Immunogenetics (2010) 62:479–486
DOI 10.1007/s00251-010-0448-x

ORIGINAL PAPER

Organization of the variable region of the immunoglobulin heavy-chain gene locus of the rat

Jacobus Hendricks · Peter Terpstra · Peter M. Dammers · Rajesh Somasundaram · Annie Visser · Maaike Stoel · Nicolaas A. Bos · Frans G. M. Kroese

Received: 9 February 2010 / Accepted: 6 April 2010 / Published online: 5 May 2010

Abstract We have mapped and annotated the variable region of the immunoglobulin heavy (IGH) gene locus of the Brown Norway (BN) rat (assembly V3.4; Rat Genomic Sequence Consortium). In addition to known variable region genes, we found 12 novel previously unidentified functional *IGHV* genes and 1 novel functional *IGHD* gene. In total, the variable region of the rat IGH locus is composed of at least 353 unique *IGHV* genes, 21 *IGHD* genes, and 5 *IGHJ* genes, of which 131, 14, and 4 are potentially functional genes, respectively. Of all species studied so far, the rat seems to have the highest number of functional *IGHV* genes in the genome. Rat *IGHV* genes can be classified into 13 IGHV families based on nucleotide sequence identity. The variable region of the BN rat spans a total length of approximately 4.9 Mb and is organized in a typical translocon organization. Like the mouse, members of the various IGHV gene families are more or less grouped together on the genome, albeit some members of IGHV gene families are found intermingled with each other. In the rat, the largest IGHV gene families are IGHV1, IGHV2, and IGHV5. The overall conclusion is that the genomic organization of the variable region of the rat IGH locus is strikingly similar to that of the mouse, illustrating the close evolutionary relationship between these two species.

Keywords Immunoglobulin locus · Rat · Immunoglobulin heavy-chain genes

Electronic supplementary material The online version of this article (doi:10.1007/s00251-010-0448-x) contains supplementary material, which is available to authorized users.

J. Hendricks · P. M. Dammers · R. Somasundaram · A. Visser · M. Stoel · N. A. Bos · F. G. M. Kroese (✉)
Immunology Section, Department of Cell Biology,
University Medical Center Groningen, University of Groningen,
A. Deusinglaan 1,
9713EZ Groningen, The Netherlands
e-mail: f.g.m.kroese@med.umcg.nl

P. Terpstra
Unit of Genetic Epidemiology and Bioinformatics,
Department of Epidemiology, University Medical Center Groningen,
University of Groningen,
A. Deusinglaan 1,
9713EZ Groningen, The Netherlands

P. M. Dammers
Institute for Life Science and Technology,
Hanze University Groningen,
Zernikeplein 11,
9747AS Groningen, The Netherlands

Introduction

The antigen recognition site of an antibody is the product of a pairing set of immunoglobulin heavy (IGH) and immunoglobulin light (IGL) chain variable domains. Both the IGH-chain variable domain and the IGL-chain variable domain are composed of conserved framework sequences that alternate with three hypervariable regions, the complementary-determining regions (CDRs), which are responsible for actual antigen recognition. The variable domains of the IGH chain are encoded by three different genes (or gene segments): variable (*IGHV*), diversity (*IGHD*), and joining (*IGHJ*) genes. The IGH gene locus of most mammalian species contains a large number of *IGHV* genes, fewer *IGHD* genes, and some *IGHJ* genes (Marchalonis et al. 1998). In many species such as humans, mice, and rats, these gene segments recombine by DNA rearrangements during B-cell genesis in the bone marrow. These recombinations result in a so-called combinatorial diversity of the variable domain of the heavy chain (Yancopoulos and Alt 1986). Similarly, recombination of

IGLV and IGLJ gene segments leads to the combinatorial diversity of IGL-chains. Imprecision of the rearrangements by addition or removal of nucleotides between the segments during the recombination process results in further enlargement of the primary repertoire of the variable domain of both IGH and IGL chains (junctional diversity). Another form of combinatorial diversity is created by the combination of IGH and IGL chains that are required to form the actual antigen recognition site.

