Published in final edited form as:

Immunogenetics. 2016 April ; 68(4): 285–293. doi:10.1007/s00251-016-0899-9.

Genomic variation in the porcine immunoglobulin lambda variable region

Xi Guo¹, John C. Schwartz^{1,2}, and Michael P. Murtaugh¹

¹ Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, 1971 Commonwealth Avenue, St. Paul, MN 55108, USA

Abstract

Production of a vast antibody repertoire is essential for protection against pathogens. Variable region germline complexity contributes to repertoire diversity and is a standard feature of mammalian immunoglobulin loci, but functional V region genes are limited in swine. For example, the porcine lambda light chain locus is composed of 23 variable (V) genes and 4 joining (J) genes, but only 10 or 11 V and 2 J genes are functional. Allelic variation in V and J may increase overall diversity within a population, yet lead to repertoire holes in individuals lacking key alleles. Previous studies focused on heavy chain genetic variation, thus light chain allelic diversity is not known. We characterized allelic variation of the porcine immunoglobulin lambda variable (IGLV) region genes. All intact IGLV genes in 81 pigs were amplified, sequenced, and analyzed to determine their allelic variation and functionality. We observed mutational variation across the entire length of the IGLV genes, in both framework and complementarity determining regions (CDRs). Three recombination hotspots were also identified, suggesting that non-allelic homologous recombination is an evolutionarily alternative mechanism for generating germline antibody diversity. Functional alleles were greatest in the most highly expressed families, IGLV3 and IGLV8. At the population level, allelic variation appears to help maintain the potential for broad antibody repertoire diversity in spite of reduced gene segment choices and limited germline sequence modification. The trade-off may be a reduction in repertoire diversity within individuals that could result in increased variation in immunity to infectious disease and response to vaccination.

Keywords

Antibody repertoire; light chain; porcine; allelic variation; *Sus scrofa*; immunology

Introduction

A diverse antibody repertoire is critical for producing specific antibodies against a massively wide range of pathogens and establishing effective immunity (Baumgarth et al. 2005; Maizels 2005; Ohlin and Zouali 2003). While the porcine immunoglobulin heavy chain and lambda and kappa light chain loci have been described, only the lambda locus is completely

Corresponding author: M. P. Murtaugh, murta001@umn.edu.

²Present address: Livestock Viral Diseases Programme, The Pirbright Institute, Ash Road, Pirbright, Surrey GU24 0NF, UK

characterized (Schwartz et al. 2012a; Schwartz et al. 2012b; Yerle et al. 1997). Within the porcine lambda locus, there are a total of 23 variable (IGLV) genes, 4 joining (IGLJ) genes, and 3 constant region (IGLC) genes. Of these, more than half of the IGLV genes are pseudogenes, and only two IGLJ and two IGLC genes are functional (Schwartz et al. 2012b). The apparent lack of diversity between the two nearly identical functional IGLJ genes limits combinatorial diversity. Also, terminal deoxynucleotidyl transferase (TdT) activity is below detectable levels in B cells undergoing light chain rearrangement, further reducing light chain junctional diversity (Wertz et al. 2013). To compensate for these deficiencies, we hypothesize that there is a substantial amount of allelic variation amongst IGLV genes at the population level, thus providing broad population immunity under conditions of increased individual immune risk. Allelic variation in the lambda locus is significant since approximately 50% of porcine immunoglobulins contain lambda light chains (Butler et al. 2005).

Studies of the porcine immunoglobulin repertoire have been limited at the individual haplotype level and little is known about the germline variation of light chain genes at the population level, which is critical for understanding the genetic basis of herd immunity. The previous characterization of the porcine lambda locus was in one adult Duroc and inferred from 239 fetal and neonatal lambda mRNA sequences (Wertz et al. 2013). Using a high-throughput sequencing approach, we found substantial allelic variation in IGLV genes among 81 pigs. All but two genes previously defined as functional were found to possess non-functional variants, while none of the previously defined pseudogenes possessed functional copies.

