002; Loemba et al. 1996; Nelson et al. 1994).

Disease severity is also widely variable between individual animals on farms infected with a single PRRSV strain. Thus, individual variation in both the adaptive humoral immune response and disease severity suggests a genetic component of disease resistance. Indeed, there are substantial differences noted between pig breeds in their response to PRRSV (Lewis et al. 2007). A recent study found that viremia in Petrain pigs is prolonged (>72 days) compared with Wiesenauer miniature pigs (up to 35 days) and that viral replication in the latter breed was approximately 100-fold less than in the former (Reiner et al. 2010). This reduced viral burden was correlated with a stronger antibody response in miniature pigs. Other studies have found reduced viral transcripts in Landrace compared to four other commercial breeds (Ait-Ali et al. 2007), and more lung-lesions in Hampshire pigs compared to Meishan pigs who possessed fewer PRRSV-positive macrophages in their lungs (Halbur et al. 1998).

The characterization of the porcine antibody repertoire in PRRSV-infected animals revealed several highly abundant light and heavy chain sequences that were quite rare among naïve animals (Chapters 4 and 5). Interestingly, these highly abundant clones were largely represented by a single animal (out of two animals and three animals, respectively for heavy and light chain). It was found that this particular animal lacked the otherwise highly expressed IGLV3-1-1 gene segment and is hypothesized that this genetic abnormality resulted in a shift in gene segment usage in favor of IGLV3-3, which dominated this particular animal's light chain repertoire (Chapter 4). Given the lack of junctional diversity apparent among light chains, it is possible that these abundant light chain clones are the result of increased usage of individual gene segments resulting from such a genetic aberration and not PRRSV-specificity. Unfortunately, given the high degree of allelic variation in the antibody loci, genetic controls are nearly impossible without cloning. It is therefore, not known whether these highly abundant clones are PRRSV-specific or the result of a genetic artifact. However, heavy chain expression also revealed several highly abundant clones among this and another PRRSV-infected animal which possessed CDR3 sequences that were exceptionally rare in non-infected animals (Chapter 5). This observation makes it very likely that the most abundant light chains are paired to these individual heavy chain clones and that they are PRRSV-specific. Additional work is needed to determine whether this is the case. If true, it this would reproduce in pigs what has previously been reported in immunized mice (Reddy et al. 2010).

With the completion of the swine genome (Groenen et al. 2013) and functional

annotation of the swine immunome (Dawson et al. 2013), efforts are underway to identify markers for disease resistance in swine. Indeed, these efforts have identified a region of interest on SSC4 near a group of innate immune response genes associated with reduced viral load and increased weight gain when challenged with PRRSV (Rowland et al. 2012). By studying the highly diverse immunoglobulin repertoire, and, more generally, the immunogenetic loci, and their variation between individuals and between breeds, there is a clear opportunity for identifying genetic components of disease resistance (Hsu et al. 2006; Sanchez-Mazas et al. 2011; Middleton and Gonzalez 2009).

1.10 - Conclusions and hypotheses

Antibody responses to infection are critical for disease control and prevention. Prior to the work described in this thesis, there was little known regarding the diversity and complexity of the porcine antibody repertoire, either in its germline state, or in its expressed state. There was likewise no information concerning individual allelic variation in the porcine antibody loci. Despite studies that suggest genetic involvement in the disease resistance to the major swine pathogen, PRRSV, it is not known what genetic loci are involved. In order to facilitate this and future endeavors, our characterization of the porcine antibody repertoire is timely.

This study was designed to obtain a first glimpse of the dynamics of the antibody repertoire as a result of infection with PRRSV. The lack of apparent genomic diversity as indicated by the restricted use of a few IGHD and IGHJ gene segments and a relatively small number of V gene segment families, suggests that this repertoire is either reduced

in size compared to other species, that diversity is driven by germline and somatic mutations in the CDRs, and/or diversity is driven through extensive junctional alterations during V(D)J recombination. It is hypothesized that animals undergoing an immune response to PRRSV will have generated highly abundant clonal populations of B-cells, which will be detectable against the background of the antibody repertoire. Further, the expressed antibody repertoire, with respect to gene segment usage, will differ between individuals due to variation in their germline sequence.

Pigs, like humans, utilize the two light chains in nearly equal amounts. In humans, the kappa and lambda loci appear to be similarly diverse. In species which are restricted to one light chain or the other, the lesser used locus is substantially reduced in size and complexity. It is hypothesized that lambda and kappa usage in pigs is associated with the total diversity and richness of the two light chain repertoires.

Using our newly gleaned information, it should be possible to generate recombinant single chain fragment variable (scFv) constructs using the most abundant heavy and light chain sequences obtained from PRRSV infected pigs. It is hypothesized that these scFvs will be PRRSV-specific. If successful, this would represent a novel means of generating antigen-specific antibodies from immunized or infected swine. Through characterization of the light chain loci, and through the pre-existing information on the heavy chain locus, it will also be possible to develop an antibody display library, expressed in yeast, that will form the basis for future investigations of the porcine antibody response to PRRSV. Further adaptations of this technology will also enable

future swine immunologists to develop recombinant antibodies specific for any conceivable antigen of interest.

CHAPTER II

ORGANIZATION, COMPLEXITY, AND DIVERSITY OF THE PORCINE IMMUNOGLOBULIN KAPPA LOCUS

Schwartz JC, Lefranc M-P, and Murtaugh MP (2012) Evolution of the porcine (*Sus scrofa domesticus*) immunoglobulin kappa locus through germline gene conversion. *Immunogenetics*. 64(4):303-11. doi:10.1007/s00251-011-0589-6

Immunoglobulin (IG) gene rearrangement and expression are central to disease resistance and health maintenance in animals. The IG kappa (IGK) locus in swine (*Sus scrofa domestica*) contributes to approximately half of all antibody molecules, in contrast to many other Cetartiodactyla, whose members provide the majority of human dietary protein, and in which kappa locus utilization is limited. The porcine kappa locus variable genes are 27.9 kb upstream of five IG kappa J genes (IGKJ) which are separated from a single constant gene (IGKC) by 2.8 kb. Fourteen variable genes (IGKV) were identified, of which nine are functional and two are open reading frame (ORF). Of the three pseudogenes, IGKV3-1 contains a frameshift and multiple stop codons, IGKV7-2 contains multiple stop codons, and IGKV2-5 is missing exon 2. The nine functional IGKV genes are phylogenetically related to either the human IGKV1 and IGKV2 subgroups. IGKV2 subgroup genes were found to be dominantly expressed. Polymorphisms were identified on overlapping BACs derived from the same individual such that 11 genes contain amino acid differences. The most striking allelic differences are present in IGKV2 genes, which contain as many as 16 amino acid changes between alleles, the majority of which are in complementarity determining region (CDR) 1. In addition, many IGKV2 CDR1 are shared between genes but not between alleles, suggesting extensive diversification of this locus through gene conversion.

INTRODUCTION

The adaptive humoral immune response must be able to generate a nearly infinite arsenal of antibodies to provide protection against an incredibly diverse array of pathogens and toxins. In pigs, the three immunoglobulin (IG) loci comprise the IG heavy (IGH) locus on chromosome 7q25-q26, the IG lambda (IGL) locus on 14q16-q21, and the IG kappa (IGK) locus on 3q12-q14 (Yerle et al. 1997). The variable domain of each IG chain is encoded by a variable (V) gene, a diversity (D) gene (heavy chain-only) and a joining (J) gene which rearrange through recognition of recombination signal (RS) sequences by the RAG1 and RAG2 complex and subsequent double strand break repair (McBlane et al. 1995; Kim and Oettinger 2000). The highly variable complementarity determining regions (CDR) 1 and 2 are encoded by the V region and differ in sequence between genes. CDR3 is the result of junctional diversity between genes which rearrange and is largely generated from exonuclease trimming of the gene ends and random nucleotide addition via terminal deoxynucleotidyl transferase (TdT) during gene rearrangement early in B cell development. Thus, the combined three CDR of the V heavy domains (VH) and three CDR of the V light domains (VL) provide antibodies with an immensely diverse antigen-binding repertoire (Wu and Kabat 1970; Lefranc and Lefranc 2001).

The use of either kappa or lambda light chain is first determined by the somatic rearrangement of the kappa locus. However, if both homologous chromosomes fail to produce functional antibody, they are ablated through recombination between a 3' kappa deleting element (KDE) and a recombining element in the J-C intron. The lambda locus

is then able to undergo rearrangement until the B cell either produces a functional light chain or is deleted (Siminovitch et al. 1985). While most members of Cetartiodactyla rely heavily on the lambda locus, pigs are unusual in that their expressed repertoires are nearly 1:1 kappa:lambda (Butler et al. 2005). This variation between species might be due to 1) the ability of either locus to produce functional antibody (i.e. locus complexity), or 2) regulation of kappa locus ablation. However, while the kappa joining (IGKJ) and constant (IGKC) genes are previously described in pigs, the variable (IGKV) gene organization and complexity are unknown (Butler et al. 2006b). Here, we interrogated available porcine genetic and genomic sequence data (Schook et al. 2005; Humphray et al. 2007; Archibald et al. 2010) to more fully characterize the organization, complexity, and expression of the IGK locus. Our findings suggest that the locus is organized similarly to, but is more complex than, other mammalian species that rely primarily on the lambda locus for light chain expression. Interestingly, the most highly expressed IGKV subgroup (IGKV2) is highly polymorphic between alleles from the same animal, with diversity due largely to variation within CDR1. However, while many CDR1 differed between alleles, they were shared between genes, a phenomenon that is best explained by evolution through homologous recombination between genes (i.e. gene conversion). This phenomenon was not, however, observed among members of the similarly sized population of IGKV1 genes which, by comparison, are poorly expressed.

MATERIALS AND METHODS

Identification and sequencing of the bacterial artificial chromosomes

Sus scrofa genome build 9 was queried to identify the bacterial artificial chromosomes (BACs) containing IGKV sequences using the Basic Local Alignment Search Tool (BLAST) within the Ensembl database (Altschul et al. 1990; Hubbard et al. 2002). Nucleotide sequences for each BAC containing the region of interest were downloaded for further analysis from GenBank. The CHORI (Children's Hospital Oakland Research Institute)-242 BAC library used was derived from a single Duroc sow (http://bacpac.chori.org/porcine242.htm). Two BAC clones (CH242-221I5 and CH242-227G10) were acquired, expanded overnight, and BAC DNA was purified using the Qiagen Plasmid Midi Prep with Qiagen-tip 500 columns. Purified DNA was submitted to the University of Minnesota Biomedical Genomics Center for library preparation and paired-end sequencing using the Illumina GAIIx platform.

Characterization of the porcine IGK locus

Approximately 20 million high quality reads were sorted by molecular tag to differentiate samples and assembled using a combination of ABySS and Velvet (Simpson et al. 2009; Zerbino and Birney 2008). Generated contigs were assembled against the existing BAC sequences using Sequencher 4.10.1 (Gene Codes Corporation). The resulting complete BAC sequences were manually annotated and interrogated for immunoglobulin features such as RS (i.e. heptamers and nonamers), promoters (i.e.

octamers), and gene structure using the annotation software Artemis (Rutherford et al. 2000).

The sequences of CH242-221I5, CH242-227G10, CH242-148A13 were acquired from GenBank (accession numbers: CU694848, FP312898, and CU928807, respectively) and assessed for IGKV, IGKJ, and IGKC genes using BLAST. Phylogenetic analyses were performed in CLC Sequence Viewer (CLC Bio) and Dendroscope (Huson et al. 2007) using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) with 1000 bootstrap iterations. Genes were annotated according to IMGT®, the international ImMunoGeneTics information system® (Lefranc et al. 2009). Translated amino acid sequences of the IGKV genes were compared and CDR and framework (FR) boundaries were annotated according to the IMGT unique numbering for V domain (Lefranc et al. 2003). IGKC gene translation was annotated according to the IMGT unique numbering for C domain (Lefranc et al. 2005). Expression of germline IGKV genes was compared using 41 BLAST hits obtained from 398,837 porcine expressed sequence tags (ESTs) in GenBank and deposited at: http://pigest.ku.dk/index.html (Gorodkin et al. 2007) using an E-value threshold of 10^{-12}.

Nomenclature

The porcine IGKV genes are named according to IMGT nomenclature (Lefranc and Lefranc 2001, Lefranc 2007, Lefranc 2008), using human V gene subgroup nomenclature to maintain consistency with porcine heavy chain, light chain, and cattle lambda light chain nomenclature (Eguchi-Ogawa et al. 2010; Butler et al. 2005; Butler et al. 2006b; Pasman et al. 2010). The porcine IGKV, IGKJ and IGKC genes were

submitted to IMGT/GENE-DB (Giudicelli et al. 2005). IgBLAST was used to organize IGKV genes into subgroups using a 75 percent identity threshold. IGKV genes are described based on IMGT nomenclature rules. Briefly, genes were deemed pseudogenes if they contained a truncation, stop codon, frameshift, or a defective initiation codon. Additionally, IGKV genes were described as ORF (open reading frame) if they were missing one (or more) key amino acids (1st-CYS 23, CONSERVED-TRP 41, hydrophobic 89, 2nd-CYS 104). RS were deemed non-canonical if the heptamer was anything other than “CACAGTG” ” for V-HEPTAMER (or “CACTGTG” for J-HEPTAMER) or if the nonamer contained at least two nucleotides which were each present in less than 10 percent of all RS described by Ramsden et al. (1994).

RESULTS

Organization of the porcine immunoglobulin kappa locus

The porcine IGK locus contains a single constant gene (IGKC) 2.8 kb downstream from five IGKJ genes. The IGKV locus begins 27.9 kb upstream from IGKJ1. We identified 14 distinct IGKV genes spanning approximately 89 kb (Figure 2.1). Two additional IGKV genes were identified on the BAC CH242-148A13 and are most likely orphons as the BAC maps to a different region of chromosome 3. BLAST failed to identify additional IGKV genes in the porcine genome or on other shotgun sequenced CH242 BACs (not shown). Three of the 14 identified IGKV genes within the locus are pseudogenes; IGKV3-1 contains a frameshift and multiple stop codons, IGKV7-2 contains multiple stop codons, and IGKV2-5 is missing the V-EXON (Table 2.1).

The kappa deleting element (KDE) was identified approximately 23.2 kb downstream from IGKC and contains a canonical RS that is identical to the cattle KDE RS (Das et al. 2009). Likewise, the recombining element in the J-C intron is comprised of a canonical heptamer (CACAGTG). The conservation of this system between pigs and other members of Cetartiodactyla suggests that ease of kappa locus ablation does not explain preferential lambda locus usage (Das et al. 2009).

Phylogenetic analysis of IGKV genes

The first four C-proximal IGKV genes are related to the human IGKV subgroups IGKV3, IGKV7, and IGKV5. The fifth gene, IGKV2-5 is missing the V-EXON and its membership in the IGKV2 subgroup is inferred here based solely on its leader sequence. The remaining nine genes are split among the IGKV1 and IGKV2 subgroups (Figure 2.2). In general, this pattern of gene organization is mirrored in humans (Kawasaki et al. 2001; Lefranc and Lefranc 2001). Compared to humans, the swine IGKV3-1, IGKV7-2, IGKV5-4 genes are closest to the human C-proximal IGKV3-7, IGKV7-3, and IGKV5-2 genes, respectively. However, the remaining porcine genes forming the IGKV1 and IGKV2 subgroups are all phylogenetically similar to only one or two human IGKV genes (IGKV1-9, IGKV1-27, IGKV2-28, and IGKV2-40). This suggests that substantial IGKV repertoire expansion and contraction of specific IGKV subgroups has occurred during the evolutionary divergence of swine and humans.

Allelic variation and gene conversion

The sequenced BACs CH242-221I5 and CH242-227G10 overlap across much of the characterized kappa locus. Thirteen of the fourteen IGKV genes are represented on both BACs as are all IGKJ genes and IGKC (Figure 2.1). Nucleotide identity across this overlap varies from approximately 99 percent in the IGKJ-IGKC locus to 97 percent in the IGKV locus. Differences of one or more nucleotides between BACs were identified in 11 IGKV genes (Table 2.1) and amino acid changes were found between each of these alleles. A total of 62 amino acid changes were identified and ranged from 2 (IGKV7-2 and IGKV5-4) to 16 (IGKV2-8) per gene (Table 2.2). Interestingly, all intact IGKV2

subgroup genes except IGKV2-12 were highly polymorphic in CDR1 (Table 2.2). The CDR1 of IGKV2-6, IGKV2-8, IGKV2-10, and IGKV2-13 contained 4, 7, 6, and 3 amino acid changes, respectively, between alleles. Interestingly, many of these CDR1 are shared between genes, while the upstream and downstream regions remain largely identical between alleles (Table 2.2). For example, IGKV2-13 and IGKV2-6 on CH242-221I5 possess the same CDR1 sequence which is different from their alleles on CH242-227G10 (Table 2.3). This suggests that the IGKV2 repertoire has diversified CDR1 through gene duplication and gene conversion. Thus, the allelic diversity of the IGKV2 subgroup may be quite large.

In contrast to the IGKV genes, the IGKJ genes were all identical at the amino acid level between BACs (not shown). IGKC contains a single nucleotide polymorphism resulting in a Ser-122 (IGKC*01) to Asn-122 (IGKC*02) amino acid change (Table 2.4). Comparison with Minnesota Miniature swine cDNA sequence (GenBank: M59321.1) reveals the presence of 10 amino acid differences from IGKC*01 described here, indicating substantial variation between pig breeds even within the relatively conserved constant region (Lammers et al. 1989). The IGKC Cys-126 is the terminal amino acid in cattle (GenBank: BC151500.1) and other described species in the IMGT database, including humans, mice, rats, and rabbits (Emorine and Max 1983, Giudicelli et al. 2005, Hieter et al. 1980, Sheppard and Gutman 1981). Interestingly, in both the Minnesota Miniature and Duroc pig breeds IGKC contains two additional amino acids downstream from Cys-126 (Table 2.4; Lammers et al. 1989).

Expression of IGKV genes

In addition to the presence of stop codons and frameshifts, functionality is further restricted in some IGKV genes by non-canonical RS necessary for recombination and octamers necessary for transcription. Non-canonical RS were found in three of the four IGKV1 subgroup genes (IGKV1-7, IGKV1-9, and IGKV1-14) (Table 2.5). The remaining member, IGKV1-11, contained a non-canonical RS on CH242-221I5 but a canonical sequence on CH242-227G10. Additionally, non-canonical RS were also located in the first two pseudogenes (IGKV3-1 and IGKV7-2). Non-canonical octamers (i.e. other than ATTTGCAT) were present upstream of IGKV3-3 and two of the four IGKV1 subgroup genes (IGKV1-7 and IGKV1-14). Non-canonical 5' splice sites (i.e. other than GT) were also observed for several genes, most notably among IGKV1 subgroup genes (Table 2.5). Combined, these data suggest that the functional repertoire of the kappa variable region is largely restricted to members of the IGKV2 subgroup.

Indeed, BLAST analysis of a porcine EST database revealed that all five functional IGKV2 subgroup genes (IGKV2-6, IGKV2-8, IGKV2-10, IGKV2-12, and IGKV2-13) were expressed, with the most highly expressed genes being IGKV2-10 and IGKV2-13) (Figure 2.3). Relatively low level expression (a combined 19% of EST hits) of IGKV1-9 and IGKV1-11 were also observed. They are the only two IGKV1 subgroup genes possessing both a canonical RS and promoter octamer. These findings agree with an earlier report that IGKV2 represents approximately 90% of the expressed porcine pre-immune repertoire (Butler et al. 2004). The gene IGKV5-4 was not found in the database despite having an intact open reading frame and both canonical RS and octamer (Table

5). This confirms that the amino acid changes (C23>R and W41>C), that assign the functionality of that gene to ‘ORF’ instead of ‘functional’, have a detrimental effect on the variable domain structure.

Table 2.1 IGKV alleles

IMGT Subgroup	IMGT gene name	IMGT allele name	Fct	Clone names	Clone accession numbers
IGKV1	IGKV1-7	IGKV1-7*01	F	CH242-227G10	FP312898
		IGKV1-7*02	F	CH242-221I5	CU694848
	IGKV1-9	IGKV1-9*01	F	CH242-227G10	FP312898
		IGKV1-9*02	F	CH242-221I5	CU694848
	IGKV1-11	IGKV1-11*01	F	CH242-227G10	FP312898
		IGKV1-11*02	F	CH242-221I5	CU694848
	IGKV1-14	IGKV1-14*01	F	CH242-227G10	FP312898
IGKV2	IGKV2-5	IGKV2-5*01	P(1)	CH242-227G10	FP312898
				CH242-221I5	CU694848
	IGKV2-6	IGKV2-6*01	F	CH242-227G10	FP312898
		IGKV2-6*02	F	CH242-221I5	CU694848
	IGKV2-8	IGKV2-8*01	F	CH242-227G10	FP312898
		IGKV2-8*02	F	CH242-221I5	CU694848
	IGKV2-10	IGKV2-10*01	F	CH242-227G10	FP312898
		IGKV2-10*02	F	CH242-221I5	CU694848
	IGKV2-12	IGKV2-12*01	F	CH242-227G10	FP312898
		IGKV2-12*02	F	CH242-221I5	CU694848
	IGKV2-13	IGKV2-13*01	F	CH242-227G10	FP312898
		IGKV2-13*02	F	CH242-221I5	CU694848
	IGKV2/OR3-1	IGKV2/OR3-1*01	P(2)	CH242-148A13	CU928807
IGKV3	IGKV3-1	IGKV3-1*01	P(3)	CH242-227G10	FP312898
				CH242-221I5	CU694848
	IGKV3-3	IGKV3-3*01	ORF(4)	CH242-227G10	FP312898
		IGKV3-3*02	ORF(5)	CH242-221I5	CU694848
	IGKV3/OR3-2	IGKV3/OR3-2*01	P(6)	CH242-148A13	CU928807
IGKV5	IGKV5-4	IGKV5-4*01	ORF(7)	CH242-227G10	FP312898
		IGKV5-4*02	ORF(7)	CH242-221I5	CU694848
IGKV7	IGKV7-2	IGKV7-2*01	P(8)	CH242-227G10	FP312898
		IGKV7-2*02	P(8)	CH242-221I5	CU694848

Fct, functionality; F, functional; P, pseudogene; ORF, open reading frame
(1) V-EXON is missing
(2) Orphon, stop codon at position 109
(3) Frameshift in CDR1-IMGT
(4) L89>R
(5) L89>C
(6) Orphon, frameshifts in FR1-IMGT and FR2-IMGT
(7) 1stCYS C23>R, CONSERVED-TRP W41>C
(8) Stop codons at positions 44, 55, and 67

Table 2.2 Protein display of in-frame IGKV genes using IMGT unique numbering

Next Page

	FR1-IMGT (1-26)		CDR1-IMGT (27-38)	FR2-IMGT (39-55)		CDR2-IMGT (56-65)	FR3-IMGT (66-104)				CDR3-IMGT (105-117)	
	A (1-15)	B (16-26)	BC (27-38)	C (39-46)	C' (47-55)	C'C" (56-65)	C" (66-74)	D (75-84)	E (85-96)	F (97-104)	FG	
	1 15	16 26	27 38	39 46	47 55	56 65	66 74	75 84	85 96	97 104	105 111	
IGKV1-7*01	AIQLTQSPASLAASL	GDTVSITCRAH	QTI......SSY	LAWYQQQP	GKPPKLLLC	DA.......C	TLQSGVP.C	GFKGSG..SG	THFTLTISGLQA	EDVATYYC	QQLNNAP	
IGKV1-7*02	AIQLTQSPASLAASL	GDTVSITCRAR	QSI......SSY	LAWYQQQP	GKPPKLLLC	DA.......C	TLQSGVP.C	GFKGSG..SG	THFTVTISGLQA	EDVATYYC	QQLNNAP	
IGKV1-9*01	AIQLTQSPASLAASL	GDTVSITCRAS	QSV......SNN	LAWYQQQA	GKPPKLLIY	WA.......S	ALQSGVP.S	RFKGSV..SG	TDFTLTISGLQA	EDVATYYC	QQLNSAP	
IGKV1-9*02	AIQMTQSPASLAASL	GDTVSITCRAS	QSV......SNN	LAWYQQQA	GKAPKLLIY	WA.......S	TLQSGVP.S	RFKGSV..SG	TDFTLTISGLQA	EDVATYSC	QQLNSAP	
IGKV1-11*01	AIQLTQSPASLAASL	GDTVSITCRAS	QSI......NKW	LAWYQQQA	GKAPKLLIY	SA.......S	TLQSGVP.S	RFKGSG..SG	TDFTLTISGLQA	EDVATYYC	QQHHSAP	
IGKV1-11*02	AIQLTQSPASLAASL	GDTVSITCRAS	QSI......NKW	LAWYQQQP	GKAPKLLIY	SA.......S	TLQSGVP.C	GFKGSG..SG	TDFTLTISGLQA	EDVATYYC	QQHHSAP	
IGKV1-14*01	AIQLTQSPASLAASL	GDTVSITCRAR	QSI......SSY	LAWYQQQP	GKTPKLLFY	DA.......C	TLQSGVP.C	GFKGSG..SG	THFTVTISGLQA	EDVATYYC	QQLNNAP	
IGKV2-6*01	AIVLTQSPLSLSVSP	GAPASISCRSS	QSLES..YSYNF	LSWYQQKP	GQSPRLLIY	FA.......T	NKASGVP.D	RFSGSG..SG	TDFTLKISRVEA	EDAGVYYC	QQNKESL	
IGKV2-6*02	AIVLTQTPLSLSVSP	GEPASISCRSS	QSLEE..YGSNL	LSWYQQKP	GQSPQLLIY	GG.......T	NRASGVP.D	RFSGSG..SG	TDFTLKISRVEA	EDAGVYYC	QQNKESP	
IGKV2-8*01	AIVLTQTPLSLSVSP	GEPASISCRSS	QSLEI..YGSNF	LSWYQQKP	GQSPQLLIY	EA.......T	NRASGVP.D	RFSGSG..SG	TDFTLKISRVEA	EDAGVYYC	QQHKESP	
IGKV2-8*02	AIVLTQTPLSLSVSP	GEPASISCRSS	QSLVDS.DGDSL	LHWYLQKP	GQSPRLLFY	FA.......T	NRASWVP.E	RFSGSG..SG	TDFTLKISRVEA	EDAGVYYC	QQYKEFP	
IGKV2-10*01	AIVLTQTPLSLSVSP	GEPASISCRSS	QSLVDS.DGDSL	LHWYLQKP	GQSPQLLIY	EA.......T	NRASGVP.D	RFSGSG..SG	TDFTLKISRVEA	EDVGVYYC	FQALQSP	
IGKV2-10*02	AIVLTQTPLSLSVSP	GEPASISCRSS	QSLLHT.DGKNY	LNWYLQKP	GQSPQRLIY	QA.......T	NRDTGVP.D	RFSGSG..SG	TDFTLKISRVEA	EDVGVYYC	FQALQSP	
IGKV2-12*01	AIVLTQTPLSLSVSP	GEPASISCRST	QSLRGS.YGKNY	LNWYQQKP	GQSPKLLIY	WA.......T	NRASGVP.D	RFSGSR..SG	TDFTLKIIRLEA	EDAGVYSC	LQDIQSP	
IGKV2-12*02	AIVLTQTPLSLSVSP	GEPASISCRST	QSLRGS.YGKNY	LNWYLQKP	GQSPKLLIY	WA.......T	NRASGVP.G	RFSGSR..SG	TDFTLKIIRLEA	EDAGVYYC	LQDIQSP	
IGKV2-13*01	AIVLTQTPLSLSVSP	GEPASISCRSS	QSLEI..YGNNF	LSWYQQKP	GQSPQLLIY	EA.......T	NRASGVP.D	RFSGSG..SG	TDFTLKISRVEA	EDAGVYYC	QQNKESL	
IGKV2-13*02	AIVLTQTPLSLSVSP	GEPASISCRSS	QSLEE..YGSNL	LSWYQQKP	GQSPQLLIY	EA.......T	NRASGVP.D	RFSGSG..SG	TDFTLKISRVEA	EDAGVYYC	QQFKEFP	
IGKV2/OR3-1*01	AIVLTQTPLPLSVSP	GEPVSISCKSS	QSLLHR.GRNNY	LHWYLQKP	GQSLQNLIY	YA.......T	NTASGVP.D	RFSGSG..LG	TDFTLKISSMEA	EDVGVHYC	QQSR	
IGKV3-3*01	EIVLTQSAAPKAVSQ	EESVIITCNGS	PGVS.....TNK	LHWYQLKT	GAPPRLLIY	ST.......S	SLAFWVP.T	RFSGSG..SG	TSYSRTISSVAA	QDAADYYC	QQSSSFP	
IGKV3-3*02	EIVLTQSAASKAVSQ	GENVIITCNGS	PGVS.....TNK	LHWYQLKT	GAPPRLLIY	SK.......S	SLAFWVP.T	RFSGSG..SG	TSYSCTISSVAA	QDAADYYC	QQSSSFP	
IGKV5-4*01	ETTVTQSPAFVSATP	GDKVNITRKAS	QDI......DDD	IMCYQQKP	GEAPKLLIK	YA.......S	IHITGVP.T	RFSGSG..YG	TDFTLTIGNMIS	EDATYYFC	QQDDNVP	
IGKV5-4*02	ETTLTQSPAFVSATP	GDKVNITRKAS	QDI......DDD	IMCYQQKP	GEAPRLLIK	YA.......S	IHITGVP.T	RFSGSG..YG	TDFTLTIGNMIS	EDATYYFC	QQDDNVP	
IGKV7-2*01	DIVLTQSPASQIASP	GEQVTISCRAR	QSVCGTWGIISC	VNRYQ*KP	GQPPKLLV*	AA.......S	S*ESEVS.A	RFSGSE..SG	THFTLAIHPMEA	DDAANYFC	QQNQESP	
IGKV7-2*02	DIVLTQSPASQIASP	GEQVTISCRAR	QSVCGTWGIISC	VNRYQ*KP	GQPPKLLV*	AA.......S	S*ESEVS.A	RFSGSE..SG	THFTLTIHPMEA	DDAANYFF	QQNQESP	

Table 2.3 Alignment of porcine IGKV CDR1-IMGT

```
IMGT positions   27 28 29 30 31 32 33 34 35 36 37 38
                 |..|..|..|..|..|..|..|..|..|..|..|..
nt positions            100                 120
                        |                   |
IGKV3-1*01       CAAAGCCTGCTTTACTAA...GATAGGAAGCACTAC
IGKV7-2*01,*02   CAGAGTGTCTGTGGTACTTGGGGCATTATTAGCTGT
IGKV3-3*01,*02   CCAGGTGTAAGC..............ACCAACAAA
IGKV5-4*01,*02   CAAGACATT.................GATGATGAT
IGKV2-6*01       CAGAGCCTTGAGAGT......TATAGTTACAATTTC
IGKV2-6*02       CAGAGCCTTGAGGAA......TATGGAAGCAATTTG
IGKV1-7*01       CAGACCATT.................AGCAGTTAT
IGKV1-7*02       CAGAGCATT.................AGCAGTTAT
IGKV2-8*01       CAGAGCCTTGAGATA......TATGGAAGCAATTTC
IGKV2-8*02       CAGAGCCTCGTAGACAGT...GATGGAGACTCGCTA
IGKV1-9*01,*02   CAGAGTGTT.................AGCAATAAT
IGKV2-10*01      CAGAGCCTCGTAGACAGT...GATGGAGACTCGCTA
IGKV2-10*02      CAGAGCCTCCTACACACT...GATGGAAAGAATTAT
IGKV1-11*01,*02  CAGAGCATT.................AACAAGTGG
IGKV2-12*01,*02  CAGAGCCTCCGAGGTAGT...TATGGAAAGAATTAT
IGKV2-13*01      CAGAGCCTTGAGATA......TATGGAAACAATTTT
IGKV2-13*02      CAGAGCCTTGAGGAA......TATGGAAGCAATTTG
IGKV1-14*01      CAGAGCATT.................AGCAGTTAT
```

Table 2.4 Protein display of IGKC genes using IMGT unique numbering

```
                    A        AB      B          BC        C     CD      D          DE          E      EF   F          FG          G
                  (1-15)           (16-26)   (27-38)   (39-45)       (77-84)               (85-96)     (97-104)   (105-117)   (118-128)
                 ————————————————>  ——————————>          ——————>       ——————>             ——————————>  ——————>               ——————————>
                 1            15   16       26 27      38 39   45       77    84             85        96  97   104 105       117 118
               87654321|............|123|.........| |........| |.....|1234567|......|12345677654321|..........|12|......| |...........| |.........|
IGKC*01          ADAKPSVFIFPPSKEQLE...TQTVSVVCLLN SFFP..REVN VKWKVDGVVQSS..GILDSVTEQDSKD.....STYSLSSTLSLPTSQY..LSHNLYSC EVTHKT..LASPL VKSFSRNECEA
IGKC*02          ADAKPSVFIFPPSKEQLE...TQTVSVVCLLN SFFP..REVN VKWKVDGVVQSS..GILDSVTEQDSKD.....STYSLSSTLSLPTSQY..LSHNLYSC EVTHKT..LASPL VKSFNRNECEA
```

Table 2.5 Genomic features of the porcine IGKV genes

										RS		
IGKV gene	**Fct**	**Octamer (promoter)**	**nt**	**ini.**	**L-PART1 (exon 1) (nt)**	**5' splice site**	**intron (nt)**	**3' splice site**	**V-EXON (exon 2) (nt)**	**V-HEPTAMER**	**V-SPACER (nt)**	**V-NONAMER**
IGKV3-1*01	P	ATTTGCAT	110	ATG	49	GT	180	AG	311	CACAGTG	12	ATATTAACT
IGKV7-2*01	p	ATTTGCAT	94	ATG	49	GC	234	AG	316	CACAATG	10	CAAAAAACC
IGKV3-3*01	ORF	ATTTGCAC	98	ATG	49	GT	208	AG	301	CACAGTG	12	ACAAAAACC
IGKV5-4*01	ORF	ATTTGCAT	94	ATG	49	GT	420	AG	301	CACAGTG	12	ACAAAAACT
IGKV2-5*01	p	ATTTGCAT	98	ATG	49	GT	-	-	-	-	-	-
IGKV2-6*01	F	ATTTGCAT	98	ATG	49	GT	368	AG	310	CACAGTG	12	ACACAAACC
IGKV1-7*01	F	ATCTGCTT	93	ATG	49	GG	196	AG	298	CAGAGTG	12	ACCTAACCC
IGKV2-8*01	F	ATTTGCAT	98	ATG	49	GT	383	AG	313	CACAGTG	12	ACACAAACC
IGKV1-9*01	F	ATTTGCAT	93	ATG	49	GT	193	AG	298	CACAGTG	12	ACAAACCCC
IGKV2-10*01	F	ATTTGCAT	98	ATG	49	GT	383	AG	313	CACAGTG	12	ACACAAACC
IGKV1-11*01	F	ATTTGCAT	93	ATG	49	GT	194	AG	298	CACAGTG	12	CCAAAAACC
IGKV1-11*02	F	ATCTGCTT	93	ATG	49	GG	194	AG	298	CACAGTG	12	CCAAAAACC
IGKV2-12*01	F	ATTTGCAT	98	ATG	49	GG	375	AG	313	CACAGTG	12	ACACAAACC
IGKV2-13*01	F	ATTTGCAT	98	ATG	49	GT	367	AG	310	CACAGTG	12	ACACAAACC
IGKV1-14*01	F	ATCTGCTT	93	ATG	49	GG	196	AG	298	CAGAGTG	12	ACCTAACCC
IGKV2/OR3-1	P	ATTTGCAT	98	ATG	49	GT	377	AG	313	CACAGTG	12	ACACAAACC
IGKV3/OR3-2	p	GTTTGCAT	91	ATG	49	GT	194	AG	273	CACAGTG	12	ACAGAAACC

Fct, functionality; F, functional; P, pseudogene; ORF, open reading frame
IMGT standardized labels are written in capital letters

FIGURE 2.1 Organization of the porcine IGK locus. Overlapping BACs spanning the region are displayed as grey and black bars to indicate sequence heterogeneity. Genes are displayed along a scaffold line as vertical bars representing functional genes (long bars), pseudogenes (short bars), or ORF (intermediate bars). A nearly intact LINE-1 insertion present on CH242-221I5 but not on CH242-227G10 is depicted as a grey box. Genes above the scaffold line are transcribed left to right. Genes below the line are expressed in the opposite orientation

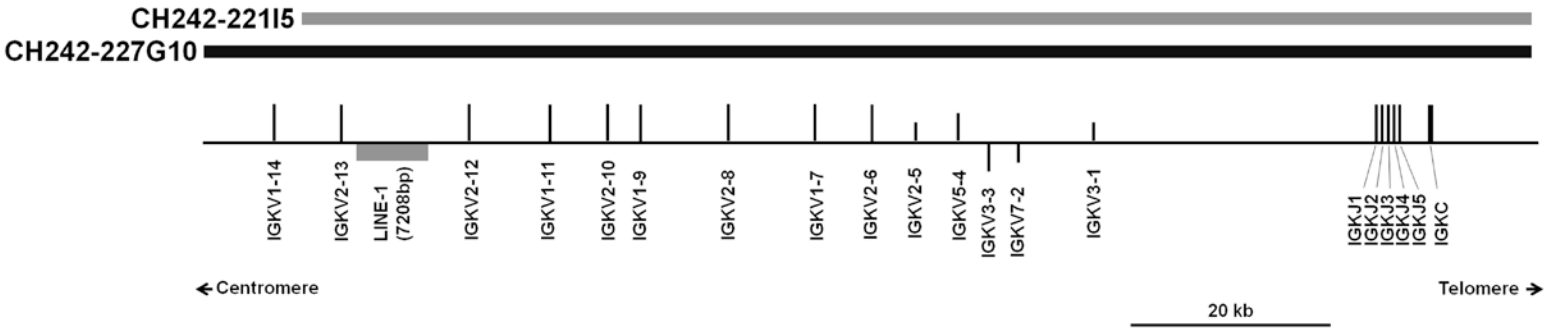


Figure 2.2 Phylogenetic analysis of porcine IGKV nucleotide sequences using UPGMA with 1000 bootstrap iterations. Nucleotide sequences are derived from the V-EXON. Thus, IGKV2-5 is not represented as it is missing this region. Most porcine IGKV genes cluster with either of two subgroups, IGKV1 or IGKV2. Nodes are labeled with bootstrap values. Scale bar represent the number of nucleotide substitutions per site.

