Published in final edited form as:

Immunogenetics. 2014 August ; 66(0): 507–511. doi:10.1007/s00251-014-0785-2.

Characterization of a polymorphic IGLV gene in pigs (*Sus scrofa*)

John C. Schwartz and Michael P. Murtaugh

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

Abstract

Swine, unlike other artiodactyls, but similar to humans, utilize both lambda and kappa light chain isotypes almost equally in the generation of their antibody repertoire. The porcine antibody light chain loci have previously been characterized in a single Duroc sow in which was seen extensive allelic variation between light chain genes on homologous chromosomes. However, the extent of variation between individuals is completely unknown. Using deep sequencing of cDNA-derived amplicons from five pigs, we report the identification and characterization of an IGLV gene that is functional and highly expressed in some animals, yet completely absent in others. Our findings provide a possible rationale for the known individual-to-individual variation in antibody responses to vaccination, infectious challenge and subsequent disease outcome.

Keywords

immunoglobulin; light chain; IGL; lambda; allele; *Sus scrofa*

Brief Communication

The ability to mount antibody responses to an almost limitless array of antigens is fundamentally important to effector and memory mechanisms of disease resistance in vertebrate animals. Despite this, total antibody repertoire diversity is constrained by the genetic complexity of the antibody loci and the error-prone recombinatorial mechanisms involved during V(D)J rearrangement. It is reasonable to hypothesize, then, that extensive allelic variation would have evolved in order to maximize this diversity among populations. Indeed, extensive variation in the induction of effectively protective responses to natural infections and vaccinations is present within outbred populations of humans and animals thus complicating disease control efforts. Knowledge of the allelic repertoire of the immune loci is therefore critical to understanding the response capacity of populations and to allow for better informed disease control efforts and animal breeding programs.

The current characterizations of the porcine immunoglobulin heavy (IGH) locus on chromosome 7 include 15 IGHV genes, four IGHD genes, five IGHJ genes, and the constant genes (Eguchi-Ogawa et al. 2012; Eguchi-Ogawa et al. 2010). It is likely however, that

additional IGHV genes exist upstream from the fifteen that are characterized based on cDNA evidence (Eguchi-Ogawa et al. 2010). In addition, we previously characterized the genomic organization of the porcine kappa (IGK) and lambda (IGL) light chain loci in a single animal on chromosomes 3 and 14, respectively (Schwartz et al. 2012a; Schwartz et al. 2012b). All of the identified IGK and IGL variable (V), joining (J), and constant (C) genes were entered in IMGT/GENE-DB (Giudicelli et al. 2005). The IGK locus contains at least 14 IGKV genes, 5 IGKJ genes, and a single IGKC gene. However, it is plausible that the kappa locus is also incompletely characterized due to the lack of flanking gene information (Schwartz et al. 2012a). The IGL locus contains 22 annotated IGLV genes, 3 IGLJ-IGLC cassettes, and a fourth IGLJ with no associated IGLC. In contrast to the IGH and IGK loci, the IGL locus is completely delimited in that flanking upstream genes and 445 kb of contiguous upstream sequence have been analyzed (Schwartz et al. 2012b). This greatly eases the ability to correctly associate lambda cDNA sequences with their respective genes. Thus, the lambda locus is the most amenable antibody locus in pigs for investigating antibody allelic variation using transcriptomic data.

Despite the apparent completeness of the porcine IGL locus, a recent report identified transcripts obtained from pigs of mixed breeds that was clearly IGLV3-like, yet distinct from other known IGLV3 subgroup members (Wertz et al. 2013). The gene from which these transcripts may have arisen was putatively designated IGLV3-6. It was further observed that these IGLV3-6 transcripts accounted for approximately 20 percent of all IGL transcripts (Wertz et al. 2013). In the present report, we provide additional transcriptomic and genomic evidence for IGLV3-6, including its genomic context and its variability among commercial swine.

To assess light chain diversity, oligonucleotide primers were designed for the light chain leader and constant regions such that all known light chain variable region genes could be amplified from cDNA (e.g. for IGLV3 subgroup genes: IGLV3 forward, 5'-CTGGAYCCCTCTCCTGCTC; IGLC reverse, 5'-CCTTCCAGGTCACCGTCA). RNA was extracted from lymphoid tissues of five 8 to 10 week old animals from a commercial source herd leveraged from a previous study (Klinge et al. 2009), reverse transcribed and PCR-amplified. The resulting amplicons were pooled in equimolar amounts from each animal and sequenced using Roche Titanium 454 pyrosequencing at the W. M. Keck Center for Functional Genomics at the University of Illinois at Urbana-Champaign. Molecular barcode tags of 10 bp were included on the 5' end of each forward primer in order to differentiate between individual animals. A total of 372,140 full-length (>350 bp, mean of 510 bp), in-frame reads were obtained and were approximately evenly distributed between animals (17 to 24 percent of all reads for each of five animals). These reads were compared to the annotated porcine IGLV genes using BLAST (Altschul et al. 1990). Amino acid numbering and nomenclature is based on IMGT®, the international IMunoGeneTics information system® (IMGT), <http://www.imgt.org> (Lefranc 2007; Lefranc 2011a; Lefranc 2011b; Lefranc et al. 2003).

