

RESEARCH ARTICLE

Open Access

Gene number determination and genetic polymorphism of the gamma delta T cell co-receptor WC1 genes

Chuang Chen¹, Carolyn TA Herzig^{1,4}, Leeson J Alexander², John W Keele³, Tara G McDanel³, Janice C Telfer¹ and Cynthia L Baldwin^{1*}

Abstract

Background: WC1 co-receptors belong to the scavenger receptor cysteine-rich (SRCR) superfamily and are encoded by a multi-gene family. Expression of particular WC1 genes defines functional subpopulations of WC1⁺ $\gamma\delta$ T cells. We have previously identified partial or complete genomic sequences for thirteen different WC1 genes through annotation of the bovine genome Btau_3.1 build. We also identified two WC1 cDNA sequences from other cattle that did not correspond to sequences in the Btau_3.1 build. Their absence in the Btau_3.1 build may have reflected gaps in the genome assembly or polymorphisms among animals. Since the response of $\gamma\delta$ T cells to bacterial challenge is determined by WC1 gene expression, it was critical to understand whether individual cattle or breeds differ in the number of WC1 genes or display polymorphisms.

Results: Real-time quantitative PCR using DNA from the animal whose genome was sequenced ("Dominette") and sixteen other animals representing ten breeds of cattle, showed that the number of genes coding for WC1 co-receptors is thirteen. The complete coding sequences of those thirteen WC1 genes is presented, including the correction of an error in the WC1-2 gene due to mis-assembly in the Btau_3.1 build. All other cDNA sequences were found to agree with the previous annotation of complete or partial WC1 genes. PCR amplification and sequencing of the most variable N-terminal SRCR domain (domain 1 which has the SRCR "a" pattern) of each of the thirteen WC1 genes showed that the sequences are highly conserved among individuals and breeds. Of 160 sequences of domain 1 from three breeds of cattle, no additional sequences beyond the thirteen described WC1 genes were found. Analysis of the complete WC1 cDNA sequences indicated that the thirteen WC1 genes code for three distinct WC1 molecular forms.

Conclusion: The bovine WC1 multi-gene family is composed of thirteen genes coding for three structural forms whose sequences are highly conserved among individual cattle and breeds. The sequence diversity necessary for WC1 genes to function as a multi-genic pattern recognition receptor array is encoded in the genome, rather than generated by recombinatorial diversity or hypermutation.

Keywords: Bovine, WC1, $\gamma\delta$ T cells

Background

Workshop cluster 1 (WC1) co-receptors belong to group B of the scavenger receptor cysteine-rich (SRCR) superfamily, as do CD163, CD5, CD6, and Sp α , all of which are expressed in immune system cells [1]. We have shown that WC1 is a member of the CD163 multigene

family whose other members are CD163A, CD163b and CD163c- α [2]. WC1 co-receptors are composed of up to eleven extracellular SRCR domains with interdomain homology, organized in the domain pattern of a-[b-c-d-e-d]-[b-c-d-e-d] according to the nomenclature of Sarrrias et al. [1]. The greatest difference among WC1 genes occurs in the most distal SRCR domain ("a" pattern) with identities as low as 50%, contrasting with other SRCR domains which have identities of approximately 90% with like domains [3]. WC1 and CD163c- α have the

* Correspondence: cbaldwin@vasci.umass.edu

¹Department of Veterinary and Animal Sciences, University of Massachusetts, Amherst, MA 01003, USA

Full list of author information is available at the end of the article

most similar extracellular SRCR domain organization [3,4] and it has been proposed that the human and murine homologs of ruminant WC1 are CD163c- α (known as SCART1 and SCART2 and also expressed on $\gamma\delta$ T cells in mice) [2,5,6]. We have shown that a multigenic array of WC1/CD163c- α homologues is conserved over evolutionary time including in the prototherian mammal duck-billed platypus and in the sauropsid chicken [2].