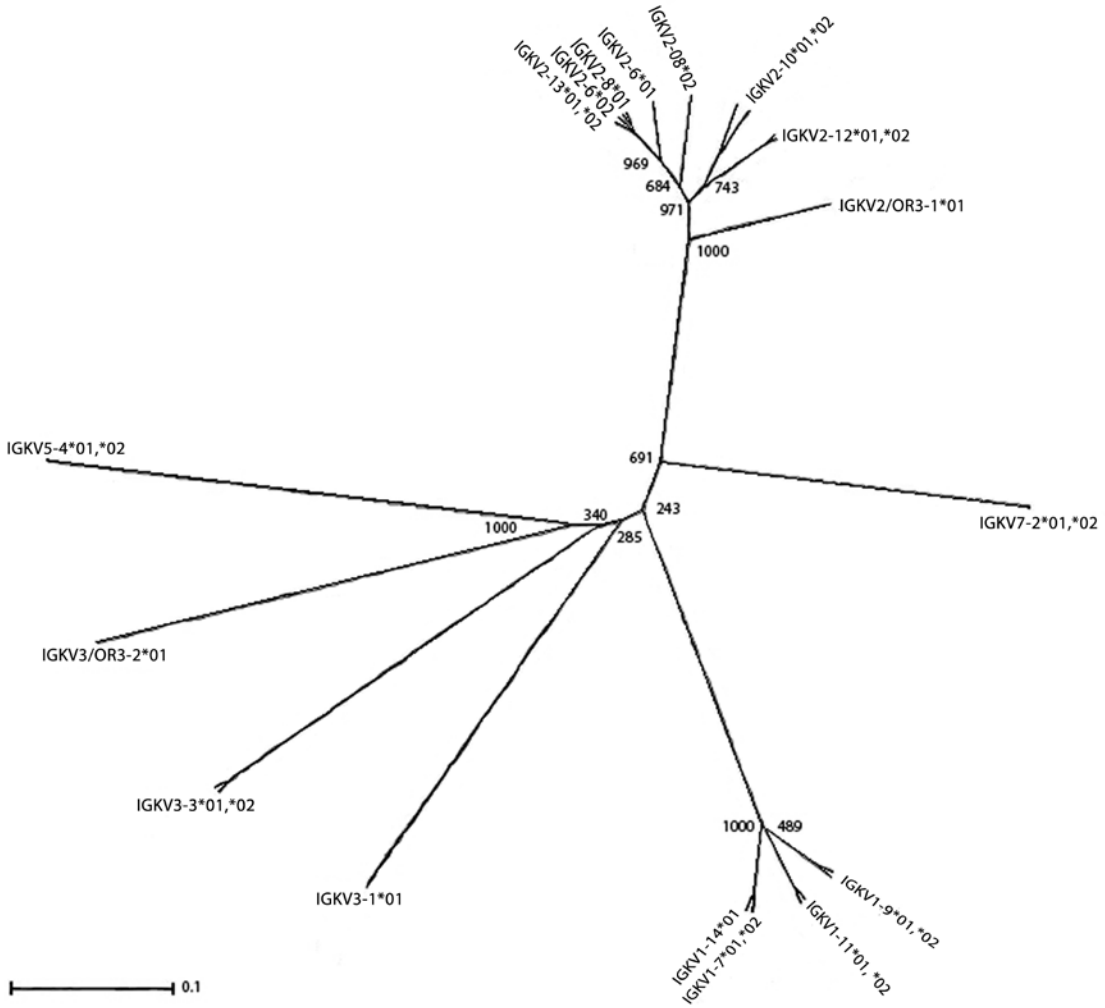
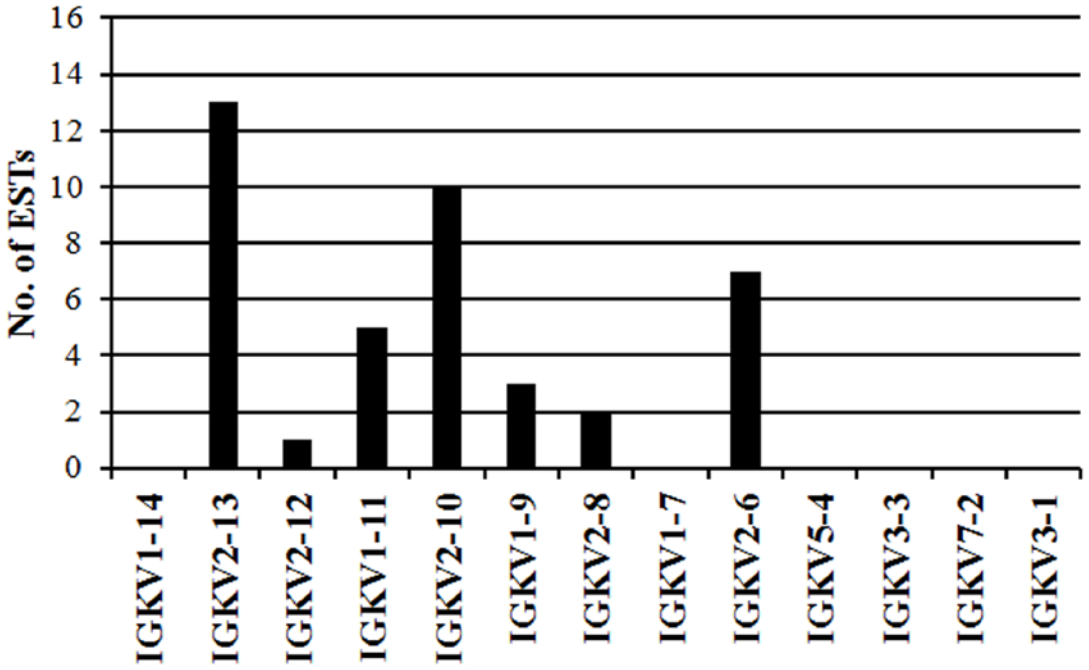


Figure 2.3 Expression of IGKV genes. BLAST analysis revealed 41 hits from 398,837 expressed sequence tags in GenBank and deposited at http://pigest.ku.dk/index.html (Gorodkin 2007). The x-axis is organized by gene order in the locus with the most 5' (most C-distal) IGKV genes on the left. The IGKV2 subgroup dominates (~80 percent) the expression profile.



DISCUSSION

The general organization of the porcine kappa locus is typical of mammals. Functional diversity of the locus is more complex than in cattle, in which only 8 of 24 IGKV genes are functional (Ekman et al. 2009). It is similar to the human kappa locus, which also has high functional diversity and equivalent expression of IGK and IGL (Kawasaki et al. 2001; Lefranc and Lefranc 2001; Butler et al. 2005). The high level of functional diversity in swine and humans may facilitate efficient IGK expression, thus avoiding KDE-dependent ablation of IGKC. Conversely, low functional diversity in cattle may result in a high rate of KDE-dependent IGKC ablation and low relative IGK usage in B cells.

IGKV2 subgroup genes are dominantly expressed as previously described (Butler et al. 2004). However, we identified only five complete IGKV2 genes in the first 89 kb of the IGKV locus, which is substantially less than an estimated 61 IGKV2 genes in 250 kb of sequence identified by Southern hybridization and sequencing of IGKV2-specific PCR clones (Butler et al. 2004). The previous estimate may be an overestimation of kappa locus complexity since the density of genes across the entire locus would need to be at least twice that observed by directly sequencing BACs. The IGKV gene density observed here is similar to that of the human C-proximal IGKV cluster (75 IGKV genes per 542 kb) (Kawasaki et al. 2001, Lefranc and Lefranc 2001). The presence of additional upstream IGKV genes in the porcine locus has not been ruled out, although a BLAST

search of shotgun sequenced genomic BACs did not reveal any additional upstream sequence specific to the kappa locus.

The presence of extensive IGKV allelic differences, up to 10% of sequence variation in IGKV2-8, occur primarily within the CDR1 of IGKV2 subgroup genes. Further, CDR1 appears to be shared by some IGKV2 subgroup genes, but not necessarily between alleles. Interestingly, a similar phenomenon was observed in the porcine IGH locus, in which many IGHV genes share individual CDR (Butler et al. 2006c; Eguchi-Ogawa et al. 2010). The pattern of variation may be due to evolution through non-crossover homologous recombination events, namely gene conversion. Evidence for germline gene conversion was also observed in the human IGKV (Bentley and Rabbitts 1983) and mouse IGHV (Cohen and Givol 1983) genes and in the human IGHC genes (Flanagan et al. 1984, Lefranc et al. 1986, Huck et al; 1988). IG gene conversion is generally presented as a mechanism of somatic antibody diversification, especially in chickens and rabbits which utilize a diverse array of pseudogenes as templates for functional IG gene assembly during B cell development to diversify limited functional germline repertoires (recently reviewed by Kurosawa and Ohta 2011). In swine, however, there is little evidence supporting somatic diversification through templated gene conversion (Butler et al. 2006b). Thus, the porcine kappa locus appears to have achieved repertoire diversity through gene duplication and germline gene conversion to increase allelic variation in CDR1.

CHAPTER III

ORGANIZATION, COMPLEXITY, AND DIVERSITY OF THE PORCINE IMMUNOGLOBULIN LAMBDA LOCUS

Schwartz JC, Lefranc M-P, and Murtaugh MP (2012) The organization, complexity and allelic diversity of the porcine (*Sus scrofa domesticus*) immunoglobulin lambda locus. *Immunogenetics*. 64(5):399-407. doi:10.1007/s00251-011-0594-9

We have characterized the organization, complexity, and expression of the porcine (*Sus scrofa domestica*) immunoglobulin lambda (IGL) light chain locus, which accounts for about half of antibody light chain usage in swine, yet is nearly totally unknown. Twenty-two IGL variable (IGLV) genes were identified that belong to seven subgroups. Nine genes appear to be functional. Eight possess stop codons, frameshifts, or both, and one is missing the V-EXON. Two additional genes are missing an essential cysteine residue and are classified as ORF (open reading frame). The IGLV genes are organized in two distinct clusters, a constant (C)-proximal cluster dominated by genes similar to the human IGLV3 subgroup, and a C-distal cluster dominated by genes most similar to the human IGLV8 and IGLV5 subgroups. Phylogenetic analysis reveals that the porcine IGLV8 subgroup genes have recently expanded, suggesting a particularly effective role in immunity to porcine-specific pathogens. Moreover, expression of IGLV genes is nearly exclusively restricted to the IGLV3 and IGLV8 genes. The constant locus comprises three tandem cassettes comprised of a joining (IGLJ) gene and a constant (IGLC) gene, whereas a fourth downstream IGLJ gene has no corresponding associated IGLC gene. Comparison of individual BACs generated from the same individual revealed polymorphisms in IGLC2 and several IGLV genes, indicating that allelic variation in IGLV further expands the porcine antibody light chain repertoire.

INTRODUCTION

Adaptive humoral immunity is mediated by a diverse array of genes which recombine during B cell development to generate a vast repertoire of antigen-binding immunoglobulin (IG) proteins. Antigen binding is carried out using the variable domains of both a heavy chain and a light chain. The antibody variable domain is encoded by a variable (V) gene, a diversity (D) gene (heavy chain-only) and a joining (J) gene which rearrange via recognition and cleavage of flanking recombination signal (RS) sequences by the RAG1 and RAG2 complex and subsequent double strand break repair whereby the intervening sequence is excised from the genome (McBlane et al. 1995; Kim and Oettinger 2000). Three complementarity determining regions (CDR) within the variable domain of each heavy and light chain exhibit a large degree of sequence variability. These CDR provide antibodies with an incredibly diverse antigen-binding repertoire (Wu and Kabat 1970, Lefranc and Lefranc 2001).

A large body of knowledge concerning the pig (*Sus scrofa domestica*) IG heavy (IGH) locus has been accumulated. Its germline organization on chromosome 7q25-q26 (Yerle et al. 1997) of the constant (IGHC) genes and the fifteen C-proximal V (IGHV) genes have been mapped, and its expression has been characterized (Eguchi-Ogawa et al. 2010). However, the two light chain loci, kappa (IGK) and lambda (IGL), on chromosomes 3q12-q14 and 14q16-q21, respectively (Yerle et al. 1997), are only partially characterized in the vicinity of their constant genes, and no information is available on the organization of their variable genes (Butler et al. 2006). A search of

GenBank and IMGT/LIGM-DB (Giudicelli et al. 2006) reveals the presence of approximately 100 cDNA sequences for the porcine kappa locus and only two partial lambda cDNA sequences, thus highlighting the lack of information on swine light chain loci. Here, we interrogated available porcine genome sequence (Schook et al. 2005; Humphray et al. 2007; Archibald et al. 2010), as well as additional sequence generated by our lab and available in GenBank and IMGT/LIGM-DB to completely characterize the organization, complexity, and expression of the porcine light chain lambda (IGL) locus. Our findings show that the locus organization is typical of mammalian species, but is substantially modified so that expression is dominated by V genes that have undergone a recent expansion.

MATERIALS AND METHODS

Identification and sequencing of bacterial artificial chromosomes

The *Sus scrofa* genome build 9 was queried to identify bacterial artificial chromosomes (BACs) containing IGLV sequences using the Basic Local Alignment Search Tool (BLAST) within the Ensembl database (Altschul et al. 1990; Hubbard et al. 2002). Nucleotide sequences in the region of interest for each BAC were downloaded from GenBank for further analysis. The CHORI (Children's Hospital Oakland Research Institute)-242 BAC library used was derived from a single Duroc sow (http://bacpac.chori.org/porcine242.htm). Four BACs were identified that span the IGL locus (CH242-141B5, CH242-524K4, CH242-158O5, CH242-288F14) and two additional BAC clones (CH242-298L14 and CH242-82N3) were found to overlap and extend upstream of the locus (Figure 3.1A). Of these BACs, three represent completely sequenced contigs (CH242-158O5, CH242-298L14, and CH242-82N3) and the remaining three are only partially sequenced. CH242-288F14 is fragmented into five contigs; however, its content almost completely overlaps CH242-158O5 and CH242-298L14. Additionally, CH242-141B5 is fragmented into seven contigs and CH242-524K4 is fragmented into 19 contigs (not shown). Due to a number of gaps in the overlap between these two BACs, it was uncertain if the entire IGL locus was represented. We re-sequenced these two BAC clones. They were expanded overnight, and BAC DNA was purified using the Qiagen Plasmid Midi Prep with Qiagen-tip 500 columns. Purified

DNA was submitted to the University of Minnesota Biomedical Genomics Center for library preparation and paired-end sequencing using the Illumina GAIIx platform.

Characterization of the porcine IG lambda locus

Approximately 20 million high quality reads were sorted by molecular tag to differentiate samples and assembled using a combination of the software programs ABySS and Velvet (Simpson et al. 2009; Zerbino and Birney 2008). Generated contigs were assembled against the existing BAC sequences from GenBank using Sequencher 4.10.1 (Gene Codes Corporation). The insert of CH242-141B5 retained a 651 bp gap in a highly repetitive region of the IGLC locus and CH242-524K4 retained several gaps, primarily in the IGLC locus and further downstream. The entire J-C region was re-sequenced using primer walking, PCR, and chain-termination sequencing to resolve the gaps. It included using primers specific for each IGLJ paired with either a reverse IGLJ specific-primer or with a conserved IGLC primer for PCR amplification and chain-termination sequencing. The extent of the locus re-sequenced for the present study thus included the entire constant region and the first six IGLV genes. However, no new IGLV genes were identified as a result of our next generation sequencing, as these were all present on previously shotgun sequenced contigs. The complete locus sequences were manually annotated and interrogated for immunoglobulin features such as RS (ie. heptamers and nonamers), promoters (ie. octamers), and gene structure using the annotation software Artemis (Rutherford et al. 2000).

The sequences of CH242-141B5, CH242-524K4, CH242-158O5, CH242-288F14, CH242-298L14, and CH242-82N3 were acquired from GenBank (accession numbers:

CU467669, CU467599, CU468977, CU468665, CT827879, and CU062407, respectively) and assessed for IGLV, IGLJ, and IGLC genes using BLAST. Phylogenetic analyses were performed in CLC Sequence Viewer (CLC Bio) and Dendroscope (Huson et al. 2007) using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) with 1000 bootstrap iterations. Genes were annotated according to IMGT®, the international ImMunoGeneTics information system® (Lefranc et al. 2009) (http://www.imgt.org). Translated amino acid sequences of the IGLV genes were compared and CDR and framework (FR) boundaries were annotated according to IMGT unique numbering for V region and V domain (Lefranc et al. 2003). IGLC gene translations were annotated according to IMGT unique numbering for C domain (Lefranc et al. 2005). Expression of germline IGLV genes was deduced using 116 BLAST hits from 398,837 porcine expressed sequence tags (ESTs) obtained from GenBank and deposited at: http://pigest.ku.dk/index.html (Gorodkin et al. 2007) using an E-value threshold of 10^{-12} and $\geq$ 98 percent identity.

Nomenclature

The porcine IGLV genes are named according to IMGT nomenclature (Lefranc and Lefranc 2001; Lefranc 2007; Lefranc 2008), using human V gene subgroup nomenclature to maintain consistency with porcine heavy chain, light chain, and cattle lambda light chain nomenclature (Eguchi-Ogawa et al. 2010; Butler et al. 2005; Butler et al. 2006; Pasman et al. 2010). The porcine IGLV, IGLJ and IGLC genes were submitted to IMGT/GENE-DB (Giudicelli et al. 2005). BLAST at NCBI was used to organize IGLV genes into subgroups using a 75 percent identity threshold according to IMGT

criteria (Lefranc and Lefranc 2001; Lefranc 2007; Lefranc 2008). Genes were deemed pseudogenes if they contained a truncation, stop codon, frameshift, or a defective initiation codon. Additionally, genes were described as ORF (open reading frame) if they were missing one (or more) key amino acids (1st-CYS 23, CONSERVED-TRP 41, hydrophobic 89, and 2nd-CYS 104). RS were deemed non-canonical if the heptamer was anything other than “CACAGTG” for V-HEPTAMER (or “CACTGTG” for J-HEPTAMER) or if the nonamer contained at least two nucleotides which were each present in less than 10 percent of all RS described by Ramsden et al. (1994).

RESULTS

Organization of the porcine immunoglobulin lambda locus

The porcine IG lambda (IGL) locus spans approximately 229 kb on chromosome 14 (Figure 3.1). The IGLV locus is organized in two distinct clusters containing 22 IGLV genes, fewer as compared to humans and cattle which have three clusters containing a total of 52 and 25 IGLV genes, respectively (Frippiat et al. 1995; Pasman et al. 2010). The C-proximal cluster contains 6 IGLV genes, of which one is a pseudogene. The C-distal cluster, approximately 65.1 kb upstream, contains 16 IGLV genes, of which 8 are pseudogenes. Two of the pseudogenes possess premature stop codons (IGLV3-1 and IGLV8-21), four have mutated start sites (IGLV(III)-8, IGLV1-15, IGLV1-20, and IGLV5-22), five are frameshifted (IGLV(III)-8, IGLV5-11, IGLV5-17, IGLV1-20, and IGLV5-22), and one (IGLV1-12) is missing the V-EXON (Table 3.1). In addition, multiple codons are deleted from both IGLV3-1 and IGLV5-11. Two additional genes are classified as ORF (IGLV7-7 and IGLV7-9) since they lack the highly conserved cysteine residue at IMGT position 104 (2nd-CYS), which is critical for intrachain disulfide bond formation with the cysteine at position 23 (1st-CYS) and are therefore unlikely to be functional (Tables 3.1 and 3.2) (Lefranc et al. 2003; Bergman and Kuehl 1979). IGLV8-16 contains an insertion within CDR3, creating a premature stop codon at the 3' end of the gene. However, due to exonuclease and terminal deoxynucleotidyl transferase (TdT) activity in the V-J junction during recombination, it is plausible for this stop codon to be altered and generate a functional V region. The intervening region between the IGLV

clusters contains the genes *PRAME* and *ZNF280B*, syntenic with cattle (Pasman et al. 2010).

The presence of additional IGLV clusters farther upstream was ruled out by analyzing contiguous, overlapping sequence represented by the fully sequenced BACs CH242-298L14 and CH242-82N3 (Figure 3.1A). These BACs extend approximately 445 kb upstream from the most upstream IGLV gene, IGLV5-22. Flanking the IGL locus upstream from IGLV5-22 on these BACs are the genes *SLC5A4*, *SLC5A1*, *YWHAH*, and *DEPCD5* (ordered from IGL-proximal to distal), syntenic with the cattle upstream flanking genes (UCSC Genome Browser, assembly: bos_taurus_UMD_3.1/bosTau6). This effectively rules out the possibility of additional upstream IGLV clusters and makes the germline porcine IGL locus more compact than that of cattle.

In contrast to the mouse IGL locus which is organized in tandem V-J-C cassettes (Sanchez et al. 1991), the organization of the porcine IGL locus is similar to most other mammals having tandem J-C cassettes downstream of the IGLV genes (Figure 3.1D). Approximately 91 kb of sequence on CH242-524K4 lies downstream of IGLJ4, the most 3’ IGLJ gene. However, no corresponding IGLC gene was identified, based on both chain-termination and next generation sequencing. Likewise, PCR failed to amplify a product when using an IGLJ4-specific forward primer when paired with either of two conserved constant region reverse primers, despite positive results with the other IGLJ-specific primers (data not shown). The canonical amino acid motif W/F-G-X-G was present in the IGLJ genes save IGLJ1, which also has a non-canonical RS, indicating that

swine only possess two functional J-C cassettes, IGLJ2-IGLC2 and IGLJ3-IGLC3 (Table 3.3).

Phylogenetic analysis of IGLV genes

The C-proximal IGLV cluster contains 5 genes (4 functional and one pseudogene) all belonging to either the IGLV3 or IGLV2 subgroups. The second cluster is comprised of 16 genes, six of which belong to the IGLV8 subgroup, four belong to IGLV5, three belong to IGLV1, two belong to IGLV7, and one, which most closely resembles members of IGLV clan III yet, is distinct from any defined subgroup. This organization differs from cattle, which possess four IGLV2 subgroup genes and four IGLV3 subgroup genes in the most C-proximal IGLV cluster and exclusively contain IGLV1 genes in their second and third IGLV clusters (Pasman et al. 2010). Interestingly, IGLV8 subgroup genes are at least 92.4% identical to each other compared to IGLV1, IGLV3, and IGLV5 genes which are only 84.5%, 74.2%, and 79.5% identical, respectively. This suggests that porcine IGLV8 genes result from a recent expansion by gene duplication (Figure 3.2).

Allelic variation

BACs at the 5' and 3' ends of the IGLV locus overlap with approximately 99 percent identity. All eight IGLV genes in these overlaps differ by at least a single nucleotide (Table 3.1). IGLV3-3, IGLV3-4, IGLV8-19, and IGLV8-21 differ by 2, 4, 6, and 8 amino acids between alleles, respectively (Table 3.2). A gene order polymorphism was found in the second IGLV cluster at the 5' end of the locus, where the last four IGLV genes have a

different order (Figures 3.1B and 3.1C). However, it is not certain if it is real or is an artifact due to whole genome shotgun sequencing and assembly error.

The CH242-141B5 insert terminates downstream of IGLJ3 and is therefore missing IGLC3 and IGLJ4. There are no nucleotide substitutions in any of the first three IGLJ genes or their respective RS between BACs. Sequence analysis revealed, however, that IGLC2 on CH242-524K4 contains a non-synonymous SNP resulting in an A1>T (using IMGT numbering) amino acid substitution. Other than this polymorphism, the constant region exons are all identical to each other (Table 3.4).

Expression of IGLV genes

Only IGLV3 and IGLV8 subgroup genes are predicted to be expressed based on functionality and RS (Table 3.5), in agreement with a previous report (Butler et al. 2006). BLAST analysis of a porcine EST database revealed that all functional IGLV8 subgroup genes are expressed, with IGLV8-13 and IGLV8-18 expressed most abundantly, but only two IGLV3 subgroup genes were expressed (Figure 3.3). In contrast, IGLV3 family members were reported to dominate the expressed pre-immune lambda repertoire in neonatal pigs (Butler et al. 2006). Additionally, low level expression of IGLV5-14 was observed despite the presence of non-canonical heptamer and nonamer sequences (Table 3.5).

A canonical octamer (ATTTGCAT) necessary for transcription was present upstream of all IGLV genes except for six (IGLV2-6, IGLV-7, IGLV5-11, IGLV5-14, IGLV1-15, and IGLV5-17). The octamer was located 106 bp upstream of the seven most highly expressed IGLV genes (six IGLV8 subgroup genes and IGLV3-3). The mean distance

among all IGLV genes was 94 bp and ranged from 82 to 108 bp (Table 3.5), which is typical of mammals (Parslow et al. 1984).

