Tandem repeats modify the structure of the canine *CD1D* gene

F. A. Looringh van Beeck*, P. A. J. Leegwater†, T. Herrmann‡, F. Broere*, V. P. M. G. Rutten*§, T. Willemse*† and I. Van Rhijn*

*Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584 CL, Utrecht, The Netherlands. †Department of Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 108, 3584 CM, Utrecht, The Netherlands. ‡Institute for Virology and Immunobiology, Julius Maximilians University Würzburg, Versbacher Straβe 7, 97078, Würzburg, Germany. §Department of Tropical Veterinary Diseases, Faculty of Veterinary Science, University of Pretoria, Onderstepoort, Private Bag X04, Pretoria, 0110, South Africa.

Summary

Among the CD1 proteins that present lipid antigens to T cells, CD1d is the only one that stimulates a population of T cells with an invariant T-cell receptor known as NKT cells. Sequencing of a 722 nucleotide gap in the dog (*Canis lupus familiaris*) genome revealed that the canine *CD1D* gene lacks a sequence homologous to exon 2 of human *CD1D*, coding for the start codon and signal peptide. Also, the canine *CD1D* gene contains three different short tandem repeats that disrupt the expected gene structure. Because canine *CD1D* cDNA lacks sequences homologous to human exon 2 and 3, the functionality of canine CD1d protein may be affected, and this could have consequences for the development and activation of canine NKT cells.

Keywords CD1, dog, microsatellite, NKT cells, simple sequence repeat

The CD1 family is a group of non-polymorphic glycoproteins that present lipid antigens to T cells (Porcelli & Modlin 1999; Brigl & Brenner 2004; Barral & Brenner 2007). Like MHC class I proteins, CD1 proteins are heterodimers of $\beta 2$ microglobulin and a heavy chain consisting of three extracellular α domains, a transmembrane region and a cytoplasmic tail, each encoded by a separate exon, preceded by an exon that encodes the start codon and signal peptide, and sometimes an additional exon that consists of 5′ UTR only.

Large variation exists in the number of *CD1A*, *CD1B*, and *CD1C* genes between mammalian species, whereas one or two *CD1D* genes are present in all mammalian species studied to date (Dascher *et al.* 1999; Hayes & Knight 2001; Eguchi-Ogawa *et al.* 2007; Looringh van Beeck *et al.* 2008, 2009). CD1d is crucial for the selection (Bendelac *et al.* 1995) and activation (Kawano *et al.* 1997) of a subset of T cells known as natural killer T (NKT) cells. The canine CD1 locus is located on chromosome 38 and contains one *CD1D* gene, eight *CD1A* genes, one *CD1C*, one *CD1B* and one *CD1E* gene (Fig. 1a) (Looringh van Beeck *et al.* 2008). In the canine genome (CanFam 3.1) and in the sequence of BAC clone XX-14K12 AC183576.27, the canine *CD1D* gene is incomplete due to an internal gap of unknown length. Upstream from this gap we detected a sequence with 62% nt identity to exon 1 of human *CD1D* and downstream from the gap a sequence with 68% identity to human exon 3 (ENST00000368171; www.ensembl.org) (Fig. 1b). No canine homolog of the human *CD1D* exon 2 was found. Putative full-length canine exons were found for *CD1D* exon 4, 5, 6 and 7, with identities of respectively 76%, 84%, 58% and 56% to the corresponding human exons.

To be able to fill the gap in the genomic DNA sequence between exon 1 and exon 3 of canine *CD1D*, we performed a two-step digestion on BAC clone XX-14K12 DNA using restriction endonuclease HpaI followed by HindIII and NotI (Fig. S1). The two restriction fragments containing the gap were subsequently cloned and sequenced. The complete DNA sequence of the gap was obtained and consists of 722 nt (GenBank GU930707) (Fig. 1c). No canine homolog of the human *CD1D* exon 2 was found in this DNA sequence.