Materials and Methods

Tissue preparation

Spleen tissues from 78 pigs and bronchial lymph node tissues from 3 pigs were obtained from a genetically uniform, cross-bred commercial source herd leveraged from a previous study (Klinge et al. 2009) (see Supplemental Table 1 for sample collection and treatment information). Tissues were homogenized using a rotor-stator homogenizer and genomic DNA was purified using the DNeasy Blood & Tissue Kit (Qiagen). Quality and concentration of all genomic DNA samples was determined by gel electrophoresis and absorbance at 260 nm (Nanodrop Technologies), respectively.

Quantitative real-time PCR for the amplification of porcine IGLV genes

Primers specific for entire IGLV gene families or individual genes were designed with Primer-BLAST and used to amplify IGLV gene segments in 12 individual PCR reactions for each of 81 animals as detailed in Supplemental Table 2 (Ye et al. 2012). Amplification primers were targeted to 5' intron, 5' framework and 3' intergenic regions to solely amplify unrearranged V genes. Of the 22 known IGLV genes, IGLV1-12 is missing the V-exon and was excluded from analysis (Schwartz et al. 2012b). SYBR Green real-time quantitative PCR was performed using the Stratagene Mx3000p and Stratagene Mx3005p platforms. The temperature profile was denaturation at 95°C for 1 min, followed by 40 cycles of 95°C (4 s) and 60°C (40 s), then 95°C (1 min), 70°C (30 s) and 88°C (30 s). The specificity of the 972

PCR products was confirmed by melting temperature analysis and, in some cases, but gel electrophoresis. Failed reactions were repeated. PCR products were purified using the QIAquick PCR Purification kit (Qiagen).

Sequencing and bioinformatic analysis

Purified PCR products were pooled and submitted to the University of Minnesota Genomics Center for paired-end (250bp × 250bp) sequencing using Illumina MiSeq. Approximately 6 million reads were generated, with 91% of reads having quality scores above 30. Overlapping paired-end reads were merged by Fast Length Adjustment of Short Reads (FLASH) (Magoc and Salzberg 2011), resulting in 75.17% successfully combined reads. Allele identification was performed using BLAST and CD-HIT (Altschul et al. 1990; Li and Godzik 2006). After trimming all of the reads to the V region, there were 275,453 unique sequences in total. Fig. 1 shows the quality filtering process used to identify and differentiate true alleles from artifacts. Briefly, filtered alleles were analyzed by Geneious 6.1 software after removal of singletons and doubletons (Kearse et al. 2012), based on the annotation of Schwartz et al (Schwartz et al. 2012b) and, with minor exceptions, alleles with a frequency <0.6% (1/162, the theoretically minimum allele frequency) were excluded. Translated amino acid sequences were aligned within each gene cluster to confirm predicted translational fidelity. Alleles with a frameshift and/or a premature stop codon were categorized as non-functional. Alleles with synonymous mutations were categorized as silent alleles and those with nonsynonymous mutations were categorized as functional alleles. Mutation identification and calculation of mutation frequency were performed using Geneious 6.1 and the R statistical package (R Development Core Team 2014).

Results

Allelic variation and functionality of IGLV genes

635,678 reads were included in the final analysis, accounting for 20.9% of all initial reads (Fig. 1). Among 81 diploid pigs, a maximum of 162 alleles is conceivably possible for each gene. However, after quality filtering, there were more than 162 clusters for each gene, indicating artifacts or errors introduced during PCR and sequencing. To compensate, a theoretical threshold of 0.6% was established to allow measurement of the relative abundance of putative alleles present in the sample population. As shown in Fig. 2A, the allelic variation and the functionality of IGLV alleles varied substantially among genes. The two dominantly expressed gene segment families, IGLV3 and IGLV8, showed more variation than other V segments, suggesting that the abundantly expressed gene have undergone selection pressure to expand their allelic repertoires. To compare synonymous and non-synonymous mutation frequencies, the number of V gene segment allele variants was determined at the nucleotide and amino acid sequence levels. As shown in Fig. 2A, the majority of nucleotide differences between alleles were non-synonymous. Indeed, for 10 of 12 gene segments in the IGLV3 and IGLV8 families, 73% of alleles were derived from non-synonymous mutations compared to 31% expected by random chance ($p < 0.0001$, χ^2 -test). IGLV8-13, which has the richest allelic variation, has 35 amino acid alleles out of 47 total nucleotide sequence variants.