All mammals use combinatorial diversity (and junctional diversity) of *IGHV*, *IGHD*, and *IGHJ* genes to form a diverse primary preimmune H-chain repertoire (Marchalonis et al. 1998). The extent of this combinatorial diversity varies, however, significantly between different species (Flajnik 2002; Marchalonis et al. 1998). Mammalian cross-species comparisons have demonstrated considerable divergence in the number and/or expression of *IGHV*, *IGHD*, and *IGHJ* genes (Das et al. 2008; Flajnik 2002; Marchalonis et al. 1998). For example, the number of potentially functional germline *IGHV* genes may vary from only 20 in pigs (Butler et al. 2006) to >100 in rats and mice (Das et al. 2008; Johnston et al. 2006). Some mammalian species such as rabbit, sheep, and cow use only a very limited number of possible *IGHV*, *IGHD*, and *IGHJ* genes (Dufour et al. 1996; Gontier et al. 2005; Mage et al. 2006; Saini et al. 1997). In chickens, even only one unique functional *IGHV* gene is present in the heavy-chain locus (Reynaud et al. 1995). Chickens and mammals that use only very few *IGHV* genes must therefore rely on additional mechanisms to compensate for the presence of a relatively limited combinatorial preimmune repertoire of their IGH chains. The strategies used to form a diverse primary IGH chain repertoire in these species include gene conversion in chickens and rabbits (Mage et al. 2006; Reynaud et al. 1995); hypermutation in chickens, sheep, and rabbits (Dufour et al. 1996; Gontier et al. 2005; Kothapalli et al. 2008; Mage et al. 2006; Reynaud et al. 1995); and extra-long H-CDR3 regions in cows (Saini et al. 1999).

For a better understanding of the generation of the primary antibody repertoire during B-cell development and the changes (somatic hypermutations) that occur in this repertoire during humoral immune responses, detailed knowledge of the germline *IGHV*, *IGHD*, and *IGHJ* genes and organization of the IGHVDJ locus are of critical importance. This information is also essential in giving insight into how various species have evolved different mechanisms to create a diverse preimmune specificity repertoire of their antibodies. With the unraveling of the mouse and human genomes, a detailed complete physical annotated map of the IGHVDJ locus has become available for these species (Johnston et al. 2006; Matsuda et al. 1998). One of the remarkable findings of these studies was that the number of functional *IGHV* genes in both species appeared to be much lower than previously estimated, whereas the number of nonfunctional *IGHV* genes [pseudogenes and open reading frame (ORF) genes lacking appropriate signal sequences] was relatively high. The genome of the rat has been unraveled, and there is a nearly complete sequence of the IGH locus (Gibbs et al. 2004). In rats, the exact number and location of *IGHV* genes are not known. Preliminary data suggest, however, that they are among the species with the highest number of (functional) *IGHV* genes in the genome (Das et al. 2008). Our previous studies (Dammers et al. 2000a) have indicated that, similar to the mouse, rat *IGHV* genes can be subdivided into IGHV gene families, on the basis of nucleotide sequence identity. *IGHV* genes belong to the same family when the *IGHV* genes share more than 80% of their nucleotides (Brodeur and Riblet 1984). We have detected previously the existence of at least 28 functional *IGHV* (germline) genes that belong to the IGHV5 family (PC7183) in the PVG rat strain (Dammers and Kroese 2001; Stoel et al. 2008). Here we present an annotated map of the variable region of the IGH locus of the Brown Norway (BN) rat, including not only functional and nonfunctional *IGHV* genes but also *IGHD* and *IGHJ* genes.

Materials and methods

Genomic sequence of the rat IGH locus

The genomic sequence of the BN/SsNHsdMCW rat (*Rattus norvegicus*) was generated by the Rat Genomic Sequence Consortium (RGSC) (Gibbs et al. 2004; Havlak et al. 2004). This sequence is available through the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov). Analysis of the variable region of the rat IGH gene locus, located on chromosome 6q32–33, was based on the Human Genome Sequencing Center assembly version RGSC V3.4 (November 2004 release; Baylor College of Medicine, Houston, TX, USA). Assembly RGSC V3.4 has been established in a hybrid approach combining the clone-by-clone method and the whole-genome shotgun method.

Mapping of the variable region of the BN rat

IGHV and IGHD gene sequences of the BN rat were obtained from the International Immunogenetics (IMGT) database (http://imgt.cines.fr) (Lefranc et al. 1999). *IGHJ* genes were taken from Lang and Mocikat (1991) (accession number X56791). Additional previously unreported *IGHV* genes were searched for in mapped and unmapped sequences of the rat genome. Unmapped sequences were taken from contigs in the “unplaced section” of the NCBI