Table 3.1 Porcine (*Sus scrofa domestica*) IGLV subgroups, gene and allele names. Allele functionality, clone names and accession numbers are provided

Next Page

IMGT subgroup	IMGT gene name	IMGT allele name	Fct	Clone names	Clone accession numbers
IGLV1	IGLV1-12	IGLV1-12*01	P(1)	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2
	IGLV1-15	IGLV1-15*01	P(2)	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2
	IGLV1-20	IGLV1-20*01	P(3)	CH242-158O5	CU468977.2
		IGLV1-20*02	P(3)	CH242-141B5	CU467669.2
IGLV2	IGLV2-6	IGLV2-6*01	F	CH242-141B5	CU467669.2
IGLV3	IGLV3-1	IGLV3-1*01	P(4)	CH242-141B5	CU467669.2
		IGLV3-1*02	P(4)	CH242-524K4	CU467599.3
	IGLV3-2	IGLV3-2*01	F	CH242-141B5	CU467669.2
		IGLV3-2*02	F	CH242-524K4	CU467599.3
	IGLV3-3	IGLV3-3*01	F	CH242-141B5	CU467669.2
		IGLV3-3*02	F	CH242-524K4	CU467599.3
	IGLV3-4	IGLV3-4*01	F	CH242-141B5	CU467669.2
		IGLV3-4*02	F	CH242-524K4	CU467599.3
	IGLV3-5	IGLV3-5*01	F	CH242-141B5	CU467669.2
IGLV5	IGLV5-11	IGLV5-11*01	P(5)	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2
	IGLV5-14	IGLV5-14*01	F	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2
	IGLV5-17	IGLV5-17*01	P(6)	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2
	IGLV5-22	IGLV5-22*01	P(7)	CH242-158O5	CU468977.2
		IGLV5-22*02	P(7)	CH242-288F14	CU468665.2
IGLV7	IGLV7-7	IGLV7-7*01	ORF(8)	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2
	IGLV7-9	IGLV7-9*01	ORF(9)	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2
IGLV8	IGLV8-10	IGLV8-10*01	F	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2
	IGLV8-13	IGLV8-13*01	F	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2
	IGLV8-16	IGLV8-16*01	F	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2
	IGLV8-18	IGLV8-18*01	F	CH242-158O5	CU468977.2
	IGLV8-19	IGLV8-19*01	F	CH242-158O5	CU468977.2
		IGLV8-19*02	F	CH242-288F14	CU468665.2
	IGLV8-21	IGLV8-21*01	P(10)	CH242-158O5	CU468977.2
		IGLV8-21*02	P(10)	CH242-288F14	CU468665.2
IGLV(III)	IGLV(III)-8	IGLV(III)-8*01	P(11)	CH242-158O5	CU468977.2
				CH242-141B5	CU467669.2

Fct, functionality; F, functional; P, pseudogene; ORF, open reading frame

(1) V-EXON is missing

(2) INIT-CODON replaced by Val (ATG>GTG)

(3) INIT-CODON replaced by Arg (ATG>CGG); frameshift at position 98; 2nd-CYS replaced by Val; STOP-CODON

(4) STOP-CODON in L-PART1, codons 80-87 are deleted

(5) 14 nt deletion and frameshift from codons 2-6, frameshift at position 45

(6) Frameshift at position 57

(7) L-PART1 and V-EXON are fused, INIT-CODON is replaced by Leu (ATG>CTG), frameshift in FR3-IMGT

(8) 2nd-CYS replaced by Tyr

(9) 2nd-CYS replaced by Ser

(10) STOP-CODON in L-PART1

(11) INIT-CODON replaced by Lys (ATG>AAG); frameshift in FR1-IMGT

Table 3.2 IMGT protein display of inframe porcine (*Sus scrofa domestica*) IGLV genes. The protein display is shown using IMGT header (IMGT Repertoire, http://www.imgt.org) and IMGT unique numbering for V region and V domain (Lefranc et al. 2003)

Next Page

```
                    FR1-IMGT          CDR1-IMGT       FR2-IMGT       CDR2-IMGT                  FR3-IMGT                  CDR3-IMGT
                     (1-26)           (27-38)         (39-55)        (56-65)                    (66-104)                  (105-117)
                        A            B          BC         C        C'        C'C"       C"        D           E          F        FG
                     (1-15)       (16-26)    (27-38)    (39-46)  (47-55)    (56-65)    (66-74)   (75-84)     (85-96)    (97-104)
                  ————————————> ———————————> ————————> ————————> ————————> ————————> ————————> ————————> ————————————> ————————>
                  1           15 16       26 27        38 39    46 47     55 56      65 66     74 75      84 85        96 97   104 105     113
                  |............| |.........| |.........| |......| |.......| |........| |.......| |........| |.........| |......| |.......|
IGLV1-15*01       QAVLTQPPS.ESGSL GQRVTLSCTGS SSNIG...GGNS VNWSQPLP GKVPRSVFT YA.......N LMAIAAP.D QFSGFK..SG SSGTLTITGLQA EDDAEYYC TAGGDSLDG
IGLV1-20*01       QAVLTQLLS.TSGSL GQRVTLSCTGS SSNIG...SGNT VNWSQQLP GKAPRTIID GN.......S NRPSGVP.D QFSGSK..SG SSGTLTITGLQA ELRLSITV *PGRTASMV
IGLV2-6*01        QSALTQPPS.VSRNL KEMETISCAGT SSDI.....GGY VSWYQQHP GLAPKFLIY YV.......N TRASGIP.D GFCGSK..SG NTASLTISGLQA EDEADYYC SSPRSGGTL
IGLV3-1*01        SCELTQPPS.LSVSL GQMARITCGGN NIG......YKH TFWYQQKV GQAPVLVMY SD.......S HQLSGIP.E LFSGS..... ...TLTISRARA EDEEDYYC AVADGKGSS
IGLV3-2*01        SYEVTQPPS.VSVNP GQRASITCEGN NIG......RKD IQWYQQKP GQAPVLFIY ED.......T NRPSGIP.E RFSASK..SG NTATLTISGAQA EDEADYYC QSYDDSYTP
IGLV3-3*01        SYELTQPSS.ESVAL GNTAKITCSGD LLD......EKY TQWYQQKP GQAPLLLIY KD.......S ERPSGIP.E RFSGSS..SG KTATLTITGAQA EDEADYYC QSADSIDNA
IGLV3-3*02        SYELTQPSS.ESVAL GNTAKITCSGD LLD......EKY TRWYQQKP GQAPLLLIY KD.......S ERPSGIP.E RFSGSS..SG KTATLTITGAQA EDEADYYC QSADSSDNA
IGLV3-4*01        SSKLTQPPG.VSVSL GGTASIACQGD NFG......SYY AHWYQQKP GQSPMLVIY EY.......N KRASGIP.D RFSGSK..SG NTATLTISGAQA EDEADYYC QSSDDSYTP
IGLV3-4*02        SSQLTQPPG.VSVSL GGTASITCQGG NFG......SYY AHWYKQKP GQSPMLVIY EY.......N KRASGIP.D RFSGSK..SG NTATLTISGAQA EDEADYYC QSSDDSYTP
IGLV3-5*01        SSKLTQPPG.VSVSL GGTASITCQGA NFG......SYY AHWYQQKP GQSPELVIY EY.......P EIFLGFL.E RFSVSR..TG DTATLTISGAQA EDEADYYC QVYDGGYHV
IGLV5-14*01       QAVLTQPPS.LSASP GPSARLPCTLS SGSSV...GSYH ISWYQRKP GRPPWYLLR FHFAS.SKDQ GSGVPSC.F SGDKDA..SA HAGLLLISGLQP EDKADCDC LNWQSSAS
IGLV7-7*01        SQMVTQEIS.LSTTS GETVTLTCGSS AGAVT...TSNY ASWVQQKP YEVLWGLIG GT.......S TQIPGVP.A RFSGSL..LG DKAALTRSGAQS KDEADYYY VLWFSNV
IGLV7-9*01        PSCVTQEPS.KTVSP GVTVTFTCGSS MGVVT...IGHY PCWFQQKP GQAPRTLIY DT.......N SKLSWTP.A WFSGSF..LG DKAALILSRAQP QDQVEYYS GAQ
IGLV8-10*01       SQTVIQEPA.MSVSP GGTVTLTCAFS SGSVT...TSNY PSWFQQTP GQPPRLLIY RT.......N NRPTGVP.S RFSGAI..SG NKAALTITGAQA NDEADYFC TLYKSSAN
IGLV8-13*01       SQTVIQEPA.MSVSP GGTVTLTCAFS SGSVT...TSNY PSWFQQTP GQPPRQLIY ST.......N NRPTGVP.S RFSGAI..SG NKAALTITGAQA EDEADYFC ALYKSSAN
IGLV8-16*01       SQTVIQEPA.MSVSL GGTVTLTCAFS SGSVT...SSNY PSWYQQTP GQPPRQLIY ST.......N SRPTGVP.S RFSGAI..SG NKATLTITGAQA EDEADYFC ALYKGSGTYT*
IGLV8-18*01       SQTVIQEPA.MSVSL GGTVTLTCAFS SGSVT...SSNY PGWFQQTP GQPPRTVIY ST.......N SRPTGVP.S RFSGAI..SG NKATLTITGAQA EDEADYFC ALYKSCTN
IGLV8-19*01       SQTVIQEPA.MSVSL GGTVTLTCAFS SGSVT...SSNN PGWFQQTP GQPPRTVIY QT.......N NRPTGVP.S RFSGAI..SG NKATLTITGAQA EDEADYFC ALGKSCTN
IGLV8-19*02       SQTVIQEPA.MSVSL GGTVTLTCAFS SGSVT...SSNY PGWYQQTP GQPPRQLIY ST.......N SRPTGVP.S RFSGAI..SG NKATLTITGAQA EDEADYFC ALYKSCTN
IGLV8-21*01       SQTVIQEPA.MSVSP GGTVTLTCAFS SGSVT...TSNY PGWYQQTP GQPPRQLIY QT.......N SRPTGVP.S RFSGAI..SG NKATLTITGAQA EDEADYFC ALEKSSAN
IGLV8-21*02       SQTVIQEPA.MSVSP GGTVTLTCAFS SGSVT...TSNY PSWFQQTP GQPPRQLIY QT.......N SRPTGVP.S RFSGAI..SG NKATLTITGAQA EDEADYFC ALEKSSAN
```

Table 3.3 Porcine (*Sus scrofa domestica*) IGLJ nucleotide (nt) and amino acid (AA) sequences and associated J-RS

IGLV gene	Fct	J-NONAMER	J-SPACER	J-HEPTAMER	J-REGION nt and AA sequence
IGLJ1	P	GGGTTTTGT	TCGAGCCTCAGT	CAGCGTG	GTATTTTTCCGGCAGCGAGACTCAGTTCACCGCCTTAG Y F S G S E T Q F T A L
IGLJ2	F	GGTTTATGT	TTGAGGCTGTAT	CACTGTG	TAATATTTTCGGCGGTGGGACCCATCTGACCGTCCTCG N I F G G G T H L T V L
IGLJ3	F	GGTTTATGT	TTGAGGCTGTAT	CACTGTG	TGTTCCCTTCGGCGGTGGGACCCATCTGACCGTCCTCG V P F G G G T H L T V L
IGLJ4	F	GGTTTATGT	TTGAGGCTGTAT	CACTGTG	CGATAGGTTCGGCCGAGGGACCCGTCTAAGTGTCCTCC D R F G R G T R L S V L

Fct, functionality; F, functional; P, pseudogene

Table 3.4 IMGT protein display of porcine (*Sus scrofa domestica*) IGLC genes. The protein display is shown using IMGT header (IMGT Repertoire, http://www.imgt.org) and IMGT unique numbering for C domain (Lefranc et al. 2005)

```
                    A       AB      B          BC        C     CD      D          DE          E      EF   F          FG          G
                  (1-15)          (16-26)    (27-38)   (39-45)       (77-84)                (85-96)     (97-104)   (105-117)   (118-128)
                 ————————————————>  ——————————>          ———————>      ———————>            ——————————>  ———————>               ——————————>
                 1            15   16       26 27      38 39   45       77    84             85        96  97   104 105       117 118       128
          87654321|.............|123|.........| |........| |.....|1234567|......|12345677654321|..........|12|......| |...........| |.........|
IGLC1*01     QPKAAPTVNLFPPSSEELG...TNKATLVCLIS DFYP..GAVT VTWKAGGTTVTQ..GVETTKPSKQSN......NKYAASSYLALSASDW..KSSSGFTC QVTHE....GTIV EKTVTPSECA.
IGLC2*01     QPKAAPTVNLFPPSSEELG...TNKATLVCLIS DFYP..GAVT VTWKAGGTTVTQ..GVETTKPSKQSN......NKYAASSYLALSASDW..KSSSGFTC QVTHE....GTIV EKTVTPSECA.
IGLC2*02     QPKATPTVNLFPPSSEELG...TNKATLVCLIS DFYP..GAVT VTWKAGGTTVTQ..GVETTKPSKQSN......NKYAASSYLALSASDW..KSSSGFTC QVTHE....GTIV EKTVTPSECA.
IGLC3*01     QPKAAPTVNLFPPSSEELG...TNKATLVCLIS DFYP..GAVT VTWKAGGTTVTQ..GVETTKPSKQSN......NKYAASSYLALSASDW..KSSSGFTC QVTHE....GTIV EKTVTPSECA.
```

Table 3.5 Genomic features of the porcine (Sus scrofa domestica) IGLV genes

										V-RS		
IGLV gene	**Fct**	**octamer (promoter)**	**nt**	**ini.**	**L-PART1 (nt)**	**5' splice site**	**intron (nt)**	**3' splice site**	**V-EXON (nt)**	**V-HEPTAMER**	**V-SPACER (nt)**	**V-NONAMER**
IGLV3-1*01	P	ATTTGCAT	96	ATG	46	GT	358	AG	282	CCCAGTG	37	ACACAAACT
IGLV3-2*01	F	ATTTGCAT	96	ATG	46	GT	370	AG	299	CACAGTG	23	ACACAAACT
IGLV3-2*02	F	ATTTGCAT	96	ATG	46	GT	363	AG	299	CACAGTG	17	ACACAAACT
IGLV3-3*01	F	ATTTGCAT	106	ATG	46	GT	143	AG	299	CACAGTG	23	ACACAAACC
IGLV3-4*01	F	ATTTGCAT	82	ATG	46	GT	151	AG	299	CACAGTG	23	ACACAAACC
IGLV3-5*01	F	ATTTGCAT	82	ATG	46	GT	151	AG	299	CACAGTG	23	ACACAAACT
IGLV3-5*02	F	ATTTGCAT	82	ATG	46	GT	151	AG	299	CACAGTG	23	ACACAAACC
IGLV2-6*01	F	ATTTGTAT	101	ATG	46	GT	116	AG	302	TACAGTG	23	ACACAAACC
IGLV7-7*01	ORF	ATCTGCAT	101	ATG	46	GT	97	AG	301	CACCGTG	23	ACATGAGCC
IGLV(III)-8	P	ATTTGCAT	93	AAG	47	GT	111	AG	267	CACAGTG	21	ACTCAAACC
IGLV7-9*01	ORF	ATTTGCAT	108	ATG	43	GT	85	AG	288	CACAGTG	23	GACACAAAG
IGLV8-10*01	F	ATTTGCAT	106	ATG	46	GT	97	AG	304	CACAGTG	23	ACCCAAACC
IGLV5-11*01	P	ACTGGCAT	87	ATG	46	GT	109	TG	303	CACAATG	23	AAAAAACAT
IGLV1-12*01	P	ATTTGCAT	108	ATG	46	GA	-	-	-	-	-	-
IGLV8-13*01	F	ATTTGCAT	106	ATG	46	GT	97	AG	304	CACAGTG	23	ACCCAAACC
IGLV5-14*01	F	ACTGGCAT	88	ATG	46	GT	121	AG	323	CACTGTG	23	AGAAGAATC
IGLV1-15*01	P	ATTGGCAT	107	GTG	46	GT	106	AG	310	CACAGTG	23	ACCAAAACC
IGLV8-16*01	F	ATTTGCAT	106	ATG	46	GT	97	AG	312	CACAGTG	23	ACCCAAACC
IGLV5-17*01	P	ACTGGCAT	87	ATG	46	GT	121	AG	302	CACTGTA	23	TGGGAACGG
IGLV8-18*01	F	ATTTGCAT	106	ATG	46	GT	97	AG	304	CACAGTG	23	ACTCAAACC
IGLV8-19*01	F	ATTTGCAT	106	ATG	46	GT	97	AG	304	CACAGTG	23	ACCCAAACC
IGLV8-19*02	F	ATTTGCAT	105	ATG	46	GT	98	AG	304	CACAGTG	23	ACCCAAACC
IGLV1-20*01	P	ATTTGCAT	88	CGG	46	GT	108	AG	309	CACAGTG	23	ACCAAAACC
IGLV1-20*02	P	ATTTGCAT	91	CGG	46	GT	108	AG	309	CACAGTG	23	ACCAAAACC
IGLV8-21*01	P	ATTTGCAT	106	ATG	46	CC	96	AG	304	CACAGTG	23	ACCCAAACC
IGLV5-22*01	P	ATTTGCAT	103	CTG	-	-	-	-	354	CACAGTG	23	AGAAGAATC

Fct, functionality; F, functional; P, pseudogene; ORF, open reading frame
IMGT standardized labels are written in capital letters.

Figure 3.1 Organization of the porcine (*Sus scrofa domestica*) immunoglobulin lambda (IGL) locus. (A) BACs used to determine locus organization. Grey and black bars represent allelic variation. (B) IGL locus gene organization. Genes are displayed along a scaffold line as vertical bars representing functional genes (long bars), pseudogenes (short bars), or ORF (intermediate bars). (C) Rearranged gene order in IGLV1-19 to IGVL1-22 in BAC CH242-288F14. (D) Organization of the IGL J-C locus. The non-Ig related genes *ZNF280B, PRAME*, and *SLC5A4* are shown for genomic context

Next Page

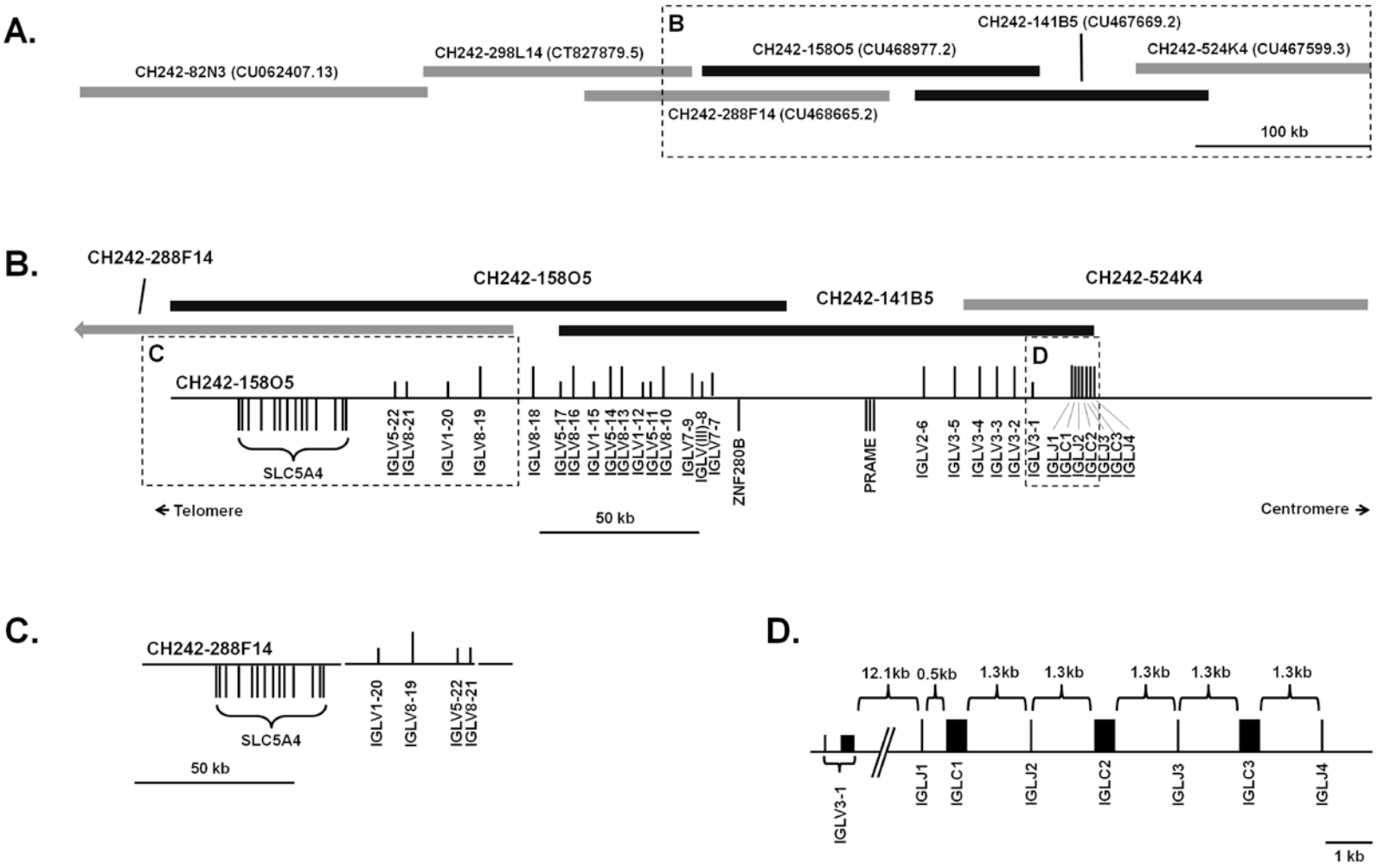

A.
B
CH242-141B5 (CU467669.2)
CH242-298L14 (CT827879.5)
CH242-158O5 (CU468977.2)
CH242-524K4 (CU467599.3)
CH242-82N3 (CU062407.13)
CH242-288F14 (CU468665.2)
100 kb
B.
CH242-288F14
CH242-158O5
CH242-141B5
CH242-524K4
C
D
CH242-158O5
SLC5A4
IGLV5-22
IGLV8-21
IGLV1-20
IGLV8-19
IGLV8-18
IGLV5-17
IGLV8-16
IGLV1-15
IGLV5-14
IGLV8-13
IGLV1-12
IGLV5-11
IGLV8-10
IGLV7-9
IGLV(III)-8
IGLV7-7
ZNF280B
PRAME
IGLV2-6
IGLV3-5
IGLV3-4
IGLV3-3
IGLV3-2
IGLV3-1
IGLJ1
IGLC1
IGLJ2
IGLC2
IGLJ3
IGLC3
IGLJ4
Telomere
50 kb
Centromere
C.
CH242-288F14
SLC5A4
IGLV1-20
IGLV8-19
IGLV5-22
IGLV8-21
50 kb
D.
12.1kb
0.5kb
1.3kb
1.3kb
1.3kb
1.3kb
1.3kb
IGLV3-1
IGLJ1
IGLC1
IGLJ2
IGLC2
IGLJ3
IGLC3
IGLJ4
1 kb

Figure 3.2 Phylogenetic analysis of porcine (*Sus scrofa domestica*) IGLV nucleotide sequences using UPGMA with 1000 bootstrap iterations. Genes comprising the most highly expressed subgroup, IGLV8, are very tightly clustered, suggestive of a relatively recent expansion. Nodes are labeled with bootstrap values

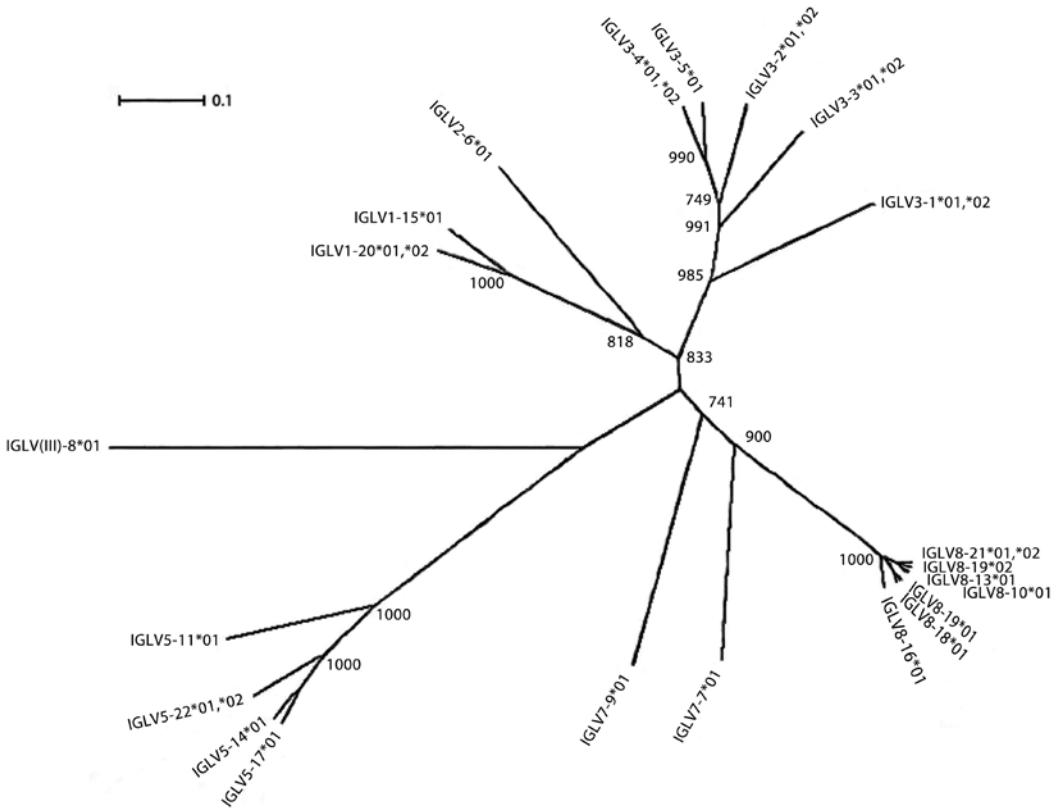
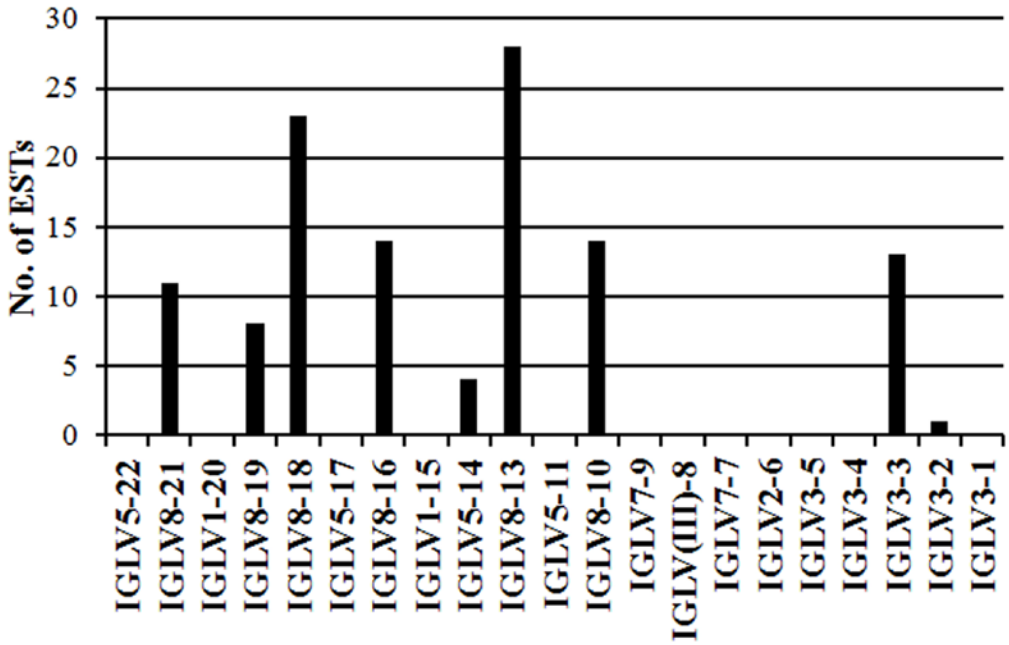


Figure 3.3 Expression of porcine (*Sus scrofa domestica*) IGLV genes. BLAST analysis revealed 116 hits to germline IGLV sequences from 398,837 expressed sequence tags in GenBank and deposited at http://pigest.ku.dk/index.html (Gorodkin 2007). The x-axis is organized by gene order in the locus with the most 5' (most C-distal) IGLV genes on the left. Genes belonging to the IGLV8 subgroup dominate the expression profile



DISCUSSION

The porcine immunoglobulin lambda (IGL) locus organization is typical among mammalian species, containing multiple V genes followed by tandem J-C cassettes. Among described species, it is most similar in organization to, but is more compact than cattle, another member of the order Cetartiodactyla. The capacity for recombinatorial diversity is significantly limited since more than half of the V genes are pseudogenes or ORF and one of four J genes lacks a paired C gene. The lack of IGLC4 downstream of IGLJ4 and a non-canonical heptamer and mutated W/F-G-X-G motif associated with IGLJ1 suggest that swine use only two J-C cassettes (IGLJ2-IGLC2 and IGLJ3-IGLC3). The only porcine lambda cDNA previously described is identical to IGLJ3 (Lammers et al. 1991).

Screening of EST databases indicates that the vast majority of expressed IGLV genes belong to the IGLV8 subgroup, followed by IGLV3 and IGLV5. The IGLV8 subgroup is the result of a recent expansion, presumably by gene duplication, suggesting that its products have provided a strong selective advantage. Our EST analysis differs from an earlier report which showed that the C-proximal IGLV3 is the most highly expressed IGLV subgroup in the porcine pre-immune repertoire (Butler et al. 2006). The preferential use of C-proximal IGLV2 and IGLV3 subgroup genes has been reported in humans (Frippiat et al. 1995). Similarly, C-proximal IGHV gene preference has been reported in mice (Yancopoulos et al. 1984; Malynn et al. 1990), rabbits (Knight and Becker 1990), and humans (Berman et al. 1991). However, in mice, IGHV gene usage

shifts upstream with animal age, possibly due to an increased presence of peripheral B cells bearing more distal genes as a result of environmental stimulation, since the number of pre-B cells possessing C-proximal rearrangements remains unchanged in the bone marrow of adult mice (Malynn et al. 1990; Jung et al. 2006). It is plausible that swine may similarly shift IGLV usage upstream with age and peripheral antigenic stimulation. In addition, several IGLV3 genes containing canonical RS were not detected in the EST database, due possibly to poor reactivity to exogenous antigen so that the B cells expressing these genes are rare, or the B cells may have been deleted due to reactivity to self-antigens.

The size of the expressed repertoire may be increased by allelic variation. Polymorphisms, comprising from 2 to 7 amino acid differences, were identified in 4 of the 8 IGLV genes that were present in overlapping BACs from the single animal used to construct the BAC library. Thus, allelic variation in swine V regions may significantly increase the complexity of the lambda repertoire.

The two most C-proximal human IGLV clusters are dominated by 34 IGLV genes largely belonging to IGLV2 and IGLV3. The reduced complexity of C-proximal IGLV in pigs compared to cattle suggests that this cluster has undergone contraction and expansion during the divergent evolution of these two species. The third human IGLV cluster is dominated by IGLV1, IGLV5, and IGLV7 which correspond to the swine IGLV genes of the distal cluster. IGLV8, however, only occurs as a single gene in the most distal region of the third human IGLV cluster, suggesting that significant gene order reshuffling as well as expansion and/or contraction has occurred specifically during suid

evolution. Thus, the immunogenetic loci may represent ideal regions for identifying intra-species gene order and gene content polymorphisms.

CHAPTER IV

THE EXPRESSED PORCINE ANTIBODY LIGHT CHAIN REPERTOIRE AND RESPONSE TO PRRSV INFECTION

In preparation for *The Journal of Immunology*

Antibody responses are fundamentally important to effector and memory mechanisms of disease resistance. Knowledge of antibody repertoire diversity is a prerequisite for the elucidation of protective immunity to viral infections. Therefore, the antibody light chain repertoire diversity was characterized in swine and marked, specific changes were observed following infection of pigs with the major swine pathogen, porcine reproductive and respiratory syndrome virus (PRRSV). Deep sequencing of >516,000 light chain VJ variable regions showed that, similar to humans, swine utilize both lambda and kappa loci equivalently, but V and J gene segment usage was highly restricted in both loci. Greater than 99% of lambda light chains were IGLV3 and IGLV8 family members, and 100% of kappa locus transcripts were IGKV1 and IGKV2. J region usage was restricted to IGLJ2 and IGLJ3 in the lambda locus, and >95% IGKJ2 in the kappa locus. Nevertheless, total diversity richness estimates were 2.3×10^5 for lambda and 1.5×10^5 for kappa. A striking example of lambda locus allelic variation also was noted; 4 of 5 pigs possessed a previously undescribed functional gene segment (IGLV3-1-1). Infection by PRRSV reduced total richness due to expression of several highly abundant clonal populations. Importantly, functional antibody repertoires differed substantially among individuals, thus illustrating the variation in immune response in outbred populations that is a challenge to development of broadly effective vaccines.

INTRODUCTION

Lymphocyte receptor diversity is generated from multiple genetic loci containing many copies of similar genes that can recombine to yield an astronomical array of potential antigen-binding structures. The antigen-binding domain of the B cell receptor is encoded by a variable (V) gene, a diversity (D) gene (heavy chain-only) and a joining (J) gene. Within this variable region exist three hypervariable or complementarity-determining regions (CDRs) which may differ considerably in sequence between genes and which constitute the antigen-binding loops of the antibody molecule (Wu and Kabat 1970; Lefranc and Lefranc 2001). In pigs, these genes are organized in three loci, a heavy chain locus and two light chain (kappa and lambda) loci and are found on separate chromosomes (7, 3, and 14, respectively) (Yerle et al. 1997).

Recently, the organization of the kappa and lambda immunoglobulin loci was characterized allowing for more detailed analyses of their expression and diversity (Schwartz et al. 2012a, 2012b). To date, there have been fragmentary investigations of the expressed porcine immunoglobulin light chain repertoire (Butler et al. 2005, 2009a; Lammers et al. 1991). The present study attempts to fill this gap via deep sequencing of an antibody light chain amplicon library generated from five pigs. Approximately 516,000 pyrosequencing reads were generated and analyzed for their diversity and richness. Analyses reveal the existence of a novel immunoglobulin lambda V (IGLV) gene segment that exists as either a functional allele or a non-functional one that is highly expressed in animals possessing a functional copy. Changes in the abundance distribution

of antibody sequences to favor several highly abundant clones was also observed during experimental infection with porcine reproductive and respiratory syndrome virus (PRRSV), suggestive of clonally-amplified PRRSV-specific antibody light chain sequences.

MATERIALS AND METHODS

Sample preparation

Lymphoid tissues (spleen, palatine tonsil, inguinal lymph node, and bronchial lymph node) collected from five animals in a previous study (Klinge et al. 2009) were utilized in this study. All five animals were of commercial source with genetically similar backgrounds and of the same age at time of challenge (approximately 3 weeks of age). To identify pathogen-associated light chain transcripts, three pigs were infected with PRRSV strain JA142 and the remaining two were sham-inoculated at three weeks of age. All five animals were euthanized at 62 days post-inoculation, at which point tissues were harvested and stored in RNAlater. Tissues were subsequently homogenized and total RNA extracted using the RNeasy Mini Kit (Qiagen) and reverse transcribed using the QuantiTect Reverse Transcription Kit (Qiagen). Genomic DNA was separately extracted using the QIAamp DNA Mini Kit (Qiagen).

Library generation and sequencing

Reverse primers were designed against the conserved constant region exon sequences using Primer3 set to default parameters (Rozen and Skaletsky 2000). This eased our approach as the kappa locus contains a single constant region exon (IGKC) and the lambda locus contains three nearly identical constant region exons. V gene family-specific forward primers were developed using the conserved leader region sequences derived from previous annotations of the porcine kappa and lambda loci (Schwartz et al. 2012a, 2012b) (Table 4.1). The Roche 454 Titanium fusion adapter A (5'-

CGTATCGCCTCCCTCGCGCCATCAG) was attached to the 5' end of each forward primer and fusion adapter B (5'-CTATGCGCCTTGCCAGCCCGCTCAG) was attached to the 5' end of each reverse primer. To differentiate between animals, a 10 base molecular ID tag was added to the forward primer between the fusion adapter and the template-specific sequence. Each PCR was performed separately in order to reduce any potential bias. Products were pooled according to pig, quantified by spectrophotometry, and pooled in equimolar amounts so that the final library contained approximately the same number of amplicons from each animal. This product was then both gel purified and column purified using the QIAquick PCR purification kit (Qiagen). The sample was sequenced at the W. M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champagne. Roche Titanium 454 pyrosequencing was performed from the forward fusion adapter A so that all of the reads were of the same orientation and more likely to span the variable region.

Sequence analysis

Pyrosequencing results were imported into the database software FileMaker Pro 11 for parsing. Reads were translated into putative amino acid sequences using EMBOSS within the Galaxy framework (Rice et al. 2000; Blankenberg et al. 2007). Short reads (< 350 bp), reads with a missing or aberrant molecular ID tag, and reads containing stop codons and/or ambiguous amino acids were excluded from further analyses. IGKV and IGLV gene segment usage was determined using BLAST (Altschul et al. 1990). The molecular ID tag as well as IGKJ and IGLJ gene segment usage were determined by identifying conserved nucleotide sequences and annotating within FileMaker. For

diversity analysis, amino acid sequences were first clustered using cd-hit at an identity threshold of 100% (Li and Godzik 2006). The clustered results were used to generate abundance curves within Excel. Cd-hit output and Chao1 were used to calculate lower-bound estimates of repertoire richness (Colwell and Coddington 1994; Chao 1984; Chao and Lee 1992).

RESULTS

Sequencing results

A total of 516,097 pyrosequencing reads was obtained with a mean length of approximately 510 bases and a median length of 521 bases. Of these, 79,172 reads that that did not span the entire variable region (< 350 bases) were excluded from downstream analyses. An additional 11,739 reads with aberrant molecular ID tags and 53,046 reads containing premature stop codons and/or ambiguous nucleotides were also excluded from downstream analyses. Investigation of these latter reads revealed that many were frame-shifted in homopolymer regions, suggesting that they are artifacts of 454 pyrosequencing, a known drawback of the platform (Shendure and Hanlee 2008). Thus, a total of 372,140 full-length, in-frame reads were analyzed for V and J gene segment usage. The percentage of reads from individual pigs were approximately equal (pig 1, 19.0%; pig 2, 16.9%; pig 3, 20.3%; pig 4, 20.0%; pig 5, 23.8%). Furthermore, the number of reads obtained from kappa and lambda-containing transcripts were similar (42.9% vs. 57.1%, respectively). As no attempt was made to equalize these during library preparation, this ratio may reflect the actual light chain usage in swine as it is similar to the 1:1 ratio previously reported (Hood et al. 1967; Butler et al. 2004).