All of the functional alleles were in the IGLV3 family, the IGLV8 family, IGLV2-6, and IGLV5-14 (Fig. 2A), in agreement with the previously annotated locus (Schwartz et al. 2012b). A substantial number of alleles for these gene segments were non-functional due to frameshifts, premature stop codons, or both. Alleles of IGLV7-7, IGLV7-9 and one allele in IGLV3-6 encoded non-functional ORFs since they lacked the conserved cysteine residue at IMGT position 104 (2nd-CYS) (Schwartz et al. 2012b). To quantify the abundance of functional versus non-functional alleles, the percentage of predicted functional and non-functional products for each gene was calculated (Fig. 2B). Of twelve V gene segments possessing functional alleles, only IGLV3-2 was entirely functional. Another five genes (IGLV3-3, IGLV5-14, IGLV8-13, IGLV8-16, and IGLV8-18) were functional at a frequency less than 50 percent. Interestingly, while less than half of the identified alleles for IGLV3-6 were functional, the alleles were present at a frequency of more than 50% in the tested population. Indeed, this gene was recently identified as being completely missing in some animals, yet highly expressed in others (Schwartz and Murtaugh 2014). Meanwhile, IGLV3-4, IGLV3-5, IGLV3-6, and IGLV8-19 had a higher frequency of functional alleles.

Mutations in IGLV genes

Non-synonymous and synonymous mutations were observed in functional alleles across the V gene segment, in both framework and CDRs, especially in families containing a higher abundance of functional alleles, i.e. IGLV3 and IGLV8 (Fig. 3). By contrast, IGLV5 and IGLV7 families had a very limited number of mutations. The IGLV5 family contained only a single functional gene, IGLV5-14, which is expressed at extremely low levels (Schwartz et al. 2012b), and the IGLV7 family, which contains two pseudogenes. V gene segment families with higher levels of mutational variation, such as IGLV3 and IGLV8, showed Poisson frequency distributions, with one or two dominant alleles and substantial numbers of minor alleles (Supplementary Table 4). Analysis of read frequencies showed that V gene segments proximal to J were not underrepresented in Illumina reads, indicating that allelic variants were not artifacts accrued from sequencing of rearranged gene segments in which intervening DNA was excised (data not shown). It also is possible that minor allele artifacts might have arisen from inadvertent PCR amplification of rearranged V gene segments present in the lymphoid tissues used in this study. To address this question, we compared by Sanger sequencing functional V gene segments known to rearrange (3-2 and 3-3) and nonfunctional genes (3-1, 7-7, and 7-9) that were PCR-amplified from lymphoid (spleen) and non-lymphoid (lung) tissue in each of 5 independent pigs. In all cases, the V gene segment sequence was identical in lung and spleen. Even if rearranged V gene segments somehow were amplified from intergenic primer binding sites, arithmetic shows that the contribution to diversity would be a rare event. Only one V gene segment is present in a rearranged form in any single differentiated B cell. Thus, in a pure sample of differentiated B cells containing 22 amplifiable V gene segments, the rearranged V gene segment would contribute only about 2 to 3% of variation, having been diluted out by non-rearranged copies in the same cell and other cells. Not all B cells in lymphoid tissue are differentiated, and B cells constitute a minority of total tissue cells. Given that the PCR strategy was designed to exclude rearranged genes, it is reasonable to conclude that the allelic variation we report is due solely to analysis of germline DNA.