Based on reactivity with specific monoclonal antibodies (mAbs) using WC1-transfected cells, WC1 bearing $\gamma\delta$ T cells (WC1⁺ $\gamma\delta$ T cell) were defined as WC1.1⁺, WC1.2⁺, and WC1.3⁺ wherein the WC1.3⁺ population is a subpopulation of WC1.1⁺ cells [7]. The WC1.1⁺ and WC1.2⁺ mAb-defined subpopulations are largely non-overlapping and may be functionally distinct subsets of WC1⁺ $\gamma\delta$ T cells since they have different cytokine production and cellular proliferation in response to stimulation [8,9]. For example, *ex vivo* WC1.1⁺ $\gamma\delta$ T cells, but not WC1.2⁺ $\gamma\delta$ T cells, proliferate well to the $\gamma\delta$ T cell antigens of *Leptospira*, and produce IFN- γ in response to either antigen or IL-12 [8,9]. However, WC1.2⁺ $\gamma\delta$ T cells respond to the rickettsiales bacteria *Anaplasma* [10]. It was also notable that WC1.1⁺ cells decreased steadily with aging, while the WC1.2⁺ cells did not, suggesting their different functional roles [9]. Since $\gamma\delta$ TCR gene usage is not different between WC1.1⁺ and WC1.2⁺ $\gamma\delta$ T cells [11], this may suggest that expression of particular WC1 family members directs the antigen-specific activation of $\gamma\delta$ T cells.

Based on Southern blot analysis, it was predicted that there were over fifty WC1 (also known as T19) ovine genes [12,13], and nineteen WC1 bovine genes [14]. To better characterize the WC1 co-receptor family, we annotated the WC1 regions in the bovine genome Btau_3.1 assembly, identifying partial or complete sequences of thirteen WC1 genes distributed between two regions on chromosome 5 [3]. The annotated number of WC1 genes is consistent with our previous study that identified thirteen different WC1 intracytoplasmic tail transcripts [15] but was fewer than the nineteen genes predicted by Southern blot analysis. In addition, we had also identified two additional Domain A transcript sequences, *WC1-nd1* and *WC1-nd2*, derived from a different breed of cattle than that used for the genome sequencing [3]. The missing genomic evidence for *WC1-nd1* and *WC1-nd2* in the genome of the animal "Dominette" could be due to gene number variation, polymorphisms among individual cattle or alternatively gaps in the assembled genome. Thus, the complexity of the WC1 multi-gene family remained unresolved including gene number and potential sequence polymorphisms; more recent assemblies have not ameliorated these problems.

Real-time quantitative PCR (Q-PCR) is highly sensitive and allows quantification of very small changes in

sequence and rare transcripts [16,17]. Real-time Q-PCR has evolved to increase the accuracy and efficiency of the nucleic acid quantification process, making Q-PCR a reliable and powerful tool [18]. For example, Q-PCR has successfully quantified viral copy number and gene number in transgenic animals and measured oncogene amplification in tumor cells [19-23]. In relative quantification methods, the amount of target gene in a sample is presented relative to a calibrator which contains both target and reference genes at a constant ratio [24]. In this study, we adapted it to determine the gene number of WC1 genes in bovine genomes.

Methods

PBMC

Cattle of the Belted Galloway and Holstein breeds were 12–24 months of age. Blood was collected into heparin by venipuncture of the jugular vein. Peripheral blood mononuclear cells (PBMC) were isolated from blood via density gradient centrifugation over ficoll-hypaque (Ficoll-Paque, LKB-Pharmacia Biotechnology, Piscataway, NJ) using standard techniques and viable cell concentrations determined by trypan blue exclusion. PBMC were cultured at 2.5×10^6 cells/ml with Concanavalin A (ConA; 1.0 μ g/ml; Sigma-Aldrich, St. Louis, MO) or leptospira antigen ([9], 0.5 μ g/ml; sonicated whole cells of *L. borgpetersenii* serovar hardjo clone RZ33) in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (HyClone, Logan, UT), 2 mM L-glutamine, 50 μ M 2-mercaptoethanol and 50 μ g/ml gentamicin at 37°C with 5% CO₂ in air for six days. All animal use complied with federal guidelines and had Institutional Animal Care and Use Committee (IACUC) approvals.

Genomic DNA extraction and RNA isolation

Genomic DNA of seven cattle from two different breeds (5 Belted Galloway and 2 Holstein) was extracted from whole blood using FlexiGene DNA Kit (50) (Qiagen, Valencia, CA) according to the manufacturer's protocol at the University of Massachusetts. To isolate RNA, pelleted *ex vivo*, ConA-activated, and *Leptospira*-activated PBMC, as well as sorted WC1.1⁺ $\gamma\delta$ T cells, were resuspended in TRIzol (Invitrogen, Carlsbad, CA) and RNA was isolated according to the manufacturer's protocol. Reverse transcription (RT) was performed using 1 μ g of total RNA, oligo dT primers and AMV reverse transcriptase (AMV RT kit; Promega, Madison, WI). Genomic DNA and cDNA from the Herford Dominette, the animal used for the current bovine genome sequencing and annotation project [25,26], were also obtained with total RNA isolated using a LeukoLOCK kit (Ambion, Austin TX) at USDA-ARS Fort Keogh, while genomic DNA from Red Angus, Angus, Charolais, Limousin, Brahman crossed with Angus, Gelbvieh, and Angus crossed with Hereford were obtained from semen or