Using Tandem Repeat Finder, three different types of tandem repeats were identified in the sequence. Tandem repeats occur more frequently in the canine genome compared to other mammalian genomes as a result of an impaired DNA replication and repair mechanism in dogs

Address for correspondence

I. Van Rhijn, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584 CL Utrecht, The Netherlands.
E-mail: i.vanrhijn@uu.nl

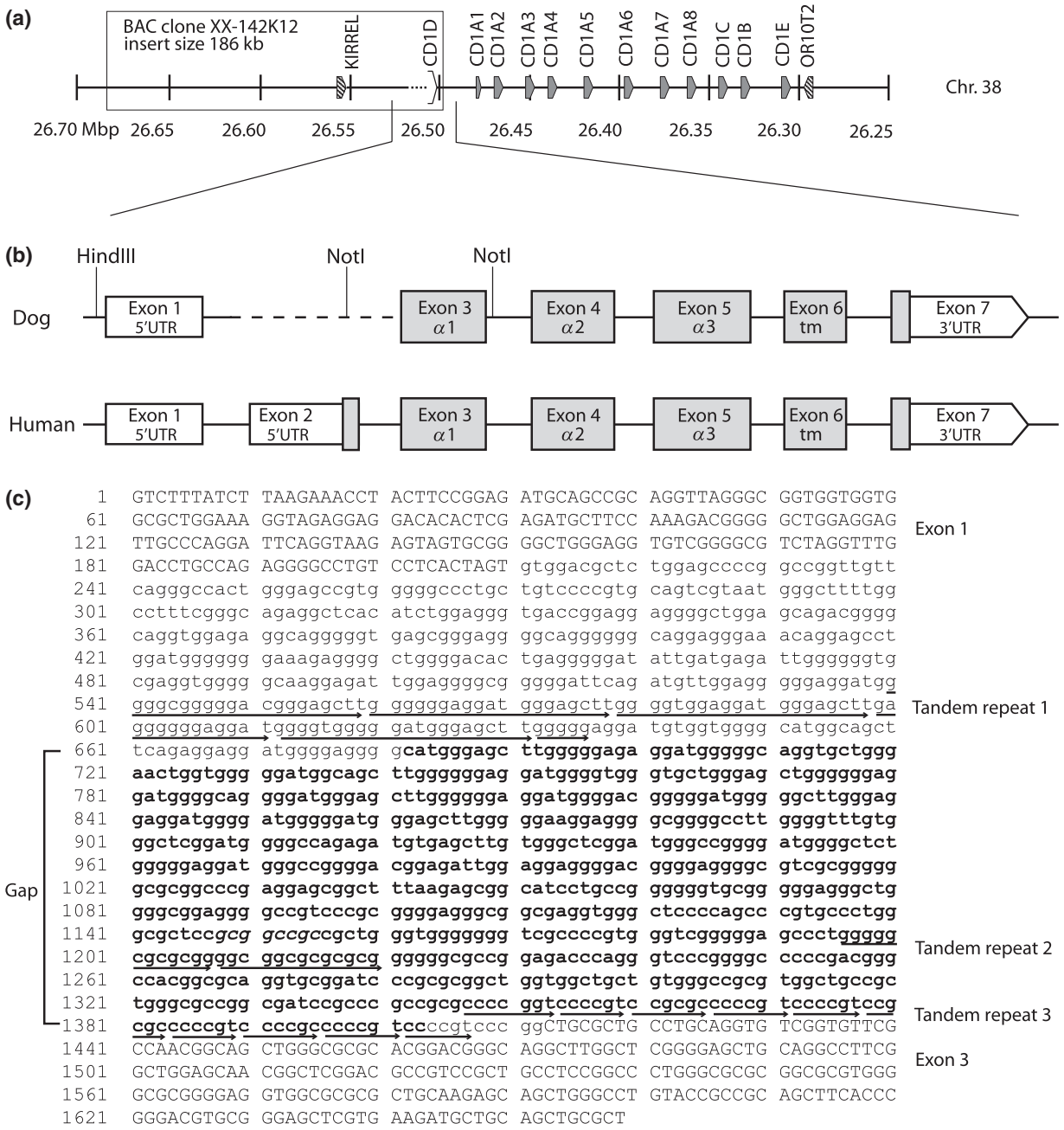

Figure 1 Canine CD1 locus and *CD1D* gene. (a) Map of the canine CD1 locus containing eight *CD1A* genes, one *CD1C* gene, one *CD1B* gene, one *CD1E* gene and one *CD1D* gene as previously published (Looringh van Beeck *et al.* 2008). The gap in the genomic sequence is indicated by a dashed line. BAC clone XX-142K12 that covers the gap and the *CD1D* gene is indicated by a box. (b) Schematic representation of the alignment of canine and human *CD1D* showing exons (boxes) and introns (line) and the gap (dashed line). Restriction sites for HindIII and NotI in the canine *CD1D* gene are shown. (c) Partial genomic sequence of canine *CD1D*, including: gap sequence (bold, lower case), canine homologs of human exon1 and 3 sequence (upper case), three types of tandem repeats (arrows) and the NotI-site within the gap (italic).