Recombination hotspots within the immunoglobulin lambda locus

Bioinformatic analysis revealed that three low-frequency alleles were products of recombination. Two of the alleles contained sequences from genes belonging to the IGLV3 family, and the third was a chimera consisting of IGLV2-6 and IGLV8-16 fragments (Fig. 4 and Table 1). All three chimeric alleles had recombination breakpoints in framework 3, upstream from CDR3, and a conserved sequence motif (CCA/TCC/TCTGACCAT) in close proximity to the apparent 5' recombination sites in framework 3 known to mediate homologous recombination during meiosis (Fig. 4) (Baudat et al. 2010; Berg et al. 2010; Grey et al. 2011). The motif identified in IGLV3 family members differs by only a single nucleotide from the 13-bp motif (CCNCCNTNNCCNC) previously described in the porcine genome as well as in the human and mouse genomes (Baudat et al. 2010; Berg et al. 2010; Tortereau et al. 2012).

Discussion

Diversity of the primary antibody repertoire in individuals is limited by the germline repertoire, which is then expanded by combinatorial rearrangement and N-nucleotide addition by TdT in the gene segment junctions (Alt et al. 1987; Blackwell and Alt 1989). However, combinatorial diversity is highly restricted within the lambda locus, as more than half of the variable genes are non-functional and the functional gene segments belong almost exclusively to only two families (Schwartz et al. 2012b). In addition, the lambda locus exhibits low levels of somatic hypermutation and junctional diversity during B cell development (Butler et al. 2006; Wertz et al. 2013). Because of these limitations, germline allelic diversity may contribute importantly to the population repertoire at the potential expense of individual antibody diversity. Such an outcome could increase the variation in response to infection or vaccination depending on the ability of individuals to produce key protective antibody responses.

Allelic variation was observed in all lambda V gene segment families. Surprisingly, a large majority of alleles were due to nonsynonymous mutations resulting in amino acid sequence variants, both in functional and non-functional alleles, as shown in Figures 2 and 3. Functional alleles were restricted primarily to the IGLV3 and IGLV8 families, indicating strong selection pressure acting in these highly expressed families (Butler et al. 2006; Wertz et al. 2013). The IGLV8 family, in particular, appears to have undergone recent evolutionary expansion and is actively transcribed early in B cell development in yolk sac (Schwartz et al. 2012b, Wertz et al. 2013). In contrast, the poorly expressed gene segments and pseudogenes in IGLV5 and IGLV7 appear to be under relatively little selective pressure, presenting minimal non-synonymous mutation frequencies (Fig. 3).

The patterns of nonsynonymous and synonymous mutations shown in Fig. 3, in which mutations are distributed across the entire ORF, in both framework and CDRs, were previously observed, and differ from human V_H gene segments, in which nonsynonymous mutations are concentrated in CDR1 and CDR2 hotspots (Chang and Casali 1994). Similar to our observations, the same study found that nonsynonymous mutations were not clustered in IGLV CDRs (Chang and Casali 1994). These findings indicate that framework regions as well as CDR are involved in antigen binding, but do not resolve if the bias toward

nonsynonymous mutations reflects positive selection for diversity or a potentially inherent mechanism favoring nonsynonymous changes (Butler and Wertz 2012). The surprising paucity of synonymous allelic variants might be due to the non-random sample population of 81 individuals from the same genetic stock or to the selection program of the breeding company, or to other reasons. However, none of the possibilities would appear to explain how synonymous alleles would not be generated, or be selected against if present.