database or from newly established bactigs of the BN rat genome (Baylor College of Medicine; http://www.hgsc.bcm.tmc.edu/projects/rat) not yet present in assembly RGSC V3.4. Nucleotide alignments were carried out using the NCBI BLASTN program (Altschul et al. 1990). New *IGHV* gene sequences were manually analyzed for the presence of an ORF of the coding region and for the presence of functional recombination signal sequences (RSS) and leader sequences using V-QUEST alignment software (http://www.imgt.cines.fr) (Giudicelli et al. 2004). Previously unreported *IGHD* genes were identified by searching manually in the genomic assembly RGSC V3.4 with sets of rat nonamer and heptamer sequences. The relative positions of the *IGHV*, *IGHD*, and *IGHJ* genes on the chromosomal map were determined by aligning the encoding parts of the *IGHV*, *IGHD*, and *IGHJ* genes (functional and nonfunctional) against the BN rat genome using the Genome Browser and BLAT programs (http://www.genome.ucsc.edu) (Karolchik et al. 2008; Kent 2002). Because of the relatively small size of *IGHD* genes, we included the IGHD 5′ and 3′ flanking RSS regions in the alignment. *IGHD* genes and flanking RSS were obtained from the IMGT database (accession numbers AABR03049813, AABR03051895, and M13798). The location of other predicted genes in the IGH locus was taken from NCBI annotation. The complete physical map and annotation of the genomic IGHVDJ region were drawn using the software package Genvision (Dnastar, Madison, WI, USA).

Nomenclature of rat IGHV, IGHD, and IGHJ genes

IGHV, *IGHD*, and *IGHJ* gene nomenclature and classification (functional, nonfunctional, ORF gene, and pseudogene) were adopted from the IMGT (Lefranc et al. 1999). Briefly, nonfunctional *IGHV*, *IGHD*, and *IGHJ* genes are either genes with an intact ORF but erroneous regulatory sequences (“ORF genes”) or genes lacking a correct ORF (pseudogenes).

Results and discussion

Organization of the variable region of the IGH locus of the BN rat

Recently, DNA sequencing resulted in elucidation of the vast majority of the genomic nucleotide sequence of the BN rat, including the IGH locus located on chromosome 6q32–33 (Gibbs et al. 2004). The current IMGT database (Lefranc et al. 1999) contains 342 *IGHV* genes (120 functional and 222 nonfunctional), 20 *IGHD* genes (13 functional and 7 nonfunctional), and 5 *IGHJ* genes (4 functional and 1 nonfunctional). In order to establish a detailed chromosomal map of this part of the IGH locus containing the exact chromosomal location and orientation of the individual genes, we aligned the coding sequences of all known *IGHV*, *IGHD*, and *IGHJ* genes (functional and nonfunctional) from the IMGT database to the rat genome assembly RGSC V3.4. As depicted in Fig. 1, the variable region of the IGH locus of the BN rat spans a total length of approximately 4.9 Mb and is organized in a typical translocon organization (many *IGHV* genes, a dozen *IGHD* genes, and a few *IGHJ* genes) similar to mice and humans. The locus has a telomeric to centromeric orientation and runs from the distally located *IGHV7S16* gene towards the proximally located *IGHJ4* gene. The upstream boundary of the IGHV region is marked by the non-IGH *zinc-finger-protein type 386* gene, similar to the situation in mice (Johnston et al. 2006).

The *IGHV* genes of the BN rat can be classified into 13 IGHV families based on nucleotide sequence identity (IGHV1, IGHV2, IGHV3, IGHV4, IGHV5, IGHV6, IGHV7, IGHV8, IGHV9, IGHV10, IGHV11, IGHV12, and IGHV15) (Table 1). In comparison to rats, the *IGHV* genes in humans and mice can be grouped together into 7 and 16 IGHV families, respectively. Most BN rat IGHV families are composed of various members, except for the IGHV15 family, which is composed of only one gene. In rats and mice, the various members (both functional and nonfunctional genes) of these IGHV families are more or less clustered together on the genome. The order of various IGHV families on the genome appears to be well preserved between rats and mice. Similar to mice (Johnston et al. 2006), the *IGHV* genes that belong to the IGHV1 and IGHV8 families are the most telomeric *IGHV* genes, whereas the members of the IGHV2 and IGHV5 families are located centromeric and closest to the *IGHD* genes. IGHV family members of the rat (and also of the mouse) are not completely spatially separated, and members of various IGHV families are frequently found intermingled with each other (e.g., *IGHV2/IGHV5* and *IGHV1/IGHV8* genes). The members of the IGHV1 family are more widely distributed over the locus and are mixed with members of various other IGHV gene families (such as IGHV7, IGHV8, IGHV11, etc.). In comparison to rat and mouse, the human IGHV family gene members are more extensively interspersed and less clustered on the IGH locus (Matsuda et al. 1998). The almost identical distribution pattern of IGHV gene families between mouse and rat strongly suggests a close evolutionary relationship shared between these species.