(Laidlaw *et al.* 2007). The repeat fragment size of the first tandem repeat (TR1) is 20 nt and was found five times. The second tandem repeat (TR2) has a fragment size of 12 nt and was repeated twice. The region containing TR1 and TR2 is characterized by a high guanine content (59%), which may explain why this part had not been sequenced before. The motif TR3 consists of a hexanucleotide repeat of the consensus DNA sequence CCCCGT. This tandem repeat was present 10 times and, based on alignment with human *CD1D*, is located at the intron–exon boundary of canine *CD1D* homolog of human *CD1D* exon 3. To determine the presence and the copy number of the tandem repeats in dog breeds other than the Boxer from which BAC clone XX-142K12 was derived, PCR was performed on genomic DNA

of three Beagles, three Labrador Retrievers and a wolf. Using a primer set located outside TR3 (primer set 1, Appendix S1), PCR products were generated and sequenced (Fig. S2). We found a $(CCCCGT)_6$ sequence in wolf genomic DNA (GenBank GU930708). In the PCR products of the three Beagles, we only detected the $(CCCCGT)_8$ sequence (GenBank GU930709). In one Labrador Retriever we found both $(CCCCGT)_8$ and $(CCCCGT)_{10}$, whereas in the other two Labrador retrievers only $(CCCCGT)_{10}$ was found (GenBank GU930710). These findings indicate that TR3 is a variable number tandem repeat with at least three different alleles. This suggests that TR3 is still an active site with tandem duplication events before and after the divergence of domestic dogs from the gray wolf between 15 000 and 100 000 years ago (Lindblad-Toh *et al.* 2005).

Using primer set 5 (Appendix S1), we observed differential transcription of *CD1D* in cDNA from various Beagle tissues (Fig. 2a). The highest transcription levels were found in thymus, lymph node, spleen, and PBMC, showing that tissue distribution of transcription of *CD1D* in dogs is comparable to other species and not hampered by the presence of the tandem repeats. Sequencing of *CD1D* PCR products generated from Beagle thymus-derived cDNA, generated with primer set 6, located outside the full-length coding sequence of human CD1d, revealed that canine *CD1D* transcript contained exon 1 and exon 4–7 (Genbank JX139112). Longer PCR products were not obtained using this primer set. Figure 2b shows the structure of the human (NM_001766.3; coding sequence 1008 bp) and canine *CD1D* cDNA sequences (Genbank JX139112; coding sequence 726 bp). Consistent with the absence of an exon 2 homolog in canine genomic DNA, the canine transcript lacks a sequence similar to human exon 2 encoding the signal sequence of CD1d. However, an alternative start codon in the CD1d reading frame was present in exon 1. No signal peptide could be identified. Exon 3, encoding the $\alpha 1$ domain, was also absent in the cDNA. The deletion of nucleotides flanking a repeat is common, and it is likely that TR3 is responsible for the deletion of the first part of canine *CD1D* exon 3 and its acceptor splice site, explaining the absence of exon 3 in the cDNA. The introns between exon 4–7 were spliced out as expected, and no additional splice variants were found. Because a homolog of exon 3, encoding the $\alpha 1$ domain, was absent, it is not expected that translation of this transcript will lead to a functional CD1d protein. Even though only Beagle CD1d cDNA was analyzed and we cannot formally rule out that other breeds have normal CD1d transcripts, the Beagle CD1d cDNA is consistent with the Boxer *CD1D* genomic sequence, and we have confirmed the presence of TR3 in Labrador and wolf. Therefore we expect that other dog breeds show the same aberrant CD1d transcript. Together, our findings suggest that, despite gene transcription, expression of the canine CD1d protein is possibly absent or altered *in vivo*. Consistent with this, a panel of anti CD1d antibodies, raised against different species, did not stain canine thymocytes (Fig. S3). Our findings raise the question of how the recently described canine NKT cells (Yasuda *et al.* 2009) can be positively selected and activated in the absence of CD1d protein. One possibility is that one of the many canine *CD1A*, *CD1B*, or *CD1C* genes might encode a restriction element for these cells.