Due to gene duplication, paralogous V gene segments have similar structures and are organized between thousands of bases of flanking region. This genetic flexibility enables non-allelic recombination in the absence of special enzymatic machinery (Seidman et al. 1978). Our observation of non-allelic recombination shows that germline diversity of IGLV can be enhanced by recombination between gene segments in the same family or between two separate families. Indeed, there also is evidence of diversification by non-crossover homologous recombination (i.e. gene conversion) within the porcine kappa locus (Schwartz et al. 2012a). Because the amplicons in the present study were generated from diverse, yet similar sequences, it could be argued that the recombinants may have been generated during PCR. However, this is highly unlikely. Recombination was observed specifically between IGLV3 gene segments 3-6*01 and 3-1, and 3-2 and 3-3, and between 2-6 and 8-16. For the 3-6*01–3-1 recombinant, 274 identical reads were generated in total from 81 individual PCR reactions that were pooled for sequencing, but no recombinants were detected between 3-6*02 and 3-1, even though the reverse primer was the same for both 3-6 alleles (Supplementary Table 2). The 3-2–3-3 recombinant was revealed in 158 identical reads; there was no similarity between the primers for each gene segment. Furthermore, there is no homology between the primer binding sites used to amplify the IGLV8 family members and IGLV2-6, making the generation of such a chimera virtually impossible since the products were amplified independently. Of all three recombinant alleles, only IGLV3-6 and IGLV3-1 shared a common reverse primer, making the recombinatorial mechanism ambiguous (i.e. either crossover or non-crossover homologous recombination). Internal cross-overs during PCR between similar V gene segments of a highly related family like IGLV8, resulting in chimeric molecules, were not observed. Therefore, We conclude that the observed recombinants were due to recombination events in the genomes investigated.

Allelic variation in human immunoglobulin heavy chains shows that antibody repertoire diversity among humans may exhibit individual genetic variation and raises the likelihood that variation in the capacity for immunoglobulin diversity may be a general vertebrate immune response feature (Boyd et al. 2010; Kidd et al. 2012). Allelic polymorphisms also have been described in other non-MHC antigen receptor molecules in swine, including the CDR2-like region of CD4, which is homologous to the HIV gp120 receptor site of human CD4 (Gustafsson et al. 1993).

The Illumina sequencing and bioinformatic analysis generated a large number of allelic artifacts that exceeded the theoretical maximum of 162 alleles per V gene segment in many cases. Genotyping artifacts and errors have been noted by others for non-model organisms (Babik et al. 2009) (Sommer et al. 2013). As have others, we expected that artifacts would be relatively rare compared to true alleles, and so eliminated alleles with low reads, including all singletons and doubletons (Babik et al. 2009). It is possible that somatic

hypermethylation acting on un-rearranged germline lambda V gene segments generated artifactual alleles (Weiss and Wu 1987). However, in addition to previously described low levels of somatic hypermutation and junctional diversity during porcine B cell development (Butler et al. 2006; Wertz et al. 2013), germline mutations have been reported primarily from tumor cell lines and individuals with autoimmune diseases (Chezar et al. 2008; Kenny et al. 2001; Riboldi et al. 2003; Weiss and Wu 1987). For these reasons we conclude that the data represent allelic variation in germline lambda V gene segments.

The 13-bp motif “CCNCCNTNNCCNC” has been studied extensively in humans and is known to bind a histone methyltransferase, PRDM9, facilitating the formation of double strand breaks (DSB) and thereby increasing the recombination frequency (Baudat et al. 2010; Berg et al. 2010; Myers et al. 2010). PRDM9 shows rapid evolution in different species including human, mouse, rat, macaque, and orangutan, exhibiting variations in the zinc-finger domain (Myers et al. 2010). One such motif discussed presently differs by only one base pair compared with the 13-bp motif found in the pig genome (Tortoreau et al. 2012). This difference could be due to an allelic difference in pig breed, since the genome was derived from a single Duroc breed female (Groenen et al. 2012). Nevertheless, breed differences are expected overall to be unsubstantial due to extensive admixture over time between wild boars and domesticated pigs, and among diverse domesticated pig breeds of Asian and European populations (Ramirez et al. 2009; White 2011). The base similarities between these motifs and the motifs in human genes indicate conserved similarities in the antibody development of mammals (Table 1).

