Published in final edited form as:
Immunogenetics. 2013 October ; 65(10): 749–762. doi:10.1007/s00251-013-0722-9.

Accuracy and coverage assessment of *Oryctolagus cuniculus* (Rabbit) Genes Encoding Immunoglobulins in the Whole Genome Sequence Assembly (OryCun2.0) and Localization of the *IGH* Locus to Chromosome 20

E. Michael Gertz[1], **Alejandro A. Schäffer**[1], **Richa Agarwala**[1], **Amélie Bonnet-Garnier**[2,3], **Claire Rogel-Gaillard**[4], **Hélène Hayes**[4], and **Rose G. Mage**[5,*]

[1]National Center for Biotechnology Information, NLM, NIH, Bethesda, MD USA, 20894

[2]INRA, UMR1198 Biologie du Développement et Reproduction, F-78352 Jouy-en-Josas, France

[3]ENVA, F-94704 Maisons Alfort, France

[4]INRA, UMR 1313 Génétique Animale et Biologie Intégrative, F-78352 Jouy-en-Josas, France

[5]Laboratory of Immunology NIAID, NIH, Bethesda, MD, USA

Abstract

We report analyses of genes encoding immunoglobulin heavy and light chains in the rabbit 6.51x whole genome assembly. This OryCun2.0 assembly confirms previous mapping of the duplicated *IGK1* and *IGK2* loci to chromosome 2 and the *IGL* lambda light chain locus to chromosome 21. The most frequently rearranged and expressed *IGHV1* that is closest to *IG DH* and *IGHJ* genes encodes rabbit VHa allotypes. The partially inbred Thorbecke strain rabbit used for whole-genome sequencing was homozygous at the *IGK* but heterozygous with the *IGHV1a1* allele in one of 79 *IGHV*-containing unplaced scaffolds and *IGHV1a2, IGHM, IGHG* and *IGHE* sequences in another. Some *IGKV, IGLV* and *IGHA* genes are also in other unplaced scaffolds. By fluorescence *in situ* hybridization, we assigned the previously unmapped *IGH* locus to the q-telomeric region of rabbit chromosome 20. An approximately 3 Mb segment of human chromosome 14 including *IGH* genes predicted to map to this telomeric region based on synteny analysis could not be located on assembled chromosome 20. Unplaced scaffold chrUn0053 contains some of the genes that comparative mapping predicts to be missing. We identified discrepancies between previous targeted studies and the OryCun2.0 assembly and some new BAC clones with *IGH* sequences that can guide other studies to further sequence and improve the OryCun2.0 assembly. Complete knowledge of gene sequences encoding variable regions of rabbit heavy, kappa and lambda chains will lead to better understanding of how and why rabbits produce antibodies of high specificity and affinity through gene conversion and somatic hypermutation.

Keywords

Rabbit; Immunoglobulin Genes; Heavy Chains; Fluorescence *in situ* hybridization; Chromosome 20; Light Chains

[*] *To whom correspondence should be addressed:* Phone: +1 301 496 6113 Fax: +1 301 496 0222 rmage@niaid.nih.govhttp://www3.niaid.nih.gov/labs/aboutlabs/li/mage.htm.

Introduction

The whole genome sequence of DNA from a single female rabbit of the partially inbred Thorbecke rabbit strain was first made public on October 20, 2009 (OryCun2.0; Accession AAGW02000000). The annotated assembly with 6.51x coverage is now available at NCBI, UCSC and Ensembl. The rabbit chosen by the Broad Institute for sequencing was obtained from Covance in 2004, and was shown to have a low heterozygosity rate compared to non-inbred NZW (New Zealand White) rabbits from the same source (Lindblad-Toh et al. 2011; Hubisz et al. 2011). DNA of the same animal was used both for the ~7x coverage sequencing project analyzed here, and for a previous low-coverage ~2x sequencing project (Accession AAGW01000000). The assembly has 2.24 Gbp in 21 autosomes and X chromosome and 489 Mbp in 3219 unplaced scaffolds including mitochondria. The methods for obtaining the OryCun2.0 whole Genome Sequence were similar to those described in 2005 for the dog (Lindblad-Toh et al., 2005) except that BACs used for anchoring were not from the same Thorbecke strain rabbit. Although the level of coverage indicated for the complete set of sequence traces is 7.48x, only reads of quality Q20 or higher are included in the assembly. Unfortunately, in January 2005, all rabbits of this strain were lost in a fire at the facility of Covance Research Products, Inc. where they were housed. To improve annotations, RNASeq data were recently obtained from NZW rabbits.

The work presented here aims to identify and improve annotation of regions of the OryCun2.0 assembly containing genes encoding rabbit immunoglobulin (Ig) heavy and light chains (reviewed in Mage et al. 2006). We evaluated the assembly of each immunoglobulin gene locus by comparing with previously reported data from targeted analyses. Complete knowledge of the gene sequences encoding variable regions of rabbit heavy chains (*IGHV)*, kappa light chains (*IGKV)*, and lambda light chains (*IGLV*) and their map locations is important for understanding the steps of gene conversion (Becker and Knight 1990; Allegrucci et al. 1990) and somatic hypermutation that lead to the maturation and clonal expansion of B-cells producing highly specific antibodies of high affinity that are a hallmark of rabbits' immune responses (Sehgal et al. 1998; Schiaffella et al. 1999; Sehgal et al. 2000).

The *IGH* locus is not detected in any of the OryCun2.0 assembled chromosomes, and to date, the rabbit *IGH* locus has not been mapped completely. It has been shown that fluorescence *in situ* hybridization analyses (FISH) with BAC clones containing known genes and whole-chromosome probes can be used to compare human and rabbit gene maps and to construct the cytogenetic rabbit map (Korstanje et al. 1999; Hayes et al. 2002; Chantry-Darmon et al. 2003 and 2005a,b). Based on FISH analyses, we report the localization of the unmapped *IGH* locus to rabbit chromosome 20. In addition, for several genes that have not been detected in the sequence assembly OryCun2.0, but which, based on synteny comparisons, are predicted to belong to the region of rabbit chromosome 20 around the *IGH* locus, we identified sequences either among the unplaced scaffolds or the whole-genome shotgun (WGS) traces aided by alignment of representative transcripts from rabbit RNAseq data.

We report allotyping data to clarify some of the expected immunoglobulin sequence content of the OryCun2.0 rabbit sequence assembly, the chromosomal location of *IGH* by FISH, sequence data for genes present in OryCun2.0 expected to be adjacent to *IGH*, and a series of evaluations of the sequences in OryCun2.0 for the heavy (*IGH),* kappa *(IGK)*, and lambda (*IGL)* loci, in that order. We identify unplaced scaffolds (chrUn's) containing *IGH, IGK*, or *IGL* segments that should eventually be placed on chromosomes. In some specific examples, we compare the order or content of *IGH, IGK*, and *IGL* sequences to data from previously sequenced and mapped rabbit genes.

Methods

Serum testing for allotypes

In 1995, sera of 11 Thorbecke inbred rabbits that Dr. G.J. Thorbecke had sent to Dr. T.J. Kindt, NIAID, NIH were obtained from Dr. Kindt and typed. Typing methods were as described by Roux and Mage (1996). *IGKC1* -encoded allotypes b4, b5, b6, b9, and *IGHV1*-encoded allotypes al, a2 and a3 were typed by standard gel diffusion methods; allotypes encoded by *IGKC2* bas1, *IGHG* hinge region d11 and d12 and the *IGHG* CH2 regions e14 and e15 were typed by hemagglutination inhibition.

Bioinformatic analyses

We used BLAST (Altschul et al. 1997) and BLAT (Kent 2002) for identification and analyses of available rabbit immunoglobulin gene sequences, and examined maps and annotations in the OryCun2.0 sequence assembly at the websites of NCBI (http://www.ncbi.nlm.nih.gov/genome?term=oryctolagus%20cuniculus), UCSC (http://genome.ucsc.edu/cgi-bin/hgGateway?org=Rabbit&db=oryCun2) and Ensembl (http://ensembl.org/Oryctolagus_cuniculus/Info/Index). We also used the command-line version of BLAST, including the BLASTN, TBLASTN (Gertz et al. 2006), and MegaBLAST (Zhang et al. 2000) modules. We used IMGT/V-Quest (Brochet et al. 2008) to identify V-regions within several rearranged heavy chain sequences. MacVector versions 11.2 and 12.6 were used for some sequence comparisons and annotations. One set of alignments was performed using Splign (Kapustin et al. 2008).

As direct alignment of human transcripts to the sequences of the OryCun2.0 assembly did not identify some of the orthologous genes in the region of the assembled rabbit chromosome 20 adjacent to the *IGH* locus, we adopted a two-stage approach, using assembly of the rabbit RNASeq data, kindly provided by Dr. Jessica Alföldi (Broad Institute), as a stepping stone. The RNASeq data were obtained from ten different tissues/cell types of a non-inbred NZW rabbit plus testes and pooled ovaries from other NZW rabbits. The genes on human chromosome 14 between coordinate 100 Mbp and the q-telomere in NCBI build 37.3 were obtained using NCBI MapViewer (Sayers et al. 2012). The list was filtered to retain only those genes that had a protein product in RefSeq (Pruitt et al. 2012) and a “PROVISIONAL”, “REVIEWED” or “VALIDATED” RefSeq status. The gene *RD3L*, a recent addition to RefSeq, was added to the list. To search for RNA that corresponds to genes syntenic to the telomeric region of rabbit chromosome 20q, we aligned 67 genes encoding 128 proteins to rabbit RNASeq transcripts from 10 tissue types using TBLASTN. For each of the 66 genes with an alignment of at least one protein product to one RNASeq transcript with E-value 7×10^{-13} or less, we took the highest-scoring match from the database of RNASeq sequences as the representative transcript for that gene, and aligned the transcript to the OryCun2.0 rabbit genome assembly using Splign. Whenever Splign produced two alignments to the same region, we chose the alignment with the best coverage (defined as the percentage of the query transcript that is aligned to an exon with at least 95% sequence identity). In addition, alignments to the sequence assembly OryCun2.0 and the rabbit WGS traces in the NCBI Trace Archive (Sayers et al. 2012) were performed using MegaBLAST to position transcripts that had not been placed by Splign.