Kappa locus gene expression

A total of 159,721 reads covering the entire variable region were obtained corresponding to the immunoglobulin kappa locus. According to BLAST analysis, 101,714 (64%) of these matched most closely with genes from the IGKV1 family and 58,007 matched most closely to members of the IGKV2 family (Figure 4.1a). This is in

contrast to previous reports which suggest that IGKV2 family members are more abundant, at least among pre-immune piglets (Butler et al. 2004; Schwartz et al. 2012a). Similar to previous reports, however, IGKJ2 accounted for the vast majority of J gene usage (Figure 4.1a). In contrast, IGKJ3 and IGKJ5 were unexpressed while IGKJ1 and IGKJ4 were expressed at very low levels (each approximately 1% of IGKJ usage).

In order to characterize gene segment usage, IGKV sequence identities were compared to the annotated gene segments. A distribution of reads of varying putative amino acid identity was observed such that the median number of mismatches varied from 3 (IGKV2-6) to 10 (IGKV1-11) (Figure 4.2a). There were no major differences noted between individual pigs in terms of IGKV gene segment usage. Of the nine putatively functional IGKV gene segments, two were unexpressed (IGKV1-7 and IGKV1-14) in all five pigs.

Lambda locus gene expression

A total of 212,419 full-length lambda variable region reads were obtained. IGLV gene BLAST results were very similar to what we previously reported using an EST database (Schwartz et al. 2012b) (Figure 4.1b). IGLV8 family members were dominantly expressed (67.6%), followed by IGLV3 family members (32.4%). We also confirmed the low level expression of IGLV5-14, as we detected a total of 71 reads from this gene segment. As predicted by the annotation, neither IGLJ1 nor IGLJ4 were found to be utilized. Additionally, usage of IGLJ2 is slightly favored over the more V-distal IGLJ3 cassette (59% vs. 41%, respectively) (Figure 4.1b). Similar to what was found with regard to the kappa locus, there was no association between specific IGLV and IGLJ

gene segments. A single, otherwise apparently functional gene segment (IGLV2-6), showed no expression in any of the five animals. This is despite the use of a conserved, specific forward primer directed against the IGLV2 leader region (Table 4.1) and post-PCR gel purification of the expected size range and inclusion in the sequenced pool.

Characterization of the novel IGLV gene segment, IGLV3-1-1

Sequenced IGL transcripts typically contained three to six putative amino acid mismatches on average compared to the annotated IGLV genes (Figure 4.3b). Interestingly, a large number of reads matching most closely to IGLV3-4*01 contained about 20 mismatches (not shown). Analysis of representative reads from this group revealed that they match most closely with an unannotated truncated pseudogene that is flanked by IGLV3-1 and IGLV3-2 and in the same orientation as the rest of the locus (Table 4.2). The pseudogenic allele (Hereafter: IGLV3-1-1*01) of this gene segment is present on the bacterial artificial chromosome (BAC) 242-141B5 (GenBank: CU468977.2) (Schwartz et al. 2012b). The 5' end of IGLV3-1-1*01 is deleted such that it becomes recognizable starting in framework region 2. IGLV3-1-1*01 is positioned approximately 182 bp downstream from the recombination signal sequence of IGLV3-2 (Figure 4.4). PCR, using forward primers directed against either the CDR-L2 of IGLV3-2 or the conserved IGLV3 leader sequence and a reverse primer directed against the CDR-L2 of IGLV3-1-1*01, and Sanger sequencing identified both the pseudogene, IGLV3-1-1*01 (180 bp), and its functional variant, IGLV3-1-1*02 (296 bp) in all four animals expressing the functional allele (data not shown). No transcripts from this gene segment were produced by pig 1(Figure 4.3b, Figure 4.5). In addition, neither set of primers

amplified product specific to either of the alleles from this animal, suggesting that IGLV3-1-1 is entirely absent from pig 1. Interestingly, among the C-proximal IGLV3 gene segment family members, IGLV3-3 is almost exclusively expressed by this pig (Figure 4.3b).

While we identified IGLV3-1-1*01 on the BAC CH242-141B5, we sought to see if it was present on the other annotated BAC, CH242-524K4. (GenBank: CU467599.3). Using full-length cDNA sequence we identified a small 381 bp contig generated from our previous Illumina re-sequencing of this BAC which contains the entire functional allele of IGLV3-1-1*02. Given the contig's small size and thus lack of genomic context, the contig likely assembled erroneously to the IGLV3-4 region, given the high degree of similarity between these regions (Schwartz et al. 2012b).

Characterization of individual pig gene segment usage revealed that pig 2 does not express the prototypic functional allele of IGLV3-1-1 (*02). Instead, pig 2 expresses a range of transcripts that are most similar to, yet are 8-12 percent different from IGLV3-1-1*02, suggesting the existence of either a novel gene segment, or more likely another allele of IGLV3-1-1, which will be referred to as IGLV3-1-1*03 (Figure 4.5, Table 4.2).

CDR-L3 characterization

As expected due to the known lack of TdT expression during light chain rearrangement, we found that light chain CDR-L3 lengths largely correspond to the known CDR-L3 lengths encoded by the V and J gene segments (Figure 4.6). Interestingly, however, we detected a distribution of functional transcripts with CDR-L3 lengths that ranged from 4 to 16 amino acids in length for kappa and 4 to 23 amino acids

in length for lambda, despite a germline restriction for 9 amino acids (IGK) and 10 to 11 amino acids (IGL). Approximately one-eighth (12.5%) of all IGK CDR-L3s and one-third (33.5%) of all IGL CDR-L3s differed from the expected lengths.

Diversity and richness of the porcine immunoglobulin light chain repertoire

Using putative amino acid sequences for the full length V-J region from each animal, we obtained lower bound richness estimates of 1.6×10^5 to 2.8×10^5 molecules for the lambda repertoire and 1.1×10^5 to 1.7×10^5 molecules for the kappa repertoire (Figure 4.7). Abundance distributions for these reads reveal a power-law distribution where the vast majority of sequences are exceptionally rare while a small number are very common, similar to what has been reported in zebra fish (Weinstein et al. 2009) (Figure 4.8).

The anti-PRRSV response

The three pigs infected with PRRSV all appear to have shifted their light chain abundance distributions such that they favor fewer, yet more abundant light chain sequences. Several of these clonal populations particularly dominate - accounting for between 1-3 percent of their respective repertoires (Figure 4.8). This analysis suggests that these highly abundant clones are PRRSV-specific. Our data also reveal that many of the most abundant clones are not shared between pigs, even between PRRSV-infected pigs. IGLV8-16 was highly abundant among pigs 1-4, although the amino acid sequences in pigs 3 and 4 differed from pigs 1 and 2 (Tables 4.3 and 4.4). This was the only sequence shared between PRRSV-infected animals. IGLV3-3 dominated the expressed lambda repertoire of pig 1 and a single sequence from this gene segment was also highly

abundant in pigs 2 and 3, but not the PRRSV-uninfected pigs (Tables 4.3 and 4.4). Among kappa transcripts, many of the highly abundant clones in the infected animals were also highly abundant in uninfected pigs, although the their proportional usage was generally much higher in the infected animals (Tables 4.5 and 4.6). These data suggest that at least for light chains, the more frequently used gene segments are more likely to be involved in a primary immune response. Although, much of this diversity has most likely arisen through somatic hypermutation and recombinatorial error (i.e. junctional diversity), our data here seem to indicate that allelic variation between individual animals substantially impacts the expression of the antibody repertoire.

Table 4.1 Primers used in the current study.

Target	Template-specific sequence
IGKV1 (F)	5'-GCCTCYTGCTGCTCTGG
IGKV2 (F)	5'-TTCCCTGCTCAGCTCCTG
IGLV2 (F)	5'-CCTTGTCACCCTCCTCACTC
IGLV3 (F)	5'-CTGGAYCCCTCTCCTGCTC
IGLV5 (F)	5'-GGACTCCTCTGCTGATCGTG
IGLV8 (F)	5'-CCTGGACGGTGCTTCTGAT
IGLV8-21 (F)	5'-CTTCTGCTTTGCTCCTCGTT
IGKC (R)	5'-GACCACCCCATCCACTTTC
IGLC (R)	5'-CCTTCCAGGTCACCGTCA

F, forward; *R*, reverse

Table 4.2 Deduced amino acid sequences of the novel functional alleles of IGLV3-1-1. Shown according to IMGT unique numbering (Lefranc et al. 2003)

```
                                FR1-IMGT          CDR1-IMGT        FR2-IMGT       CDR2-IMGT                 FR3-IMGT                  CDR3-IMGT
                                 (1-26)             (27-38)         (39-55)        (56-65)                   (66-104)                  (105-117)
                           A              B            BC          C         C'        C'C"        C"        D          E          F          FG
                        (1-15)         (16-26)      (27-38)    (39-46)   (47-55)    (56-65)    (66-74)   (75-84)    (85-96)    (97-104)
                        ——————————————> ——————————>              ———————> ————————>            ————————> —————————> ———————————> ———————>
                        1            15 16       26 27        38 39    46 47      55 56      65 66     74 75      84 85        96 97   104 105
                        |.............| |.........| |..........| |......| |.......| |........| |.......| |........| |..........| |......| |.......|
IGLV3-1-1*02            SSQLTQVPG.VSVSL GGTASIACLGD NFV.......YA ANWYQQKP GQAPILVIY GG.......S LRPLGIP.E RFSGSS..SG NSATLTISGAQA EDEADYYC QSADSSDNA
IGLV3-1-1*03            SSQLTQPPG.VSVSL GGTASIACLGD NFI.......YA ANWYQQKP GQAPILVIY DG.......S RRPSGIP.E RFSGSS..SG KTATLTITGAQA EDEADYYC QSADSSENA
```

Table 4.3 Frequency of the five most abundant lambda light chain sequences in five individual pigs. Actual amino acid sequences (coded in column 3) are presented in table 4.4

	Abundance Rank	Sequence	V	J	aa mismatches (V region)	Percent of repertoire by pig				
						1	2	3	4	5
IGL Pig 1	1	A	IGLV8-16	IGLJ3	1	3.07	0.42	4.5E-03	0	0
	2	B	IGLV3-3	IGLJ3	0	2.65	0.32	0.20	0.11	4.9E-02
	3	C	IGLV3-3	IGLJ3	2	1.72	0.17	4.5E-03	2.5E-03	0
	4	D	IGLV3-3	IGLJ2	0	0.86	0.20	8.9E-03	5.1E-03	1.4E-02
	5	E	IGLV8-10	IGLJ2	0	0.77	0.25	0.25	0.07	0.26
IGL Pig 2	1	F	IGLV3-3	IGLJ2	1	0	1.07	0	2.5E-03	0
	2	G	IGLV8-18	IGLJ3	6	2.6E-03	0.42	0	0	0
	3	A	IGLV8-16	IGLJ3	1	3.07	0.42	4.5E-03	0	0
	4	H	IGLV8-19	IGLJ3	8	0	0.39	0	0	0
	5	I	IGLV8-18	IGLJ3	7	0	0.34	2.2E-03	0	0
IGL Pig 3	1	J	IGLV8-10	IGLJ2	3	0	0	0.80	0	0
	2	K	IGLV8-16	IGLJ3	0	0	0.01	0.73	0	0
	3	L	IGLV3-1-1	IGLJ3	1	0	0	0.49	0.28	0.17
	4	M	IGLV3-1-1	IGLJ2	0	0	0	0.40	0.26	0.85
	5	N	IGLV3-3	IGLJ2	2	0	0	0.36	0	0
IGL Pig 4	1	O	IGLV8-13	IGLJ3	3	0	0	0	0.56	0
	2	L	IGLV3-1-1	IGLJ3	1	0	0	0.49	0.28	0.17
	3	P	IGLV8-16	IGLJ3	1	0	0	0	0.28	0
	4	Q	IGLV3-1-1	IGLJ3	0	0	0	0.20	0.27	0.09
	5	M	IGLV3-1-1	IGLJ2	0	0	0	0.40	0.26	0.85
IGL Pig 5	1	M	IGLV3-1-1	IGLJ2	0	0	0	0.40	0.26	0.85
	2	R	IGLV3-5	IGLJ2	0	0.08	0.15	0.11	0.21	0.65
	3	S	IGLV3-5	IGLJ2	2	0	0	0	0.17	0.47
	4	T	IGLV3-1-1	IGLJ2	0	0	0	0.20	0.11	0.46
	5	E	IGLV8-10	IGLJ2	0	0.77	0.25	0.20	0.07	0.26

Table 4.4 Most abundant lambda light chain sequences as described in previous table (Table 4.3). Shown according to IMGT unique numbering (Lefranc et al. 2003)

	FR1-IMGT (1-26)			CDR1-IMGT (27-38)		FR2-IMGT (39-55)		CDR2-IMGT (56-65)			FR3-IMGT (66-104)						CDR3-IMGT (105-117)	FR4 (J region)
	A (1-15)		B (16-26)		BC (27-38)	C (39-46)	C' (47-55)	C'C" (56-65)		C" (66-74)	D (75-84)		E (85-96)		F (97-104)	FG		
	1	15	16	26	27 38	39 46	47 55	56	65	66 74	75	84	85	96	97 104	105	113	
A	SQTVIQEPA	.MSVSL	GGTVTLTCAFS		SGSVT...SSNY	PSWYQQTP	GQPPRQLIY	ST.....N		SRPTGVP.S	RFSGAI..SG		NKATLTITGAQA		EDEADYFC	ILYKGSGTY..		TPFGGGTHLTVL
B	SYELTQPSS	.ESVAL	GNTAKITCSGD		LLD.....EKY	TQWYQQKP	GQAPLLLIY	KD.....S		ERPSGIP.E	RFSGSS..SG		KTATLTITGAQA		EDEADYYC	QSADSIDN...		VPFGGGTHLTVL
C	SYELTQPSS	.ESVAL	GNTAKITCSGD		LLD.....EKY	TQWYQQKP	GQAPLLLIY	KD.....S		ERPSGIP.E	RFSGSS..SG		KTATLTITGAQA		EDEADYYC	QSTNSIDN...		VPFGGGTHLTVL
D	SYELTQPSS	.ESVAL	GNTAKITCSGD		LLD.....EKY	TQWYQQKP	GQAPLLLIY	KD.....S		ERPSGIP.E	RFSGSS..SG		KTATLTITGAQA		EDEADYYC	QSADSID....		NIFGGGTHLTVL
E	SQTVIQEPA	.MSVSP	GGTVTLTCAFS		SGSVT...TSNY	PSWFQQTP	GQPPRLLIY	RT.....N		NRPTGVP.S	RFSGAI..SG		NKAALTITGAQA		NDEADYFC	TLYKSSA....		NIFGGGTHLTVL
F	SYELTQPSS	.ESVAL	GNTAKITCSGD		LLD.....EKY	TQWYQQKP	GQAPLLLIY	KD.....S		ERPSGIP.E	RFSGSS..SG		KTATLTITGAQA		EDEADYYC	QSADSIEDA..		IIFGGGTHLTVL
G	SQTVIQEPA	.MSVSL	GGTVTLTCAFS		SGSVT...TSNY	PSWFQQTP	GQPPRTVIY	NT.....N		SRPTGVP.S	RFSGAI..SG		NKAALTITGAQA		EDEADYFC	ALSKSCSN...		VPFGGGTHLTVL
H	SQTVIQEPA	.MSVSL	GGTVTLTCAFS		SGSVT...SSNY	PSWYQQTP	GQPPRQLIY	RA.....N		VRPTGVP.G	RFSGAI..SE		NKAALTITGAQA		EDEADYFC	ALYKSGTNG..		VPFGGGTHLTVL
I	SQTVIQEPA	.MSVSL	GGTVTLTCAFS		SGSVT...TSNY	PSWFQQIS	GQPPRTVIY	NT.....N		SRPTGVP.S	RFSGAI..SG		NKAALTITGAQA		EDEADYFC	ALSKSCTN...		VPFGGGTHLTVL
J	SQTVIQEPA	.MSVSP	GGTVTLTCAFS		SGSVT...TSNY	PGWFQQTP	GQPPRLLIY	RT.....I		NRPTGVP.S	RFSGAI..SG		NKAALTITGAQA		NDEADYFC	TLYKSSTN...		AIFGGGTHLTVL
K	SQTVIQEPA	.MSVSL	GGTVTLTCAFS		SGSVT...SSNY	PSWYQQTP	GQPPRQLIY	ST.....N		SRPTGVP.S	RFSGAI..SG		NKATLTITGAQA		EDEADYFC	ALYKGSGN...		VPFGGGTHLTVL
L	SSQLTQVPG	.VSVSL	GGTASIACLGD		NFV.....YA	ANWYQQKP	GQAPILVIY	GG.....S		LRPLGIP.E	RFSGSS..SG		NSATLTISGAQA		EDEADYYC	QSADSSDN...		VPFGGGTHLTVL
M	SSQLTQVPG	.VSVSL	GGTASIACLGD		NFV.....YA	ANWYQQKP	GQAPILVIY	GG.....S		LRPLGIP.E	RFSGSS..SG		NSATLTISGAQA		EDEADYYC	QSADSSDN...		AIFGGGTHLTVL
N	SYELTQPSS	.ESVAL	GNTAKITCSGD		LLD.....EKY	TRWYQQKP	GQAPLLLIY	KD.....S		ERPSGIP.E	RFSGSS..SG		KRATLTITGAQA		EDEADYYC	QSPDSSKR...		ATFGGGTHLTVF
O	SQTVIQEPA	.MSVSP	GGTVTLTCAFS		SGSVT...TSNY	PSWFQQTP	GQPPRQLIY	ST.....D		NRPTGVP.S	RFSGAI..SG		NTAALTITGAQA		EDEADYFC	ALYKSSGN...		APFGGGTHLTVL
P	SQTVIQEPA	.MSVSL	GGTVTLTCAFS		SGSVT...SNNY	PSWYQQTP	GQPPRQLIY	ST.....N		SRPTGVP.S	RFSGAI..SG		NKATLTITGAQA		EDEADYFC	ALYKGSGTAAE		VPFGGGTHLTVL
Q	SSQLTQVPG	.VSVSL	GGTASIACLGD		NFV.....YA	ANWYQQKP	GQAPILVIY	GG.....S		LRPLGIP.E	RFSGSS..SG		NSATLTISGAQA		EDEADYYC	QSADSSDN...		APFGGGTHLTVL
R	SSKLTQPPG	.VSVSL	GGTASITCQGA		NFG.....SYY	AHWYQQKP	GQSPELVIY	EY.....P		EIFLGFL.E	RFSVSR..TG		DTATLTISGAQA		EDEADYYC	QVYDGGYH...		VIFGGGTHLTVL
S	SSQLTQPPG	.VSVSL	GGTASITCQGA		NFG.....SYY	AHWYKQKP	GQSPELVIY	EY.....P		EIFLGFL.E	RFSVSR..TG		DTATLTISGAQA		EDEADYYC	QVYDGGYH...		VIFGGGTHLTVL
T	SQTVIQEPA	.MSVSP	GGTVTLTCAFS		SGSVT...TSNY	PSWFQQTP	GQPPRLLIY	RT.....N		NRPTGVP.S	RFSGAI..SG		NKAALTITGAQA		NDEADYFC	TLYKSSA....		NIFGGGTHLTVL

Table 4.5 Frequency of the five most abundant kappa light chain sequences in five individual pigs. Actual amino acid sequences (coded in column 3) are presented in table 4.6

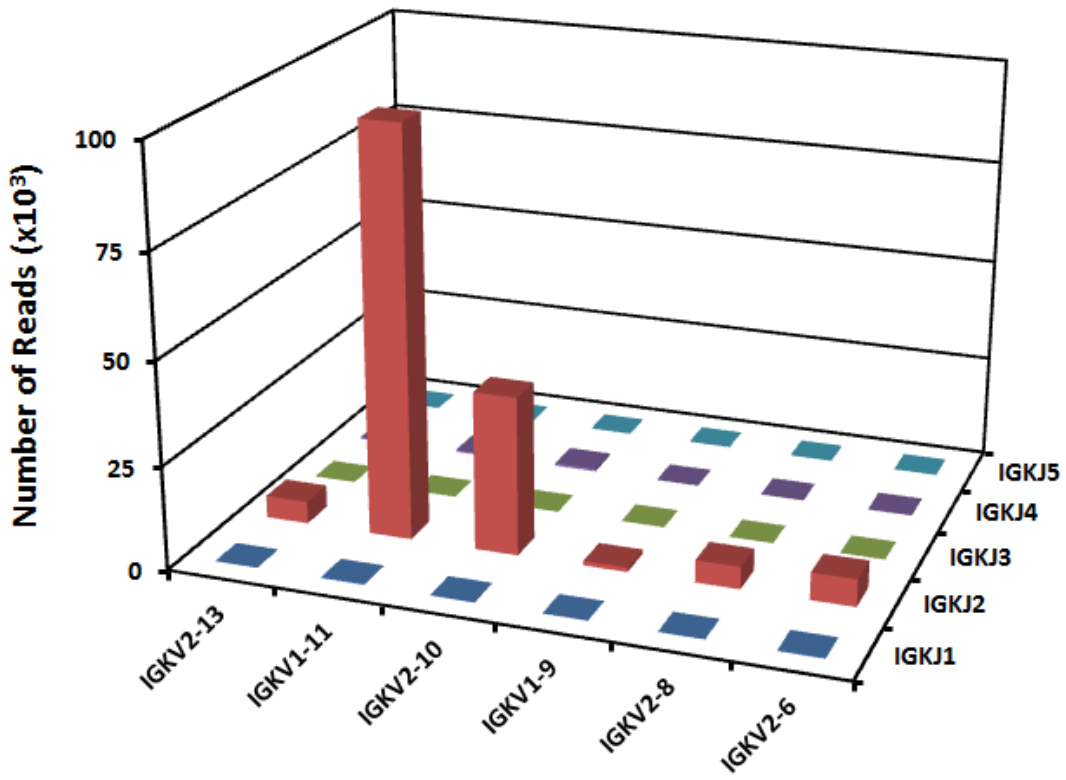
	Abund. Rank	Sequence	V	J	aa mismatches (V region)	Percent of repertoire by pig				
						1	2	3	4	5
	1	**A**	IGKV1-11	IGKJ2	1	2.43	1.71	0.52	0.01	1.11
IGK	**2**	**B**	IGKV2-10	IGKJ2	0	1.54	0.14	0.01	0	0
Pig 1	**3**	**C**	IGKV1-11	IGKJ2	2	0.79	0.07	0	0	0
	4	**D**	IGKV1-11	IGKJ2	0	0.78	0.01	0	0	0
	5	**E**	IGKV2-10	IGKJ2	0	0.81	1.35	0.70	0.34	0.57
	1	**A**	IGKV1-11	IGKJ2	1	2.43	1.71	0.52	0.01	1.11
IGK	**2**	**E**	IGKV2-10	IGKJ2	6	0.81	1.35	0.70	0.34	0.57
Pig 2	**3**	**F**	IGKV2-8	IGKJ2	1	0.07	0.39	0.1	0.08	0.11
	4	**G**	IGKV1-11	IGKJ2	8	0.47	0.37	0.16	0.04	0.22
	5	**H**	IGKV2-8	IGKJ2	7	0.23	0.31	0.11	0	0.27
	1	**E**	IGKV2-10	IGKJ2	3	0.81	1.35	0.70	0.34	0.57
IGK	**2**	**A**	IGKV1-11	IGKJ2	0	2.43	1.71	0.52	0.01	1.11
Pig 3	**3**	**I**	IGKV1-11	IGKJ2	1	0	0	0.41	0	0
	4	**J**	IGKV1-11	IGKJ2	0	0	0	0.34	0	0
	5	**K**	IGKV1-11	IGKJ2	2	0	0	0.31	0	0
	1	**L**	IGKV1-11	IGKJ2	3	0	0	0	0.99	0
IGK	**2**	**M**	IGKV1-11	IGKJ2	1	0	0	0.25	0.44	0.02
Pig 4	**3**	**N**	IGKV1-11	IGKJ2	1	0	0	0	0.34	0
	4	**O**	IGKV1-11	IGKJ2	0	0	0	0.22	0.34	0
	5	**E**	IGKV2-10	IGKJ2	0	0.81	1.35	0.70	0.34	0.57
	1	**A**	IGKV1-11	IGKJ2	0	2.43	1.71	0.52	0.01	1.11
IGK	**2**	**E**	IGKV2-10	IGKJ2	0	0.81	1.35	0.70	0.34	0.57
Pig 5	**3**	**P**	IGKV1-11	IGKJ2	2	0	0	0	0	0.35
	4	**Q**	IGKV2-10	IGKJ2	0	0.28	0.29	0.16	0.11	0.30
	5	**R**	IGKV1-11	IGKJ2	0	0	0	0.01	0	0.29

Table 4.6 Most abundant lambda light chain sequences as described in previous table (Table 4.5). Shown according to IMGT unique numbering (Lefranc et al. 2003)

	FR1-IMGT (1-26)			CDR1-IMGT (27-38)			FR2-IMGT (39-55)			CDR2-IMGT (56-65)			FR3-IMGT (66-104)			CDR3-IMGT (105-117)			FR4 (J region)
	A (1-15)		B (16-26)	BC (27-38)		C (39-46)	C' (47-55)		C'C" (56-65)		C" (66-74)	D (75-84)		E (85-96)	F (97-104)		FG			
	1	15	16 26	27	38	39 46	47	55	56	65	66 74	75	84	85	96	97 104	105	113		
A	AIQMTQSPASLAASL		GDTVSITCRAS	QSI.....SSY		LAWYQQQP	GKAPKLLIY		AA.....S		SLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		QQHNSAP....		YGFGAGTKLELK
B	AIVLTQTPLSLSVSP		GEPASISCRSS	QSLLHT.DGKNY		LNWYLQKP	GQSPQLLIY		YA.....T		IRDTGVP.D	RFTGSG..SG		TDFTLKISRVEA	GDVGVYYC		SQALQSP....		FGFGAGTKVELK
C	AIQMTQSPASLAASL		GDTVSITCRAS	QSI.....SNN		LAWYQQQA	GKAPKLLIY		AA.....S		SLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		QQHNSVP....		LGFGAGTKLELK
D	AIQMTQSPASLAASL		GDTVSITCRAS	QSI.....SSY		LAWYQQQA	GKAPKLLIY		WA.....S		VLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		QQHNSAP....		YGFGAGTKLELK
E	AIVLTQTPLSLSVSP		GEPASISCRSS	QSLLHT.DGKNY		LNWYLQKP	GQSPQLLIY		YA.....T		NRDTGVP.D	RFTGSG..SG		TDFTLKISRVEA	EDVGVYYC		FQALQSP....		YGFGAGTKLELK
F	AIVLTQTPLSLSVSP		GEPASISCRSS	QSLVDS.DGDSL		LHWYLQKP	GQSPRLLFY		FA.....T		NRASGVP.D	RFSGSG..SG		TDFTLKISRVEA	EDAGVYYC		QQNKESP....		YGFGAGTKLELK
G	AIQLTQSPASLAASL		GDTVSITCRAS	QSI.....SSY		LAWYQQQP	GKAPKLLIY		AA.....S		SLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		LQHSSAP....		YGFGAGTKLELK
H	AIVLTQSPLSLSVSP		GEPASISCRSS	QSLEI..YGSNL		LSWYQQKP	GQSPRLLIY		FA.....T		NKASGVP.D	RFSGSG..SG		TDFTLKISRVEA	EDAGVYYC		QQHKESP....		YGFGAGTKLELK
I	AIQMTQSPASLAASL		GDTVSITCRAS	QSI.....SSY		LTWYQQQP	GKAPKLLIH		AA.....S		TLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		QQYNSAP....		LGFGAGTKLELK
J	AIQLTQSPASLAASL		GDTVSITCRAS	QSI.....NKW		LAWYQQQP	GKAPKLLIY		TA.....S		TLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		FDYSSAP....		NGFGAGTKLELK
K	AIQLTQSPASLAASL		GDTVSITCRAS	QSI.....SSY		LGWYQQQP	GKAPKLLIY		AA.....S		SLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVASYYC		LQYNSAP....		YGFGAGTKLELK
L	AIQLTQSPASLAASL		GDTVSITCRAS	QSI.....SSY		LGWYQQQP	GKAPKLLIY		KA.....S		SLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		LQHNTTP....		YGFGAGTKLELK
M	AIQLTQSPASLAASL		GDTVSITCRAS	QSI.....SSY		LAWYQQQP	GKAPKLLIY		AA.....S		SLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		LQHSSAP....		YGFGAGTKLELK
N	AIQLTQSPASLAASL		GDTVSITCRAS	QSI.....SSN		LAWYQQQP	GKAPKLLIY		KA.....S		SLQTGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		FQHNSAP....		YGFGAGTKLELK
O	AIQLTQSPASLAASL		GDTVSITCRAS	QSI.....SSY		LGWYQQQP	GKAPKLLIY		AA.....S		SLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		LQHNSAP....		YGFGAGTKLELK
P	AIQMTQSPASLAASL		GDTVSITCRAS	QSI.....SSY		LAWYQQQP	GKAPKLLIY		DA.....S		SLQSGVP.S	RFKGSG..SG		TDFTLTISGLQA	EDVATYYC		QQHNSAP....		HSFGAGTKLELI
Q	AIVLTQSPLSLSVSP		GEPASISCRSS	QSLLHS.DGASL		LYWYQQKP	GQSPRLLIY		YA.....T		NRATGVP.D	RFTGSG..SG		TDFTLKISRVEA	EDVGVYYC		QQIIHSP....		YGFGAGTKLELK
R	AIQLTQSPASLAASL		GDTVSITCRAS	QTI.....SSY		LAWYQQQP	GKAPKLLIY		KA.....S		TLQSGVP.S	RFKGSG..SG		TDYTLTISGLQA	EDVATYYC		LQHSYAP....		YGFGAGTKLELK

Figure 4.1 Combinatorial light chain V-J usage in swine. (A) IGK usage. (B) IGL usage. Results pooled from five animals. The two horizontal axes enumerate the V (x-axis) and J (z-axis) gene segments and are ordered based on their genomic positions. Only V gene segments that were expressed are displayed on the x-axis. All J gene segments are displayed, however. There was no expression of IGKJ3, IGKJ5, IGLJ1, and IGLJ4.

A



Next Page

B

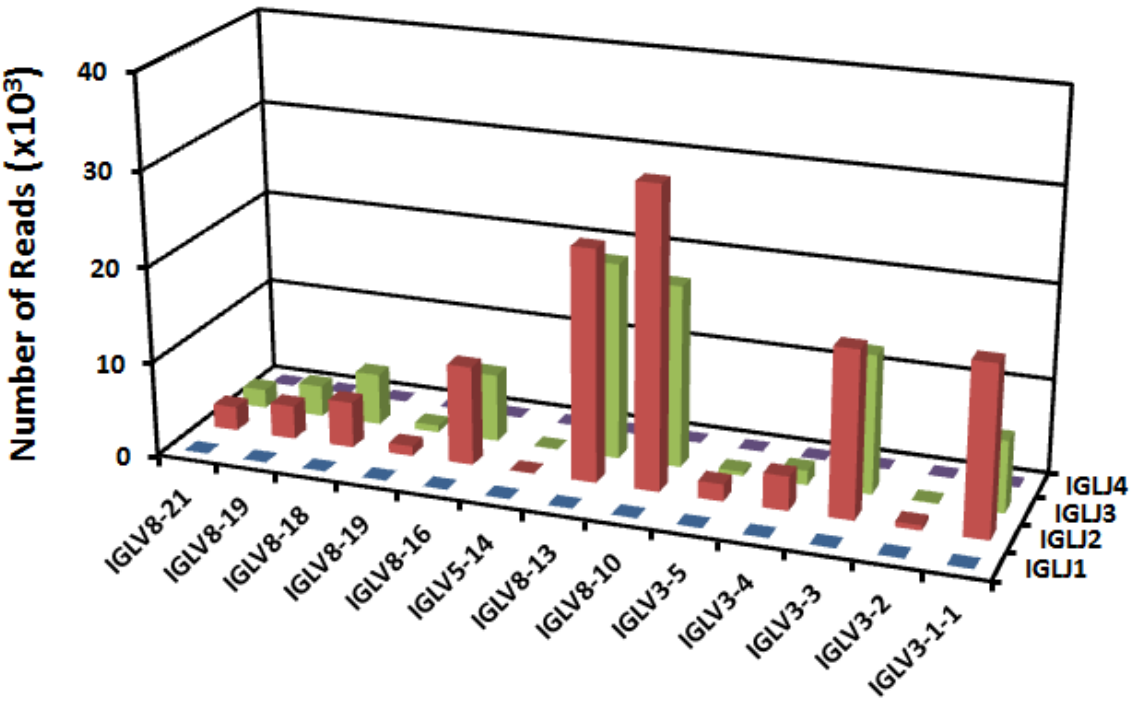
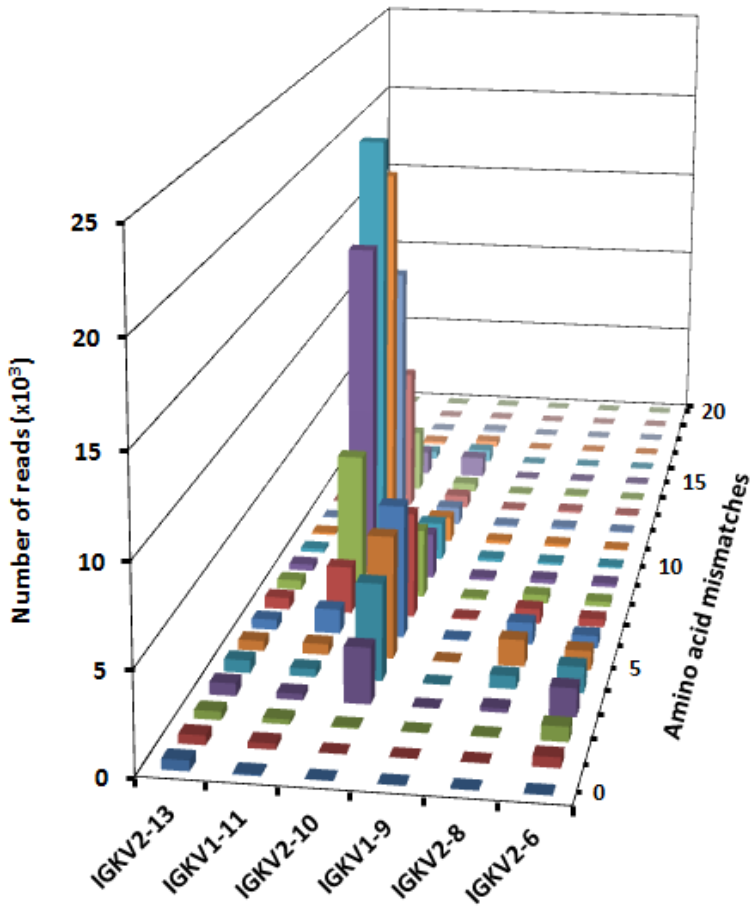

Number of Reads (x10³)
40
30
20
10
0
IGLV8-21
IGLV8-19
IGLV8-18
IGLV8-19
IGLV8-16
IGLV5-14
IGLV8-13
IGLV8-10
IGLV3-5
IGLV3-4
IGLV3-3
IGLV3-2
IGLV3-1-1
IGLJ4
IGLJ3
IGLJ2
IGLJ1

Figure 4.2 Number of amino acid mismatches in sequenced transcripts compared to the published annotations. (A) IGKV. (B) IGLV. Results pooled from five animals. The x-axis shows V gene segments that showed some level of expression and are arranged based on their genomic context.