To summarize, we have discovered that the allelic variation of IGLV segment genes was substantial, and functional alleles were highest in the most highly expressed families, IGLV3 and IGLV8. At the population level, allelic variation appears to help maintain the potential for broad antibody repertoire diversity in spite of reduced gene segment choices and limited germline sequence modification. The trade-off may be a reduction in repertoire diversity within individuals that could result in increased variation in immunity to infectious disease and response to vaccination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Kent Reed for critically reading the manuscript and helpful suggestions. Funding was provided by the National Pork Board grant 10-139 (J.C.S. and M.P.M.). J.C.S. was supported by the Molecular Virology Training Grant T32 AI83196 from the National Institutes of Health and a Doctoral Dissertation Fellowship from the University of Minnesota. X.G. was supported in part by a University of Minnesota MnDRIVE Global Foods Venture fellowship.

Bibliography & References Cited

- Alt FW, Blackwell TK, Yancopoulos GD. Development of the primary antibody repertoire. *Science*. 1987; 238:1079–87. [PubMed: 3317825]
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215:403–10. [PubMed: 2231712]

- Babik W, Taberlet P, Ejsmond MJ, Radwan J. New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Mol Ecol Resour.* 2009; 9:713–9. [PubMed: 21564729]
- Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science.* 2010; 327:836–40. [PubMed: 20044539]
- Baumgarth N, Tung JW, Herzenberg LA. Inherent specificities in natural antibodies: a key to immune defense against pathogen invasion. *Springer Semin Immunopathol.* 2005; 26:347–62. [PubMed: 15633017]
- Berg IL, Neumann R, Lam KW, Sarbajna S, Odenthal-Hesse L, May CA, Jeffreys AJ. PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat Genet.* 2010; 42:859–63. [PubMed: 20818382]
- Blackwell TK, Alt FW. Mechanism and developmental program of immunoglobulin gene rearrangement in mammals. *Annu Rev Genet.* 1989; 23:605–36. [PubMed: 2694945]
- Boyd P, Hudgens E, Loftus JP, Tompkins D, Wysocki M, Kakach LT, Labresh J, Baldwin CL, Lunney JK. Expressed gene sequence and bioactivity of the IFN γ -response chemokine CXCL11 of swine and cattle. *Vet Immunol Immunopathol.* 2010; 136:170–5. [PubMed: 20347492]
- Butler JE, Sinkora M, Wertz N, Holtmeier W, Lemke CD. Development of the neonatal B and T cell repertoire in swine: implications for comparative and veterinary immunology. *Vet Res.* 2006; 37:417–41. [PubMed: 16611556]
- Butler JE, Wertz N. The porcine antibody repertoire: variations on the textbook theme. *Front Immunol.* 2012; 3:153. [PubMed: 22754553]
- Butler JE, Wertz N, Sun J, Wang H, Lemke C, Chardon P, Piumi F, Wells K. The pre-immune variable kappa repertoire of swine is selectively generated from certain subfamilies of V κ 2 and one J κ gene. *Vet Immunol Immunopathol.* 2005; 108:127–37. [PubMed: 16112743]
- Chang B, Casali P. The CDR1 sequences of a major proportion of human germline Ig V H genes are inherently susceptible to amino acid replacement. *Immunology Today.* 1994; 15:367–373. [PubMed: 7916950]
- Chezar I, Lobel-Lavi L, Steinitz M, Laskov R. Ongoing somatic hypermutation of the rearranged VH but not of the V-lambda gene in EBV-transformed rheumatoid factor-producing lymphoblastoid cell line. *Mol Immunol.* 2008; 46:80–90. [PubMed: 18718665]
- Grey C, Barthes P, Chauveau-Le Fric G, Langa F, Baudat F, de Massy B. Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol.* 2011; 9:e1001176. [PubMed: 22028627]
- Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens HJ, Li S, Larkin DM, Kim H, Frantz LA, 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, Bosse M, Botti S, Bujie Z, Bystrom M, Capitanu B, Carvalho-Silva D, Chardon P, Chen C, Cheng R, Choi SH, Chow W, Clark RC, Clee C, Crooijmans RP, Dawson HD, Dehais P, De Sapio F, Dibbits B, Drou N, Du ZQ, Eversole K, Fadista J, Fairley S, Faraut T, Faulkner GJ, Fowler KE, Fredholm M, Fritz E, Gilbert JG, Giuffra E, Gorodkin J, Griffin DK, Harrow JL, Hayward A, Howe K, Hu ZL, Humphray SJ, Hunt T, Hornshoj H, Jeon JT, Jern P, Jones M, Jurka J, Kanamori H, Kapetanovic R, Kim J, Kim JH, Kim KW, Kim TH, Larson G, Lee K, Lee KT, 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 SJ, Onteru S, Panitz F, Park EW, et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature.* 2012; 491:393–8. [PubMed: 23151582]
- Gustafsson K, Germana S, Sundt TM 3rd, Sachs DH, LeGuern C. Extensive allelic polymorphism in the CDR2-like region of the miniature swine CD4 molecule. *J Immunol.* 1993; 151:1365–70. [PubMed: 8335933]
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012; 28:1647–9. [PubMed: 22543367]