To search for BACs derived from the *IGH, IGK*, or *IGL* regions among the BAC clones in the BAC Library of the Children's Hospital Oakland Research Institute (CHORI) LBNL-1: White Rabbit (Oryctolagus cuniculus) http://bacpac.chori.org/library.php?id=159; (Ros et al. 2004, 2005), we used MegaBLAST to search traces created by the Broad Institute by partial sequencing of BACs from the CHORI library. The Broad Institute deposited these *O.cuniculus* traces from the CHORI library in NCBI's Trace Archive with the

"center_project" attribute set to G1348. The methods we used to search for BACs with MegaBLAST are described in more detail as part of the Supplementary online resource in which the BACs are listed.

Cytogenetic and fluorescence *in situ* hybridization analyses (FISH) on RBP-banded chromosomes

Ros et al. (2004) described the preparation of a BAC library and reported the sequence of a BAC (Genbank AY386696) that encompasses genes encoding IgM, IgG, IgE and four IgA isotypes (Figure 1). Dr. Josef Platzer, Roche Diagnostics, Penzberg, Germany generously provided a stab culture containing this BAC clone 27N5 (AY386696). The BAC DNA was purified and labeled to carry out FISH on rabbit metaphase spreads as described (Hayes et al. 2002). To obtain prometaphase and early metaphase R-banded chromosome spreads, rabbit fibroblast cell cultures were synchronized with an excess of thymidine and treated with 5-bromodeoxyuridine (BrdU) during the second half of S phase (Hayes et al., 1991). RBP bands (R for R-banding, B for BrdU incorporation and P for Propidium iodide) were revealed according to Lemieux et al. (1992). Chromosomes are numbered according to the standardized reference rabbit karyotype (Hayes et al. 2002).

Results

Immunoglobulin allotypes in the donor strain

Results of typing sera from 11 Thorbecke inbred rabbits obtained about 10 years before the OryCun donor rabbit was selected for sequencing support the allotype sequences found in the OryCun2.0 assembly (Table 1). Table 1A shows the serum typing results for *IGHV1*- and *IGHG*-encoded allotypes. Two rabbits were homozygous for *VH1a1, d11, e15* (probably haplotype C) and four for *VH1a2, d12, e15* (probably haplotype E, F, or M) (Mage et al. 2006). Even though the animals had been inbred to accept allografts, five rabbits were heterozygous for the *IGHV1a* alleles *IGHV1al* and *IGHV1a2*, and for *IGHG* hinge region alleles *IGHG d11* and *d12*. The alleles *d11* ATG Met and *d12* ACG Thr distinguish the two haplotypes *VH1a1, d11, e15* and *VH1a2, d12, e15* that were present in the strain in 1995. Although the same Thorbecke strain of rabbit used for the OryCun2.0 assembly had undergone nine more years of inbreeding, it was still heterozygous at the *IGH* locus with two *IGHV1a* alleles *IGHV1al* and *IGHV1a2* (detailed below). Table 1B shows the typing results for *IGKC1*- and *IGKC2*-encoded allotypes in sera from the same 11 rabbits. Two of these 11 animals were heterozygous for the *IGKC1* b4 and *IGKC1* b5 types and none had the *IGKC2* bas1 type. Although the b4 and b5 allotype-encoding alleles at the *IGKC1* locus were not fixed in the strain in 1995, it appears that the rabbit used for the OryCun2.0 assembly was homozygous for *IGKC1* type b5. We identified the *IGKC1* gene sequence encoding the b5 allotype and the *IGKC2* encoding bas2 on chromosome 2 of the OryCun2.0 assembly (detailed below).

Placement of *IGH* on an unassembled region of chromosome 20 by FISH

The *IGH* locus could not be detected on any of the OryCun2.0 assembled chromosomes. Based on synteny data of genes conserved in man, mouse, rat, and dog obtained by accessing Homologene and MapViewer links via NCBI's Entrez Gene database (Maglott et al. 2011), the rabbit *IGH* region was predicted to be localized at the telomeric region of rabbit chromosome 20. FISH analyses were performed (Figure 2) using rabbit BAC 27N5 (AY386696; see Figure 1) to map the rabbit *IGH* locus. Figure 2 shows a rabbit metaphase spread with R-banded chromosomes stained in red by propidium iodide and yellow-green fluorescent hybridization signals on the telomeric regions of both chromosome 20 homologues, as predicted. Figure 2c shows a diagrammatic representation (Ideogram) of the R bands on rabbit chromosome 20.

Identification of parts of the *IGH* locus in unplaced scaffolds and BACs

We evaluated whether any of the difficulties in assembling the rabbit *IGH* locus extended to a larger contiguous segment of chromosome 20, further from the telomere. In human, mouse, rat and dog, *CKB* (creatine kinase, brain), which has been mapped to rabbit chromosome 20 (Rogel-Gaillard et al. 2009), and *IGH* are localized near the same chromosome telomere. Rabbit unplaced scaffold chrUn0053 (NW_3159377.1) contains the gene *CKB*. It was not possible to join chrUn0053 to the assembled rabbit chromosome 20, or to extend chrUn0053 further to include the *IGH* locus. The most telomeric human gene that could be placed on chrUn0053 is *ASPG*. A number of genes that lie between *ASPG* and *IGH* on human chromosome 14 could not be placed anywhere in the OryCun2.0 assembly (Table 2 and Supplementary Table S1), despite the presence of identifiable ortholog transcripts in rabbit RNASeq data. Well-characterized genes such as *SIVA1* and *AKT1* are missing, as is *TMEM121*, the non-*IGH* human protein-coding gene nearest to *IGH*. For several of these unplaced genes, it was possible to align representative transcripts from the rabbit RNASeq data to the OryCun2.0 WGS traces with high percent identities. Additional sequencing of BACS is needed to anchor the *IGH* locus to the current rabbit genome assembly. BACs with *IGHV* genes in the libraries described by Ros et al. (2004), Rogel-Gaillard et al. (2001), and in the CHORI LBNL-1 described below could contribute to this effort.

We tried to identify parts of the *IGH* locus in OryCun2.0, including unplaced scaffolds. Rabbit *IGHV, IG DH, IGHJ, IGHM, IGHG, IGHE* and *IGHA* genes encode immunoglobulin heavy chains (reviewed in Mage et al. 2006). Rabbits have typical IgM, no IgD, and only one isotype each of IgG and IgE, but 13 different IgA isotypes (Burnett et al. 1989; Volgina et al. 2005)

Variable (*IGHV*), Diversity (*DH*) and Joining (*JH*) region genes

The allelic *IGHV1a* genes *IGHV1a1, IGHV1a2* or *IGHV1a3* that are closest to the *DH* and *JH* gene clusters are the most frequently rearranged and expressed. The donor rabbit appears to have been heterozygous for the IGHV1a1 and IGHV1a2 allotypes, consistent with the serum typing of some members of the donor strain (Table 1A). The unplaced scaffold chrUn0742 (NW_003160066.1) has an exact match to a published sequence encompassing *IGHV1a1* (M93171.1), but the match is surrounded by other *VH* genes, rather than occurring in the expected location nearest to *DH* and *JH* gene clusters (Becker et al. 1989; Becker and Knight 1990; Ros et al. 2004). Scaffold chrUn0439 has a sequence that is 99% identical to the previously reported promoter region of *VH1a2* (U26530.1; Tunyaplin and Knight, 1995) and an exact match to the published sequence encompassing *VH1a2* (M93172.1). These are followed as expected by *DH* and *JH* gene clusters, and by constant-region genes extending as far as those encoding Cγ and Cε (Figure 1).

Most other *IGHV* genes that do rearrange are collectively referred to as a-negative because their protein products do not react with the anti-VH1a allotype sera used to define the allelic *IGHV1a* gene products. Generally 10-30% of immunoglobulins with sequences reflecting their origin from rearranged a-negative genes are found in normal rabbits. Some selected high affinity antibodies recovered by phage display have been found to have rearranged expressed a-negative genes (R Mage unpublished). None of the 16 *VH* in unmapped BAC AY386694 (Figure 1) have sequences for x, y, or z genes (Ros et al, 2004) so the genomic context of these genes remains unknown. They are thought to have been conserved for response to particular pathogens. By sequence analysis, we found 79 unplaced scaffolds containing as few as one and as many as 65 IGHV genes (Online resource Supplementary Table S2). Scaffold chrUn0439 contains *IGVH1a2* and a 322 bp fragment of the *IGVH2a2* gene (M93175.1) that extends from position 177140-177730, 5 bps short of the end of the

scaffold, leaving no room for other *IGHV* genes. *VH3a2* and *VH4a2* are found on chrUn0493 (NW_003159817.1). BLAST searches with a y33 sequence (M77083.1 positions 274-560) yielded 4 hits with 96-99% identity on chrUn0195, two matches on chrUn0789 (NW_003160113.1) and single hits with 96- 99% identity on 11 other unplaced scaffolds; x32 sequence L03900.1 is also was located on chrUn0195 as well as two other chrUn (2408 and 0740), and a z sequence was on yet another scaffold (chrUn1072).