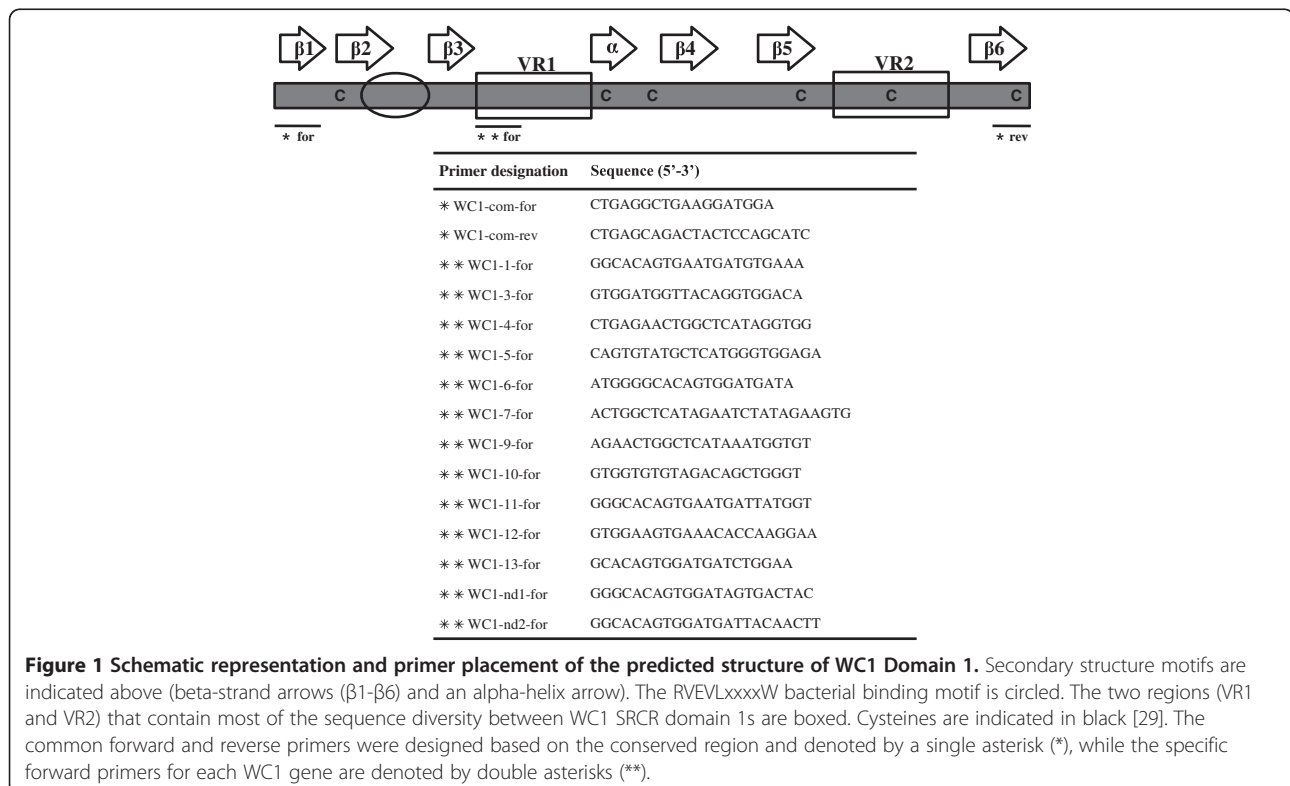
leukocytes using standard isolation methods at USDA-ARS Clay Center.

Real time quantitative-PCR

Genomic DNA from sixteen cattle of ten different breeds was used as template. Real time Q-PCR amplification was done with primers against consensus bovine sequences (i.e. common primers) for all members of each gene family as follows: WC1 SRCR "domain 1" (see Figure 1 for sequence), IFNA (forward 5' ATGGCCCC AGCCTGGTCCTTCC, reverse 5' TCAGTCCTTTCTC CTGAAACTC), IFNB (forward 5' ATGACCTACCGGT GCCTCCTCC, reverse 5' TCAGTCACGGACGTAACC TG), IFNE (forward 5' ATGATTAACAAGGCTTTCTT TG, reverse 5' GCTTTTAAAGCCTGCAGTCCG), and IF NW (forward 5' ATGGCCTTCATGCTCTCTCTAC, reverse 5' TCAAGGTGAGTTCAGGTCTCCATC). T cell receptor delta joining gene 1 (TRDJ1) (forward 5' CCTC AACCACAAGAGTCTGTAC, reverse 5' CCAGCTGGG AGTCTGAGATC) and Glyceraldehyde-3-phosphate dehydrogenase (*GAPD*) (forward 5' TTCAACGGCACA GTCAAGG, reverse 5' ACATACTCAGCACCAGCATC AC). The WC1 common primer pair (WC1-com-for and WC1-com-rev) was designed to amplify all WC1 "do-main 1" sequences based on the conserved region annotated in a previous study [3] (Figure 1). The GenBank accession numbers of the sequences used for designing