The vast majority (>90%) of rat *IGHV* genes are orientated in the direction of the IGHD cluster, whereas a small number of *IGHV* genes have an inverted orientation (in the direction of the telomere). These inverted *IGHV* genes are grouped together on four inverted regions of the chromosomal map. Three of these regions are inverted repeats (positions 3.74, 3.79, and 4.76 Mb on the map of Fig. 1) containing eight pairs of 100% identical *IGHV*

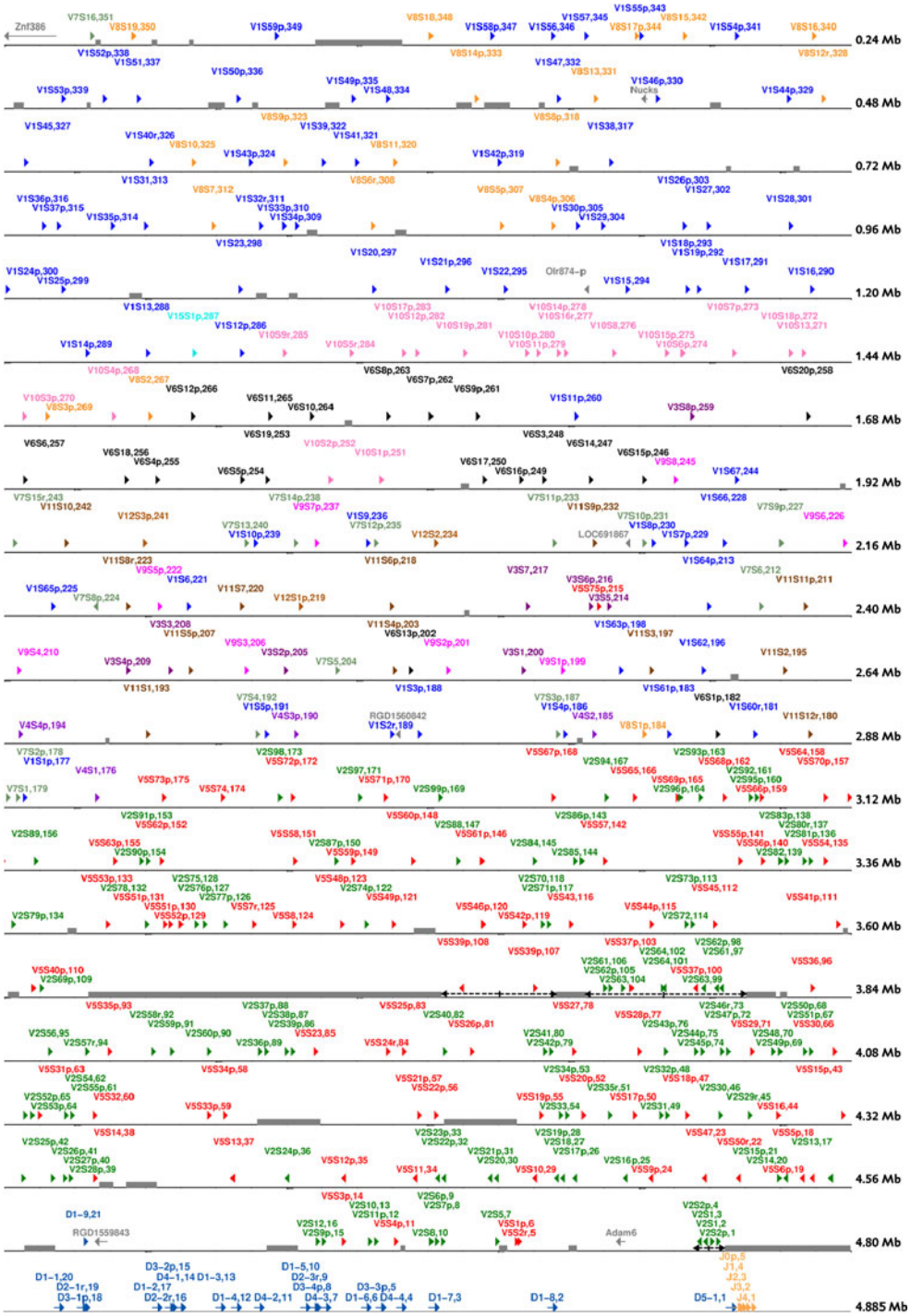

Fig. 1 Chromosomal map of the variable region of the IGH locus of the BN rat. Shown is the IGHV region from chromosome 6 ranging from gene *Znf386* to IGHJ4 (RGSC V3.4: 138,451,833–143,326,393 bp). The map was established on the basis of the IMGT database (http://imgt.cines.fr), as indicated in "Materials and Methods," and does not contain the 12 newly identified *IGHV* genes. Genes and their orientation are indicated by an *arrow point*. The *IGHV*, *IGHD*, and *IGHJ* genes are numbered according to the IMGT nomenclature (Lefranc et al. 1999). Members of the same IGHV family share identical colors (non-*IGHV* genes are shown in *gray*). Nonfunctional genes are indicated with a "p" for pseudogene or with an "r" for ORF gene after the family number. The last number in the gene name is the rank number of the gene in the locus, starting at the centromeric end. *Black dotted lines* indicate near-perfect inverted repeats. Gaps in assembly V3.4 are marked by *solid gray bars*. This map is also available as an MS Excel file (Online Resource 1) and as a "bed"-type file (Online Resource 2) that can be projected on the current rat genome version V3.4 at the UCSC genome website (www.genome.ucsc.edu). These files can be found in "Supplementary Material"

genes. The majority of the inverted *IGHV* genes are, however, located on a large inverted region of the IGHV locus (between the genes *IGHV2S28p,39* and *IGHV2S12,16*) close to the centromeric end of the IGHV region of the locus. This area of the locus also contains a gene, *IGHV5S47*, which is identical to the noninverted gene *IGHV5S8*. These two genes are approximately 1.1 Mb separated from each other. We further noted the presence of another pair of identical *IGHV* genes (*IGHV5S51p*) located only 1,183 bp apart (Fig. 1; 3.41 Mb). In contrast to the other repeated pairs of genes, the latter pair of *IGHV* genes has the same (normal) orientation on the chromosome. If the rat VDJ locus indeed contains inverted regions, then this would imply that the VDJ recombination mechanism for these genes would use inversion instead of deletion. The presence of inverted repeats and completely 100% identical *IGHV* genes has not