ChrUn0439, the scaffold on which *VH1a2* is located, and chrUn0499 both have sequences corresponding to the region of BAC 219D23 (AY386695) that contains genes encoding DH regions (Figure 1). The differences between *JH* gene sequences in previously known haplotypes are mixed in the assembly of the JH genes present in chrUn0439. However, chrUn0499 is misassembled. A 22.3 kb gap in the chrUn0499 assembly includes part of the region where *DH* genes would have been assembled. Moreover, at least eleven *VH* sequences found in chrUn0499 do not correspond to the order expected from previously ordered *VH* genes in cosmid or BAC clones from VHa1 or VHa2 haplotypes (Becker et al. 1989; Becker and Knight, 1990; Ros et al. 2004).

Constant region (*IGHC*) genes

Table 3 summarizes comparisons of constant region exon sequences and positions of *IGHM, IGHG* and *IGHE* in BAC 27N5 (AY386696) with those in chrUn0439.

IGHM—*IGHM* encoded constant domains 1- 4 are found between positions 72286 and 74083 on the reverse strand of chrUn0439. BAC 27N5 (AY386696; Fig.1; Ros et al. 2004), sequenced from a strain of rabbit with a recombinant haplotype F-I (Mage et al. 1971; Newman et al. 1991), aligns well to chrUn0439, but there are some differences as summarized in Table 3A. The most troubling of these differences introduces a premature stop codon, and although it is likely to be a sequencing error, the trace archive contain three traces encoding Ser and four encoding the stop codon TCA.

IGHG—*IGHG* exons encoding the CH1, hinge region, CH2 and CH3 are found between positions 22868 and 24231 on the reverse strand of ChrUn0439. However, the *IGHG* hinge region assembled in chrUn0439 encodes the d11 (ATG-Met) allotype predicted from the typing results to be associated with the VH1a1 haplotype and not the d12 (ACG-Thr) allotype associated with VH1a2 (Table 3B). Of three traces found corresponding to the hinge region, codons for the ATG-Met were present in two and ACG-Thr was present in the third. A silent change, from ACA to ACG, in the codon for Ala formed by the splicing of the CH1 domain to the hinge region was also present in the sequences of all three traces (Table 3B). ACG has been found in more than 40 sequenced hinge regions from domestic breeds and other leporid species (Esteves et al. 2006) but ACA is found in rabbits of the F-I haplotype [BAC 27N5 (AY386696) and the cDNA sequence K00752.1 of Bernstein et al. (1983)].

IGHE—The sequence in chrUn0439 that contains the four exons encoding rabbit IgE is 99% similar to that of BAC clone 27N5 (AY386696). Table 3C shows the three silent and one replacement change found in exons 1 and 4. These differences are all supported by sequences in the trace archives.

IGHA, Negative Regulatory Elements and the 3’ enhancer region of the *IGH* locus

As shown in Figure 1, Ros et al. (2004) mapped the genes encoding Cα4, Cα5b, Cα1, and Cα2 within BAC clone 27N5 (AY386696) as closest to the gene encoding Cε. They found the Cα13 gene, located at the 3’ end of the *IGH* locus, and the enhancer sequences (hs) for the locus within a Fosmid15B (AY386698; Kingzette et al. 1998). Volgina et al. (2005) later

described conserved 300 bp negative regulatory elements (NRE) associated with 8 of the 13 functional Cα. Although absent from chromosome 20, some *IGHA* constant regions and NRE sequences were found in unplaced scaffolds. Supplementary Table S4 summarizes locations of *IGHA*, NRE and hs1, 2 sequences found in the BAC clone 27N5 (AY386696), Fosmid15B (Figure 1), and partial matches in unplaced scaffolds chrUn1855 (NW_003161178.1) with Cα13, NRE and hs1, 2; chrUn2352 (NW_003161675.1) with Cα9; chrUn1490 (NW_003160813.1) with Cα11/Cα12; and chrUn2570 (NW_003161893.1) with Cα7 NRE.

An *IGHV* orphon on chromosome 7

According to the definition in IMGT (International ImMunoGeneTics Information System; Lefranc and Lefranc, 2001; http://www.imgt.org/IMGTindex/orphon.html), “orphons” are genes similar to immunoglobulin genes found outside the *IGH, IGK* or *IGL* loci. A single *IGHV* orphon on chromosome 7 is non-functional, both because it has missing coding sequence near the 3’ end and because its location lacks *DH, JH* and *CH* genes.

Immunoglobulin Light chain genes

***IGK1* and *IGK2* light chains—**The *IGK* locus was predicted to be on rabbit chromosome 2, based on synteny with human chromosome 2 (Korstanje et al. 1999). The *IGKC* genes (Figure 3A) are duplicated in the rabbit and the two adjacent regions are denoted *IGKC1* and *IGKC2* (reviewed in Mage et al. 2006). The *IGK2* locus is normally poorly expressed (McCartney-Francis et al. 1987). The differential expression is partly due to the fact that (compared to *IGK1*) the intronic enhancer of *IGK2* and surrounding matrix-associated region has undergone deletions (Emorine and Max 1983b; Sperry et al. 1989). In OryCun2.0, the *IGKC1* region is located at 98.92 Mb on the forward strand of chromosome 2. Part of the *IGKV* and *IGKJ* region associated with *IGKC1* was sequenced by Ros et al. (2005) (Figure 3b). The *IGKC2* region is also found on assembled chromosome 2, at 98.22 Mb, but in the opposite transcriptional orientation (negative strand). The ~0.8 Mb span of the duplicated *IGK* genetic region on chromosome 2 shown may be less than the true size because of inaccurate estimates for gaps in the assembly.

Table 4 shows locations of *IGKC1, IGKC2, IGKJ* (Jκ), enhancers, and the range over which identified *IGKV* (Vκ) genes are found on chromosome 2. Supplementary Tables S5A and S5B provide more detailed information about Vκ genes, pseudogenes and gene fragments found on chromosome 2 and unplaced scaffolds. Despite the presence of the b4 constant region allotype in some rabbits of the donor strain nine years earlier (Table 1B), we found that the OryCun2.0 rabbit was homozygous at *IGKC1* for the b5 allotype and the *IGKC2* sequence corresponds to the bas2 allotype as inferred from the typing of sera (Table 1B). Supplementary Table S5B lists 21 Vκ genes or pseudogenes present in three unplaced scaffolds.

In rabbits of b4, b5 and b6 types, there is an extra Cys at position 80 (IMGT position 96) that forms an interdomain disulfide bond with Cys 171 of the *IGKC1* constant region. Such an inter-domain bond in addition to the usual intradomain disulfide bond between Cys at positions 23 and 88 (IMGT position 104) cannot form with the *IGKC2* constant region and is not found in orders other than Lagomorpha. Vκ lacking Cys 80 are found in rabbits of b9 allotype (McCartney-Francis et al. 1984; Popkov et al. 2003) but are rare in those of b5 type (Sehgal et al. 1999). It was of interest to know where there were Vκ genes with and without Cys at position 80 in OryCun2.0, since the donor rabbit was of the b5 allotype. The column headed Cys 80 in Table 4 shows whether the particular Vκ gene sequence or group of sequences contains a codon for Cys 80. Most of the *IGKV* identified and mapped between *IGCK1* and *IGCK2* include the Cys 80. However, 14 Vκ genes or pseudogenes that

surround *IGKC2* and associated *IGKJ1, IGKJ2-4* and *IGKJ3*, encode either Ala or Pro at position 80 and only two encode Cys. The 3' enhancer, *IGKC2* and associated Jκ are in the opposite orientation from most of these Vκ. Most Vκ are also encoded on the reverse strand at higher genetic coordinates between the two *IGKC*. However, interspersed Vκ encoded on the forward strand are present. The last column in Table 4 shows intervals where gaps in the assembly of more than 200 bp are present in the region of interest on chromosome 2. Assignment of individual numbers to Vκ must be provisional because of these and additional gaps in the three chrUn containing Vκ gene sequences.

Lambda light chains

Table 5 shows the positions of *IGLC* and *IGLJ* gene sequences found on chromosome 21 (NC_013689.1) and several unplaced scaffolds (chrUn). New information about *IGL* genes of rabbit, including four *IGLJ* and 43 *IGLV* genes, was added to IMGT during 2012 (http://imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=rabbit&group=IGLV). IMGT lists 7 subgroups of *IGLV* genes in OryCun2.0 on chromosome 21. Positions shown for OryCun2.0 chromosome 21 by IMGT are numbered as found in genomic scaffold NW_003159316.1. Because chromosome 21 starts with an arbitrary offset of 3 Mb to allow for an unmapped telomeric segment, the chromosome 21interval 3000001-4922929 corresponds to 1-1922929 on scaffold NW_003159316.1. As of January 9, 2013, 21 of the 43 *IGLV* are listed as pseudogenes, 20 as functional, and two as ORFs, potentially functional open reading frames. In addition to the initial 3 Mb gap, the scaffold containing the *IGL* genes consists of more than 100 contigs with gaps between them. Rabbit lambda constant region genes, *IGLC5* paired with *IGLJ5* and *IGLC6* paired with *IGLJ6* are known to be functional (Hayzer et al. 1990; Jaton et al. 1990).