A



Next Page

B

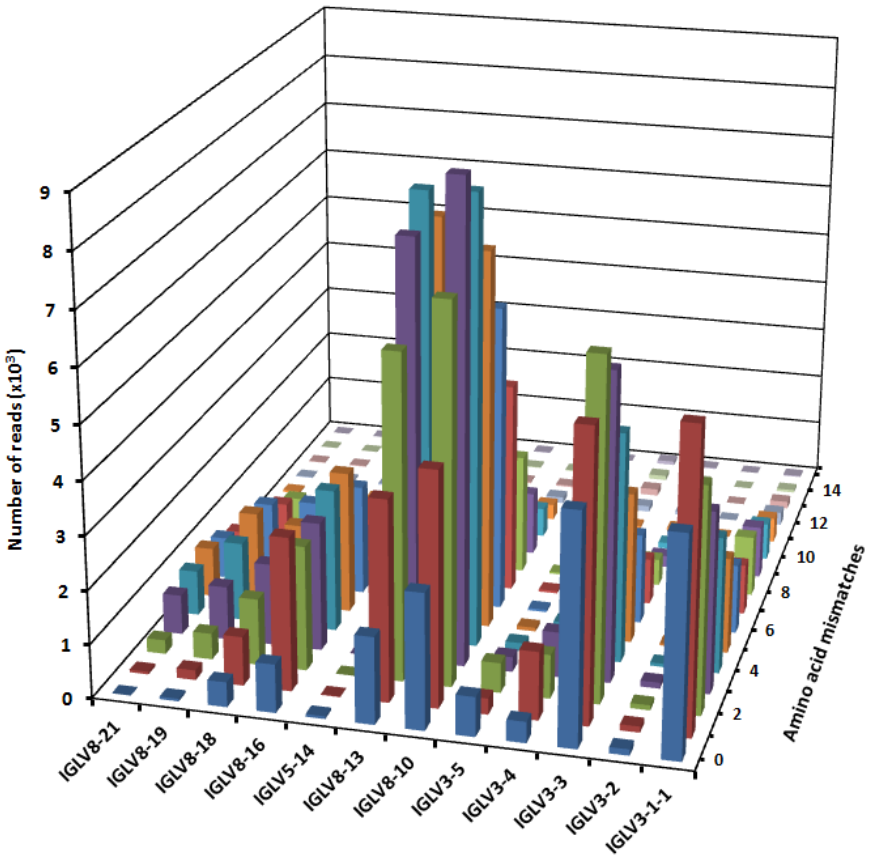
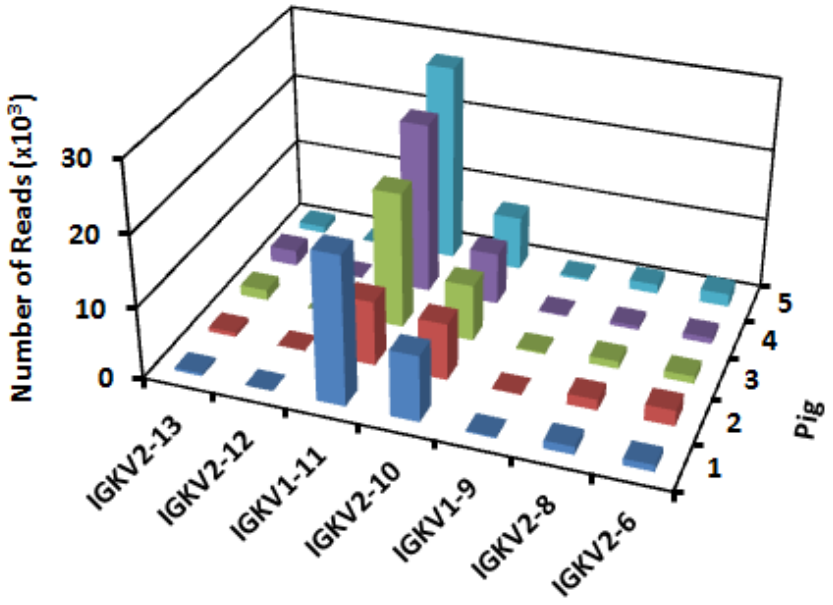

Number of reads (x10³)
0
1
2
3
4
5
6
7
8
9
IGLV8-21
IGLV8-19
IGLV8-18
IGLV8-16
IGLV5-14
IGLV8-13
IGLV8-10
IGLV3-5
IGLV3-4
IGLV3-3
IGLV3-2
IGLV3-1-1
Amino acid mismatches
0
2
4
6
8
10
12
14

Figure 4.3 Light chain V gene segment usage in individual pigs. (A) IGK. (B) IGL. The x-axis shows V gene segments that showed some level of expression and are arranged based on their genomic context.

Next Page

A


Number of Reads (x10³)
30
20
10
0
IGKV2-13
IGKV2-12
IGKV1-11
IGKV2-10
IGKV1-9
IGKV2-8
IGKV2-6
5
4
3
2
1
Pig

B

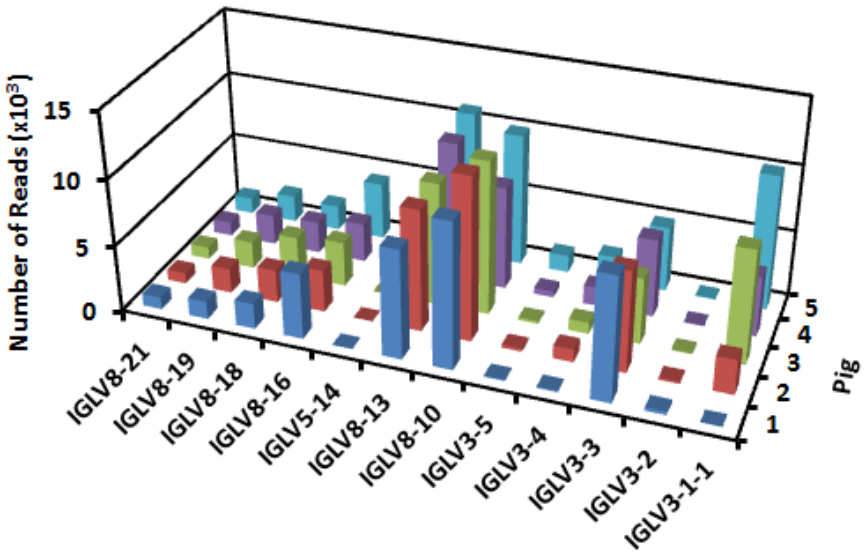
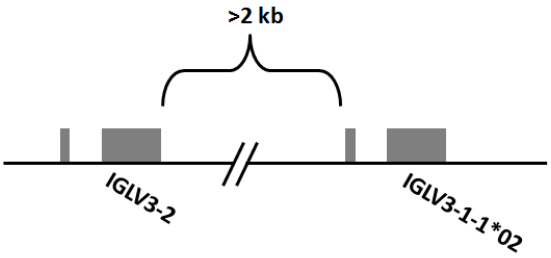

Number of Reads (x10³)
15
10
5
0
IGLV8-21
IGLV8-19
IGLV8-18
IGLV8-16
IGLV5-14
IGLV8-13
IGLV8-10
IGLV3-5
IGLV3-4
IGLV3-3
IGLV3-2
IGLV3-1-1
5
4
3
2
1
Pig

Figure 4.4 The genomic context of the novel IGLV3-1-1 gene segment in its (A) functional (*02) and (B) non-functional (*01) allelic forms.

A



B

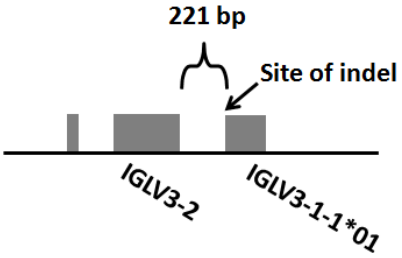


Figure 4.5 Number of amino acid mismatches in sequenced transcripts of each of the five pigs that BLAST to the novel IGLV3-1-1 gene segment. Z-axis value of zero indicates identity with the prototypic functional allele IGLV3-1-1*02. Pig 1expresses no transcripts from this gene segment, while pig 2 expresses transcripts that differ markedly from the prototypic allele.

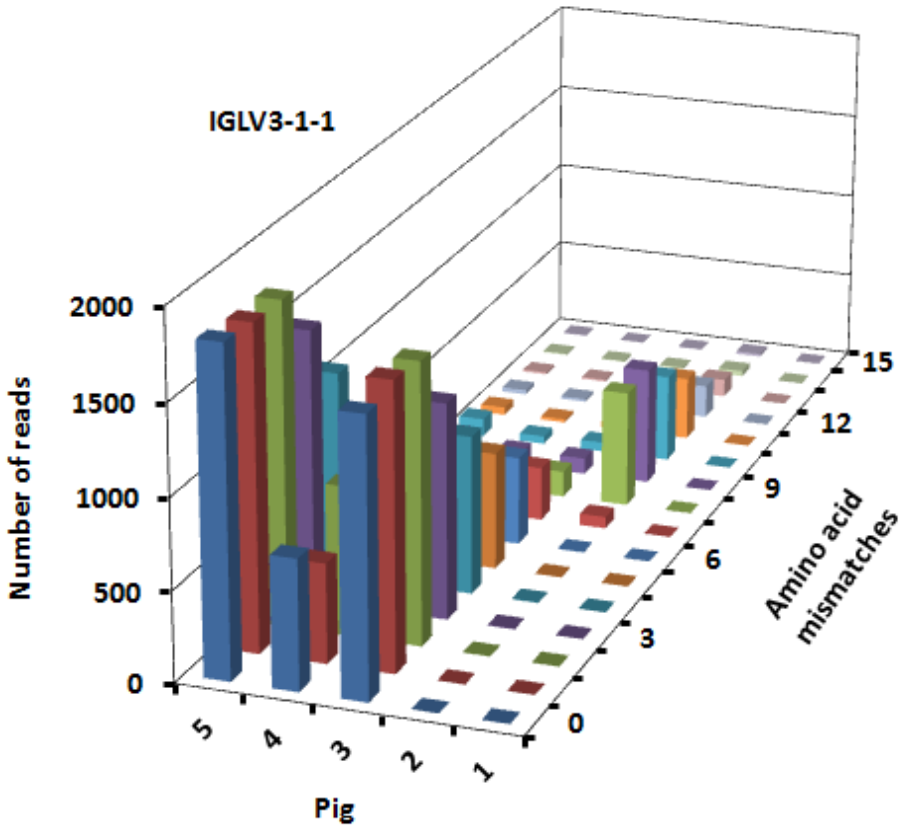
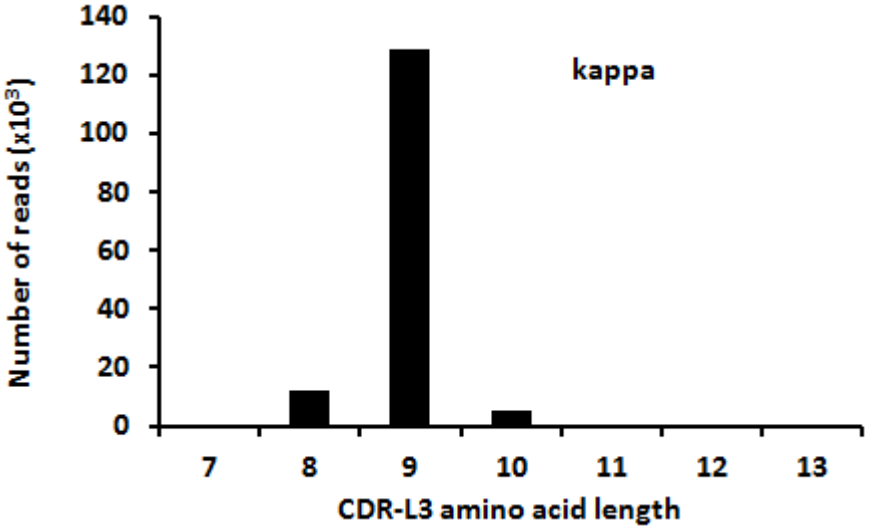


Figure 4.6 CDR-L3 length distributions. (A) Distribution of 187,810 IGL CDR3 regions. (B) Distribution of 147,189 IGK CDR3 regions.

A



B

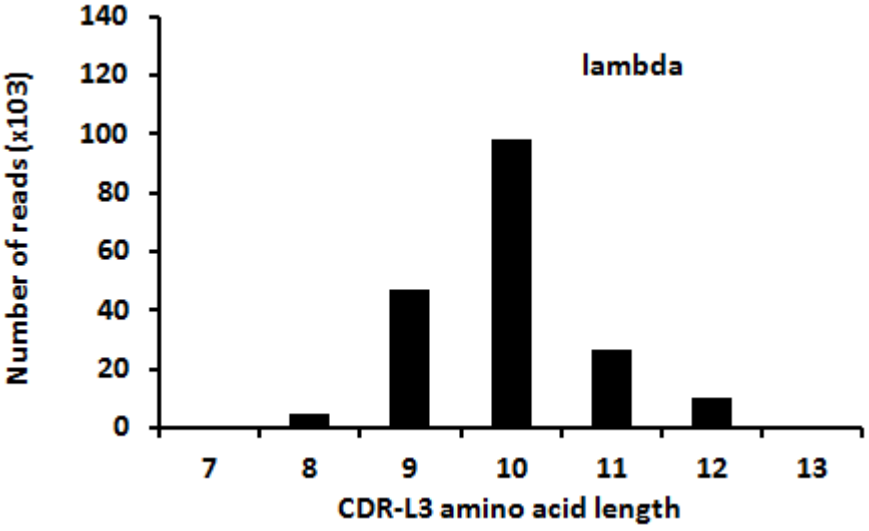


Figure 4.7 Lower-bound richness estimates for the total size of the light chain repertoire among individual pigs. Lambda: average, $2.3x10^5$; standard deviation: $4.9x10^4$. Kappa: average, $1.5x10^5$; standard deviation, $4.9x10^4$

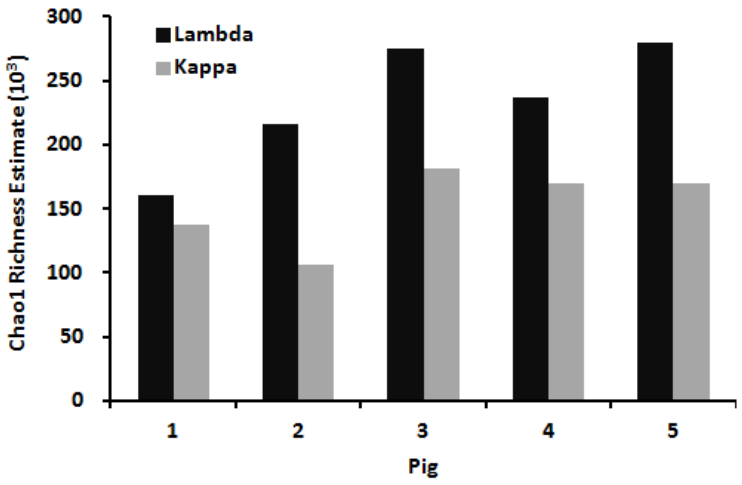
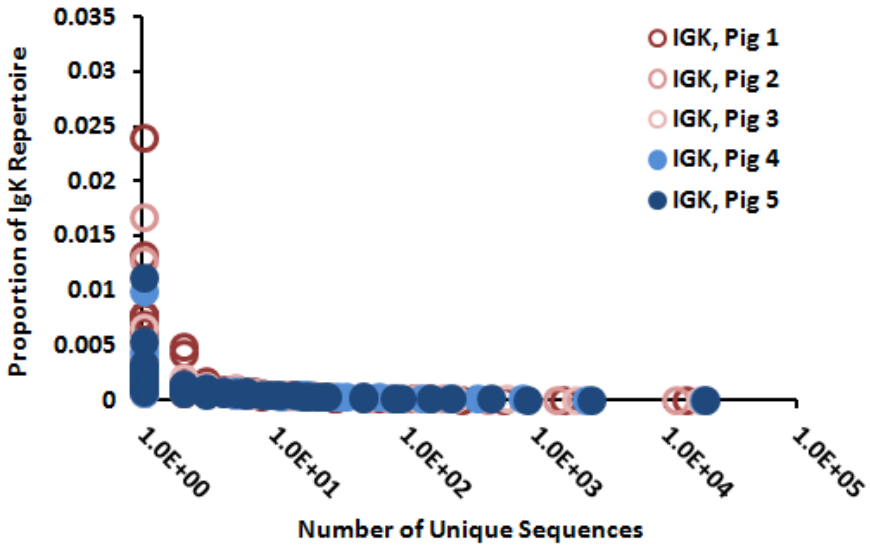


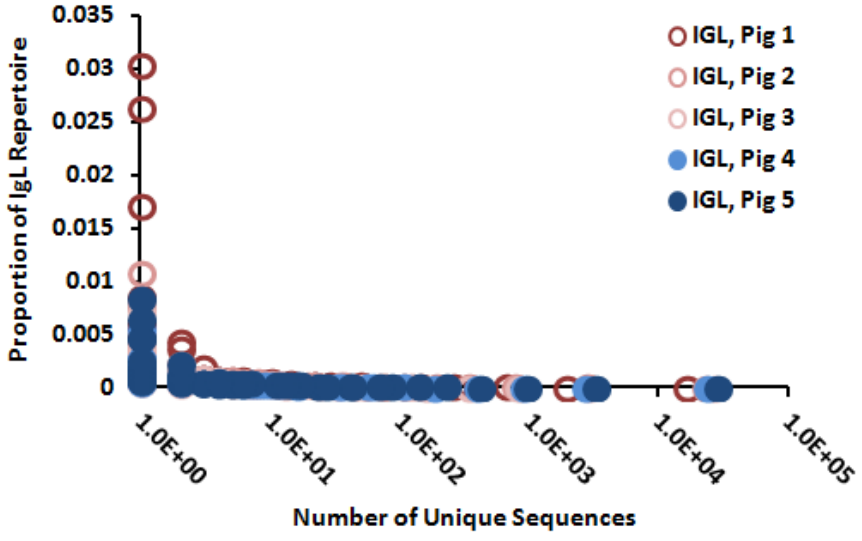
Figure 4.8 Light chain abundance distributions for five individual pigs. (A) IGK. (B) IGL. Open circles represent sequences derived from PRRSV-infected animals and closed circles represent uninfected animals. The y-axis is displayed on a linear scale to emphasize the differences between the infected and uninfected pigs. When displayed on a log-log plot, the curve forms a straight line, characteristic of a power-law distribution.

Next Page

A


0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Proportion of IgK Repertoire
1.0E+00
1.0E+01
1.0E+02
1.0E+03
1.0E+04
1.0E+05
Number of Unique Sequences
IGK, Pig 1
IGK, Pig 2
IGK, Pig 3
IGK, Pig 4
IGK, Pig 5

B


0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Proportion of IgL Repertoire
1.0E+00
1.0E+01
1.0E+02
1.0E+03
1.0E+04
1.0E+05
Number of Unique Sequences
IGL, Pig 1
IGL, Pig 2
IGL, Pig 3
IGL, Pig 4
IGL, Pig 5

DISCUSSION

Gene segment content and usage

In neonatal pigs it has been reported that the C-proximal IGLV3 family members dominate the expressed pre-immune repertoire (Butler et al. 2006). This is in contrast to our previous findings as well as our present findings in which IGLV8 family members dominate the IGLV expression profile (Schwartz et al. 2012b). It has been reported that mouse heavy chain gene usage shifts upstream in peripheral B-cells with age and antigenic stimulation (Malynn et al. 1990; Jung et al. 2006). In swine, however, this paradigm is challenged by a recent report showing the exclusive usage of IGLV8 in porcine yolk sac and fetal liver, with IGLV8-10 comprising more than >50 percent of lambda rearrangements (Wertz et al. 2013). The IGKV repertoire is dominated by the expression of two gene segments, namely, IGKV2-10 and IGKV1-11, with minor contributions from the remainder of the known functional gene segments. This contrasts with our previous findings where these remaining functional gene segments were generally expressed at higher levels (Schwartz et al. 2012a). However, this previous analysis was based on a relatively small number of porcine-derived expressed sequence tags obtained from GenBank with no information regarding their origin. Thus, given the mixed reports for both lambda and kappa gene segment usage, it seems plausible that gene segment usage in peripheral B-cells, especially among adult animals, is heavily influenced by environmental conditions.

Three putatively functional V gene segments (IGLV2-6, IGKV1-7, and IGKV1-14) were not expressed in any of the five animals studied. Comparison with the described

genomic features reveal that IGLV2-6 is the only putatively functional gene segment in the lambda locus to lack a conserved octamer in the promoter region (ATTTGCAT -> ATTTGTAT) and a conserved heptamer in the recombination signal sequence (CACAGTG -> TACAGTG). Likewise, both IGKV1-7 and IGKV1-14 are the only known V gene segments in the kappa locus lack a canonical octamer (ATCTGCAT, both) and both possess a non-canonical 5' splice site in their intron (GT -> GG) (Schwartz et al. 2012a,b). Thus, our findings highlight the importance of these non-coding regions on the expressibility or usage of V gene segments.

As shown in Figure 2a, there is a notably high level of sequence divergence in these IGK transcripts from the annotation. However, because the annotation describes the 14 most C-proximal IGKV genes in the locus, it is quite likely that additional unannotated gene segments lie further upstream. These annotated gene segments were found to be highly variable between the two chromosomes that were sequenced, making it likely that the large amount of sequence variation among the IGK transcripts is due to both a high level of allelic variation and an incomplete locus annotation and thus, additional upstream gene segment usage (Schwartz et al. 2012a). Indeed, the average number of amino acid differences between lambda transcripts and the annotated genome were around two to five, that same number for kappa varied between six and ten amino acids. This suggests that the annotation for the kappa locus is incomplete, as it seems unlikely that somatic hypermutation would preferentially target one light chain and not the other. A recent study has also putatively identified an additional IGKV2 gene segment, although its genomic context is unknown (Butler et al. 2013).

Allelic variation

Two pigs in the present study (2 and 4) appear to possess intermediate levels of usage compared to the others (pigs 3 and 5). One interesting question raised is whether pigs 2 and 4 are heterozygous for the truncated pseudogene allele of IGLV3-1-1. Additional analyses would need to be carried out in order to confirm this. If true, it would suggest that allelic heterozygosity can have a large impact on the antibody repertoire in individuals. Indeed, gene segment deletions in individual humans have previously been shown to perturb the expressed heavy chain repertoire (Boyd et al. 2010).

The results in the present study and in others (Butler et al. 2013; Wertz et al. 2013) confirm that light chain junctional diversity is limited in swine as in other species. Thus, allelic variation and somatic hypermutation must play a major role in generating the light chain repertoire. However, inferring allelic variation from peripheral B-cells given the cloud of diversity created from somatic hypermutation is a challenge, especially in older animals that have been exposed to environmental antigens. Although evidence was seen for some allelic variation, especially with regard to the three alleles of IGLV3-1-1, there is likely additional variation that is buried in the data, as not all alleles necessarily need to be so different from one another.

The results here and in the single Duroc sequenced in the genome sequencing project suggest that allelic variation in the antibody loci is substantial (Schwartz et al. 2012a,b). In order to more accurately assess allelic variation, however, gene segments would need to be sequenced in their germline configuration from a variety of individuals. As the heavy chain locus and kappa locus are incompletely characterized, the lambda

locus is therefore the best candidate to study. As the IGLV8 gene segment family appears to have expanded very recently, it would be interesting to investigate other species in the Suidae family and perhaps Tayassuidae, to assess the evolutionary origin of this gene segment family in swine. It may be hypothesized that IGLV8 represents a relatively recent resurrection of an ancestrally defunct, or poorly utilized family that only recently developed a selective advantage in swine.

Response to PRRSV

All three PRRSV-infected animals appeared to favor more highly abundant clones in their light chain repertoires. Two of these animals show highly expressed antibody sequences such that they clearly stand out above the background level of antibody abundance, suggesting that these sequences have undergone antigen-specific clonal expansion. However, the few antibody sequences that were identified from these highly abundant clones do not account for the broad anti-PRRSV antibody responses observed against a variety of PRRSV antigens found in all three PRRSV-infected individuals via ELISA (Klinge et al. 2009). Thus, at least the majority of anti-PRRSV responses are buried in the background antibody profile.

The most abundant clones among the PRRSV-infected group were derived from pig 1. Interestingly, this animal showed a marked reduction in total estimated richness in the lambda repertoire compared to the other four individuals. This reduction is seemingly due to genetic constraints in the C-proximal IGLV3 gene segment cluster. In particular, the sequences obtained from this animal are suggestive of a high degree of IGL homozygosity. Within the C-proximal IGLV cluster, only IGLV3-3 is expressed at high

levels in this animal, while the other IGLV3 gene segments are either unused or exceedingly rare. In addition, the novel gene segment IGLV3-1-1, which is highly expressed in other animals is entirely missing from this individual. As the most abundantly expressed antibody sequence from this animal was of IGLV3-3 origin, it is quite possible that the reduction in richness that we observed was due to such genetic constraints and not PRRSV-infection. A conceivable consequence that is suggested by our data could be that any compensatory increased usage of other gene segments may actually enhance that animal's ability to respond, if those gene segments confer greater specificity to the antigen(s) in question (namely, PRRSV). Indeed, IGLV3-3 was represented by the two most abundant light chain sequences in the PRRSV-infected animals. Further investigation is necessary to determine if the highly abundant antibody sequences found in the infected pigs are indeed PRRSV-specific.

The present study is the most comprehensive analysis to date of the light chain repertoire comparing animals infected and uninfected with a pathological agent. It is also heretofore the deepest characterization of the expressed porcine light chain repertoire. From these analyses, it is apparent that there is a high degree of genetic variation within the light chain loci of commercially bred swine. Such immunogenetic variation may be of significance to the swine industry in understanding the genetic bases of disease resistance.

CHAPTER V

THE EXPRESSED PORCINE ANTIBODY HEAVY CHAIN REPERTOIRE AND RESPONSE TO PRRSV INFECTION

In preparation for *The Journal of Immunology*

This study characterized the expressed immunoglobulin heavy chain (IGH) repertoire in pigs infected with porcine reproductive and respiratory syndrome virus (PRRSV) as well as in control sham-inoculated pigs. More than 445,000 high quality reads were generated from an amplicon library spanning the IGH variable domain via 454 pyrosequencing. Reads from PRRSV-infected and uninfected animals were analyzed for repertoire diversity and richness and the total porcine IGH repertoire was estimated to contain approximately $3x10^5$ unique antigen binding sites, based on the diversity of the complementarity determining regions (CDRs). The IGH repertoire was dominated by a relatively small number of transcripts. This effect is more pronounced in the PRRSV-infected pool, presumably due to an anti-PRRSV response. Furthermore, inspection of CDR3 lengths revealed an abnormal profile in PRRSV-infected tissues compared to sham-inoculated tissues. These differences were largely accounted for by a small number of unique sequences which were very common in the infected group, yet exceedingly rare in the uninfected group.

INTRODUCTION

Adaptive humoral immunity is mediated by a diverse array of gene segments which recombine during B cell development to generate a vast repertoire of antigen-binding immunoglobulin molecules. During B cell development, 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 immunoglobulin heavy chain variable region is generated from the recombination of a variable gene segment (V), a diversity gene segment (D), and a joining gene segment (J). Diversity in the immunoglobulin light chains is generated in a similar manner except that light chains lack a D gene segment. Further diversity is generated in B cells by undergoing class switch recombination to a different constant region gene segment and 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.

Of the three CDRs in both light and heavy chains, the first two CDRs are germline encoded and variable between V gene segments. In contrast, the third CDR of the heavy chain is largely the product of exonuclease and terminal deoxynucleotidyl transferase activity during V(D)J recombination. This typically results in an unrecognizable D gene segment and highly variable amino acid sequence and variable

CDR3 length. Thus, the potential antibody repertoire size is many orders of magnitude due to the number of theoretically possible V(D)J combinations, junctional diversity (i.e. CDR3) between gene segments, and the combination of different heavy and light chains.

A model of porcine reproductive and respiratory syndrome virus (PRRSV) infection offers an interesting opportunity to investigate the adaptive humoral response to infection in swine. PRRSV is of health and welfare concern to the swine industry and is of major economic significance to U.S. swine producers. The porcine antibody light chain repertoire has previously been investigated. From that study, highly abundant clones in PRRSV infected animals were identified, consistent with an anti-PRRSV response (Chapter 4). In the present study, we sought to identify highly abundant heavy chain clones in the same PRRSV infected animals as a means of better understanding the dynamics of the antibody response to PRRSV.

There has been no study to date investigating the total diversity and richness of the expressed porcine antibody heavy chain repertoire. Pigs, unlike humans, are highly restricted in their germline heavy chain repertoire as they possess only a single highly conserved IGHV gene segment family (Eguchi-Ogawa et al. 2010). As a result, it is believed that swine must compensate for this by relying more heavily on V(D)J junctional diversity (i.e. CDR3) (Butler et al. 2000). Here we report that swine possess a rich CDR3 repertoire, yet no richer than humans. Despite this, due to total CDR diversity and framework diversity likely resulting from somatic hypermutation, the expressed porcine heavy chain repertoire is at least as rich as the human repertoire.

MATERIALS AND METHODS

Library Preparation. Porcine lymphoid tissues collected from a previous study (spleen, palatine tonsil, inguinal lymph node, and mesenteric lymph node) were used in this investigation (Klinge et al. 2009). 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. Two unique forward primers were directed against the leader region and two unique reverse primers were directed against the 3' IGHJ and 5' IGHC region of framework 4 (Table 5.1). The 454 Titanium FLX fusion adapter (forward, adapter A: 5'-CGTATCGCCTCCCTCGCGCCATCAG; reverse, adapter B: 5'-CTATGCGCCTTGCCAGCCCGCTCAG) followed by a 10 bp barcode (uninfected: ATATCGCGAG, infected: TGAACAATCG) for sample differentiation was attached to the 5’ end of each primer. Thus, a total of eight primers, four for the infected pool and four for the uninfected pool, were used. The resulting PCR amplicons were pooled by infection status before being pooled in equimolar amounts into a single aliquot. 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.

Sequence analysis. Obtained reads were separated by barcode (infected vs. uninfected) and reads containing multiple ambiguous bases were removed from analysis. The IGHV, IGHD, and IGHJ gene segments were identified using BLAST based on their previously described germline sequences (Altschul et al. 1990; Eguchi-Ogawa et al. 2010). Translated, full-length sequences were clustered based on similarity using cd-hit (Li and Godzik 2006). 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 (Gotelli and Entsminger 2011) was used to compute individual-based rarefaction curves. The Chao 1 non-parametric estimate of total diversity was calculated using the program EstimateS (Colwell 2009).

Table 5.1 Primers used to amplify the porcine antibody heavy chain transcriptome.

Target	Template-specific sequence
IGHV, leader (F)	5'- AACTGGGTGGTCTTGTTTGC
IGHV, leader (F)	5'- TCTCTTACAAGGTRTCCAGGGTG
FR4 (R)	5'- TGAGGACACGACGACTTCAA
FR4 (R)	5'- AAGATTTTGGGGCTGGTTTC

F, forward; *R*, reverse

RESULTS

Sequencing results

Approximately 604,000 reads were obtained from the 454 sequencing reaction. Of these, 445,267 had quality scores greater than or equal to 39 and lengths greater than 250 nucleotides in length, with an average length of 424 nucleotides. Whereas lengths greater than 375 bases are necessary in order to capture the entire variable region repertoire. Approximately equal representation of reads was obtained from each group of animals: 220,857 reads were derived from JA142-infected tissues and 224,410 were derived from mock-infected tissues. After sequence filtering, between 102,000 and 169,000 putative amino acid sequences were obtained for each CDR.