Table 1 Numbers of *IGHV* gene family members in the IGHV locus of the BN rat

IGHV gene family		Number of *IGHV* genes		
IMGT subgroups[a]	Previous mouse nomenclature[b]	Potentially functional genes[c]	Nonfunctional genes[c]	Total number of *IGHV* genes
IGHV1	J558	25	42	67
IGHV2	Q52	39 (+ 4[d])	60 (+ 2)	99 (+ 6)
IGHV3	36–60	4	4	8
IGHV4	X-24	2	2	4
IGHV5	7183	26 (+ 1)	53 (+ 3)	79 (+ 4)
IGHV6	J606	8	11	19
IGHV7	S107(T15)	6	9	15
IGHV8	3609	9	10	19
IGHV9	VGAM3-8	4	4	8
IGHV10	VH10	2	17	19
IGHV11	CP3	5	7	12
IGHV12	CH27	1	2	3
IGHV15	VH15A	0	1	1
Total		131 (+ 5)	222 (+ 5)	353 (+ 10)

The table also contains the 12 newly identified *IGHV* genes (see the text for further explanation).

[a] IGHV nomenclature according to IMGT.

[b] IGHV family nomenclature according to Brodeur and Riblet (1984).

[c] Functional and nonfunctional (pseudogenes or ORF genes), according to IMGT standards.

[d] The number of 100% identical genes is presented in parentheses.

been identified in humans (Matsuda et al. 1998) and mice (Johnston et al. 2006). It might well be that the presence of inverted repeats in rats in the IGH locus may reflect inconsistencies in the current genome assembly (Worley et al. 2008). Our map must be therefore be taken as tentative, and it will be interesting to see whether an "upgraded" version of the rat genome sequences confirms the current assembly in regions with inverted repeats.

In addition to *IGHV* genes, there are also six non-*IGHV* genes mapped on the IGHV gene locus (Fig. 1). A metallopeptidase domain 6 gene (*Adam6*) is found at the proximal end of this locus. Of the remaining five non-*IGHV* genes, two have reported annotations: nuclear-casein-dependent kinase substrate 1 (*Nucks1*) and the olfactory receptor *pseudogene 874*. These genes are located between IGHV1 and IGHV8 family members at the distal end of the IGHV gene locus. The other three genes are NCBI-predicted genes: homolog of the Brix domain gene *BXDC1* (RGD1560842) and two prematurely terminated fragments of potential rat homologs (RGD1559843 and LOC691867). It is unknown whether these non-*IGHV* genes are functionally expressed. If these genes are functional, their expression might well be influenced by immunoglobulin enhancers as a consequence of the VDJ recombination process, or they may even be lost during this process. For these reasons, we assume that these non-*IGHV* genes in this locus are nonfunctional genes.