Discussion

Rabbit immunogenetics is of real-world importance. The rabbit is the best or the only animal model for syphilis, tularemia, tuberculosis, and other infectious and autoimmune diseases. Rabbits are significant commercial sources of highly specific high affinity antibodies. Therefore, it is important to analyze the sequences that contribute to the rabbit antibody diversification process. Even if only a fraction of the immunoglobulin sequences can be identified, these sequences can be added to the database underlying NCBI's IgBLAST (Ye et al., in press) and hence be useful for studies of diversified rabbit antibodies. There is considerable pre-genome-sequencing literature about allelic variants and isotypes of rabbit immunoglobulins (Mage et al., 2006, Pinheiro et al. 2011)). In this study, we evaluated the status of the assembly, chromosomal assignments, and annotations of the immunoglobulin loci in OryCun2.0. To support our analysis, we report typing of sera from rabbits of the same strain as the rabbit used for sequencing; the sera were typed in 1995. By FISH, we definitively placed the heavy-chain locus (*IGH*) near the q-telomere of rabbit chromosome 20, as would be expected by synteny with human.

The heavy-chain genes are not placed on any rabbit chromosome in OryCun2.0, though we report partially assembled pieces of the *IGH* locus found among the unplaced scaffolds. Immunoglobulin genes are difficult to assemble, particularly the variable regions comprising repeats of highly similar sequences. The constant region genes of the *IGH* locus, however, are less repetitive, more highly conserved between animals, and are predicted, by synteny, to be furthest from the telomere. One might hope to join the constant heavy-chain region to the assembled chromosome 20 or to a large unplaced scaffold that might be identified as part of chromosome 20. We therefore attempted to locate within OryCun2.0 the sequences of genes that by synteny should be near the same telomere of chromosome 20. Some genes could be located in the assembly on unplaced scaffolds, notably chrUn0053, but others, particularly

those near the *IGH* locus are missing from the assembly (Table 2, Supplementary Table S1). Despite the absence of these genes in OryCun2.0, the genes were readily identified as present in RNASeq data from other rabbits.

We have therefore shown that the assembly of chromosome 20 in OryCun2.0 is incomplete. Inadequate coverage of chromosome 20 may be due to a) idiosyncrasies of rabbit chromosome 20; b) anomalies in the sequenced rabbit; or c) a larger than expected lower tail in the coverage when the genome coverage by high-quality traces is 6.51x. Such a negative result objectively justifies additional sequencing efforts. To aid these efforts, this study collects sequences from within OryCun2.0 that should be compared when future rabbit genome sequences are generated.

We identify at least 372 *IGHV* genes found scattered on at least 79 unplaced scaffolds, most prominently ChrUn0195 (Supplementary Table S2). The total number of rabbit *IGHV* genes per haploid genome is uncertain because the sequenced rabbit was heterozygous at the *IGH* locus. These *IGHV* fragments may be useful in analyzing rabbit rearranged and diversified heavy chain sequences. Unfortunately, the *IGHV* genes cannot be mapped; our sequence analysis of scaffolds chrUn0195, 0493, 0499, 0742, each of which contains at least 11 *IGHV* genes, shows their partial assemblies are not consistent with previous targeted sequencing studies of the *IGH* locus.

The most complete portion of the *IGH* locus can be found in chrUn0439. It includes *VH1* (VH1a2 allotype) and extends through the *IGHG, IGHM*, and *IGHE* genes (Figure 1 and Table 3). There is reason to suspect some misassembly in chrUn0439. To support our analysis of immunoglobulin genes, we report typing of sera from 11 *Thorbecke* rabbits in 1995. The *IGHG* hinge region assembled in chrUn0439 (NW_003159763.1) encodes the d11 (ATG-Met) predicted from the serum typing results in Table 1 to be associated with the VH1a1 haplotype and not VH1a2 (Mage et al. 2006). A haplotype with VH1a2 and d11 had never been reported until a rare recombinant [R3K (F-C)] was generated during intensive laboratory breeding (Kelus and Steinberg, 1991).

The *IGHA* locus and associated NREs are partially represented in ChrUn1490, 1855, 2352, and 2570 (Supplementary Table S4), but various previously reported Cα genes (Volgina et al. 2005) are missing in OryCun2.0. Thus, there is still no complete map of rabbit *IGHA*. Ros et al. (2004) probed their BAC library (~ 4X coverage) and the CHORI BAC library (~7X coverage) and failed to detect a BAC containing the *IGHA* region. Our search also did not detect *IGHA*-containing BACs in the CHORI library.

Both copies of the duplicated *IGK* locus are found, as expected, on chromosome 2 of the OryCun2.0 assembly. We identified 94 Vκ on chromosome 2 (Tables 4, S5A) and an additional 21 Vκ on three unplaced scaffolds (Table S5B). The 21Vκ on unplaced scaffolds may belong in the assembled chromosome 2, represent parts of the unassembled haplotype of a heterozygous rabbit, or be orphons. Previous estimates of the total number of rabbit Vκ genes included: more than 50 (Heidmann and Rougeon, 1984; Lamoyi and Mage 1985), 39 to 135 (Sehgal et al. 1999), and more than 100 (Ros et al. 2005). Although 94 is comparable to previous estimates of the total number of Vκ regions, it may be an underestimate, because there are gaps in the chromosome 2 assembly.

With the relative positions and map of Vκ regions present in OryCun2.0 tabulated in this study, potential sources of gene conversion donor sequences in clonally diversified B cells can now be located in more detail than when first attempted by Sehgal et al. (2000). During human B-cell development, rearrangement of the distal Vκ to Jκ genes occurs by inversion and retains adjacent Vκ (Weichhold et al. 1990). The rabbit *IGK* locus is more complex because opposite transcriptional orientations of Vκ genes would result in most gene

rearrangements occurring by inversion and retention of Vκ (Figure 3a and Table 4). Identification of the rabbit Vκ makes more feasible analyses of the rearranged kappa *IGKVJC* gene sequences, similar to those for human kappa (Cox et al. 1994; Collins et al. 2008).

Rabbit mAbs produced by phage display can best be obtained as chimeric chains with rabbit Vκ, and human Cκ but this leads to an unpaired Cys80 (Rader et al. 2000). Popkov et al. (2003) showed that b9 and *Basilea* rabbits, strains known to have Vκ that lack Cys80, produced higher yields of selected chimeric rabbit-human Fab displayed on phage than NZW rabbits of b4 allotype. Most likely, assembly of expressed protein occurred more efficiently when the free thiol group from an unpaired Cys80 was not present. We carried out multiple alignments and annotated the residue at position 80 in the Vκ genes (Tables 4 and S5 and Fig 3a). We found 14 Vκ genes that lack Cys80 surrounding the *IGKC2* region.

The *IGL* loci are on chromosome 21 of OryCun2.0, as expected. We summarized information on *IGL* genes in Table 5. With sequences of *IGL V* genes available, it will be possible to analyze the lambda light chain sequences recovered by phage display from *Basilea* and b9 rabbits (Popkov et al. 2003), better identify the sources of *VL* genes used for rearrangement, and evaluate whether gene conversion contributed to sequence diversification.

Concluding Remarks

The Genome Reference Consortium (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/), recently announced:

> "...plans to update the human reference assembly to GRCh38 in the summer of 2013. This revision is aimed at addressing issues found with the current model for representing genome assemblies, which uses a single, preferred tiling path to produce a single consensus representation of the genome. Subsequent analysis has shown that for most mammalian genomes a single tiling path is insufficient to represent a genome in regions with complex allelic diversity. The GRC is working to create assemblies that better represent this diversity and provide more robust substrates for genome analysis."

NCBI has established mechanisms for including alternate assemblies and patches of specific loci. As of January 2013, there are some alternate locus assemblies, including for MHC on chromosome 6, and patches available for the human genomes in MapViewer, but no alternate assemblies for the immunoglobulin loci any other loci on human chromosomes 2, 14, or 22. It will be useful if future builds of mammalian genomes also use the framework of patches in these immunoglobulin gene regions.

Rabbits of the Thorbecke OryCun2.0 strain were more susceptible to infection with several strains of *Mycobacterium tuberculosis* than outbred rabbits from the same supplier (Covance) (Dorman et al. 2004; Mendez et al. 2008). These authors noted that in addition to likely immunological differences from outbred NZW, the physical appearance of the inbred rabbits suggested that they had developmental defects. Our analyses of the OryCun2.0 genome assembly and annotations reveal a need to improve the current assembly and annotations, not only to include allelic variation, but variation due to multiple copies of genes encoding antibody *IGH* and *IGK* variable regions, variable and constant regions of lambda light chains, and *IGHA* constant regions. Genome sequences of other rabbit strains are also needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the Intramural Research Programs of the NIH, NIAID and NLM and by the Animal Genetics and Animal Physiology Divisions of INRA.