Gene segment usage

Analysis of germline gene segment usage was restricted to the limited information currently available for heavy chain gene structure in the pig. Although the IGHD and IGHJ regions have been completely characterized in the pig, only the first 15 IGHV gene segments are characterized (Eguchi-Ogawa et al. 2010). Using BLAST, approximately 100,000 reads possessed high similarity to one of the known IGHJ segments. As expected, IGHJ5 usage was nearly ubiquitous; however, novel, very low level usage of both IGHJ1 and IGHJ3 was also detected (Table 5.2). Approximately 41,000 reads possessed high similarity to one of the four known IGHD gene segments. Of these, IGHD1 and IGHD2 were both highly utilized, as expected (Eguchi-Ogawa et al. 2010). Despite their position with respect to CDR3, we found no statistically significant

differences in either IGHD or IGHJ gene segment usage between PRRSV-infected and uninfected pigs ($P>0.05$ for each gene segment).

Approximately 49,000 sequences were >98 percent identical to one of the known IGHV gene segments. Novel, low level expression (24 reads) of the previously described pseudogene, IGHV1 was detected from this analysis. 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 (Figure 5.1). Interestingly, expression of IGHV4 and IGHV11 was proportionally increased in PRRSV-infected animals, suggesting that PRRSV-specific antibodies preferentially possess these gene segments. BLAST analysis of approximately 193,000 sequences revealed a wide distribution of reads in relation to the established IGHV annotation, presumably due to factors such as missing gene segments in the annotation, allelic variation, junctional diversity resulting from the enzymatic remodeling of CDR3 during recombination, and somatic hypermutation (Figure 5.2).

Repertoire diversity

As expected, rarefaction analysis revealed that the larger CDR3 contributes more to heavy chain diversity than either CDR1 or CDR2 (Figure 5.3). Interestingly, there is a decrease in antibody diversity as a result of PRRSV infection, likely as a result of clonal selection and proliferation of a relatively few B cells with affinity for PRRSV epitopes. In addition, the Chao 1 estimate of total antibody diversity suggests that there is a heavy constraint on amino acid usage in each of the CDRs (Table 5.3), particularly CDR3. The framework regions, despite accounting for the vast majority of the variable region,

possess a similar level of richness as CDR3, or approximately $1x10^5$ unique sequences. This framework richness is likely due to a low level of extra-CDR somatic hypermutation as a large proportion (approximately 20 percent) of all framework sequences were identical to each other (not shown).

CDR3 diversity and effect of PRRSV-infection

Analysis of CDR3 length distribution (Figure 5.4) reveals that the average length in swine is 13 amino acids, two amino acids longer than recently reported in humans using a similar method (Glanville et al. 2009). Additionally, swine appear to possess an increased preference for short CDR-H3s compared to humans, as evidenced by the bimodality of Figure 5.4. In pigs infected with PRRSV, the distribution follows the same trend; however, there is disproportionately greater usage of CDR3s with amino acid lengths 5, 11, 13, 15, 17, 20, 21, and 22 in infected animals. Sorting the transcripts by CDR3 length and 100% identity threshold revealed exceptionally large clusters in the PRRSV-infected group compared to the uninfected group in these CDR3 size classes (Figure 5.5). Furthermore, the heavy chain abundance distribution shows that a small number of unique reads account for a large fraction of the antibody repertoire - a fact that is more prominent in PRRSV-infected animals (Figure 5.6). The amino acid profile for several of the CDR3 size classes also reveal that single amino acids are preferentially selected for at each position among infected pigs (Figure 5.7). Comparatively, CDR3 lengths not apparently associated with the PRRSV-response show no such difference in amino acid profile. Indeed, the CDR3 amino acid sequence ARADCYSDGGICYFFDHGVMDL accounts for approximately 36 percent of all

CDR3s of length 22 in PRRSV-infected animals (Figure 5.7D). Likewise, SRGYVYICGWACMDL, the most abundant CDR3 in the infected pool, accounts for approximately 31 percent of all length 15 CDR3s (Figure 5.7B). In the uninfected pool, however, these same sequences account for approximately 0.1 percent of CDR3s of the same respective lengths (Figures 5.7A and 5.7C).

CDR3 appears to be highly restricted as few amino acids appear to actually be tolerated. Between IMGT positions 107 and 114 of CDR3, glycine and tyrosine account for about one-third of total amino acid usage. These together with serine, alanine, and cysteine account for nearly 60 percent of amino acid usage (Figure 5.7B). The framework-proximal regions are even more restricted and largely resemble their germline sequences on IGHV and IGHJ. Despite this high degree of restriction, however, the middle portion of CDR3, encoded by IGHD, is often unrecognizable from its germline sequence, presumably due to enzymatic activity during VDJ recombination. This CDR3 amino acid usage profile strongly resembles that of humans, suggesting that this restriction is similar across species (Glanville et al. 2009).

As a result of these analyses, several PRRSV-specific CDR3 sequences can be derived and associated with full-length amino acid sequences (Table 5.4). Of the six PRRSV-specific antibody sequences described here, two (bearing the CDR3s described in Figure 7) match the translated germline sequence of IGHV4 to within a difference of 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 sequence (Butler et al. 2006c). The remaining three do not closely match any previously

described IGHV gene segment; however, they may be derived as a result of more extensive somatic hypermutation.

As the individual animals were not molecularly tagged, it is impossible given the sequencing data to ascertain whether the putatively PRRSV-specific clones are derived from a single animal or both. In order to help answer this question, quantitative real-time PCR was conducted using primers generated against four of the putatively PRRSV-specific CDR3s. This analysis suggests that the most abundant clones are present in both PRRSV-infected animals, although one animal in particular appears to have shown a greater response (Pig 6, Figure 5.8).

Table 5.2 IGHD and IGHJ gene segment usage in PRRSV-infected and uninfected animals.

Gene Segment	Uninfected	Infected
IGHD1	11,686	15,851
IGHD2	5,522	7,854
IGHD3	0	0
IGHD4	0	0
IGHJ1	0	6
IGHJ2	0	0
IGHJ3	83	37
IGHJ4	0	0
IGHJ5	46,606	53,678

Figure 5.1 Percent usage of known IGHV gene segments showing fold differences between PRRSV-infected and uninfected groups. Analysis included 45,918 reads that matched known IGHV gene segments with an identity >98 percent. All infected/uninfected pairs have $P < 0.001$, except for IGHV8, using a two-tailed χ^2 test.

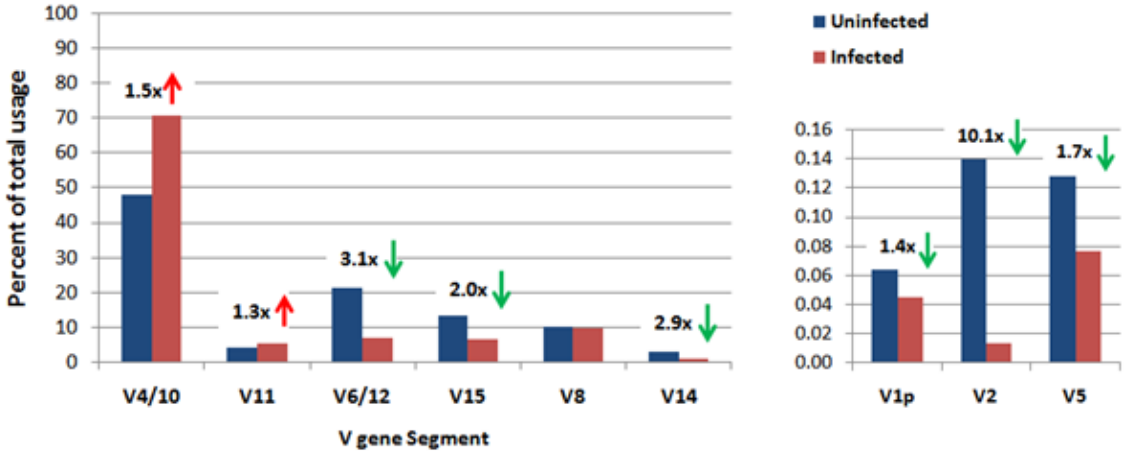
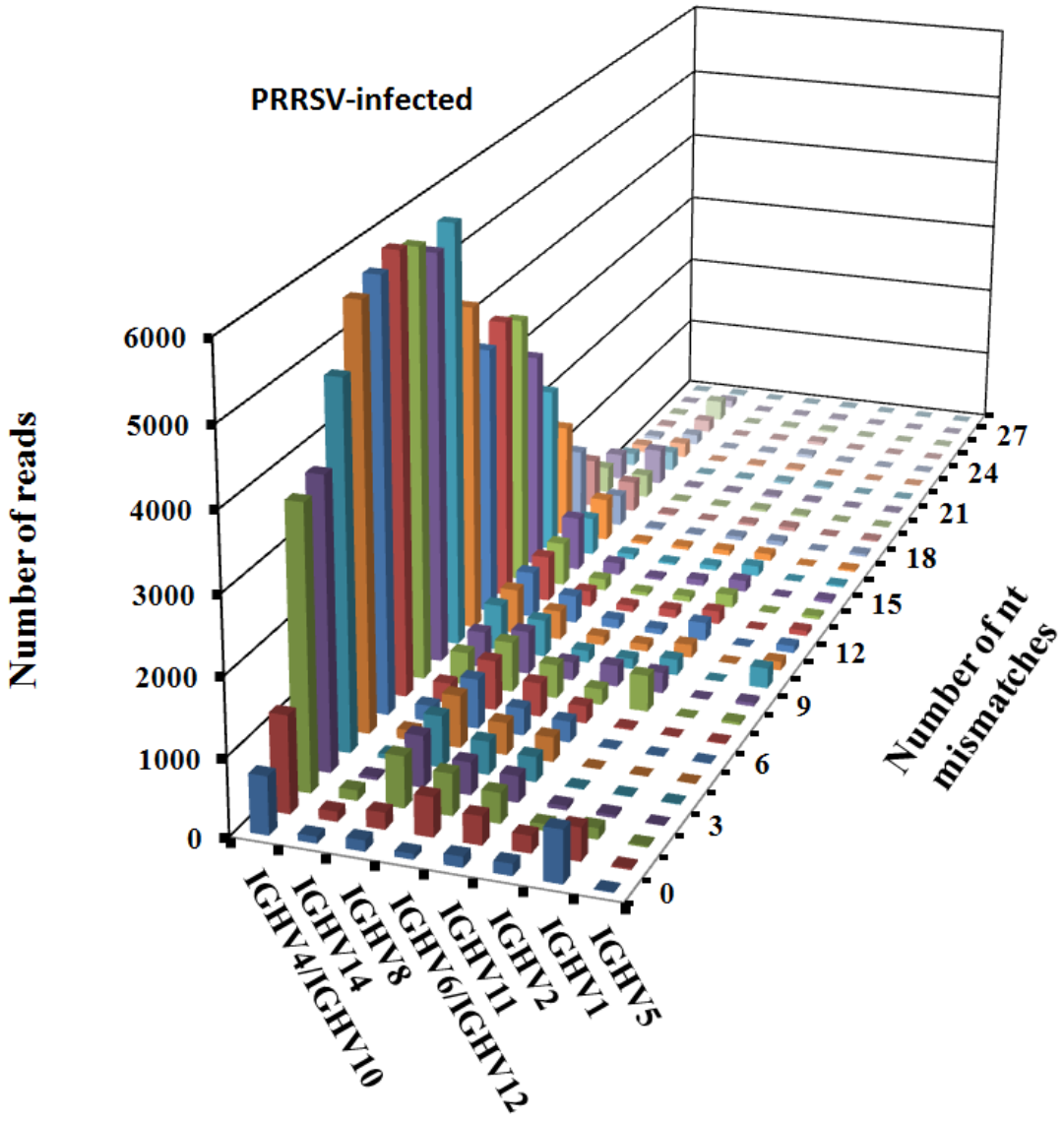


Figure 5.2 Number of nucleotide mismatches in sequences heavy chain transcripts compared to the published annotated IGHV gene segments. (A) PRRSV-infected. (B) Uninfected. Gene segments displayed on x-axis are arranged according to the most to least expressed.

A



B

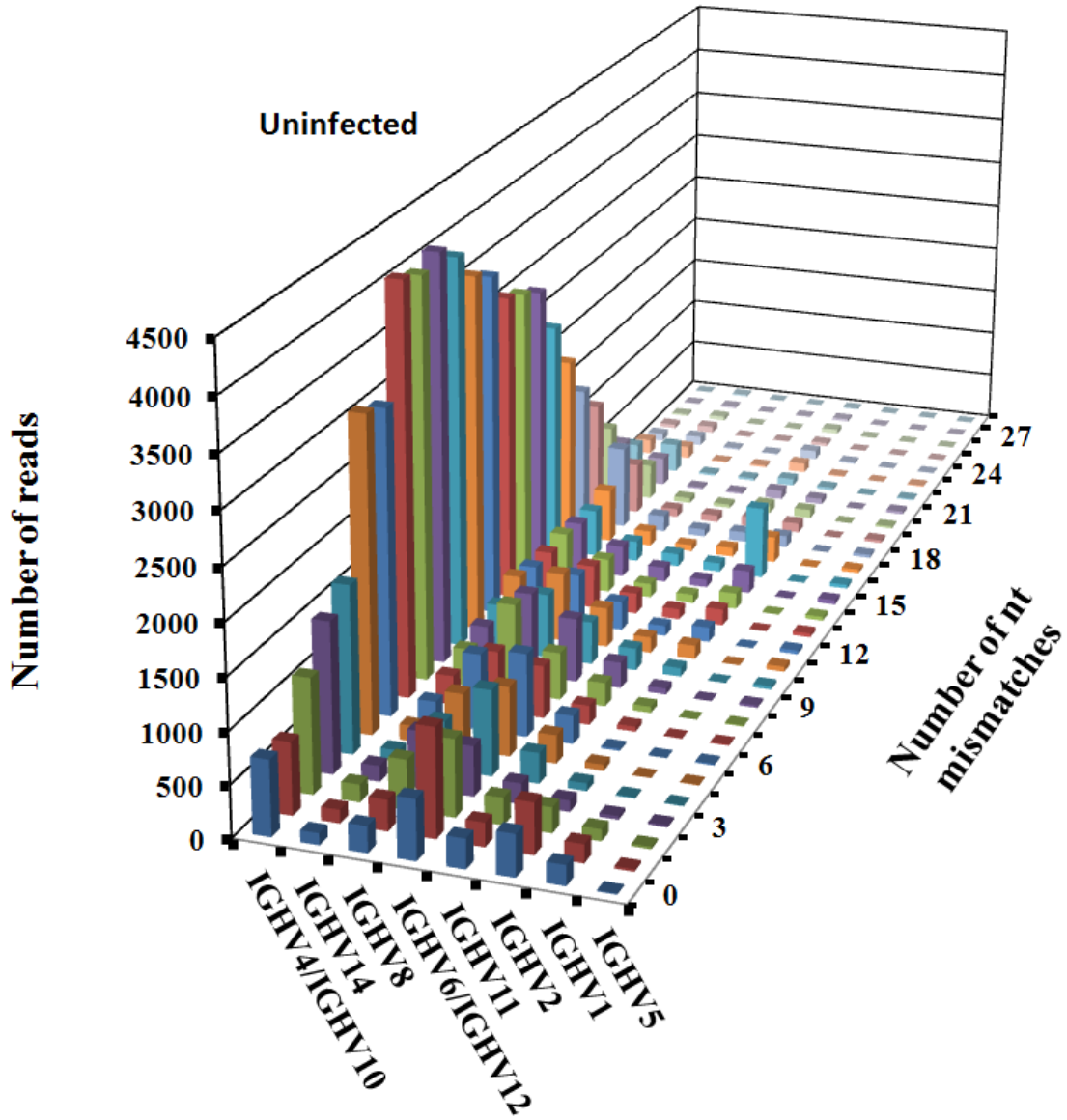

Uninfected
Number of reads
4500
4000
3500
3000
2500
2000
1500
1000
500
0
IGHV4/IGHV10
IGHV14
IGHV8
IGHV6/IGHV12
IGHV11
IGHV2
IGHV1
IGHV5
0
3
6
9
12
15
18
21
24
27
Number of nt mismatches

Figure 5.3 Individual-based rarefaction curves of all three heavy chain CDRs and concatenated CDRs. Clusters were generated using translated amino acid sequences and 100 percent identity cutoff.

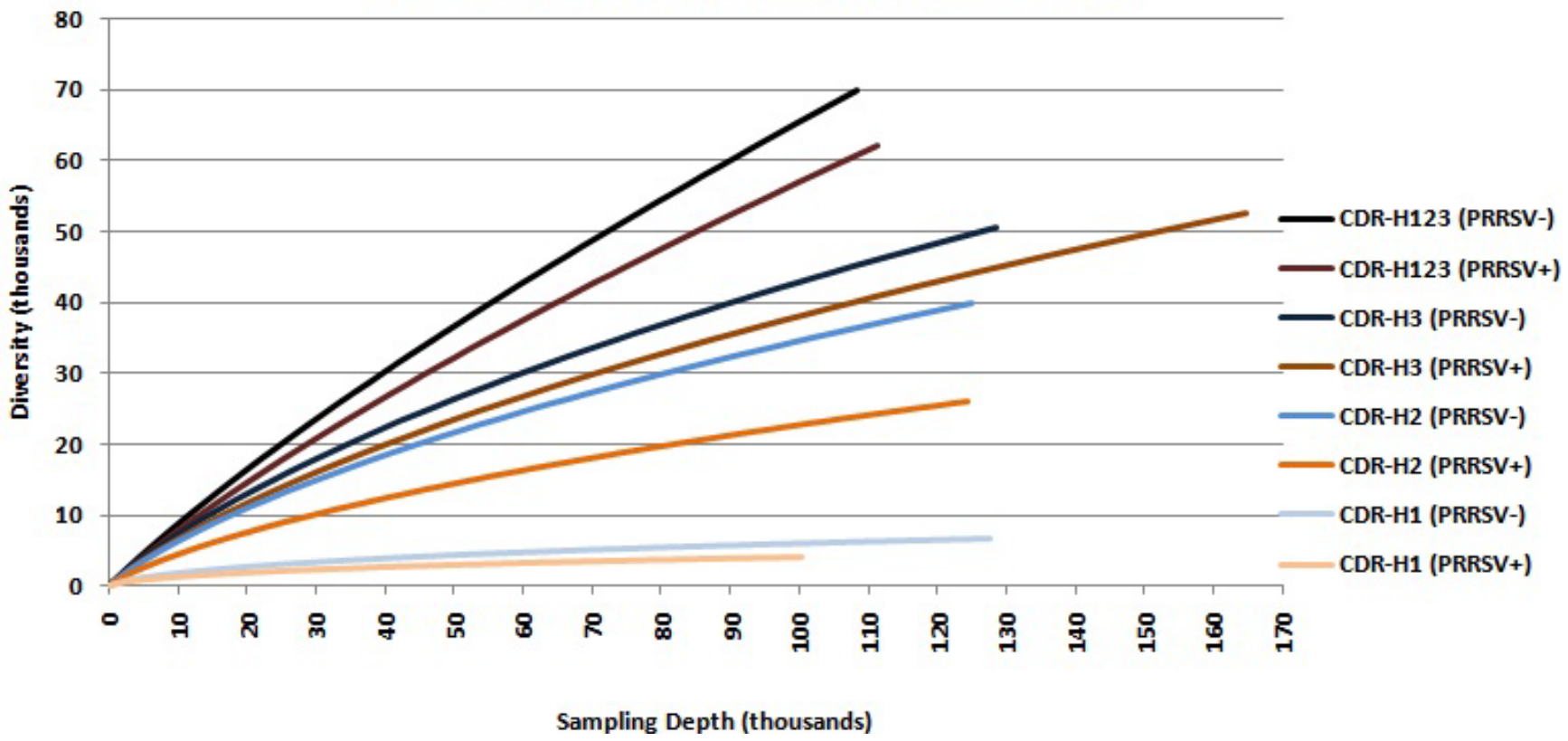


Table 5.3 Richness estimates for the antibody repertoires of pigs. The non-parametric estimators (Chao 1 and ACE) represent a lower bound of total repertoire richness. ACE: abundance-based coverage estimator. Hurlbert's PIE: probability of an interspecific encounter; that is, the probability that any two randomly chosen sequences are non-identical.

Status	CDR	Chao 1	Chao 1, 95% C.I.		ACE	Hurlburt's PIE
			Lower Bound	Upper Bound		
PRRSV+	CDR1	6,555	6,252	6,900	5,388	0.96004
	CDR2	57,861	56,478	59,307	58,650	0.9938
	CDR3	116,574	114,639	118,570	115,486	0.99888
	CDR123	246,597	241,223	252,133	295,934	0.9994
PRRSV-	CDR1	10,906	10,528	11,323	8,478	0.97203
	CDR2	90,644	88,868	92,486	90,517	0.99796
	CDR3	117,143	114,199	118,282	117,143	0.99969
	CDR123	297,429	291,128	303,910	364,137	0.99991

Figure 5.4 Observed CDR3 nucleotide length diversity in PRRSV-infected (190,045 sequences) and mock-infected (197,206 sequences) animals.

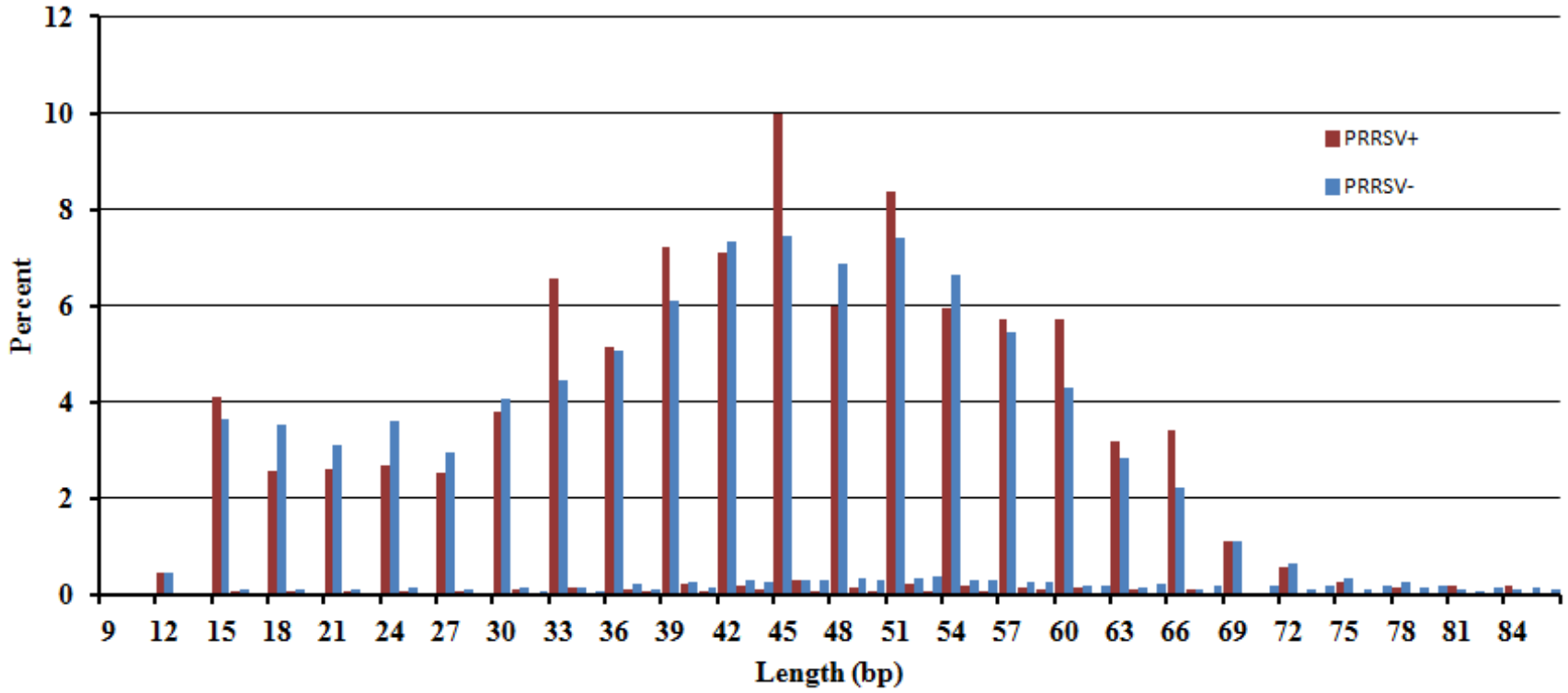


Figure 5.5 Cluster analysis of unique CDR3 sequences. Approximately 80,000 putative amino acid CDR3 sequences were clustered using cd-hit with an identity threshold of 100 percent. The PRRSV-infected group possesses several extremely abundant CDR3 sequences.

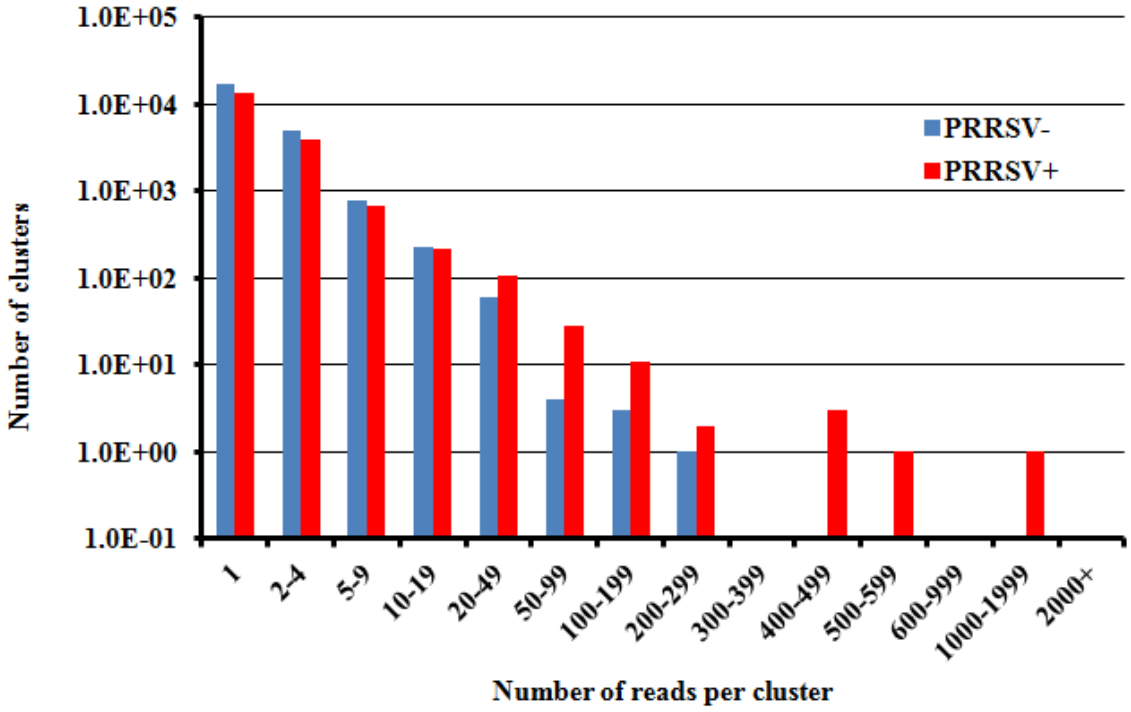


Figure 5.6 Abundance distribution of all three concatenated CDRs in the PRRSV-infected and uninfected groups. A small number of unique CDR combinations are exceptionally common, especially among PRRSV-infected animals.

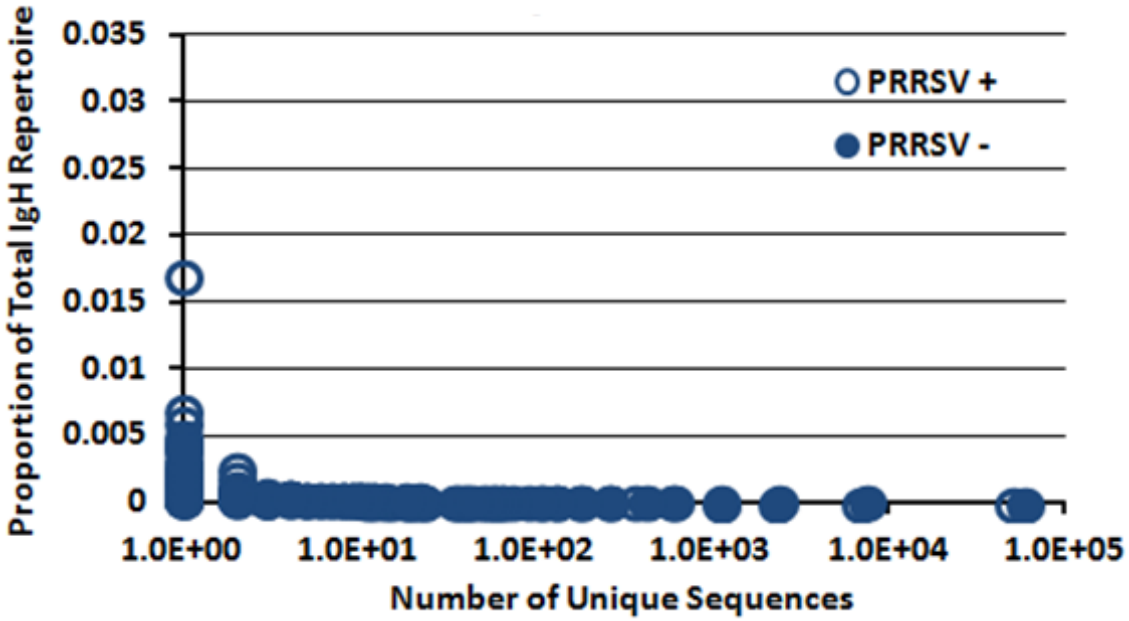
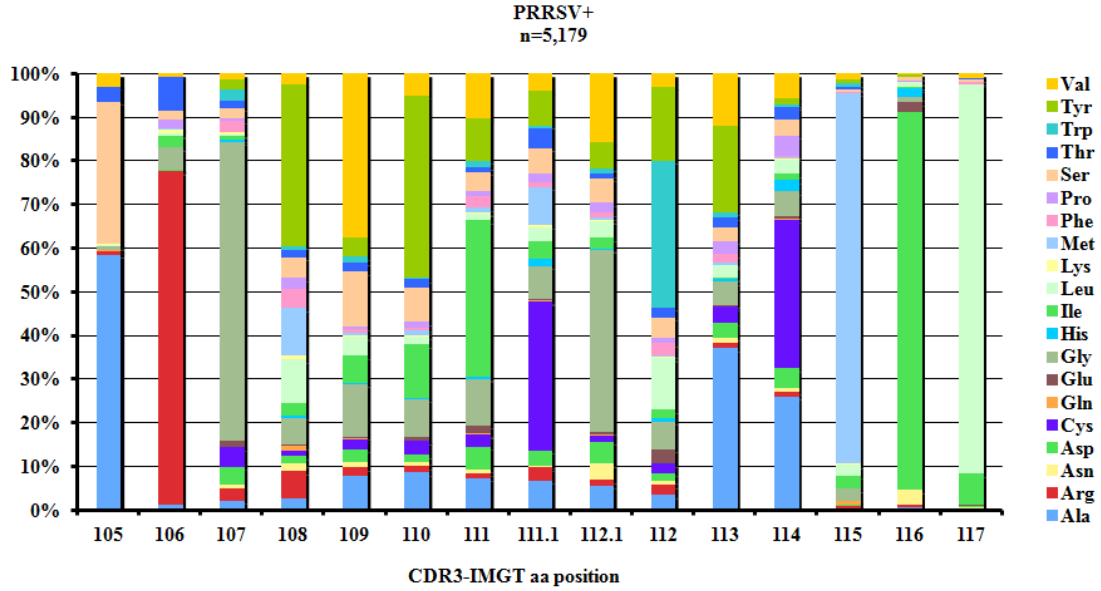


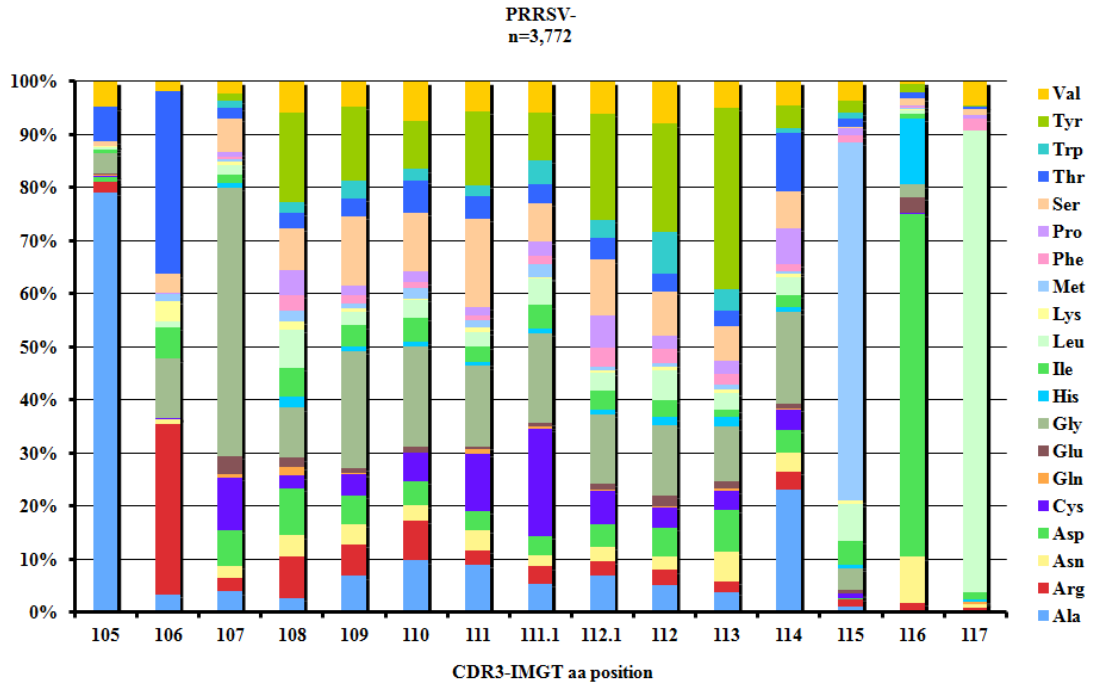
Figure 5.7 CDR3 amino acid usage. CDR3 lengths of 15 amino acids (A and B) and 22 amino acids (C and D) for both PRRSV-infected (B and D) and mock-infected (A and C) animals are represented. Amino acid numbering is based on the IMGT system (Lefranc et al. 2003).