Identification of novel IGHV genes

The current assembly (RGSC V3.4) has a number of gaps in the IGHV region of the IGH locus (~300 kb) of which the nucleotide sequence still has to be determined (Twigger et al. 2008). Approximately 7% of the variable region of the IGH locus has not been mapped yet. These regions are indicated in Fig. 1 as gap regions. The largest gap is found in the IGHV2–IGHV5 region (Fig. 1; between 3.6 and 3.7 Mb). These gaps may potentially contain novel *IGHV* genes. There are available genomic sequences of the BN rat that have not yet been incorporated into the current assembly (RGSC V3.4). These sequences are present as contigs and are grouped together as "unplaced sequences" in the NCBI database. To explore the presence of unidentified *IGHV* genes, we used BLASTN to align all IMGT-listed rat *IGHV* genes to the unplaced contigs NW_047922.1 and NW_047772.1. Contig NW_047922.1 contains genomic sequences that are not yet assigned to any specific chromosome, whereas contig NW_047772.1 contains genomic sequences that are specific for chromosome 6. This search resulted in the identification of 18 *IGHV*

genes (17 genes in contig NW_ 047922.1 and 1 gene in contig NW_047772.1). All these 18 *IGHV* genes share 100% identity to an *IGHV* gene already present in the IMGT database. All other potential IGHV homologs (i.e., sequences with ≥80% identity with a known IGHV sequence in the IMGT database) found in unplaced sequences did not comply with the IMGT criteria for functional *IGHV* genes (Lefranc et al. 1999). Thus, the unplaced genomic sequences did not reveal any new previously unidentified rat *IGHV* genes.

In addition to the unplaced sequences mentioned in the previous paragraph, there are also available bactig sequences (Baylor College of Medicine) that are not included in assembly RGSC V3.4. These bactigs are composed of overlapping bacterial artificial chromosome (BAC) sequences that may contain sequences that could map on the gap regions of this assembly. We analyzed two nonoverlapping bactigs that span the entire variable region of the IGH locus, including the gap regions in assembly V3.4 (Fig. 1), for the presence of additional unreported *IGHV* genes. One of these bactigs (gpwy_grzy) (7.5 Mb) includes 71 BACs and extends into the major gap region. The other bactig (kdyb_kdzq; 1.65 Mb) is located centromeric to the largest gap region and consists of 18 overlapping BACs and probably also extends into the major gap region. Bactig gpwy_grzy did not reveal any previously unknown *IGHV* genes, although this bactig partially overlaps the major gap region. Because the nucleotide sequences of these BACs are currently not complete and also lack sufficient accuracy, this does not, however, imply that the major gap region does not contain any *IGHV* gene. On the other hand, 11 novel IGHV sequences that meet the IMGT criteria for functional germline *IGHV* genes (preliminary third-party annotation accession numbers BN001223–BN001233) were found in bactig kdyb_kdzq. These criteria include the appropriate sequence length, at least one ORF, a proper leader sequence, and a functional RSS (Lefranc et al. 1999). Furthermore, the flanking intron regions of these genes were not identical to the flanking regions of previously established *IGHV* genes already listed in the IMGT database. Based on sequence identity, five of these novel *IGHV* genes belong to the IGHV5 family (designated *IGHV5-1* to *IGHV5-5*), and six belong to the IGHV2 family (designated *IGHV2-1* to *IGHV2-6*). The finding that these novel *IGHV* genes all belong to either the IGHV2 family or the IGHV5 family is consistent with the notion that most gap regions are found in the area of the IGHV locus where members of these two IGHV families are located. To reveal whether these 11 newly identified *IGHV* genes are also functionally expressed in rearranged IGHVDJ transcripts, we aligned the newly identified IGHV sequences to the *R. norvegicus* nucleotide collection database of the NCBI. Two of these *IGHV* genes (*IGHV2-3* and *IGHV5-1*) share a 100% identity with rearranged IGHVDJ BN rat complementary DNA sequences (accession numbers L07402 and X78897, respectively). In a recent study, we looked at the expression of *IGHV5* genes in rat B-cell subsets (Hendricks et al., manuscript in preparation). We found the expression of 100% identical *IGHV5-1* and *IGHV5-2* genes in mature B cells. In addition, we detected another previously unidentified *IGHV5* gene (named *IGHV5-6*). This *IGHV* gene is also 100% identical to the *IGHV* gene expressed in BN hybridoma Hg16 (Dammers et al. 2001) (accession number Z75899). These findings indicate that at least some of these novel germline genes are also functionally used.