We thank Dr. Josef Platzer, Roche Diagnostics, Penzberg, Germany for the gift of the 27N5 BAC, Dr. Jessica Alföldi of the Broad Institute for rabbit RNASeq data, Drs. Christoph Rader, Nancy McCartney-Francis and Michael Mage for suggestions to improve the manuscript, and Daniel Schäffer for assistance with the figures. The UCSC rabbit Genome Browser was produced with the following acknowledgements:

Sequencing/Assembly: The Broad Institute at MIT and Harvard, Cambridge, MA, USA

UCSC Rabbit Genome Browser (oryCun2): Hiram Clawsom and Antonio Coelho

Initial Genome Browser Annotations: UCSC Genome Bioinformatics Group, University of California, Santa Cruz, CA, USA. BLAST

References

Allegrucci M, Newman BA, Young-Cooper GO, Alexander CB, Meier D, Kelus AS, Mage RG. Altered phenotypic expression of immunoglobulin heavy-chain variable-region (VH) genes in Alicia rabbits probably reflects a small deletion in the VH genes closest to the joining region. Proc Natl Acad Sci USA. 1990; 87:5444–5448. [PubMed: 2115171]

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST - A new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]

Becker RS, Zhai SK, Currier SJ, Knight KL. Ig VH, DH, and JH germ-line gene segments linked by overlapping cosmid clones of rabbit DNA. J Immunol. 1989; 142:1351–1355. [PubMed: 2492580]

Becker RS, Knight KL. Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. Cell. 1990; 63:987–997. [PubMed: 2124176]

Bernstein KE, Alexander CB, Mage RG. Nucleotide sequence of a rabbit IgG heavy chain from the recombinant F-I haplotype. Immunogenetics. 1983; 18:387–397. [PubMed: 6313520]

Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res. 2008; 36:W503–508. [PubMed: 18503082]

Burnett RC, Hanly WC, Zhai SK, Knight KL. The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 Cα genes. EMBO J. 1989; 8:4041–4047. [PubMed: 2512120]

Chantry-Darmon C, Rogel-Gaillard C, Bertaud M, Urien C, Perrocheau M, Chardon P, Hayes H. 133 new gene localizations on the rabbit cytogenetic map. Cytogenet Genome Res. 2003; 103:192–201. [PubMed: 15004485]

Chantry-Darmon C, Bertaud M, Urien C, Chadi-Taourit S, Perrocheau M, Rogel-Gaillard C, Hayes H. Expanded comparative mapping between man and rabbit and detection of a new conserved segment between HSA22 and OCU4. Cytogenet Genome Res. 2005a; 111:134–139. [PubMed: 16103654]

Chantry-Darmon C, Urien C, Hayes H, Bertaud M, Chadi-Taourit S, Chardon P, Vaiman D, Rogel-Gaillard C. Construction of a cytogenetically anchored microsatellite map in rabbit. Mamm Genome. 2005b; 16:442–459. [PubMed: 16075371]

Chantry-Darmon C, Urien C, de Rochambeau H, Allain D, Pena B, Hayes H, Grohs C, Cribiu EP, Deretz-Picoulet S, Larzul C, Save JC, Neau A, Chardon P, Rogel-Gaillard C. A first-generation microsatellite-based integrated genetic and cytogenetic map for the European rabbit (*Oryctolagus cuniculus*) and localization of *angora* and *albino*. Anim Genet. 2006; 37:335–341. [PubMed: 16879342]

Collins AM, Wang Y, Singh V, Yu P, Jackson KJ, Sewell WA. The reported germline repertoire of human immunoglobulin kappa chain genes is relatively complete and accurate. Immunogenetics. 2008; 60:669–676. [PubMed: 18712520]

Cox JPL, Tomlinson IM, Winter G. A directory of human germ-line Vκ segments reveals a strong bias in their usage. Eur J Immunol. 1994; 24:827–836. [PubMed: 8149953]

Dorman SE, Hatem CL, Tyagi S, Aird K, Lopez-Molina J, Pitt MLM, Zook BC, Dannenberg AM Jr. Bishai WR, Manabe YC. Susceptibility to tuberculosis: clues from studies with inbred and outbred New Zealand White rabbits. Infect Immun. 2004; 72:1700–1705. [PubMed: 14977978]

Duvoisin RM, Hayzer DJ, Belin D, Jaton J-C. A rabbit Ig λ L chain C region gene encoding c21 allotypes. J Immunol. 1988; 141:1596–1601. [PubMed: 2842399]

Emorine L, Dreher K, Kindt TJ, Max EE. Rabbit immunoglobulin κ genes: structure of a germline b4 allotype J-C locus and evidence for several b4-related sequences in the rabbit genome. Proc Natl Acad Sci USA. 1983a; 80:5709–5713. [PubMed: 6412231]

Emorine L, Max EE. Structural analysis of a rabbit immunoglobulin κ 2 J-C locus reveals multiple deletions. Nucleic Acids Res. 1983b; 11:8888–8890.

Esteves PJ, Carmo C, Godinho R, van der Loo W. Genetic diversity at the hinge region of the unique immunoglobulin heavy gamma (IGHG) gene in leporids (*Oryctolagus, Sylvilagus and Lepus*). Int J Immunogenetics. 2006; 33:171–177. [PubMed: 16712647]

Esworthy S, Max EE. The rabbit κ1b5 immunoglobulin gene: another J region gene cluster with only one functional J gene segment? J Immunol. 1986; 136:1107–1111. [PubMed: 3079796]

Gertz EM, Yu Y-K, Agarwala R, Schäffer AA, Altschul SF. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol. 2006; 4:41. [PubMed: 17156431]

Hayes H, Petit E, Dutrillaux B. Comparison of RBG-banded chromosomes of cattle, sheep, and goats. Cytogenet Cell Genet. 1991; 57:51–55. [PubMed: 1855392]

Hayes H, Rogel-Gaillard C, Zijlstra C, de Haan NA, Urien C, Bourgeaux N, Bertaud M, Bosma AA. Establishment of an R-banded rabbit karyotype nomenclature by FISH localization of 23 chromosome-specific genes on both G- and R-banded chromosomes. Cytogenet Genome Res. 2002; 98:199–205. [PubMed: 12698004]

Hayzer DJ, Young-Cooper GO, Mage RG, Jaton J-C. cDNA clones encoding immunoglobulin chains from rabbit expressing the phenotype c7. Eur J Immunol. 1990; 20:2707–2712. [PubMed: 2125274]

Heidmann O, Rougeon F. Immunoglobulin kappa light-chain diversity in rabbit is based on the 3' length heterogeneity of germ-line variable genes. Nature. 1984; 311:74–76. [PubMed: 6433207]

Hole NJK, Harindranath N, Young-Cooper GO, Garcia R, Mage RG. Identification of enhancer sequences 3' of the rabbit immunoglobulin κL chain loci. J Immunol. 1991; 146:4377–4384. [PubMed: 1904082]

Hubisz MJ, Lin MF, Kellis M, Siepel A. Error and error mitigation in low-coverage genome assemblies. PLoS ONE. 2011; 6:e17034. [PubMed: 21340033]

Jaton J-C, Hayzer DJ, Frutiger S, Hughes GJ, Young-Cooper GO, Mage RG. Chemical and immunochemical identification of the second major rabbit immunoglobulin λ chain polypeptide bearing c7 epitopes. Eur J Immunol. 1990; 20:2713–2718. [PubMed: 1702725]

Kapustin Y, Souvorov A, Tatusova T, Lipman D. Splign: algorithms for computing spliced alignments with identification of paralogs. Biol Direct 21. 2008; 3:20.

Kelus AS, Steinberg CM. Is there a high rate of mitotic recombination between the loci encoding immunoglobulin VH and CH regions in gonial cells? Immunogenetics. 1991; 33:255–259. [PubMed: 1902823]

Kent WJ. BLAT – The BLAST-like alignment tool. Genome Res. 2002; 12:656–664. [PubMed: 11932250]

Kingzette M, Spieker-Polet H, Yam P-C, Zhai S-K, Knight KL. Trans-chromosomal recombination within the Ig heavy chain switch region in B lymphocytes. Proc Nat Acad Sci USA. 1998; 95:11840–11845. [PubMed: 9751752]

Knight KL, Becker RS. Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. Cell. 1990; 60:963–970. [PubMed: 2317867]

Korstanje R, O'Brien PCM, Yang F, Rens W, Bosma AA, van Lith HA, van Zutphen LFM, Ferguson-Smith MA. Complete homology maps of the rabbit (*Oryctolagus cuniculus*) and human by reciprocal chromosome painting. Cytogenet Cell Genet. 1999; 86:317–322. [PubMed: 10575232]

Lamoyi E, Mage RG. The lack of K1b9 light chains in Basilea rabbits is probably due to a mutation in an acceptor site for mRNA splicing. J Exp Med. 1985; 162:1149–1160. [PubMed: 3930650]

Lefranc, MP.; Lefranc, G. The Immunoglobulin Facts Book. Academic Press; San Diego: 2001.

Lemieux N, Dutrillaux B, Viégas-Péquignot E. A simple method for simultaneous R- or G-banding and fluorescence in situ hybridization of small single-copy genes. Cytogenet Cell Genet. 1992; 59:311–312. [PubMed: 1544332]

Lindblad-Toh K, Wade CM, Mikkelsenl TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ III, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, deJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin C-W, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis1 M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli K-P, Parker HG, Pollinger JP, Searle SMJ, Sutter NB, Thomas R, Webber C, Broad Sequencing Platform members. Lander ES. Genome sequence, comparative analysis and haplotype structure of the domestic Dog. Nature. 2005; 438:803–819. [PubMed: 16341006]

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, Ward LD, Lowe CB, Holloway AK, Clamp M, Gnerre S, Alföldi J, Beal K, Chang J, Clawson H, Cuff J, Di Palma F, Fitzgerald S, Flicek P, Guttman M, Hubisz MJ, Jaffe DB, Jungreis I, Kent WJ, Kostka D, Lara M, Martins AL, Massingham T, Moltke I, Raney BJ, Rasmussen MD, Robinson J, Stark A, Vilella AJ, Wen J, Xie X, Zody MC, Broad Institute Sequencing Platform and Whole Genome Assembly Team; Baldwin J, Bloom T, Chin CW, Heiman D, Nicol R, Nusbaum C, Young S, Wilkinson J, Worley KC, Kovar CL, Muzny DM, Gibbs RA, Baylor College of Medicine Human Genome Sequencing Center Sequencing Team; Cree A, Dihn HH, Fowler G, Jhangiani S, Joshi V, Lee S, Lewis LR, Nazareth LV, Okwuonu G, Santibanez J, Warren WC, Mardis ER, Weinstock GM, Wilson RK, Genome Institute at Washington University. Delehaunty K, Dooling D, Fronik C, Fulton L, Fulton B, Graves T, Minx P, Sodergren E, Birney E, Margulies EH, Herrero J, Green ED, Haussler D, Siepel A, Goldman N, Pollard KS, Pedersen JS, Lander ES, Kellis M. A high-resolution map of human evolutionary constraint using 29 mammals. Nature. 2011; 478:476–482. [PubMed: 21993624]

Mage RG, Young-Cooper GO, Alexander C. Genetic control of variable and constant regions of immunoglobulin H chains. Nature New Biology. 1971; 230:63–64.