Our analysis revealed a population of transcripts most similar to, albeit approximately 20 percent different from IGLV3-4, that was present in four of the five animals (not shown). The existence of a large population of full-length productive transcripts suggested the

existence of a functional allele of an uncharacterized IGLV3 subgroup member in at least four of the five pigs. Altogether, these transcripts accounted for approximately 13 percent of all IGLV-containing reads. Comparison to transcripts described by Wertz et al. (2013) revealed that these transcripts are derived from the same putative IGLV3-6 gene. Further analysis revealed that three of the animals expressed full-length productive IGLV3-6 transcripts of similar identity, while transcripts from pig 2 were on average 10% different, indicating the existence of a second productive allele (*02) (Figure 1, Table 1). Thus, these data and those of Wertz et al. (2013) indicate that IGLV3-6 is one of the most highly utilized light chain genes in swine.

To characterize the genomic context of IGLV3-6, the transcripts were compared to the annotated genome assembly. This revealed a small, truncated pseudogene 182 bp downstream from the 3' end of the recombination signal (RS) sequence of IGLV3-2 and 4.7 kb upstream from IGLV3-1 on the BAC CH242-141B5 (GenBank: CU467669) (Figure 1a). The 5' end of this novel pseudogene is deleted such that it becomes recognizable starting at amino acid position 44 of framework region 2. A single nucleotide polymorphism within CDR3 distinguishes the intact portion of the truncated pseudogene (*03) from the full-length *01 allele.

The presence of the truncated IGLV3-6*03 allele on CH242-141B5 raised a question about the structure of IGLV3-6 on the other annotated BAC, CH242-524K4 (GenBank: CU467599), since it was not detected when the IGL locus was characterized (Schwartz et al. 2012b). Analysis of genomic contigs revealed a small 381 bp contig that contained the full-length functional IGLV3-6*01 allele. However, due to its small size and high similarity to IGLV3-4, the contig was erroneously assembled to the IGLV3-4 region (Schwartz et al. 2012b). Although the intergenic distance between the full length IGLV3-6 and IGLV3-2 genes could not be precisely determined, comparison with sequence from CH242-524K4 suggests that it is at least 2.6 kb downstream from IGLV3-2 (Figure 2b). Furthermore, similarity with intergenic sequence between IGLV3-4 and IGLV3-5, suggests that IGLV3-6 and IGLV3-2 could be separated by as much as 7.5 kb.

Based on the organization of IGLV3-6 on CH242-141B5 and CH242-524K4, oligonucleotide primers were designed to further characterize IGLV3-6 in each of the five animals. Four separate PCR amplifications were performed using forward primers specific for either the CDR2 of IGLV3-2 or the conserved IGLV3 leader sequence and reverse primers specific for either the CDR2 of IGLV3-6, the leader region of IGLV3-1, or the intergenic region between IGLV3-1 and IGLV3-6 (Table 2). The reactions targeting both IGLV3-6 and IGLV3-1 confirmed the genomic organization of both the full length and truncated variants of IGLV3-6 (Figure 2d). The remaining reactions suggested the existence of both the truncated and full-length genes in pigs 2–5. Subsequent chain-termination Sanger sequencing identified both the truncated (180bp) and full-length (296bp) variants of IGLV3-6 in all four of these animals (Figure 2). A single Sanger read obtained from pig 3 (Figure 2d, reaction 3) confirmed the genomic organization of the truncated pseudogene 182 bp downstream from the RS of IGLV3-2. Although one of the two products indicative of the functional gene was detected for pig 1 (Figure 2d, reaction 4), Sanger sequencing revealed this product to be non-specifically amplified IGLV3-4, whereas sequencing of the

corresponding product from pigs 2–5 revealed the functional IGLV3-6 gene. No transcripts belonging to IGLV3-6 were produced by pig 1 (Figure 1) and as PCR and Sanger sequencing failed to identify the gene, it seems that IGLV3-6 is either missing entirely, or is at least non-functional and structurally different in this animal (Figure 2).

As the exact location of the full-length gene remains undetermined, and as evidence was observed for the coexistence of both the truncated and full-length IGLV3-6 genes in four of the five animals, it is plausible that the full-length and truncated forms of IGLV3-6 represent separate duplicated genes. Indeed, bioinformatic mis-assembly of two very similar genes could result in one of them getting deleted from the final assembly (Salzberg and Yorke 2005). To avoid the possibility of assembly error, individual Illumina reads from the two BACs were analyzed using BLAST for the presence or absence of reads specific to either the truncated or full-length forms of IGLV3-6. This analysis revealed that the full-length gene is missing from CH242-141B5, yet present on CH242-524K4; while the truncated gene is present on CH242-141B5, yet absent from CH242-524K4. It is therefore our interpretation that IGLV3-6 is most likely a single gene.

In conclusion, we have identified a lambda light chain gene in pigs that is highly expressed in some individuals but completely missing in others. Such variation in the antibody repertoire could potentially impact the porcine immune response and provide a rationale for observed variable immune responses to vaccination and disease challenge. Our findings will inform ongoing studies investigating the diversity of the porcine antibody repertoire and molecular mechanisms of disease resistance.