Identification of an additional IGHD gene

So far, 20 (13 functional and 7 nonfunctional) *IGHD* genes have been described by the IMGT. We manually searched genomic assembly RGSC V3.4 for the presence of additional *IGHD* genes by using available RSS from functional rat *IGHD* genes. With this approach, we found a previously unidentified member of the IGHD1 subgroup (Fig. 1). Remarkedly, this gene, named *IGHD1-9* (third-party annotation accession number pending), is located among *IGHV2/IGHV5* genes ~200 kb upstream of the IGHD gene cluster (Fig. 1). This *IGHD* gene contains an ORF and has functional RSS (12-bp spacer) flanking the gene on both sides (chromosomal coordinates can be found in supplementary files). Most probably, *IGHD1-9* is also functionally expressed, since it is used in a rearranged IGHVDJ sequence (accession number AJ286179), albeit this IGHVDJ sequence is derived from another rat strain (PVG). The total number of functional *IGHD* genes in the BN rat is therefore most likely 14.

Concluding remarks

Together, our data imply that the total number of unique *IGHV* genes in the BN rat is at least 353 (see Table 1), including the 12 newly identified *IGHV* genes. In this estimate, the pairs of identical genes (ten pairs in total) are counted as one. Of these 353 *IGHV* genes, 131 (37%) meet all criteria for functional germline *IGHV* genes and can therefore be potentially expressed. The remaining genes are nonfunctional because they either do not have at least one ORF (pseudogenes) or lack an appropriate RSS (ORF genes). Nearly all nonfunctional *IGHV* genes are pseudogenes. It should be noted here that there may be more nonfunctional *IGHV* genes because the bactigs (see the previous discussion) were only analyzed for the presence of functional *IGHV* genes. Also in humans, the number of nonfunctional *IGHV* genes exceeds the number of func-

tional *IGHV* genes (approximately one third) (Matsuda et al. 1998); however, in the mouse, there seems to be a higher number of functional (55%) *IGHV* genes than nonfunctional *IGHV* genes (Johnston et al. 2006). In general, however, there is a positive correlation between the number of functional *IGHV* genes and the number of nonfunctional *IGHV* genes (Das et al. 2008). The presence of large numbers of nonfunctional genes reflects the diversification of *IGHV* genes in evolution, as proposed before (Ota and Nei 1994). This process involves gene duplications and functional elimination after deleterious mutations (pseudogenes and ORF genes).

The rat genome harbors the highest number of functional *IGHV* genes (at least 131) of all mammalian species studied so far (Table 1). The (antigen-independent) recombination of this large number of *IGHV* genes with one of the 14 *IGHD* genes and with one of the 4 *IGHJ* genes accounts for a more diverse combinatorial IGH repertoire in rats compared to other species. In addition, this repertoire is enlarged by junctional diversity, as illustrated by the presence of TdT in rat B-cell precursor cells (Opstelten et al. 1986) and variable numbers of N insertions in sequenced rat *IGHV* genes (Dammers et al. 2000b).

Our main conclusion is that the overall organization of the variable region of the IGH locus and the distribution of IGHV family members in the rat are strikingly similar to the corresponding region in the mouse (Johnston et al. 2006), despite the fact that these two species diverged from each other 41 million years ago (Kumar and Hedges 1998). Also in the mouse, the IGHV1 (J558), IGHV2 (Q52), and IGHV5 (PC7183) gene families have the highest number of IGHV members. Both in rat and in mice, members of these families represent approximately two thirds of all functional *IGHV* genes. These genes therefore contribute the most to the available germline repertoire. In the mouse, the IGHV1 family (J588) is, by far, the largest IGHV family (almost half of all functional *IGHV* genes), whereas in the rat, the largest IGHV family is the IGHV2 (Q52) family, with 40 unique members (31% of all functional *IGHV* genes).

Conflict of interest The authors do not have conflicts of interests.