the primers in this experiment are as follows: IFNA (NM_001017411.1; *IFNA*), IFNB (NM_174350.1; *IFNB*), IFNE (NM_176891; *IFNE*), and IFNW (NM_174351.1; *IFNW*). The T cell receptor delta joining gene 1 (*TRDJ1*) primers and Glyceraldehyde-3-phosphate dehydrogenase (*GAPD*) primers were described previously [27,28]. Gene numbers among animals were evaluated for differences using ANOVA.

Real-time Q-PCR amplification and analysis were performed using a Stratagene Mx3005P instruments with software version 4.01 (Stratagene, La Jolla, CA). The Q-PCR assays were optimized in terms of Mg²⁺ concentration and the annealing temperature [30]. Q-PCR amplification mixture (25ul) was prepared by using Sybr Premix Ex Taq (TAKARA, Pittsburgh, PA): 20 ng template DNA, 2-fold concentration of premix reagent including Takara Ex Taq™ HS and SYBR® Green I, 0.5ul ROX reference dye, and 1ul of forward and reverse primers (final concentration is 0.5uM for each). Real-time PCR amplification was conducted for 35 cycles, each cycle consisting of denaturation (95°C for 5 sec), annealing (55°C for 20 sec) with a single fluorescence measurement taken at the end of the annealing step, and extension (72°C for 20 sec). After amplification, melting-curve analysis was performed by raising the temperature to 95°C for 1 min, heating the sample at 55°C for 30 sec followed by 95°C for 30 sec. The $\Delta\Delta C_T$ method was



applied for gene number determination [21]: relative amount of targets = $(1 + E)^{-\Delta\Delta C_T}$, where $\Delta\Delta C_T$: ΔC_T of the targets - ΔC_T of the calibrators, ΔC_T of the target: C_T of the targets - C_T of the reference, and ΔC_T of the calibrators: C_T of the calibrators - C_T of the reference. In this case, the 'targets' were bovine WC1 domain 1, bovine *IFNA*, bovine *IFNB*, bovine *IFNW* while the 'reference' was bovine *GAPD*, and the 'calibrators' were bovine *TRDJI* and bovine *IFNE*. Real-time PCR products were analyzed on 1% or 1.2% TAE agarose gels, visualized using SYBR Safe (Invitrogen) and cloned into the pCR2.1 vector (Invitrogen) according to the manufacturer's protocol for sequencing.

PCR amplification specific for Domain 1

For amplifying each Domain 1 sequence specifically, PCR reactions were performed with PCR Mastermix (Promega, Madison, WI) according to the manufacturer's instructions and primers were designed based on one of the most variable regions of domain 1 (Figure 1, [3]). The reverse common primer (WC1-com-rev) was designed within a conserved region of the domain 1 of all known WC1 molecules as denoted in Figure 1. The GenBank accession numbers of the expressed gene sequences used for designing the primers in this experiment are as follows: WC1-3 (previously known as archetypal WC1.1; X63723), WC1-nd1 (clone CH525; FJ031216), WC1-nd2 (clone CCnd2; JN998896), while the others (WC1-1, WC1-4, WC1-5, WC1-6, WC1-7, WC1-9, WC1-10, WC1-11, WC1-12 and WC1-13) were annotated in our previous study [3] and shown in Table 1. Cycling parameters for those reactions were 30

sec at 95°C, 1 min at 58°C and 1 min at 72°C for 30 cycles with an expected amplicon size of approximately 200 bp. PCR products were visualized using SYBR Safe (Invitrogen) on 1% or 1.2% TAE agarose gels and isolated after visualization and cloned into the pCR2.1 vector (Invitrogen) according to the manufacturer's protocol and sent for commercial sequencing (GeneWiz, South Plainfield, NJ).