Next Page

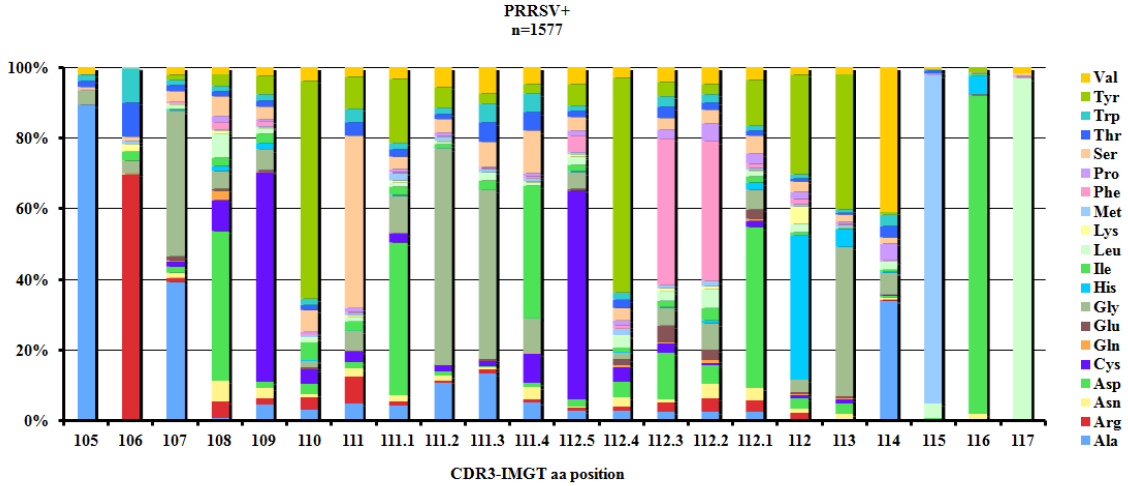
A


PRRSV+
n=5,179
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
105
106
107
108
109
110
111
111.1
112.1
112
113
114
115
116
117
CDR3-IMGT aa position
Val
Tyr
Trp
Thr
Ser
Pro
Phe
Met
Lys
Leu
Ile
His
Gly
Glu
Gln
Cys
Asp
Asn
Arg
Ala

B


PRRSV-
n=3,772
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
105
106
107
108
109
110
111
111.1
112.1
112
113
114
115
116
117
CDR3-IMGT aa position
Val
Tyr
Trp
Thr
Ser
Pro
Phe
Met
Lys
Leu
Ile
His
Gly
Glu
Gln
Cys
Asp
Asn
Arg
Ala

C


PRRSV+
n=1577
100%
80%
60%
40%
20%
0%
105 106 107 108 109 110 111 111.1 111.2 111.3 111.4 112.5 112.4 112.3 112.2 112.1 112 113 114 115 116 117
CDR3-IMGT aa position
Val
Tyr
Trp
Thr
Ser
Pro
Phe
Met
Lys
Leu
Ile
His
Gly
Glu
Gln
Cys
Asp
Asn
Arg
Ala

D

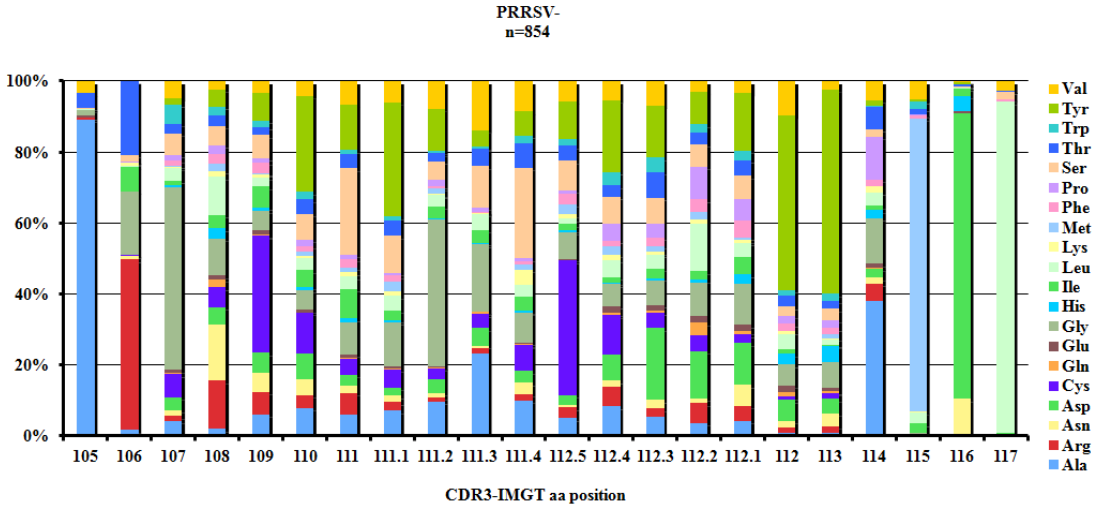
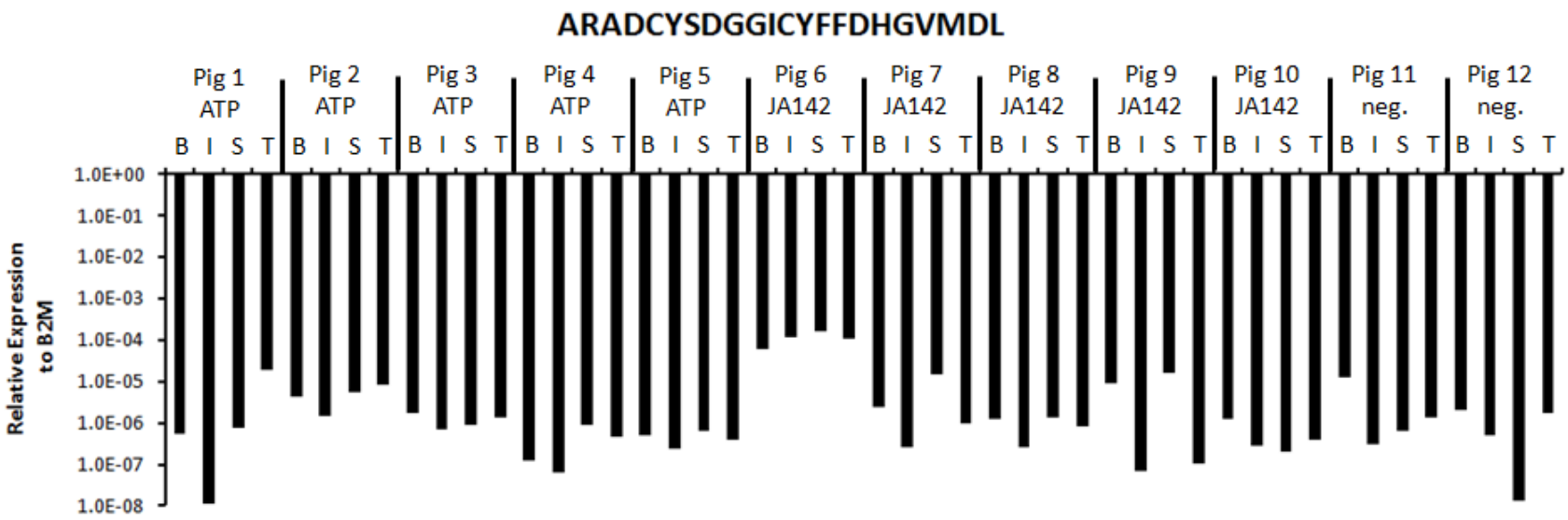

PRRSV-
n=854
100%
80%
60%
40%
20%
0%
105 106 107 108 109 110 111 111.1 111.2 111.3 111.4 112.5 112.4 112.3 112.2 112.1 112 113 114 115 116 117
CDR3-IMGT aa position
Val
Tyr
Trp
Thr
Ser
Pro
Phe
Met
Lys
Leu
Ile
His
Gly
Glu
Gln
Cys
Asp
Asn
Arg
Ala

Table 5.4 Putatively PRRSV-specific heavy chain aa sequences. Shown are the most abundant sequences for each PRRSV-specific CDR-H3. CDR/FR boundaries are based on the IMGT system (Lefranc et al. 2003).

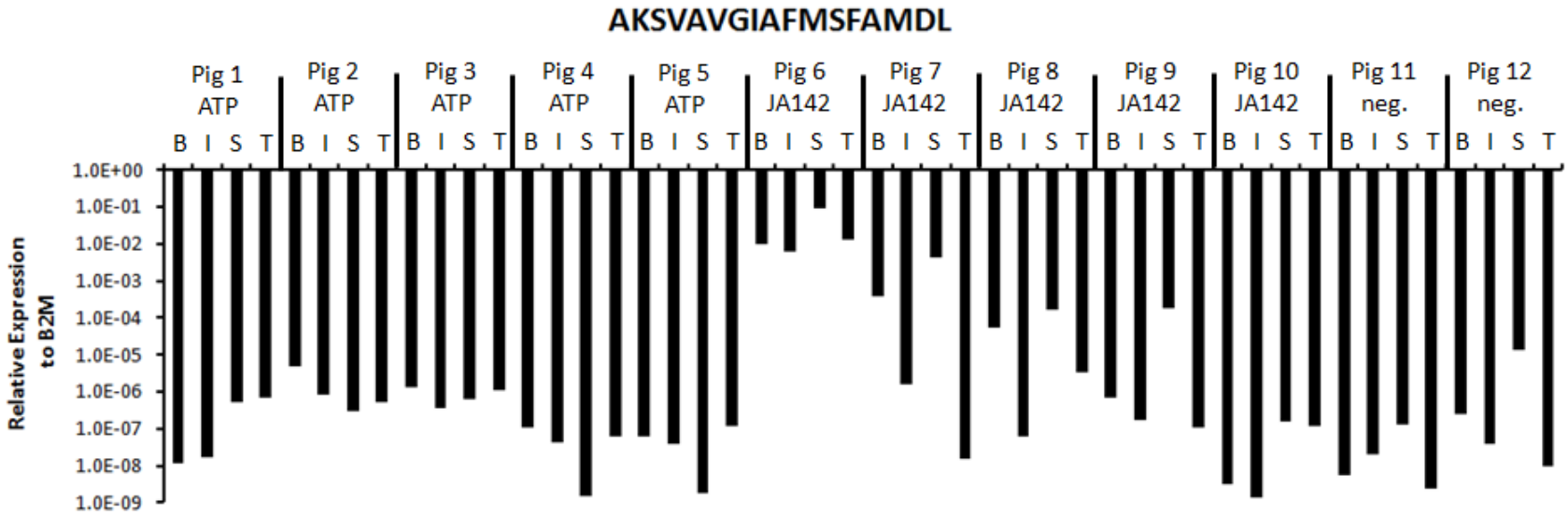
FR1	CDR1	FR2	CDR2	FR3	CDR3	FR4
EEKLVESGGGLVQPGGSLRLSCVGS	GFTFSSTY	INWVRQAPGKGLEWLAA	ISTGGGST	YYADSVKGRFTISRDNSQNTAYLQMNSLRTEDTARYYC	ARADCYSDGGICYFFDHGVMDL	WGPGVEVVVSS
EEKLVESGGGLVQPGGSLRLSCVGS	GFTFSSYA	VSWVRQAPGKGLEWLAC	IYSSGSAT	YYADSVKGRFTISRDNSQNTAYLQMNSLTTEDTARYYC	AKSVAVGIA.....FMSFAMDL	WGPGVEVVVSS
EEKLVESGGGLGQPGGSLRLSCVGS	GFAFSSSY	INWVRQAPGKGLEWLAA	ISTSGIGT	YYADSVKGRFTISSDNSQNTAYLQMNSLRTEDTARYYC	SRGYVYIC.......GWACMDL	WGPGVEVVVSS
EEKLVESGGGLVQPGGSLRLSCVGS	GFTFSSTY	INWVRQAPGKGLEWLAV	ISTDGVDT	YYADSVKGRFTISRGNSQNTAFLQMNSLRTEDTARYYC	VRGYIYG.........ASYLDL	WGPGVEVVVSS
EEKLVESGGGLVQPGGSLRLSCVGS	GFTFSSTY	INWVRQAPGKGLEWLAT	IYRSDGNT	DYEDSVKGRFTISRDNSQNTAYLQMNSLRTEDTARYYC	VRDVYP..........STMDL	WGPGVEVVVSS
EVKLVESGGGLVQPGGSLRLSCVGS	GFTFSVYN	MVWVRQRPGKGLEWLAC	ITSRGSST	YYADSVKGRFTISRDNSQNTAYLQMNSLRTEDTARYYC	ARDS.............DMDL	WGPGVEVVLSS

Figure 5.8 Quantitative real-time PCR of heavy chain CDR3s in uninfected pigs (n=2), or pigs infected with PRRSV strains ATP (n=5) or JA142 (n=5). Four tissues from each animal were examined: bronchial lymph node, B; inguinal lymph node, I; spleen, S; and tonsil, T. Results were normalized using beta-2-microglobulin (B2M) and shown based on relative abundance to B2M. Four highly-abundant putatively PRRSV-specific clones were analyzed, those with the 22 amino acid sequence (A), 17 amino acid sequence (B), 15 amino acid sequence (C), and 11 amino acid sequence (D). Pigs 6 and 7 and both uninfected pigs were used for heavy chain transcriptome sequencing.

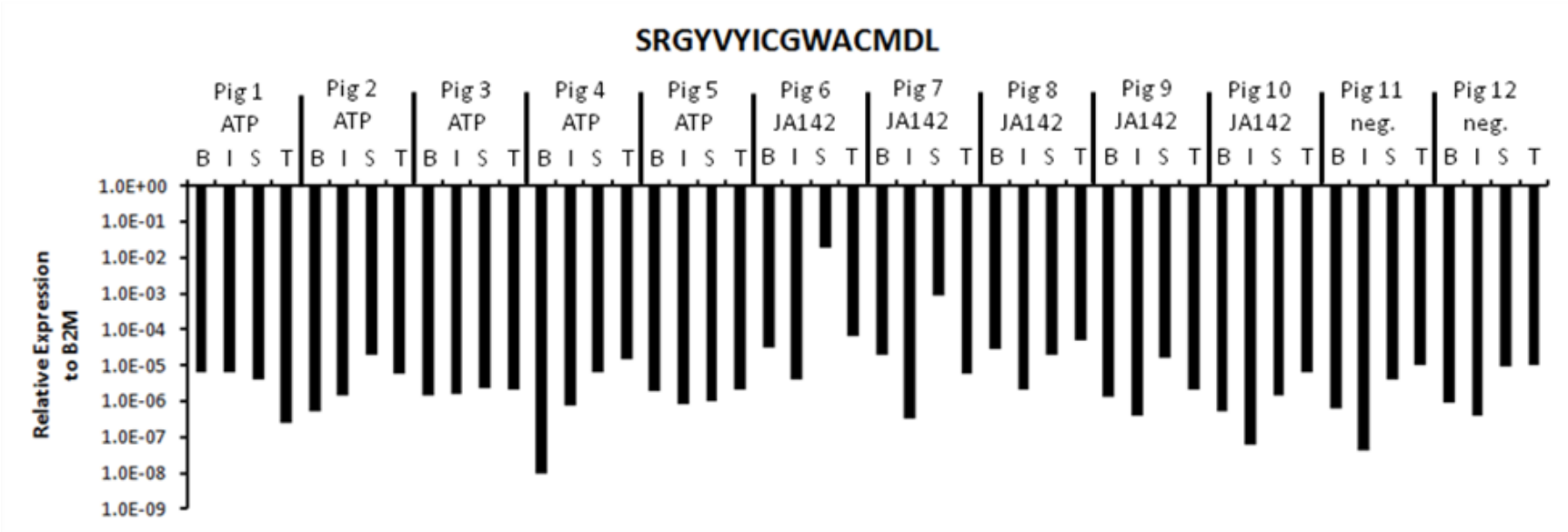
A


ARADCYSDGGICYFFDHGVMDL
Pig 1 ATP
Pig 2 ATP
Pig 3 ATP
Pig 4 ATP
Pig 5 ATP
Pig 6 JA142
Pig 7 JA142
Pig 8 JA142
Pig 9 JA142
Pig 10 JA142
Pig 11 neg.
Pig 12 neg.
B I S T
Relative Expression to B2M
1.0E+00
1.0E-01
1.0E-02
1.0E-03
1.0E-04
1.0E-05
1.0E-06
1.0E-07
1.0E-08

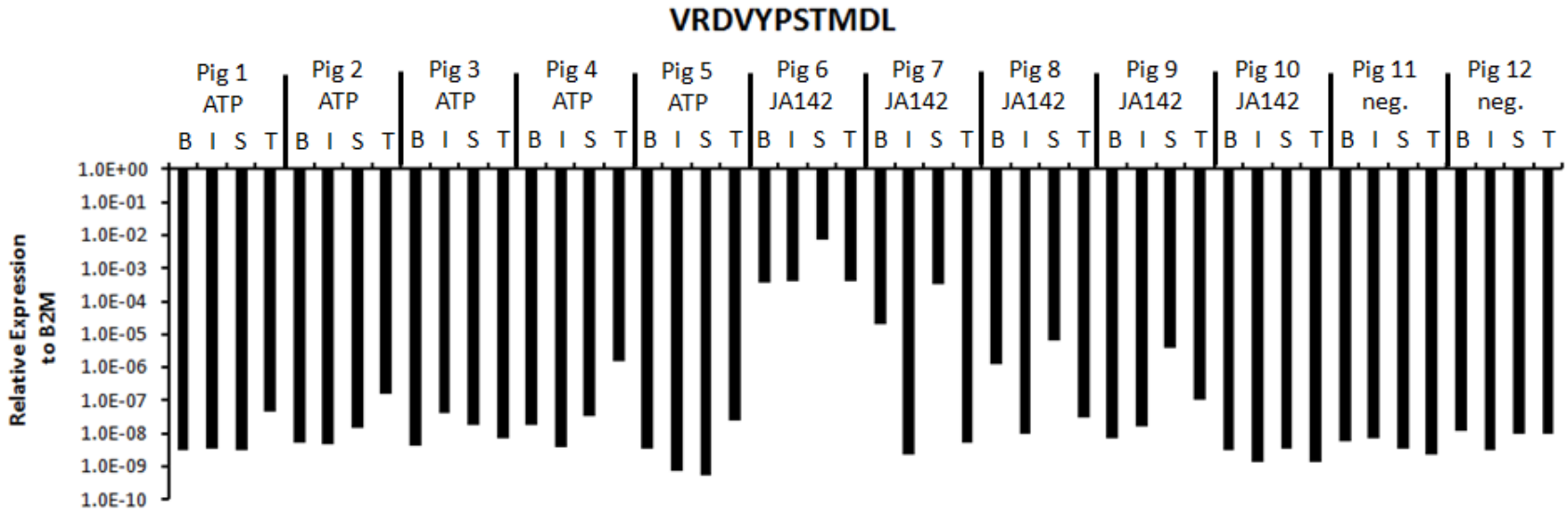
B


AKSVAVGIAFMSFAMDL
Pig 1 ATP
Pig 2 ATP
Pig 3 ATP
Pig 4 ATP
Pig 5 ATP
Pig 6 JA142
Pig 7 JA142
Pig 8 JA142
Pig 9 JA142
Pig 10 JA142
Pig 11 neg.
Pig 12 neg.
B I S T
Relative Expression to B2M
1.0E+00
1.0E-01
1.0E-02
1.0E-03
1.0E-04
1.0E-05
1.0E-06
1.0E-07
1.0E-08
1.0E-09

C


SRGYVYICGWACMDL
Pig 1 ATP
Pig 2 ATP
Pig 3 ATP
Pig 4 ATP
Pig 5 ATP
Pig 6 JA142
Pig 7 JA142
Pig 8 JA142
Pig 9 JA142
Pig 10 JA142
Pig 11 neg.
Pig 12 neg.
B I S T
Relative Expression to B2M
1.0E+00
1.0E-01
1.0E-02
1.0E-03
1.0E-04
1.0E-05
1.0E-06
1.0E-07
1.0E-08
1.0E-09

D


VRDVYPSTMDL
Pig 1 ATP
Pig 2 ATP
Pig 3 ATP
Pig 4 ATP
Pig 5 ATP
Pig 6 JA142
Pig 7 JA142
Pig 8 JA142
Pig 9 JA142
Pig 10 JA142
Pig 11 neg.
Pig 12 neg.
B I S T
Relative Expression to B2M
1.0E+00
1.0E-01
1.0E-02
1.0E-03
1.0E-04
1.0E-05
1.0E-06
1.0E-07
1.0E-08
1.0E-09
1.0E-10

DISCUSSION

The current analysis suggests that relatively few B-cell clones are highly responsive to PRRSV infection. As all individual animals responded to a variety of PRRSV antigens by ELISA, it is likely that the vast majority of antigen-specific clones are obscured in the background of the antibody repertoire (Klinge et al. 2009). However, our use of specific CDR3 class sizes allowed for greater sensitivity to detect abundant clones compared to assaying the entire repertoire, such as by analyzing the abundance distribution alone. Nevertheless, with similar appropriate light chain information (Chapter 4), it may be possible to construct scFvs to assay for the specificity of these most abundant clones.

Our observations revealed that heavy chain abundance distributions follow a similar power law curve as previously seen for zebrafish heavy chain and porcine light chain sequences (Weinstein et al. 2009; Chapter 4). This same distribution was still apparent for sub-regions, such as for the individual CDRs and the framework regions. This finding is perhaps not surprising given our understanding of the mechanisms of antibody diversification, namely, somatic hypermutation, clonal expansion, and the subsequent gradual decline in specific antibodies. The lower bound Chao 1 richness estimate of approximately $3x10^5$ unique heavy chain CDR sequences in the pig is 36 percent higher than that reported for humans ($2.2x10^5$), suggesting that swine can suitably overcome germline restriction to a single IGHV family (Glanville et al. 2009). These values suggest that the antibody repertoire is severely restricted, likely as a result of selection against auto-reactivity and structural instability. For example, if any amino acid

combination were possible, one would predict a CDR3 with 13 amino acids to possess a theoretical richness of $8x10^{16}$. Still, analyses suggest that CDR3 alone only accounts for about one-third of the heavy chain CDR repertoire, belying its own diversity.

In the present study, novel, low-level expression of IGHJ1 was observed, despite its possession of a non-canonical heptamer, as well as novel expression of IGHJ3 (Eguchi-Ogawa et al. 2010). A previous report suggested that experimental ablation of IGHJ5, previously the only known utilized IGHJ gene segment, results in a B-cell knockout animal (Eguchi-Ogawa et al. 2010). If substantiated, this would suggest that swine are incapable of compensating for that deficiency by using IGHJ3, despite this gene segment being presently described as functional and expressed and possessing a canonical recombination signal. Thus, in addition to the recombination signal, heretofore undescribed factors must also help dictate IGHJ gene segment usage.

CHAPTER VI

GENERAL DISCUSSION

Modern commercial swine are susceptible to numerous infectious diseases. Chief among these, PRRSV continues to represent a serious economic threat to pork producers. In an ideal production setting, animals would respond to disease and vaccination in a predictable manner. Unfortunately, to that end, individual responses to vaccination and disease vary greatly. Furthermore, there is limited information available on the genetic determinants of disease resistance and the immunological mechanisms that control PRRSV. The immunogenetic loci, for good reason, are perhaps the most genetically diverse parts of the mammalian genome. In order to understand how genetics control disease resistance and susceptibility, we first must have a good understanding of the capacity of the immunogenetic loci to respond to disease and vaccination. Thus, the goal of this thesis was to characterize the antibody loci and describe their response to PRRSV.

Chapters 2 and 3 described the organization and allelic diversity for both of the porcine antibody light chain loci in a single Duroc sow. Evidence was found for recent evolution within the kappa locus via non-crossover homologous recombination, whereby IGKV gene segments may acquire new CDRs by templating off other gene segments. As a result, an extensive amount of allelic variation was found among members of the IGKV2 family. Within the lambda locus, further evidence of allelic variation was found (though not to the same extent). Intriguingly, a gene segment family (IGLV8) appears to have recently expanded, likely as a result of selective pressure. Gene segments from this family appear to be the most highly expressed IGLV gene segments in adult pigs and most prominently in fetal liver and yolk sac (Wertz et al. 2013; Chapters 3 and 4). In Chapter 4, at least three alleles of the gene segment IGLV3-1-1 were identified. Within

four of the five pigs investigated in that chapter, this gene segment contributed substantially to the expressed light chain repertoire. The remaining animal possessed a homozygous mutant form of this gene segment and its expressed lambda repertoire was considerably altered and reduced in total richness and diversity. Importantly, several highly abundant heavy and light chain clones, hypothesized as PRRSV-specific, were identified within the three PRRSV-infected pigs. The present study therefore represents the first characterization of the porcine antibody light chain loci and provides the first glimpse at the dynamics of the porcine heavy and light chain antibody repertoires as a result of infection, as well as their diversity, richness, and allelic variability.

There are some limitations of the present study that are worth some thought, however. The light chain loci were characterized from a single Duroc. As this result is based on a single individual, it remains unknown how similar this particular animal is to other swine, especially to other breeds. The amount of allelic variation observed in this animal was substantial and several alleles varied by as much as 10 percent. The kappa locus contained at least two large (>3kb) insertions on one chromosome and not the other, and the lambda locus contained a possible difference in gene segment order between the two SSC 14 chromosomes. This certainly highlights the tremendous genetic variability that is possible in these loci, but also suggests that even closely related swine may possess genetics that are quite different. Indeed, as observed in Chapters 4 and 5 near-normal distributions were seen for transcripts that were on average several amino acids different from the annotations, suggesting that the animals' antibody genetics were somewhat different from the Duroc that was sequenced for the pig genome sequencing

project and represented in Chapters 2 and 3. A limitation that directly stems from this is that it is nearly impossible to control for this genetic variation when attempting to determine the effect of infection on the antibody repertoire. Incidentally, one pig in particular appeared to respond the most to PRRSV in Chapters 4 and 5, with the most highly abundant clones all derived from that animal. Interestingly, this same animal was the one pig that lacked a functional copy of IGLV3-1-1. Perhaps as a compensatory mechanism, other gene segments such as IGLV3-3 were utilized more heavily in this pig. This genetic disruption may therefore have increased the ability of this pig to generate the observed highly clonal populations. A further conceivable consequence could be that any compensatory increased usage of other gene segments may actually enhance that animal's ability to respond, if those antibodies have greater specificity to the antigen in question (namely, PRRSV). Indeed, IGLV3-3 was represented by the two most abundant light chain sequences in the PRRSV-infected animals. The animals used for assessing heavy chain usage were not barcoded, and we therefore do not know exactly which of the PRRSV-specific antibody sequences came from which animal. However, a quantitative PCR analysis, using primers directed against the CDR3s, suggested that the most abundant heavy chain sequences were also from the same animal that most of the light chain responses were observed in (Pig 6 in Figure 5.8 and Pig 1 in Chapter 4).

Based on the identification of upstream flanking genes and analysis of several hundred kilobases of upstream genomic sequence which contained no additional IGLV gene segments, it is likely that the entire lambda locus has been characterized. The kappa locus, however, is likely incomplete. Attempts to identify additional kappa-specific

sequence in the pig genome failed. Two IGKV pseudogenes were found on BACs obtained from the same chromosome (SSC 3), however these BACs contained no additional immunoglobulin-related sequences. Fourteen IGKV gene segments out of an estimated upward limit of approximately 60 were identified. Additionally, as no upstream flanking gene segments were identified, it seems plausible to conclude that there are additional, as of yet, unidentified IGKV gene segments that remain to the discovered. The light chain transcriptome analyses in Chapter 4 also suggest that this is the case. While the average number of amino acid differences between lambda transcripts and the annotated genome were around two to five, that same number for kappa varied between six and ten amino acids. This suggests that the annotation for the kappa locus is incomplete, as it seems unlikely that somatic hypermutation would preferentially target one light chain and not the other.

Identifying new alleles and new gene segments is made difficult given the cloud of diversity created from both somatic hypermutation and junctional diversity. Although evidence was seen for some allelic variation, especially with regard to the three alleles of IGLV3-1-1. There is likely additional information on this genetic variation that is buried in the data. With respect to IGLV3-1-1, two of the pigs appear to possess intermediate levels of usage compared to the others. One interesting question raised is whether these pigs are heterozygous for the truncated pseudogene allele of IGLV3-1-1. Additional analyses would need to be carried out in order to confirm this.

In order to more accurately assess allelic variation, gene segments would need to be sequenced in their germline configuration. As the heavy chain locus and kappa locus

are currently incomplete, the lambda locus is the best candidate to study. As a future direction, Xi "Cassie" Guo is currently analyzing an Illumina MySeq dataset generated using genome-specific primers directed at each of the IGLV gene segments from a population of 81 pigs. Results from this study may provide considerable insight into the genetic variation in the lambda locus. Preliminary analyses suggest that there is substantial variation between individuals. Additional, future studies may investigate the impact of domestication on the evolution of the porcine antibody repertoire by sequencing more distantly related animals, such as wild boar (*Sus scrofa*), red river hog (*Potamochoerus porcus*), and Visayan warty pig (*Sus cebifrons*). Specimens of these animals are readily available at the local Minnesota Zoological Gardens. Of particular interest in this case would be the recently expanded and highly expressed IGLV8 gene segment family, the expression of which in swine is unique among artiodactyls. It may be hypothesized that this gene segment family represents a relatively recent resurrection of an ancestrally defunct, or poorly utilized family that only recently developed a selective advantage in swine.

Chapters 4 and 5 identified several highly abundant clonal heavy and light chain populations that are putatively PRRSV-specific. As a result of this, we have synthesized three scFv constructs (Integrated DNA Technologies, Inc.) by pairing the most abundant heavy and light chain variable region sequences. In particular, two distinct heavy chain sequences (IGHV3-4 with CDR3s of 15 and 22 amino acids as discussed in Chapter 5) were paired with the most abundant light chains (either IGLV3-3 or IGLV8-16). The heavy chain sequence with the 15aa CDR3 was paired with both light chains, while the

other was paired with IGLV3-3 only. A future direction is to clone these into a suitable expression vector (such as pET-25), and the expressed and purified scFvs will be assayed for PRRSV-specificity using recombinant PRRSV proteins and ELISA. As only a few highly abundant clones were identified it would be hypothesized that these particular clones are specific against the most immuno-dominant PRRSV epitopes, such as those on non-structural protein (nsp)2 and nsp7. Ideally, all four combinations of heavy and light chains would have been generated; however, due to cost limitations, only these three were synthesized. However, these three constructs should be sufficient to provide preliminary support for the supposition that the most abundant heavy and light chain clones are PRRSV-specific.

This approach to antigen-specific scFv development could additionally be used to determine the contribution of each chain in generating antibody affinity. By synthesizing scFv combinations of light and heavy chain pairs in which either chain is held constant and measuring their affinity using a Biacore-based system, it would be possible to quantify the contribution of the heavy chain (and potentially individual CDRs) on antibody-binding. Given that camelid heavy chain isotypes IgG2 and IgG3 are incapable of heterodimerizing with light chains, the observation that a single heavy chain CDR3 is by itself capable of cross-neutralizing influenza virus (Ekiert et al. 2012), and the lack of junctional diversity (and overall diversity) in the light chain, it could be hypothesized that porcine heavy chains are capable of binding antigen on their own.

In addition, gene segment family specific primers have been designed to amplify the expressed porcine antibody repertoire for the purpose of developing scFv antibody

surface display libraries. As an additional and planned future direction, an scFv library will be generated from animals with high anti-PRRSV neutralizing antibody titers, expressed on the surface of yeast, and screened for clones that bind fluorescently-labeled PRRSV using fluorescence-assisted cell sorting (FACS). The identified PRRSV-specific clones will then be assayed for their neutralizing ability using an ELISA-based neutralization assay previously developed in our lab and their characteristics will be investigated (e.g. presence and extent of SHM, framework mutations, insertions/deletions, and size and composition of heavy chain CDR3). Furthermore, this technology has the potential to be used to generate recombinant scFvs that are able to target virtually any conceivable antigen of interest. This work is currently being carried forward by Dr. Sally Robinson.

In conclusion, this study was the first to characterize the porcine antibody light chain loci and illuminate the porcine antibody repertoire. Future endeavors to understand the role of antibody genetics in porcine disease resistance and efforts to identify mechanisms of virus neutralization will build on this current work.

REFERENCES

Ait-Ali T, Wilson AD, Wescott DG, Clapperton M, Mellencamp MA, Drew TW, Bishop SC, and Archibald AL (2007) Innate immune responses to replication of porcine reproductive and respiratory syndrome virus in isolated porcine alveolar macrophages. Viral Immunol 20:105-118.

Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410.

Andabaka T, Nickerson JW, Rojas-Reyes MX, Reuda JD, Bacic Vrca V, Barsic B (2013) Monoclonal antibody for reducing the risk of respiratory syncytial virus infection in children. Cochrane Database Syst Rev CD006602.

Arakawa H, Furusawa S, Ekino S, Yamagishi H (1996) Immunoglobulin gene hyperconversion ongoing in chicken splenic germinal centers. EMBO J 15:2540-2546.

Arakawa H, Hauschild J, Buerstedde J-M (2002) Requirement of the activation-induced deaminase gene for immunoglobulin gene conversion. Science 295:1301-1306.

Arakawa H, Saribasak H, Buerstedde J-M (2004) Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. PLoS Biol 2(7):e179.

Archibald AL, Bolund L, Churcher C, Fredholm M, Groenen MAM, Harlizius B, Lee KT, Milan D, Rogers J, Rothschild MF, Uenishi H, Wang J, Schook LB, the Swine Genome Sequencing Consortium (2010) Pig genome sequence – analysis and publication strategy. BMC Genomics 11:438.