Mage RG, Lanning D, Knight KL. B cell and antibody repertoire development in rabbits: the requirement of gut-associated lymphoid tissues. Develop Comp Immunol. 2006; 30:137–153.

Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. 2011; 39:D52–57. [PubMed: 21115458]

McCartney-Francis N, Skurla RM Jr, Mage RG, Bernstein KE. κ chain allotypes and isotypes in the rabbit: cDNA sequences of clones encoding *b9* suggest an evolutionary pathway and possible role of the interdomain disulfide bond in quantitative allotype expression. Proc Natl Acad Sci USA. 1984; 81:1794–1798. [PubMed: 6424124]

McCartney-Francis N, Young-Cooper G, Alexander C, Mage RG. Expression of K2 isotype mRNA in normal and Basilea rabbits. Molec Immunol. 1987; 24:357–364. [PubMed: 3116401]

Mendez S, Hatem CL, Kesavan AK, Lopez-Molina J, Pitt MLM, Dannenberg AM Jr, Manabe YC. Susceptibility to tuberculosis: composition of tuberculous granulomas in Thorbecke and outbred New Zealand White rabbits. Vet Immunol Immunopathol. 2008; 122:167–174. [PubMed: 18155300]

Newman BA, Young-Cooper GO, Alexander CB, Becker RS, Knight KL, Kelus AS, Meier D, Mage RG. Molecular analysis of recombination sites within the immunoglobulin heavy chain locus of the rabbit. Immunogenetics. 1991; 34:101–109. [PubMed: 1678366]

Pinheiro A, Lanning D, Alves PC, Mage RG, Knight KL, van der Loo W, Esteves PJ. Molecular bases of genetic diversity and evolution of the immunoglobulin heavy chain variable region (*IGHV*) gene locus in leporids. Immunogenetics. 2011; 63:397–406. [PubMed: 21594770]

Popkov M, Mage RG, Alexander CB, Thundivalappil S, Barbas CF III, Rader C. Rabbit immune repertoires as sources for therapeutic monoclonal antibodies: the impact of kappa allotype-correlated variation in cysteine content on antibody libraries selected by phage display. J Mol Biol. 2003; 325:325–335. [PubMed: 12488098]

Pruitt K, Tatusova T, Brown GR, Maglott DR. NCBI reference sequences (RefSeq) current status, new features and genome annotation policy. Nucleic Acids Res. 2012; 40:D130–D135. [PubMed: 22121212]

Rader C, Ritter G, Nathan S, Elia M, Gout I, Jungbluth AA, Cohen LS, Welt S, Old LJ, Barbas CF III. The rabbit antibody repertoire as a novel source for the generation of therapeutic human antibodies. J Biol Chem. 2000; 275:13668–13676. [PubMed: 10788485]

Rogel-Gaillard C, Piumi F, Billault A, Bourgeaux N, Save J-C, Urien C, Salmon J, Chardon P. Construction of a rabbit bacterial artificial chromosome (BAC) library: application to the mapping of the major histocompatibility complex to position 12q1.1. Mamm Genome. 2001; 12:253–255. [PubMed: 11252177]

Rogel-Gaillard, C.; Ferrand, N.; Hayes, H. Genome mapping in the rabbit.. In: Kole, C.; Cockett, NE., editors. Genome Mapping in Animals, Vol II Wildlife & Companion Animals. Springer-Verlag; Berlin, Heidelberg, New York, Tokyo: 2009. p. 165-230.

Ros F, Puels J, Reichenberger N, van Schooten WCA, Buelow R, Platzer J. Sequence analysis of 0.5Mb of the rabbit germline immunoglobulin heavy chain locus. Gene. 2004; 330:49–59. [PubMed: 15087123]

Ros F, Reichenberger N, Dragecic T, van Schooten W, Buelow R, Platzer J. Sequence analysis of 0.4 megabases of the rabbit germline immunoglobulin kappa1 light chain locus. Anim Genet. 2005; 36:51–57. [PubMed: 15670131]

Roux, KH.; Mage, RG. Rabbit immunoglobulin allotypes.. In: Weir, DM.; Herzenberg, LA.; Blackwell, C.; Herzenberg, LA., editors. Weir's Handbook of Experimental Immunology. 5th edition. Vol. 1. Blackwell Science Inc; Oxford, England: 1996. p. 26.1-26.17.

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Fingerman IM, Geer LY, Helmberg W, Kapustin Y, Krasnov S, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Karsch-Mizrachi I, Ostell J, Panchenko A, Phan L, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, Wilbur WJ, Yaschenko E, Ye J. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2012; 40:D13–25. [PubMed: 22140104]

Schiaffella E, Sehgal D, Anderson AO, Mage RG. Gene conversion and hypermutation during diversification of V_H sequences in developing splenic germinal centers of immunized rabbits. J Immunol. 1999; 162:3984–3995. [PubMed: 10201919]

Sehgal D, Schiaffella E, Anderson AO, Mage RG. Analysis of single B cells by polymerase chain reaction reveal rearranged V_H with germline sequences in spleens of immunized adult rabbits: implications for B cell repertoire maintenance and renewal. J Immunol. 1998; 161:5347–5356. [PubMed: 9820508]

Sehgal D, Johnson G, Wu TT, Mage RG. Generation of the primary antibody repertoire in rabbits: expression of a diverse set of *Igk-V* genes may compensate for limited combinatorial diversity at the heavy chain locus. Immunogenetics. 1999; 50:31–42. [PubMed: 10541804]

Sehgal D, Schiaffella E, Anderson AO, Mage RG. Generation of heterogeneous rabbit anti-DNP antibodies by gene conversion and hypermutation of rearranged V_L and V_H gene during clonal expansion of B cells in splenic germinal centers. Eur J Immunol. 2000; 30:3634–3644. [PubMed: 11169406]

Sperry AO, Blasquez VC, Garrard WT. Dysfunction of chromosomal loop attachment sites: illegitimate recombination linked to matrix association regions and topoisomerase II. Proc Natl Acad Sci USA. 1989; 86:5497–5501. [PubMed: 2546156]

Tunyaplin C, Knight KL. Fetal VDJ gene repertoire in rabbit: evidence for preferential rearrangement of VH1. Eur J Immunol. 1995; 25:2583–2587. [PubMed: 7589130]

Volgina V, Yam P-C, Knight KL. A negative regulatory element in the rabbit 3′ IgH chromosomal region. Int Immunol. 2005; 117:973–982. [PubMed: 16000331]

Weichhold GM, Klobeck H-G, Ohnheiser R, Combriato G, Zachau HG. Megabase inversions in the human genome as phyiological events. Nature. 1990; 347:90–92. [PubMed: 2118596]

Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. 2013 in press, doi: 10.1093/nar/gkt382.

Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol. 2000; 7:203–214. [PubMed: 10890397]

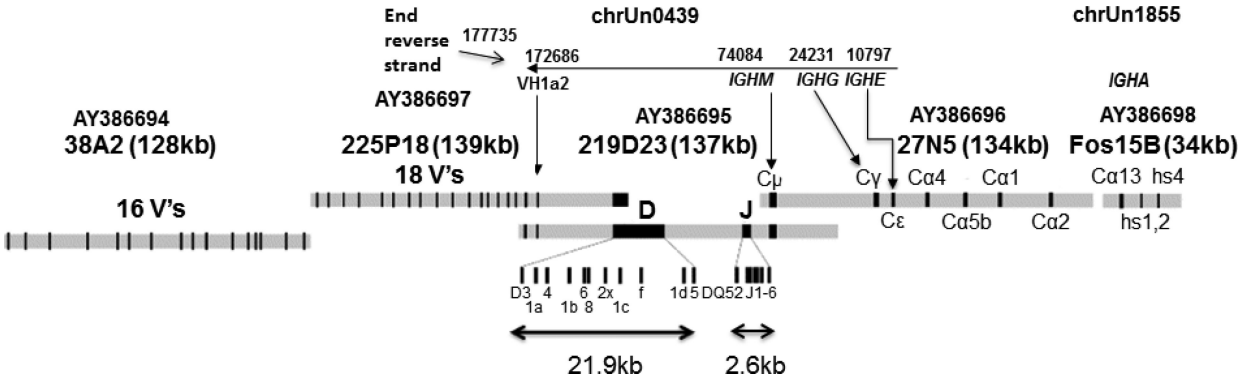

Fig 1.