Acknowledgments

We thank Juan Abrahante for assistance and advice in sequence data management. 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.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215:403–10. [PubMed: 2231712]
- Eguchi-Ogawa T, Toki D, Wertz N, Butler JE, Uenishi H. Structure of the genomic sequence comprising the immunoglobulin heavy constant (IGHC) genes from *Sus scrofa*. *Mol Immunol.* 2012; 52 :97–107. [PubMed: 22673207]
- Eguchi-Ogawa T, Wertz N, Sun XZ, Puimi F, Uenishi H, Wells K, Chardon P, Tobin G. 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.* 2010; 184:3734–3742. [PubMed: 20207992]
- Giudicelli V, Chaume D, Lefranc MP. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* 2005; 33:D256–61. [PubMed: 15608191]
- Klinge KL, Vaughn EM, Roof MB, Bautista EM, Murtaugh MP. Age-dependent resistance to Porcine reproductive and respiratory syndrome virus replication in swine. *Virol J.* 2009; 6:177. [PubMed: 19860914]
- Lefranc MP. WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report. *Immunogenetics.* 2007; 59:899–902. [PubMed: 18046549]

- Lefranc MP. IMGT Collier de Perles for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. Cold Spring Harb Protoc. 2011a; 2011:643–51. [PubMed: 21632788]
- Lefranc MP. IMGT unique numbering for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. Cold Spring Harb Protoc. 2011b; 2011:633–42. [PubMed: 21632789]
- Lefranc MP, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol. 2003; 27:55–77. [PubMed: 12477501]
- Salzberg SL, Yorke JA. Beware of mis-assembled genomes. Bioinformatics. 2005; 21:4320–1. [PubMed: 16332717]
- 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]
- 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; 55:319–328. [PubMed: 23570908]

Figure 1. IGLV3-6 expression in individual pigs. Number of amino acid mismatches in light chain transcripts that align to the IGLV3-6 gene by BLAST. Z-axis value of zero indicates identity with the prototypic functional allele IGLV3-6*01. Pig 1 produced no IGLV3-6 transcripts

Figure 2.

Genomic context and genotyping of IGLV3-6. The genomic organization for both the truncated (a) and full length (b) forms of IGLV3-6. Four separate reactions were performed using genomic DNA from each of five animals as well as from purified BAC DNA (c and d). Expected sizes of PCR products are shown in (a and b). The BAC CH242-141B5 (d) contains a copy of the truncated form of IGLV3-6 and yields PCR products consistent with this. Conversely, CH242-524K4 (c) contains a full-length copy of IGLV3-6, the PCR result for which is also consistent and is shown next to a replicate reaction 4 from pig 3. Faint reaction 2 products of the approximate expected size were visible on the gel for animals 2–5.

Table 1

Deduced amino acid sequences of the IGLV3-6 alleles

	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)					C (27-38)					C' (39-46)					C'' (47-55)					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																																	
IGLV3-6*01	SSQLTQVPG.VSVSL	GGTASIA	CLGD	NFV	YA	ANWYQ	QKP	GQAPILVIY	GG	S	LRPLGIP.E	RPSGSS..SG	NSATLTISGAQA	EDEADYYC	QSADSSDNA																																								
IGLV3-6*02	SSQLTQVPG.VSVSL	GGTASIA	CLGD	NFI	YA	ANWYQ	QKP	GQAPILVIY	GG	S	LRPLGIP.E	RPSGSS..SG	NSATLTISGAQA	EDEADYYC	QSADSSDNA																																								
IGLV3-6*03	SSQLTQVPG.VSVSL	GGTASIA	CLGD	NFI	YA	ANWYQ	QKP	GQAPILVIY	GG	S	LRPLGIP.E	RPSGSS..SG	NSATLTISGAQA	EDEADYYC	QSADSSDNA																																								

Positions differing from the *01 allele are underlined in the *02 and *03 alleles.

Sequences for the *01 and *03 alleles are based on genomic sequence derived from CH242-524K4 (CU467599) and CH242-141B5 (CU467669), respectively.

Sequence for the *02 allele is based on the most abundant IGLV3-6 cDNA from pig 2 (286 out of 1683 transcripts; 762 out of 1683 excluding CDR3).

Amino acid positions numbered according IMGT (Lefranc et al. 2003).

Table 2

Primers used to amplify and genotype IGLV3-6

Target	Template-specific sequence
IGLV3-6 (F) *	5'-CATCCTGGTCATCTATGGTGGT
IGLV3-1/3-6 intergenic region (R)	5'-AGCTGTGTCCTCCCTTGAGA
IGLV3-1 (R)	5'-GAGCAGGAGAGGGGTCTAGG
IGLV3-6 (R)	5'-CCACCATAGATGACCAGGATG
IGLV3-2 (F)	5'-GTTTCATCTATGAGGACACCAACC
IGLV3, leader (F)	5'-CTGGAYCCCTCTCCTGCTC

*
(F), forward; (R), reverse