- Kenny TP, Semrad TJ, Malyj W, Robbins DL. Mutational analysis of immunoglobulin germline derived Vlambda4B light chains in rheumatoid arthritis. *Clin Immunol*. 2001; 99:283–90. [PubMed: 11318600]
- Kidd MJ, Chen Z, Wang Y, Jackson KJ, Zhang L, Boyd SD, Fire AZ, Tanaka MM, Gaeta BA, Collins AM. The inference of phased haplotypes for the immunoglobulin H chain V region gene loci by analysis of VDJ gene rearrangements. *J Immunol*. 2012; 188:1333–40. [PubMed: 22205028]
- Klinge KL, Vaughn EM, Roof MB, Bautista EM, Murtaugh MP. Age-dependent resistance to Porcine reproductive and respiratory syndrome virus replication in swine. *Virology*. 2009; 6:177. [PubMed: 19860914]
- Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006; 22:1658–9. [PubMed: 16731699]
- Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011; 27:2957–63. [PubMed: 21903629]
- Maizels N. Immunoglobulin gene diversification. *Annu Rev Genet*. 2005; 39:23–46. [PubMed: 16285851]
- Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*. 2010; 327:876–9. [PubMed: 20044541]
- Ohlin M, Zouali M. The human antibody repertoire to infectious agents: implications for disease pathogenesis. *Mol Immunol*. 2003; 40:1–11. [PubMed: 12909126]
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: 2014. <http://www.R-project.org/>
- Ramirez O, Ojeda A, Tomas A, Gallardo D, Huang LS, Folch JM, Clop A, Sanchez A, Badaoui B, Hanotte O, Galman-Omitogun O, Makuza SM, Soto H, Cadillo J, Kelly L, Cho IC, Yeghoyan S, Perez-Enciso M, Amills M. Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. *Mol Biol Evol*. 2009; 26:2061–72. [PubMed: 19535739]
- Riboldi P, Ikematsu W, Brambilla B, Caprani C, Gerosa M, Casali P. Diversity and somatic hypermutation of the Ig VHDJH, V kappa J kappa, and V lambda J lambda gene segments in lymphoma B cells: relevance to the origin of the neoplastic B cell clone. *Hum Immunol*. 2003; 64:69–81. [PubMed: 12507816]
- Schwartz JC, Lefranc MP, Murtaugh MP. Evolution of the porcine (*Sus scrofa domestica*) immunoglobulin kappa locus through germline gene conversion. *Immunogenetics*. 2012a; 64:303–11. [PubMed: 22109540]
- Schwartz JC, Lefranc MP, Murtaugh MP. Organization, complexity and allelic diversity of the porcine (*Sus scrofa domestica*) immunoglobulin lambda locus. *Immunogenetics*. 2012b; 64:399–407. [PubMed: 22186825]
- Schwartz JC, Murtaugh MP. Characterization of a polymorphic IGLV gene in pigs (*Sus scrofa*). *Immunogenetics*. 2014; 66:507–11. [PubMed: 24934119]
- Seidman JG, Leder A, Nau M, Norman B, Leder P. Antibody diversity. *Science*. 1978; 202:11–7. [PubMed: 99815]
- Sommer S, Courtiol A, Mazzoni CJ. MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics*. 2013; 14:542. [PubMed: 23937623]
- Tortoreau F, Servin B, Frantz L, Megens HJ, Milan D, Rohrer G, Wiedmann R, Beever J, Archibald AL, Schook LB, Groenen MA. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics*. 2012; 13:586. [PubMed: 23152986]
- Weiss S, Wu GE. Somatic point mutations in unrearranged immunoglobulin gene segments encoding the variable region of lambda light chains. *EMBO J*. 1987; 6:927–32. [PubMed: 3109891]
- Wertz N, Vazquez J, Wells K, Sun J, Butler JE. 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*. 2013; 55:319–28. [PubMed: 23570908]
- White S. from globalized PIG BREEDS TO CAPITALIST PIGS: A STUDY IN ANIMAL CULTURES AND EVOLUTIONARY HISTORY. *Environmental History*. 2011; 16:94–120.

- Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012; 13:134. [PubMed: 22708584]
- Yerle M, Lahbib-Mansais Y, Pinton P, Robic A, Goureau A, Milan D, Gellin J. The cytogenetic map of the domestic pig. Mamm Genome. 1997; 8:592–607. [PubMed: 9250869]

Fig. 1.

Illumina reads filtering process. The workflow demonstrates the methods used in the study for DNA isolation and processing of Illumina sequence data. Left-side numbers indicate the reads included in each step. Percentages in parentheses indicate the number of reads at the indicated step relative to the initial number of reads. Right-side numbers indicate the discarded reads.

A

B

Fig. 2. Allelic variation within the immunoglobulin lambda locus among 81 pigs. (A) Absolute numbers of functional variants, functional invariant, and non-functional alleles at each gene segment. (B) Relative abundance of functional and non-functional alleles.

Fig. 3. Mutation frequencies of V gene alleles in the lambda locus. X-axis: Relative location of framework (black) and CDR (light gray) regions of individual IGLV. Y-axis: percent frequency ($\log_{10}$) of allelic mutations when all alleles are aligned. Non-synonymous mutations (N), synonymous mutations (S). Gaps are not counted as mutations.

Fig. 4.
Putative recombination hotspots in the porcine immunoglobulin lambda locus. The black boxes indicate the location of the recombination related motifs. The three CDRs (complementarity determining region) are indicated by red boxes. Colored nucleotides denote non-consensus bases.

Table 1

Recombination hotspots in FIGURE 4

Recombination Hotspot	Motif sequence	Distance to the Recombination Site (bp)	Reads	Relative abundance %	Comparative human IGLV ^a	Number of different nucleotides ^b
(A)	CCACCCTGACCAT	0	274	0.58	IGLV3-6	0
					IGLV3-9	0
(B)	CCACCCTGACCAT	52	156	0.20	IGLV3-21	0
					IGLV3-22	0
					IGLV3-1	1
					IGLV3-30	1
					IGLV3-4	1
					IGLV3-10	1
(C)	CCTCTCTGACCAT	36	134	0.38	IGLV2-14	1
					IGLV2-18	1
					IGLV2-34	1
					IGLV10-54	1
					IGLV2-23	2
					IGLV2-33	2

^aComparative human IGLV indicates the human IGLV gene segments that have the most similar corresponding motifs

^bNumber of different nucleotide indicates the numbers of different nucleotides between the motifs in porcine IGLV gene segments and the human IGLV homologs