References

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

Brodeur PH, Riblet R (1984) The immunoglobulin heavy chain variable region (IgH-V) locus in the mouse: I. One hundred IgH-V genes comprise seven families of homologous genes. Eur J Immunol 14:922–930

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

Dammers PM, Kroese FG (2001) Evolutionary relationship between rat and mouse immunoglobulin IGHV5 subgroup genes (PC7183) and human IGHV3 subgroup genes. Immunogenetics 53:511–517

Dammers PM, Visser A, Popa ER, Nieuwenhuis P, Bos NA, Kroese FG (2000a) Immunoglobulin V_H gene analysis in rat: most marginal zone B cells express germline encoded V_H genes and are ligand selected. Curr Top Microbiol Immunol 252:107–117

Dammers PM, Visser A, Popa ER, Nieuwenhuis P, Kroese FG (2000b) Most marginal zone B cells in rat express germline encoded Ig V_H genes and are ligand selected. J Immunol 165:6156–6169

Dammers PM, Bun JC, Bellon B, Kroese FG, Aten J, Bos NA (2001) Immunoglobulin V_H-gene usage of autoantibodies in mercuric chloride-induced membranous glomerulopathy in the rat. Immunology 103:199–209

Das S, Nozawa M, Klein J, Nei M (2008) Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. Immunogenetics 60:47–55

Dufour V, Malinge S, Nau F (1996) The sheep Ig variable region repertoire consists of a single VH family. J Immunol 156:2163–2170

Flajnik MF (2002) Comparative analyses of immunoglobulin genes: surprises and portents. Nat Rev Immunol 2:688–698

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R et al (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521

Giudicelli V, Chaume D, Lefranc MP (2004) IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V–J and V–D–J rearrangement analysis. Nucleic Acids Res 32:W435–W440

Gontier E, Ayrault O, Godet I, Nau F, Ladeveze V (2005) Developmental progression of immunoglobulin heavy chain diversity in sheep. Vet Immunol Immunopathol 103:31–51

Havlak P, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Weinstock GM, Gibbs RA (2004) The Atlas genome assembly system. Genome Res 14:721–732

Johnston CM, Wood AL, Bolland DJ, Corcoran AE (2006) Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. J Immunol 176:4221–4234

Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Harte RA, Hinrichs AS, Hsu F, Kober KM, Miller W, Pedersen JS, Pohl A, Raney BJ, Rhead B, Rosenbloom KR, Smith KE, Stanke M, Thakkapallayil A et al (2008) The UCSC genome browser database: 2008 update. Nucleic Acids Res 36:D773–D779

Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12:656–664

Kothapalli N, Norton DD, Fugmann SD (2008) Cutting edge: a *cis*-acting DNA element targets AID-mediated sequence diversification to the chicken Ig light chain gene locus. J Immunol 180:2019–2023

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920

Lang P, Mocikat R (1991) Immunoglobulin heavy-chain joining genes in the rat: comparison with mouse and human. Gene 102:261–264

Lefranc MP, Giudicelli V, Ginestoux C, Bodmer J, Muller W, Bontrop R, Lemaitre M, Malik A, Barbie V, Chaume D (1999) IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 27:209–212

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

Marchalonis JJ, Schluter SF, Bernstein RM, Shen S, Edmundson AB (1998) Phylogenetic emergence and molecular evolution of the immunoglobulin family. Adv Immunol 70:417–506

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

Opstelten D, Deenen GJ, Rozing J, Hunt SV (1986) B lymphocyte-associated antigens on terminal deoxynucleotidyl transferase-positive cells and pre-B cells in bone marrow of the rat. J Immunol 137:76–84

Ota T, Nei M (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V_H gene family. Mol Biol Evol 11:469–482

Reynaud CA, Garcia C, Hein WR, Weill JC (1995) Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. Cell 80:115–125

Saini SS, Hein WR, Kaushik A (1997) A single predominantly expressed polymorphic immunoglobulin V_H gene family, related to mammalian group, I, clan, II, is identified in cattle. Mol Immunol 34:641–651

Saini SS, Allore B, Jacobs RM, Kaushik A (1999) Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. Eur J Immunol 29:2420–2426

Stoel M, Evenhuis WN, Kroese FG, Bos NA (2008) Rat salivary gland reveals a more restricted IgA repertoire than ileum. Mol Immunol 45:719–727

Twigger SN, Pruitt KD, Fernandez-Suarez XM, Karolchik D, Worley KC, Maglott DR, Brown G, Weinstock G, Gibbs RA, Kent J, Birney E, Jacob HJ (2008) What everybody should know about the rat genome and its online resources. Nat Genet 40:523–527

Worley KC, Weinstock GM, Gibbs RA (2008) Rats in the genomic era. Physiol Genomics 32:273–282

Yancopoulos GD, Alt FW (1986) Regulation of the assembly and expression of variable-region genes. Annu Rev Immunol 4:339–368