Diagram of part of the rabbit *IGHV* and *IGHC* region described by Ros et al. (2004) adapted to annotate positions in OryCun 2.0. The donor rabbit was homozygous for the VH1a2 allotype. Arrows point to the VH1a2-encoding gene in overlapping clones 225P18 (AY386697) and 219D23 (AY386695) and the locations of *IGHM, IGHG* and *IGHE* on OryCun2.0 unplaced scaffold chrUn0439 (NW_003159763.1). Partial matches are found in unplaced scaffold chrUn1855 (NW_003161178.1) with Cα13, hs1, 2 and hs4. Note that BAC 38A2 (AY386694) with 16 VH genes does not overlap 225P18. Reprinted from Gene, Vol 330, Ros F, Puels J, Reichenberger N, van Schooten WCA, Buelow R, Platzer J, Sequence analysis of 0.5Mb of the rabbit germline immunoglobulin heavy chain locus, Pages No. 49-59, Copyright 2004, with permission from Elsevier.

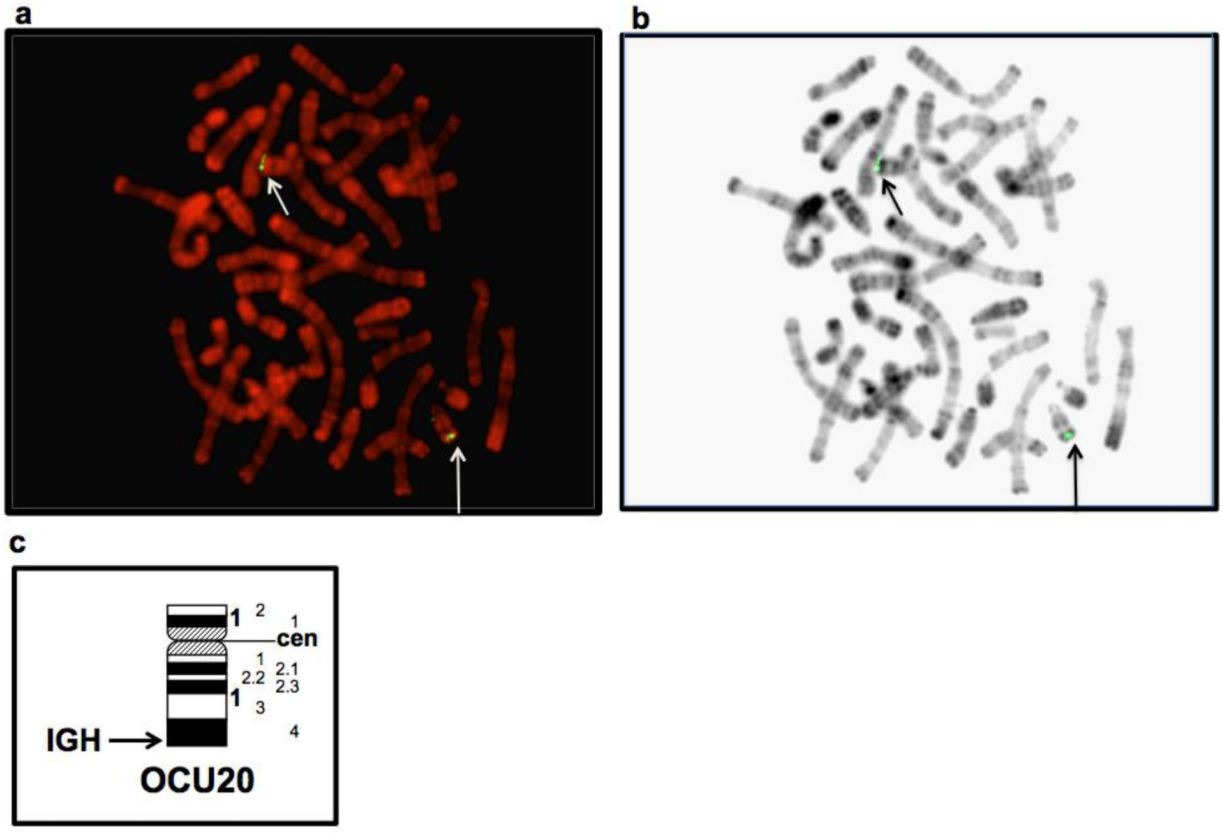

Fig 2.

Identification of rabbit *IGH* on chromosome 20. Arrows point to the hybridizing regions on rabbit Chromosome 20 seen by (**a**) fluorescence in situ hybridization and (**b**) on RBP-banded chromosomes (R-banding, with BrdU incorporation and Propidium iodide staining). Panel (**c**) shows a diagrammatic representation (Ideogram) of R bands on chromosome 20.

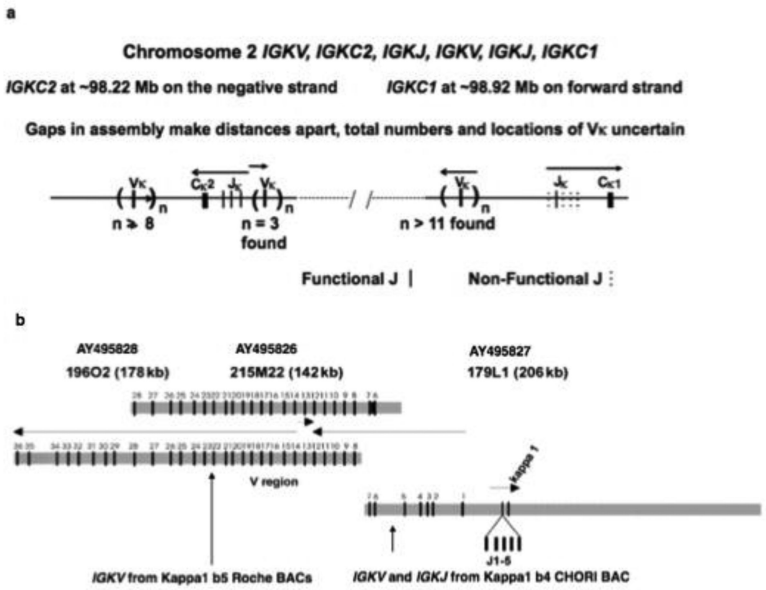

Fig 3. Diagrams of rabbit *IGKV* and *IGKC* regions

(**a**) Positions of *IGKC1* and *IGKC2* relative to *IGKV* genes. Arrows indicate transcriptional orientations as summarized in Table 4. The dashed line indicates the region between positions 98254316 and 98905144 (Table 4) where a few Vκ are present on the forward strand (+) and most are present on the reverse strand (−).

(**b**) Diagram of part of the *IGKV* and *IGKJ* regions associated with the *IGKC1* region sequenced by Ros et al. (2005). GenBank Accession numbers are shown above the BAC clones. A total of 36 Vk were identified. The rabbit donor of DNA used for generating overlapping BAC clones 19602 (AY495828) and 215M22 (AY495826) was homozygous for the *IGKC1* gene encoding the b5 allotype, as was the OryCun2.0 donor rabbit. BAC clone 179L1 (AY495827) from the LBLN-1 White Rabbit BAC library with *IGKC1* that encoded the b4 allotype, includes DNA for two Vκ that overlapped BAC 215M22, and 5 additional Vκ as well as the (*IGKJ1*) region having five distinct J segments and *IGKC1* of b4 allotype. Modified from Figure 1 in Ros F, Reichenberger N, Dragecic T, van Schooten W, Buelow R, Platzer J. (2005) Sequence analysis of 0.4 megabases of the rabbit germline immunoglobulin kappa1 light chain locus. Anim Genet 36:51-57, John Wiley & Sons Ltd and reproduced with permission of the publisher.

Table 1

Typing results for allotypes in sera of Thorbecke rabbits in 1995

A. *IGHV1*- and *IGHG-Encoded* Allotypes

Rabbit	VHa1	VHa2	VHa3	d11	d12	e14	e15
02	1	2	-	11	12	-	15
03	1	-	-	11	-	-	15
04	1	2	-	11	12	-	15
05	-	2	-	-	12	-	15
06	1	2	-	11	12	-	15
07	-	2	-	-	12	-	15
08	1	2	-	11	12	-	15
09	1	2	-	11	12	-	15
10	-	2	-	-	12	-	15
11	-	2	-	-	12	-	15
12	1	-	-	11	-	-	15

B. *IGKC1*- and *IGKC2*-Encoded Allotypes

Rabbit	Cκ1b4	Cκ1b5	Cκ1b6	Cκ1b9	Cκ2bas1
02	-	5	-	-	-
03	-	5	-	-	-
04	-	5	-	-	-
05	-	5	-	-	-
06	-	5	-	-	-
07	-	5	-	-	-
08	-	5	-	-	-
09	-	5	-	-	-
10	-	5	-	-	-
11	4	5	-	-	-
12	4	5	-	-	-

A number shows the presence of the allotype; a dash indicates absence of the allotype.