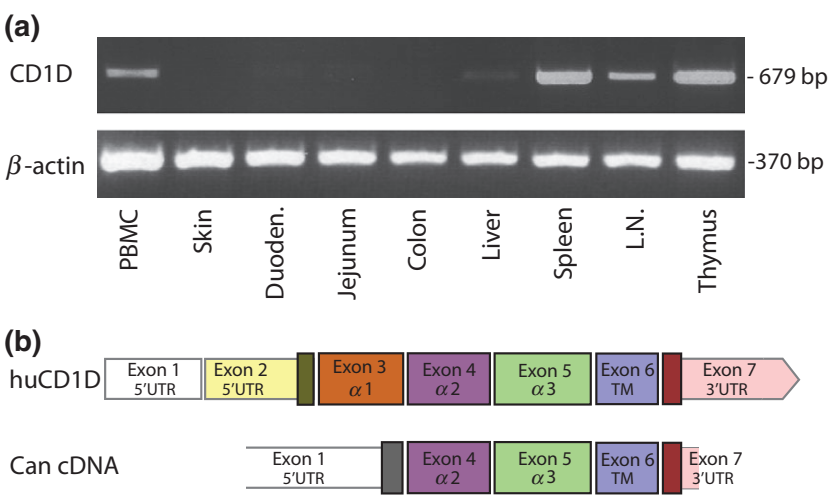

Figure 2 Canine *CD1D* transcription. (a) Transcription of canine *CD1D* in different dog tissues. Primer set 5 was used for the detection of *CD1D* transcription in different dog tissues. To compare the quantity of input of cDNA, a PCR for β-actin was performed. L.N.: Lymph node. (b) Structure of human and canine *CD1D* transcripts, with the untranslated parts represented by narrower boxes in lighter shades, and the coding sequence represented by broader boxes in darker shades. The canine transcript was obtained using primer set 6 and lacks a sequence homologous to human *CD1D* exon 2 and 3. TM: Transmembrane domain. Primer sequences are provided in Appendix S1.

Acknowledgements

IVR was supported by an NWO Meervoud subsidy; Netherlands. FALVB was supported by Royal Canin, France.

References

Barral D.C. & Brenner M.B. (2007) CD1 antigen presentation: how it works. *Nature Reviews Immunology* **7**, 929–41.

Bendelac A., Lantz O., Quimby M.E., Yewdell J.W., Bennink J.R. & Brutkiewicz R.R. (1995) CD1 recognition by mouse NK1 + T lymphocytes. *Science* **268**, 863–5.

Brigl M. & Brenner M.B. (2004) CD1: antigen presentation and T cell function. *Annual Review of Immunology* **22**, 817–90.

Dascher C.C., Hiromatsu K., Naylor J.W. *et al.* (1999) Conservation of a CD1 multigene family in the guinea pig. *Journal of Immunology* **163**, 5478–88.

Eguchi-Ogawa T., Morozumi T., Tanaka M., Shinkai H., Okumura N., Suzuki K., Awata T. & Uenishi H. (2007) Analysis of the genomic structure of the porcine CD1 gene cluster. *Genomics* **89**, 248–61.

Hayes S.M. & Knight K.L. (2001) Group 1 CD1 genes in rabbit. *Journal of Immunology* **166**, 403–10.

Kawano T., Cui J., Koezuka Y. *et al.* (1997) CD1d-restricted and TCR-mediated activation of valpha14 NKT cells by glycosylceramides. *Science* **278**, 1626–9.

Laidlaw J., Gelfand Y., Ng K.W., Garner H.R., Ranganathan R., Benson G. & Fondon J.W. 3rd (2007) Elevated basal slippage mutation rates among the Canidae. *Journal of Heredity* **98**, 452–60.

Lindblad-Toh K., Wade C.M., Mikkelsen T.S. *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–19.

Looringh van Beeck F.A., Zajonc D.M., Moore P.F., Schlotter Y.M., Broere F., Rutten V.P., Willemse T. & Van Rhijn I. (2008) Two canine CD1a proteins are differentially expressed in skin. *Immunogenetics* **60**, 315–24.

Looringh van Beeck F.A., Reinink P., Hermsen R. *et al.* (2009) Functional CD1d and/or NKT cell invariant chain transcript in horse, pig, African elephant and guinea pig, but not in ruminants. *Molecular Immunology* **46**, 1424–31.

Porcelli S.A. & Modlin R.L. (1999) The CD1 system: antigen-presenting molecules for T cell recognition of lipids and glycolipids. *Annual Review of Immunology* **17**, 297–329.

Yasuda N., Masuda K., Tsukui T., Teng A. & Ishii Y. (2009) Identification of canine natural CD3-positive T cells expressing an invariant T-cell receptor alpha chain. *Veterinary Immunology and Immunopathology* **132**, 224–31.