Arun SS, Breuer W, and Hermanns W (1996) Immunohistochemical examination of light-chain expression (lambda/kappa ratio) in canine, feline, equine, bovine and porcine plasma cells. Zentralbl Veterinarmed A 43(9):573-576.

Bain B, Brazil M (2003) Adalimumab. Nat Rev Drug Discov 2:693-694.

Barclay AN (2003) Membrane proteins with immunoglobulin-like domains – a master superfamily of interaction molecules. Semin Immuol 15:215-223.

Bayne-Jones S and Wilson DW (1922) Immunological reactions of Bence-Jones proteins. II. Differences between Bence-Jones proteins from various sources. Johns Hopk Hosp Bull 32:119.

Becker RS and Knight KL (1990) Somatic diversification of immunoglobulin VDJ genes: evidence for somatic gene conversion in rabbits. Cell 63(5):987-997.

Behring E and Kitasato S (1890) Über das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren [On the acquisition of immunity to diphtheria and tetanus in animals]. Dtsch Med Wochenschr 16:1113-1114.

Bentley DL and Rabbitts TH (1983) Evolution of V genes: evidence indicating that recently duplicated human VK sequences have diverged by gene conversion. Cell 32:181-189.

Bergman LW and Kuehl WM (1979) Formation of an intrachain disulfide bond on nascent immunoglobulin light chains. J Biol Chem 254:8869-8876.

Berman JE, Nickerson KG, Pollock RR, Barth JE, Schuurman RK, Knowles DM, Chess L, and Alt FW (1991) VH usage in humans: biased usage of the VH6 gene in immature B lymphoid cells. Eur J Immunol 21:1311-1314.

Binns RM (1982) Organisation of the lymphoreticular system and lymphocyte markers in the pig. Vet Immunol Immunopathol 3:95-146.

Blankenberg D, Taylor J, Schenck I, He J, Zhang Y, Ghent M, Veeraraghavan N, Albert I, Miller W, Makova KD, Hardison RC, and Nekrutenko A (2007) A framework for collaborative analysis of ENCODE data: making large-scale analyses biologist-friendly. Genome Res 17(6):960-4.

Boyd SD, Gaëta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, Jones CD, Simen BB, Hanczaruk B, Nguyen KD, Nadeau KC, Egholm M, Miklos DB, Zehnder JL, Collins AM (2010) Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. J Immunol 184:6986-6992.

Bransteitter R, Pham P, Calabrese P, Goodman MF (2004) Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. J Biol Chem 279:51612-51621.

Breman JG and Arita I (1980) The confirmation and maintenance of smallpox eradication. N Engl J Med 22:1267-1273.

Burton DR (2002) Antibodies, viruses and vaccines. Nat Rev Immunol 2:706-713.

Butler JE (1997) Immunoglobulin gene organization and the mechanism of repertoire development. Scand J Immunol 45(5):455-462.

Butler JE, Lager KM, Splichal I, Francis D, Kacskovics I, Sinkora M, Wertz N, Sun J, Zhao Y, Brown WR, DeWald R, Dierks S, Muyldermans S, Lunney JK, McCray PB, Rogers CS, Welsh MJ, Navarro P, Klobasa F, Habe F, and Ramsoondar J (2009a) The piglet as a model for B cell and immune system development. Vet Immunol Immunop 128:147-70.

Butler JE, Sinkora M, Wertz N, Holtmeier W, and Lemke CD (2006a) Development of the neonatal B and T cell repertoire in swine: implications for comparative and veterinary immunology. Vet Res 37:417-441.

Butler JE, Sun J, Wertz N, and Sinkora M (2006b) Antibody repertoire development in swine. Dev Comp Immunol 30:199-221.

Butler JE, Weber P, Sinkora M, Sun J, Ford SJ, and Christenson R (2000) Antibody repertoire development in fetal and neonatal piglets. II. Characterization of heavy chain complementarity-determining region 3 diversity in the developing fetus. J Immunol 165:6999-7011.

Butler JE, Weber P, and Wertz N (2006c) Antibody repertoire development in fetal and neonatal piglets. XIII. Hybrid VH genes and the preimmune repertoire revisited. J Immunol 177:5459-5470.

Butler JE and Wertz N (2012) The porcine antibody repertoire: variations on the textbook theme. Front Immunol 3:153

Butler JE, Wertz N, Deschacht N, and Kacskovics I (2009b) Porcine IgG: structure, genetics, and evolution. Immunogenetics 61:209-230.

Butler JE, Wertz N, Sun J (2013) Antibody repertoire development in fetal and neonatal piglets. XIV. Highly restricted IGKV gene usage parallels the pattern seen with IGLV and IGHV. Mol Immunol 55(3-4):329-336.

Butler JE, Wertz N, Sun J, Wang H, Lemke C, Chardon P, Piumi F, and Wells K (2005) The pre-immune variable kappa repertoire of swine is selectively generated from certain subfamilies of Vκ2 and one Jκ gene. Vet Immunol Immunop 108:127-137.

Butler JE, Wertz N, Sun J, Wang H, Chardon P, Piumi F, and Wells K (2004) Antibody repertoire development infetal and neonatal pigs.VII. Characterization of the preimmune kappa light chain repertoire. J Immunol 173:6794-6805.

Casellas R, Nussenzweig A, Wuerffel R, Pelanda R, Reichlin A, Suh H, Qin XF, Besmer E, Kenter A, Rajewsky K, Nussenzweig MC (1998) Ku80 is required for immunoglobulin isotype switching. EMBO J 17(8):2404-2411.

Centers for Disease Control and Prevention [CDC] (2012) Progress toward interruption of wild poliovirus transmission - worldwide, January 2011-March 2012. Morb Mortal Wkly Rep 61:353-357.

Chames P, van Regenmortel M, Weiss E, Baty D (2009) Therapeutic antibodies: successes, limitations and hopes for the future. J Pharmacol 157(2):220-233.

Chao A (1984) Nonparametric estimation of the number of classes in a population. Scand J Statist 11:265-70.

Chao A and Lee S-M (1992) Estimating the number of classes via sample coverage. J Am Stat Assoc 87:210-17.

Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW (2003) Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature 422:726-730.

Cohen JB and Givol D (1983) Allelic immunoglobulin VH genes in two mouse strains: possible germline gene recombination. EMBO J 2:2013-2018.

Colwell RK (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. Accessed: April 14, 2013.

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. Phil Trans R Soc Lond B 345:101-118.

Das S, Nikolaidis N, and Nei M (2009) Genomic organization and evolution of immunoglobulin kappa gene enhancers and kappa deleting element in mammals. Mol Immunol 46:3171-3177.

Dawson HD, Loveland J, Pascal G, Gilbert J, Uenishi H, Mann K, Sang Y-M, Zhang J, Carvalho-Silva D, Hunt T, Hardy M, Hu Z, Zhao S-H, Anselmo A, Sinkai H, Chen C, Badaoui B, Berman D, Amid C, Kay M, Lloyd D, Snow C, Morozumi T, RP-Y Cheng, Bystrom M, Bourneuf E, Kapetanovic R, Schwartz JC, Kataria R, Astley M, Fritz E, Steward C, Thomas M, Wilming L, Giuffra E, Archibald A, Bed'Home B, Beraldi D, Ait-Ali T, Blecha F, Botti S, Freeman T, Hume D, Lunney JK, Murtaugh MP, Reecy JM, Harrow J, Rogel-Gaillard C, and Tuggle CK (2013) Structural and Functional Annotation of the Porcine Immunome. BMC Genomics 14:332.

Di Noia JM, Neuberger MS (2007) Molecular mechanisms of antibody somatic hypermutation. Ann Rev Biochem 76:1-22.

Durandy A (2003) Activation-induced cytidine deaminase: a dual role in class-switch recombination and somatic hypermutation. Eur J Immunol 33:2069-2073.

Edelman GM, Gally JA (1962) The nature of Bence-Jones proteins. Chemical similarities to polypeptide chains of myeloma globulins and normal gamma-globulins. J Exp Med 116:207-227.

Edwards S, Fukusho A, Lèfevre P-C, Lipowski A, Pejsak Z, Roehe P, and Westergaard J (2000) Classical swine fever: the global situation. Vet Microbiol 73:103-119.

Eguchi-Ogawa T, Toki D, Wertz N, Butler JE, and Uenishi H (2012) Structure of the genomic sequence comprising the immunoglobulin heavy constant (IGHC) genes from *Sus scrofa*. Mol Immunol 52:97-107.

Eguchi-Ogawa T, Wertz N, Sun XZ, Puimi F, Uenishi H, Wells K, Chardon P, Tobin GJ, and Butler JE (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 184:3734-3742.

Ehrlich, P (1897) Die Wertbemessung des Diphterieheilserums und deren theoretische Grundlagen [The valuation of diphtheria anti-serum and its theoretical basis]. Klin Jahr 6:299-326.

Ekiert DC, Kashyap AK, Steel J, Rubrum A, Bhabha G, Khayat R, Lee JH, Dillon MA, O'Neil RE, Faynboym AM, Horowitz M, Horowitz L, Ward AB, Palese P, Webby R, Lerner RA, Bhatt RR, Wilson IA (2012) Cross-neutralization of influenza A viruses mediated by a single antibody loop. Nature 489:526-32.

Ekman A, Niku M, Liljavirta J, and Iivanainen A (2009) *Bos taurus* genome sequence reveals the assortment of immunoglobulin and surrogate light chain genes in domestic cattle. BMC Immunol 10:22.

Elsik CG, Tellam RL, Worley KC, et al. (150+ authors) (2009) The genome sequence of tuarine cattle: a window into ruminant biology and evolution. Science 324(5926):522-528.

Emorine L and Max EE (1983) Structural analysis of a rabbit immunoglobulin kappa 2 J-C locus reveals multiple deletions. Nucl Acids Res 11:8877-8890.

FAO (Food and Agriculture Organization of the United Nations), Economic and Social Development Department (2009) FAO Statistical Yearbook 2009, Vol. 4. Available at: http://www.fao.org/docrep/014/am079m/am079m00.htm#Contents_en. Accessed: April 4, 2013.

Fischer E (1894) Einfluss der Configuration auf die Wirkung der Enzyme [Effect of configuration on the action of enzymes]. Ber Dtsch Chem Ges 27:2985-2993.

Fischer N (2011) Sequencing antibody repertoires: the next generation. MAbs 3(1):17-20.

Flanagan JG, Lefranc M-P, and Rabbitts TH (1984) Mechanisms of divergence and convergence of the human immunoglobulin alpha1 and alpha2 constant region gene sequences. Cell 36:681-688.

Frippiat JP, Williams SC, Tomlinson IM, Cook GP, Cherif D, Le Paslier D, Collins JE, Dunham I, Winter G, and Lefranc M-P (1995) Organization of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2. Hum Mol Genet 4:983-991.

Frönicke L, Chowdhary BP, Schertan H, and Gustavsson I (1996) A comparative map of the porcine and human genomes demonstrates ZOO-FISH and gene mapping-based chromosomal homologies. Mamm Genom 7:285-290.

Gellert M (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. Annu Rev Biochem 71:101-132.

Gennery AR (2006) Primary immunodeficiency syndromes associated with defective DNA double-strand break repair. Br Med Bull 77-78:71-85.

Giudicelli V, Chaume D, and Lefranc, M-P (2005) IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucl Acids Res 33:D256-261.

Giudicelli V, Duroux P, Ginestoux C, Folch G, Jabado-Michaloud J, Chaume D and Lefranc M-P (2006) IMGT/LIGM-DB, the IMGT® comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. Nucl Acids Res 34:D781-784.

Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, Ni I, Mei L, Sundar PD, Day GMR, Cox D, Rajpal A, and Pons J (2009) Precise determination of the diversity of a combinatorial antibody library gives insight into the human antibody repertoire. Proc Nat Acad Sci USA 106(48):20216-20221.

Goodman DJ, Gaff C, Gerondakis S (1993) The IL-4 induced increase in the frequency of resting murine splenic B cells expressing germline Ig heavy chain gamma 1 transcripts correlates with subsequent switching to IgG1. Int Immunol 5(2):199-208.

Gorodkin J, Cirera S, Hedegaard J, Gilchrist MJ, Panitz F, Jørgensen C, Scheibye-Knudsen K, Arvin T, Lumholdt S, Sawera M, Green T, Nielsen BJ, Havgaard JH, Rosenkilde C, Wang J, Li H, Li R, Liu B, Hu S, Dong W, Li W, Yu J, Wang J, Stærfeldt H, Wernersson R, Madsen LB, Thomsen B, Hornshøj H, Bujie Z, Wang X, Wang X, Bolund L, Brunak S, Yang H, Bendixen C, and Fredholm M (2007) Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags. Genome Biol 8:R45.

Gotelli NJ and Entsminger GL (2011) EcoSim: Null models software for ecology. Version 7. Acquired Intelligence Inc. & Kesey-Bear. Jericho, VT 05465. http://garyentsminger.com/ecosim.htm. Accessed: April 14, 2013.

Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-F, Li S, Larkin DM, Kim H, Frantz LAF, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse1 M, Botti S, Bujie Z, Bystrom M, Capitanu B, Carvalho-Silva D, Chardon P, Chen C, Cheng R, Choi S-H, Chow W, Clark RC, Clee C, Crooijmans RPMA, Dawson HD, Dehais P, De Sapio F, Dibbits B, Drou N, Du Z-Q, Eversole K, Fadista J, Fairley S, Faraut T, Faulkner GF, Fowler KE, Fredholm M, Fritz E, Gilbert JGR, Giuffra E, Gorodkin J, Griffin DK, Harrow JL, Hayward A, Howe K, Hu Z-L, Humphray SJ, Hunt T, Jensen HH, Jeon JT, Jern P, Jones M, Jurka J, Kanamori H, Kapetanovic R, Kim J, Kim J-H, Kim K-W, Kim T-H, Larson G, Lee K, Lee K-T, Leggett R, Lewin HA, Li Y, Liu W, Loveland JE, Lu Y, Lunney JK, Ma J, Madsen O, Mann K, Matthews L, McLaren S, Morozumi T, Murtaugh MP, Narayan J, Nguyen DT, Ni P, Oh S-J, Onteru S, Panitz F, Park E-W, Park H-S, Pascal G, Paudel Y, Perez-Enciso M, Ramirez-Gonzalez R, Reecy JM, Rodriguez-Zas S, Rohrer GA, Rund L, Sang Y, Schachtschneider K, Schraiber JG, Schwartz JC, Scobie L, Scott C, Searle S, Servin B, Southey BR, Sperber G, Stadler P, Sweedler J, Tafer H, Thomsen B, Wali R, Wang J, Wang J, White S, Xu X, Yerle M, Zhang G, Zhang J, Zhang J, Zhao S, Rogers J, Churcher C and Schook LB (2012) Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491(7424):393-8.

Halbur PG, Rothschild MF, Thacker BJ, Meng XJ, Paul PS, and Bruna JD (1998) Differences in susceptibility of Duroc, Hampshire, and Meishan pigs to infection with a high virulence strain (VR2385) of porcine reproductive and respiratory syndrome virus (PRRSV). J Anim Breed Genet 115:181-189.

Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hammers C, Bajyana Songa E, Bendahman N, and Hammers R (1993) Naturally occurring antibodies devoid of light chains. Nature 363:446-448.

Hieter PA, Korsmeyer SJ, Waldman TA, and Leder P (1981) Human immunoglobulin kappa light-chain genes are deleted or rearranged in lambda-producing B cells. Nature 290:368-372.

Hieter PA, Max EE, Seidman JG, Maizel JV, and Leder P (1980) Cloned human and mouse kappa immunoglobulin constant and J region genes conserve homology in functional segments. Cell 22:197-207.

Holtkamp DJ, Kliebenstein JB, Neumann EJ, Zimmerman JJ, Rotto H, Yoder TK, Wang C, Yeske P, Mowrer C, Haley C (2011) Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on U.S. pork producers. *In*: Proceedings of the 92nd Annual Meeting of the Conference of Research Workers in Animal Disease. Chicago, IL. Dec. 4-6, 2011. No. 69.

Hood L, Gray WR, Sanders BG, and Dreyer WJ (1967) Light chain evolution. Cold Spring Harbor Symp Quant Biol 32:133-146.

Horzinek MC (2011) Rinderpest: the second viral disease eradicated. Vet Microbiol 149:295-297.

Hsu E, Pulham N, Rumfelt LL, and Flajnik MF (2006) The plasticity of immunoglobulin gene systems in evolution. Immunol Rev 210:8-26.

Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, and Clamp M (2002) The Ensembl genome database project. Nucleic Acids Res 30:38-41.

Huck S, Lefranc G, and Lefranc M-P (1989) A human immunoglobulin IGHG3 allele (Gmb0, b1, c3, c5, u) with an IGHG4 converted region and three hinge exons. Immunogenetics 30:250-257.

Humphray SJ, Scott CE, Clark R, Marron B, Bender C, Camm N, Davis J, Jenks A, Noon A, Patel M, Sehra H, Yang F, Rogatcheva MB, Milan D, Chardon P, Rohrer G, Nonneman D, de Jong P, Meyers SN, Archibald A, Beever JE, Schook LB, and Rogers J (2007) A high utility map of the pig genome. Genome Biol 8:R139.

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, and Rupp R (2007) Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinform 8:460.

Juneja RK, Gahne B, Rapacz J, and Hasler-Rapacz J (1986) Linkage between the porcine genes encoding immunoglobulin heavy-chain allotypes and some serum alpha-protease inhibitors: a conserved linkage in pig, mouse, and human. Anim Genet 17(3)225-233.

Jung D, Giallourakis C, Mostoslavsky R, and Alt FW (2006) Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. Annu Rev Immunol 24:541-70.

Kabat EA, Wu TT, Perry H, Gottesman K, and Foeller C (1991) Sequences of proteins of immunological interest, fifth edition. NIH publication No. 91-3242.

Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Asakawa S, Sasaki T, Klobeck HG, Combriato G, Zachau HG, and Shimizu N (2001) Evolutionary dynamics of the human immunoglobulin κ locus and the germline repertoire of the Vκ genes. Eur J Immunol 31:1017-1028.

Kawasaki K, Minoshima S, Schooler K, Kudoh J, Asakawa S, de Jong PJ, and Shimizu N (1995) The organization of the human immunoglobulin λ gene locus. Genome Res 5:125-135.

Kim DR and Oettinger MA (2000) V(D)J recombination: site-specific cleavage and repair. Mol Cells 10:367-374.

Klein F, Halper-Stromberg A, Horwitz JA, Gruell H, Scheid JF, Bournazos S, Mouquet H, Spatz LA, Diskin R, Abadir A, Zang T, Dorner M, Billerbeck E, Labitt RN, Gaebler C, Marcovecchio PM, Incesu R-B, Eisenreich TR, Bieniasz PD, Seaman MS, Bjorkman PJ, Ravetch JV, Ploss A, Nussenzweig MC (2012) HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. Nature 492:118-122.

Klinge KL, Vaughn EM, Roof MB, Bautista EM, Murtaugh MP (2009) Age-dependent resistance to porcine reproductive and respiratory syndrome virus replication in swine. Virol J 6:177.

Knight KL and Becker RS (1990) Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. Cell 60:963-970.

Kehoe JM and Capra JD (1974) Nature and significance of immunoglobulin subclasses. N Y State J Med 74:489-491

Korngold L and Lipari R (1956) Multiple myeloma proteins. III. The antigenic relationship of Bence-Jones proteins to normal gamma-globulin and multiple myeloma serum proteins. Cancer 9:262-272.

Kumar S and Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392(6679):917-920.

Kurosawa K and Ohta K (2011) Genetic diversification by somatic gene conversion. Genes 2:48-58.

Lager KM, Mengeling WL, and Brockmeier SL (1997) Duration of homologous porcine reproductive and respiratory syndrome virus immunity in pregnant swine. Vet Microbiol 58:127-133.

Lammers BM, Beaman KD, and Kim YB (1991) Sequence analysis of porcine immunoglobulin light chain cDNAs. Mol Immunol 28:877-880.

Lefranc M-P (2007) WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report. Immunogenetics 59:899-902.

Lefranc M-P (2008) WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report. Immunoglobulins and T cell receptors report August 2007, 13th International Congress of Immunology, Rio de Janeiro, Brazil. Dev Comp Immunol 32:461-463.

Lefranc M-P and Lefranc G (2001) The Immunoglobulin FactsBook. Academic Press, London, UK pp 1-458.

Lefranc M-P, Helal AN, De Lange G, Chaabani H, Van Loghem E, and Lefranc G (1986) Gene conversion in human immunoglobulin gamma locus shown by unusual location of IgG allotypes. FEBS Letters, 196, 96-102 (1986).

Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, and Lefranc G (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol 27:55-77.

Lefranc M-P, Pommié C, Kaas Q, Duprat E, Bosc N, Guiraudou D, Jean C, Ruiz M, Da Piédade I, Rouard M, Foulquier E, Thouvenin V, and Lefranc G (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Dev Comp Immunol 29:185-203.

Lefranc M-P, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, Wu Y, Gemrot E, Brochet X, Lane J, Regnier L, Ehrenmann F, Lefranc G, and

Duroux P (2009) IMGT®, the international ImMunoGeneTics information system®. Nucl Acids Res 37:D1006-1012.

Lewis CRG, Ait-Ali T, Clapperton M, Archibald AL, and Bishop S (2007) Genetic perspectives on host responses to porcine reproductive and respiratory syndrome (PRRS). Viral Immunol 20(3):343-358.

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

Loemba HD, Mounir S, and Mardassi H (1996) Kinetics of humoral immune response to the major structural proteins of the porcine respiratory and reproductie syndrome virus. Arch Virol 141:751-761.

Lohse L, Nielson J, and Erikson L (2004) Temporary CD8+ T-cell depletion in pigs does not exacerbate infection with porcine reproductive and respiratory syndrome virus (PRRSV). Viral Immunol 17:594-603.

Luo Z, Ronai D, Scharf MD (2004) The role of activation-induced cytidine deaminase in antibody diversification, immunodeficiency, and B-cell malignancies. J Allergy Clin Immunol 114(4):726-35.

Malynn BA, Yancopoulos GD, Barth JE, Bona CA, and Alt FW (1990) Biased expression of JH-proximal VH genes occurs in the newly generated repertoire of neonatal and adult mice. J Exp Med 171:843-859.

Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, Miyata T, and Honjo T (1998) The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J Exp Med 188(11):2151-2162.

McBlane JF, Van Gent DC, Ramsden DA, Romeo C, Cuomo CA, Gellert M, and Oettinger MA (1995) Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell 83:387-395.

Meier W, Wheeler J, Husmann RJ, Osario F, and Zuckermann FA (2000) Characteristics of the immune response of pigs to PRRS virus. Vet Res 31:41.

Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão TLL, Stadler T, Rabosky DL, Honeycutt RL, Flynn JJ, Ingram CM, Steiner C, Williams TL, Robinson TJ, Burk-Herrick A, Westerman M, Ayoub NA, Springer MS, and Murphy WJ (2011) Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334(6055):521-524.

Middleton D and Gonzalez F (2009) The extensive polymorphism of KIR genes. Immunology 129(1):8-19.

Molina RM, Cha SH, Chittick W, Lawson S, Murtaugh MP, Nelson EA, Christoper-Hennings J, Yoon KJ, Evans R, Rowland RR, and Zimmerman JJ (2008) Immune response against porcine reproductive and respiratory syndrome virus during acute and chronic infection. Vet Immunol Immunopathol 126:283-292.

Moore MW, Durdik J, Persiani DM, and Selsing E (1985) Deletions of kappa chain constant region genes in mouse lambda chain-producing B cells involve intrachromosomal DNA recombinations similar to V-J joining. Proc Natl Acad Sci U.S.A. 82:6211-6215.

Mulupuri P, Zimmerman JJ, Hermann J, Johnson CR, Cano JP, Yu W, Dee SA, and Murtaugh MP (2008) Antigen-specific B-cell responses to porcine reproductive and respiratory syndrome virus infection. J Virol 82(1):358-370.

Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102(5):553-563.

Muramatsu M, Sankranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of teh RNA-editing deaminase family in germinal center B-cells. J Biol Chem 274(26):18470-18476.

Murtaugh MP, Xiao Z, and Zuckermann F (2002) Immunological responses of swine to porcine reproductive and respiratory syndrome virus infection. Viral Immunol 15(4):533-547.

Nelson EA, Christopher-Hennings J, and Benfield D (1994) Serum immune response to the proteins of porcine reproductive and respiratory syndrome (PRRS) virus. J Vet Diagn Invest 6:410-415.

Niku M, Liljavirta J, Durkin K, Schroderus E, and Iivanainen A (2012) The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed. Dev Comp Immunol 37:457-461.

Pace CS, Fordyce MW, Franco D, Kao CY, Seaman MS, Ho DD (2013) Anti-CD4 monoclonal antibody ibalizumab exhibits breadth and potency against HIV-1, with natural resistance mediated by the loss of a V5 glycan in envelope. J Acquir immune Defic Syndr 62(1):1-9.

Parslow TG, Blair DL, Murphy WJ, and Granner DK (1984) Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. Proc Natl Acad Sci U.S.A. 81:2650-2654.

Pasman Y and Kaushik AK (2012) Partial organization of bovine variable-heavy chain gene locus and influence of recombination signal sequences (RSS) on various variable region gene expression. J Immunol 188:42.6 (Abstr).

Pasman Y, Saini SS, Smith E, and Kaushik AK (2010) Organization and genomic complexity of bovine λ-light chain gene locus. Vet Immunol Immunop 135:306-313.

Pham P, Bransteitter R, Petruska J, Goodman MF (2003) Processive AID-catalyzed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature 424:103-107.

Poljak RJ, Amzel LM, Avey HP, Chen BL, Phizackerley RP, and Saul F (1973) Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8-Å resolution. Proc Nat Acad Sci USA 70(12):3305-3310.

Prabakaran P, Chen W, Singarayan MG, Stewart CC, Streaker E, Feng Y, and Dimitrov DS (2011) Expressed antibody repertoires in human cord blood cells: 454 sequencing and IMGT/High V-QUEST analysis of germline gene usage, junctional diversity, and somatic mutations. Immunogenetics 64(5):337-350.

Price SA, Bininda-Edmonds ORP, and Gittleman JL (2005) A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla). Biol Rev 80:445-473.

Rada C, Ehrenstein MR, Neuberger MS, Milstein C (1998) Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. Immunity 9(1):135-141.

Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS (2002) Immunoglobulin class switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. Curr Biol 12:1748-1755.

Ramsden DA, Baetz KA, and Wu GE (1994) Conservation of sequence in recombination signal sequence spacers. Nucleic Acids Res 22:1785-1796.

Reddy ST, Ge X, Miklos AE, Hughes RA, Kang SH, Hoi KH, Chrysostomou C, Hunicke-Smith SP, Iverson BL, Tucker PW, Ellington AD, and Georgiou G (2010) Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. Nat Biotech 28(9):965-969.

Reiner, G, Willems H, Pesch S, and Ohlinger VF (2010) Variation in resistance to the porcine reproductive and respiratory syndrome virus (PRRSV) in Pietrain and miniature pigs. J Anim Breed Genet 127(2):100-106.

Rettenberger G, Bruch J, Leeb T, Brenig B, Klett C, and Hameister H (1996) Assignment of pig immunoglobulin kappa gene IGKC to chromosome 3q12-q14 by fluorescence in situ hybridization (FISH). Mamm Genome 7:324-325.

Reynaud C-A, Anquez V, Dahan A, and Weill J-C (1985) A single rearrangement event generated most of the chicken immunoglobulin light chain diversity. Cell 40:283-291.

Reynaud C-A, Anquez V, Grimal H, Weill J-C (1987) A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell 48:379-388.

Reynaud C-A, Dahan A, Anquez V, and Weill J-C (1989) Somatic hyperconversion diversifies the single VH gene of the chicken with a high incidence in the D region. Cell 59:171-183.

Rice P, Longden I, and Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16(6):276-7.

Rowland RR, Lunney J, and Dekkers J (2012) Control of porcine reproductive and respiratory syndrome (PRRS) through genetic improvements in disease resistance and tolerance. Front Genet 3:260.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *In*: Krawetz S, Mesner S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-86.

Ruiz M, Giudicelli V, Ginestoux C, Stoehr P, Robinson J, Bodmer J, Marsh SGE, Bontrop R, Lemaitre M, Lefranc G, Chaume D, and Lefranc MP (2000) IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 28:219-221.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, and Barrell B (2000) Artemis: sequence visualization and annotation. Bioinformatics 16:944-945.

Sanchez P, Nadel B, and Cazenave PA (1991) V lambda-J lambda rearrangements are restricted within a V-J-C recombination unit in the mouse. Eur J Immunol 21:907-911.

Sanchez-Mazas A, Fernandez-Viña M, Middleton D, Hollenbach JA, Buhler S, Di D, Rajalingam R, Dugoujon JM, Mack SJ, and Thorsby E (2011) Immunogenetics as a tool in anthropological studies. Immunol 133(2):143-164.

Sasso EH, van Dijk KW, Bull A, van der Maarel SM, and Milner ECB (1992) VH genes in tandem array comprise a repeated germline motif. J Immunol 149(4):1230-1236.

Schook LB, Beever JE, Rogers J, Humphray S, Archibald A, Chardon P, Milan D, Rohrer G, and Eversole K (2005) Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome. Comp Funct Genom 6:251-255.

Schwartz JC, Lefranc M-P, and Murtaugh MP (2012) The organization, complexity and allelic diversity of the porcine (*Sus scrofa domesticus*) immunoglobulin lambda locus. Immunogenetics 64(5):399-407.

Schwartz JC, Lefranc M-P, and Murtaugh MP (2012) Evolution of the porcine (*Sus scrofa domesticus*) immunoglobulin kappa locus through germline gene conversion. Immunogenetics 64(4):303-311.

Shendure J and Hanlee J (2008) Next-generation DNA sequencing. Nat Biotech 26:1135-45.

Sheppard HW and Gutman GA (1981) Allelic forms of rat kappa chain genes: evidence for strong selection at the level of nucleotide sequence. Proc Nat Acad Sci USA 78:7064-7068.

Shojaei F, Saini SS, Kaushik AK (2003) Unusually long germline D_H genes contribute to large sized CDR3H in bovine antibodies. Mol Immunol 40(1):61-67.

Siminovitch KA, Bakhshi A, Goldman P, and Korsmeyer SJ (1985) A uniform deleting element mediates the loss of kappa genes in human B cells. Nature 316:260-262.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, and Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117-1123.

Sinkora J, Rehakova Z, Samankova L, Haverson K, Butler JE, Zwart R, and Boersma W (2001) Characterization of monoclonal antibodies recognizing immunoglobulin κ and λ chains in pigs by flow cytometry. Vet Immunol Immunopath 80:79-91.

Sircar A, Sanni KA, Shi J, Gray JJ (2011) Analysis and modeling of the variable region of camelid single domain antibodies. J Immunol 186(11):6357-6367.

Sun J, Butler JE (1997) Sequence analysis of pig switch mu, C mu, and C mu m. Immunogenetics 46(6):452-460.

Sun J, Kacskovics I, Brown WR, and Butler JE (1994) Expressed swine V_H genes belong to a small V_H gene family homologous to human V_HIII. J Immunol 153:5618-5627.

Sun X, Wertz N, Lager K, Sinkora M, Stepanova K, Tobin G, and Butler JE (2012) Antibody repertoire development in fetal and neonatal piglets. XXII. λ rearrangement precedes κ rearrangement during B-cell lymphogenesis in swine. Immunol 137(2):149-59.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. Nucleic Acids Res 22(22):4673-4680.

Tiselius A (1937) Electrophoresis of serum globulin: electrophoretic analysis of normal and immune sera. Biochem J 31(9):1464-1477.

United States Department of Agriculture. Animal and Plant Health Inspection Service [USDA-APHIS] (2008) Pseudorabies (Aujeszky's disease) and its eradication: a review of the U.S. experience. Available at: http://www.aphis.usda.gov/publications/animal_health/content/printable_version/pseudo_rabies_report.pdf

Wang Y, Jackson KJ, Gäeta B, Pomat W, Siba P, Sewell WA, and Collins AM (2011) Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants. Immunogenetics 63(5):259-265.

Wang Y, Jackson KJ, Sewell WA, and Collins AM (2008) Many human immunoglobulin heavy-chain IGHV gene polymorphisms have been reported in error. Immunol Cell Biol 86(2):111-115.

Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong PD, Shaw GM (2003) Antibody neutralization and escape by HIV-1. Nature 422:307-312.

Weinstein JA, Jiang N, White RA 3rd, Fisher DS, and Quake SR (2009) High-throughput sequencing of the zebrafish antibody repertoire. Science 324:807-810.

Wertz N, Vazquez J, Wells K, Sun J, Butler JE (2013) Antibody repertoire development in fetal and neonatal piglets. XII. Three IGLV genes comprise 70% of the pre-immune repertoire and there is little junctional diversity. Mol Immunol 55:319-328.

Wu TT and Kabat EA (1970) An analysis of the sequences of the variable regions of Bence Jones proteins and their myeloma light chains and their implications for antibody complementarity. J Exp Med 132:211-250.

Xiao Z, Batista L, Dee S, Halbur P, and Murtaugh MP (2004) The level of virus-specific T cell and macrophage recruitment in porcine reproductive and respiratory syndrome virus infection in pigs is independent of virus load. J Virol 78:5923-5933.

Yancopolous GD, Desiderio SV, Paskind M, Kearney JF, Baltimore D, and Alt FW (1984) Preferential utilization of the most JH-proximal VH gene segments in pre-B-cell lines. Nature 311:727-733.

Yerle M, Lahbib-Mansais Y, Pinton P, Robic A, Goureau A, Milan D, and Gellin J (1997) The cytogenetic map of the domestic pig (*Sus scrofa domestica*). Mamm Genome 8:592-607.

Zerbino DR and Birney E (2008) Velvet: algorithms for de novo short read assembly using de bruijn graphs. Genome Res 18:821-829.