Table 2

A sample of genes syntenic with telomeric human chromosome 14q located near the telomere of rabbit 20q

Genes in *Homo sapiens* in genomic order[a]		Human Chromosome 14	Rabbit chrUn0053 Present
	O[b]		
CKB	−	103989170-103985996	Yes
TRMT61A	+	103995509-104003410	Yes
BAG5	−	104029151-104022881	Yes
APOPT1	+	104029299-104057236	Yes
KLC1	+	104095525-104167888	Yes
XRCC3	−	104163954-104181823	Yes
ZYFVE21	+	104182081-104200005	Yes
PPP1R13B	−	104200088-104313927	Yes
C14orf2	−	104378624-104387903	Yes
TDRD9	+	104394817-104519004	Yes
RDL3	−	104406761-104408645	Yes
ASPG	+	104552023-104579046	Yes
KIF26A	+	104605060-104647235	No
23 additional human genes that are not present on rabbit chromosome 20			
TMEM121	+	105992953-105996539	No
IGHA2,E, G,D,M,J,D7-27	−	106053244-106331887	NA[c]
IGHD2-2	−	106382685-106382715	NA
IGHD1	−	106385361-106385377	NA
IGHV6-1	−	106405609-106406056	NA
IGHVII-1-1	−	106410796-106411247	NA
IGHV1 ...-... IGHV3-76	−	106452669-107236538	NA

[a] Protein-encoding genes in NCBI build 37.3. More details can be found in Supplementary Table S1.

[b] O indicates transcriptional orientation in human chromosome 14

[c] NA-not applicable. However, no rabbit *IGH* genes were present.

Table 3

Comparisons of constant region exon sequences and positions of *IGHM, IGHG* and *IGHE* in BAC 27N5 (AY386696) with those in chrUn0439

A. *IGHM* in BAC 27N5 (AY386696) constant region exon positions			**chrUn0439 NW_003159763 positions**		**Number of confirming traces (yes)**	
	3302-5095			74083-72286		
Exon	Codon position			Codon position	yes	no
1	3346-3348	Gly GGT	Gly GGC	74038-74036	8	0
	3412-3414	Val GTC	Ser ATC	73975-73973	3	3
	3511-3513	Ser TCG	STOP TCA	73873-73871	3	4
	3583-3585	Ser AGC	Asp GAC	73801-73799	1	2
2	3860-3862	Leu CTG	Leu CTA	73525-73523	2	5
3	4327-4330	Ile ATC	Ile ATT	73057-73055	1	0
	4408-4410	Ser AGC	Ser AGT	72974-72972	1	0
4	4901-4903	Pro CCA	Pro CCG	72483-72481	6	0
B. *IGHG* in BAC 27N5 (AY386696) positions						
	55582-56945			24231-22868		
Exon	Codon position			Codon position	Yes	no
Hinge	56073-56074	[A]CA[a] Thr	[A]CG Met	23742-23741	3	0
Hinge	56094-56096	ACG[b]	ATG[b]	23719-23717	1	2
2	56434-56436	Thr ACG[c]	Ala GCG[c]	23383-23381	4	0
C. *IGHE* in BAC 27N5 (AY386696) positions						
	69248-70872			9172-10797		
Exon[d]	Codon position			Codon position (−)	Yes	no
1	69356-69358	Pro CCG	Pro CCA	10689-10687	4	0
	69377-69379	Ala GCC	Ala GCT	10669-10667	4	0
4	70543-70545	Thr ACC	Ala GCC	9500-9498	6	0
	70676-70678	Ser AGT	Ser AGC	9363-9361	5	0

[a]This is a silent difference in an Ala codon formed by splicing [A] at position 56107 of exon 1 to the hinge region. The CA or CG are in the second and third codon positions.

[b]These codons define the d12 allotype of the BAC donor rabbits and d11 predicted from serum typing of Thorbecke rabbits to be heterozygous d11/d12 with both codons present.

[c]These codons define the e14 allotype of the BAC donor rabbits and e15 allotype predicted from serum typing to be encoded in Thorbecke rabbits (Table 1).

[d]The sequence of BAC 27N5 (AY386696) between 69248 and 70872 is 99% identical to the reverse strand of chrUn0439 between 9172 and 10797 with seven differences (one a gap). Two silent changes are in exon 1 (69248-69594), and two changes, one replacement and one silent are in exon 4 (70532-70872) of BAC 27N5 (AY386696). Three additional differences, one a gap, are in introns.

Table 4

IGKC1 and IGKC2 regions on rabbit chromosome 2 (NC_013670.1)

	Position on Chr 2	O	Cys 80	Gaps >200 bp in assembly
3’ enhancer	98933297-98934433	+	NA	
IGKC1	98926095-98926409	+	NA	
IGKJ5	98923099-98923136	+	NA	
IGKJ4	98922804-98922839	+	NA	
IGKJ3	98922514-98922550	+	NA	
IGKJ2	98922229-98922267	+	NA	
IGKJ1	98921851-98921886	+	NA	
[a]*IGKC1+IGKJ* 1-5	98921714-98926540	+	NA	
[b]Vκ (κ–230, κ–232)	98904850-98905144	−	yes	
[b]Vκ st (κ–166)	98889234-98889535	−	yes	98898348-98898616
>9 Vκ on reverse strand				98861572-98880208
Vκ	98807842-98808123	+	yes	98849776-98850195
>10 Vκ on reverse strand				
Vκ	98739177-98739342	+		
18 Vκ on reverse strand				
Vk	98635279-98635578	−		
	98634352-98635071	+		
[c]AJ874460 OCU-INRACCDDV0118 Microsatellite (tg)12	98634812-98634836			
Vκ	98630969-98631265	−		
6 Vκ on reverse strand				
Vκ	98607292-98607587	+		
16 Vκ on reverse strand				
Vκ	98504852-98505151	+		98519157-98521310
9 Vκ on reverse strand		−		
Vκ	98461462-98461758	−	yes	98473025-98475614
Vκ	98456577-98456876	−	yes	98460230-98460797
Vκ	98453275-98453573	−	yes	
Vκ fs	98395447-98395721	+	Pro	
Vκ	98390619-98390897	+	Pro	
Vκ	98343555-98343835	+	Pro	
Vκ	98303780-98304058	−	Pro	
Vκ	98298948-98299235	−	Pro	
Vκ	98254031-98254315	+	Ala	
Vκ	98246562-98246849	+	Ala	
Vκ	98239440-98239724	+	yes	
IGKJ3	98226217-98226255	−	NA	98235051-98235308

	Position on Chr 2	O	Cys 80	Gaps >200 bp in assembly
IGKJ2-4	98225838-98225876	−	NA	
IGKJ1	98225536-98225572	−	NA	
IGKC2	98223558-98223737	−	NA	
d_3, enhancer	98217226-98216069	−	NA	
^{e}Vκ (mp054)	98176612-98176896	+	Ala	98200751-98201488
^{e}Vκ (mp057)	98171978-98172262	+	Ala	
^{f}Vκ (anti-A33)	98168180-98168470	+	Ala	
^{e}Vκ (mpl09)	98156268-98156561	+	Ala	
Vκ fs	98149319-98149599	+	Pro	
Vκ ps (409 bp insertion)	98144730-98144777	+	Ala	
	98144100-98144319	+		
^{e}Vκ (mp159)	98137161-98137448	+	yes	
^{b}Vκ (4-110 VκI–127)	98129843-98130127	+	Ala	98133687-98134398

Abbreviations: O transcriptional orientation; NA not applicable; st stop codon; fs frameshift; ps pseudogene

^{a}Emorine and Max 1983a; 1983b; Esworthy and Max, 1986

^{b}Similar to sequences in (Sehgal et al.1999)

^{c}AJ874460 was previously unplaced (Chantry-Darmon et al. 2006)

^{d}Hole et al.1991

^{e}similar to sequences in (Popkov et al. 2003)

^{f}similar to sequences in (Rader et al. 2000)

Table 5

Locations of lambda light chain genes on OryCun2.0 chromosome 21 (NC_013689.1) and possible locations on unplaced scaffolds

IMGT name[a]	Accession Number (positions)	Position if present on chromosome 21[a]	Unplaced scaffolds
IGLC			
IGLC1*01	M12388.1[b] (2-318)	Not found	chrUn0128 (NW_003159452.1) 66073-66385 (99% identity 0 gaps)
IGLC2	M12761.1[b] (2-318)	Not found	chrUn0253 (NW_003159577.1) 369309-369625 (99% identity 0 gaps)
IGLC3	M12762.1 (2-306)	5820953-5821254 97% identity 3 gaps	chrUn0895 (NW_003160219.1) 30223-30526 (98% identity 1 gap)
IGLC4*02	M12763.1 (2-318)	4880210-488052 99% identity 0 gaps	
IGLC5[c]	M23227.1 (7-323)	4870032-4870348 100% identity	
IGLC6[c]	M23229.1 (7-323)	4875953-4876269 100% identity	
IGLJ			
IGLJ1*01		4862766-4862803	
IGLJ3*01		4978942-4878979	
IGLJ5*01[c]	M23226.1 (41-78)	4868770-4868807	
IGLJ6*01[c]	M23228.1 (41-78)	4874683-4874720	
IGLV			
43 *IGLV*[d] present		4540184-4837388	

[a] New information about *IGL* genes of rabbit was added to IMGT during 2012. See: http://imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=rabbit&group=IGLC http://imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=rabbit&group=IGLJ http://imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=rabbit&group=IGLV IMGT lists position numbers for the contig NW_003159316.1. These are 3 Mb lower than position numbers on chromosome 21.

[b] Lower quality alignments to both of M12388.1 and M127621.1 *IGLC1* and *IGLC2* respectively, are also found on chrUn0351 (NW_003159675.1), chrUn0545 (NW_003159869.1) and chrUn0895 (NW_003160219).

[c] Rabbit lambda constant region genes, *IGLC5* paired with *IGLJ5*01* and *IGLC6* paired with *IGLJ6*01* are known to be functional (Hayzer et al. 1990; Jaton et al. 1990).

[d] IMGT lists 21 *IGLV* as pseudogenes, two as potentially functional open reading frames, and 20 as functional.