PCR amplification for complete coding sequence

For amplifying the complete coding sequence of WC1 genes, 2 µl of pooled cDNA was used as a template and PCR reactions were conducted using the Elongase Amplification system (Invitrogen) with a final concentration of 1.5 mM Mg²⁺. Based on previous research, forward primers in the signal sequence (WC1atg-for 5'A TGGCTCTGGGCAGACACCTCTC) and reverse (WC1-groups1,2-rev 5'TCAYGAGAAAGTCAYTGKGGATG) primers in the intracytoplasmic tail sequence were designed to amplify all known WC1 transcripts except WC1-11 which required the following primers: forward (WC1-atg-for 5'ATGGCTCTGGGCAGACACCTCTC) and reverse (WC1group3rev 5'-CTACATGGTGCTAAGCTCC ACATC) [3]. Cycling parameters were 30 sec at 94°C, 30 sec at 55°C and 5 min 30 sec at 68°C for 35 cycles for all reactions. PCR products were analyzed on 1.2% TAE agarose gels, visualized using SYBR Safe (Invitrogen) and cloned into the pCR-XL vector (Invitrogen) for sequencing.

Sequence analyses

Sequencing was performed commercially (Genewiz) to verify amplicons. Nucleotide sequences were aligned and consensus sequences were created using Bioedit version 7.0.5.3 [31]. GenBank accession numbers of annotated sequences used for comparisons in analyses are shown in Table 1 as annotated and/or reported in our previous research [3] except archetypal WC1.1 whose GenBank number is X63723. Multiple sequence alignments were performed using clustalw2 (<http://www.ebi.ac.uk/Tools/clustalw2/index.html> website; [32]) and the default parameters, but manually optimized when necessary, and were visualized using Bioedit [31]. Phylogenetic analyses were performed using deduced amino acid sequences of WC1 domain 1 as indicated. Phylogenetic trees were created using Bayesian analysis in MrBayes3.2 [33]. For Bayesian analysis, 2 runs with 3 cold chains and 1 heated chain each were done. An amino acid mixed model was used to approximate the posterior probabilities of trees. The 90-taxa SRCR domain 1 alignment was run with temperature settings of 0.2 for 830,000 generations. Trees were sampled every 100 generations and the burn-in fraction was 0.25. The convergence diagnostic used was the average standard deviation of split frequencies, which were <0.05 (0.01) for the run. Phylograms were

Table 1 Chromosomal location in Btau_3.1assembly and GenBank accession number of WC1 genes

Gene name	cDNA clone	GLEAN number	GenBank accession number
WC1-1	CH501 ^a	GLEAN_13183 ^a	FJ031186
WC1-2	CCnd1	GLEAN_13182 ^a	JN998897
WC1-3	CH534 ^a	GLEAN_13181 ^a	FJ031191 ^a
WC1-4	CH496 ^a	GLEAN_13179 ^a	FJ031202 ^a
WC1-5	CH590 ^a	GLEAN_13176 ^a	JQ900627 ^a
WC1-6	CC6	GLEAN_00457/GLEAN_00458 ^a	JN234380
WC1-7	CC7	GLEAN_00456 ^a	JN234377
WC1-8	CCnd2	GLEAN_12186 ^a	JN998896
WC1-9	CH505 ^a	GLEAN_12191 ^a	FJ031208 ^a
WC1-10	CH601 ^a	GLEAN_12192 ^a	JQ900628 ^a
WC1-11	CH486 ^a	GLEAN_12182/GLEAN_09904 ^a	FJ031209 ^a
WC1-12	CC12	GLEAN_09902 ^a	JN234378
WC1-13	CH504 ^a	GLEAN_12187 ^a	FJ031187 ^a

a. Herzig et al., 2009, *BMC Genomics*, [3].

visualized using FigTree V1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Magnetic bead cell sorting

PBMC were stained for surface markers at 4°C for 20 min in PBS with 2 mM EDTA and 0.5% BSA. The anti-WC1 mAb BAG25A (VMRD, Pullman, WA) for WC1.1 epitopes was used for sorting. Cells were then incubated with goat anti-mouse IgM-conjugated magnetic microbeads (Miltenyi Biotec, Auburn, CA, USA) at 4°C for 20 min. After washing twice, cells were applied to the column following the manufacturer's instructions. The purity of

collected fractions was assessed by flow cytometry and analyzed using FlowJo (Tree Star, Ashland, OR, USA).

Results

The WC1 family is composed of thirteen genes

Due to gaps in the bovine genome Btau_3.1 assembly [3,26], we were uncertain whether we had identified the total complement of WC1 genes present. Moreover, the possibility existed that gene number variation occurs among breeds of cattle or individuals within a breed. To address this we adapted Q-PCR to determine WC1 gene numbers in the Hereford Dominette, the reference/

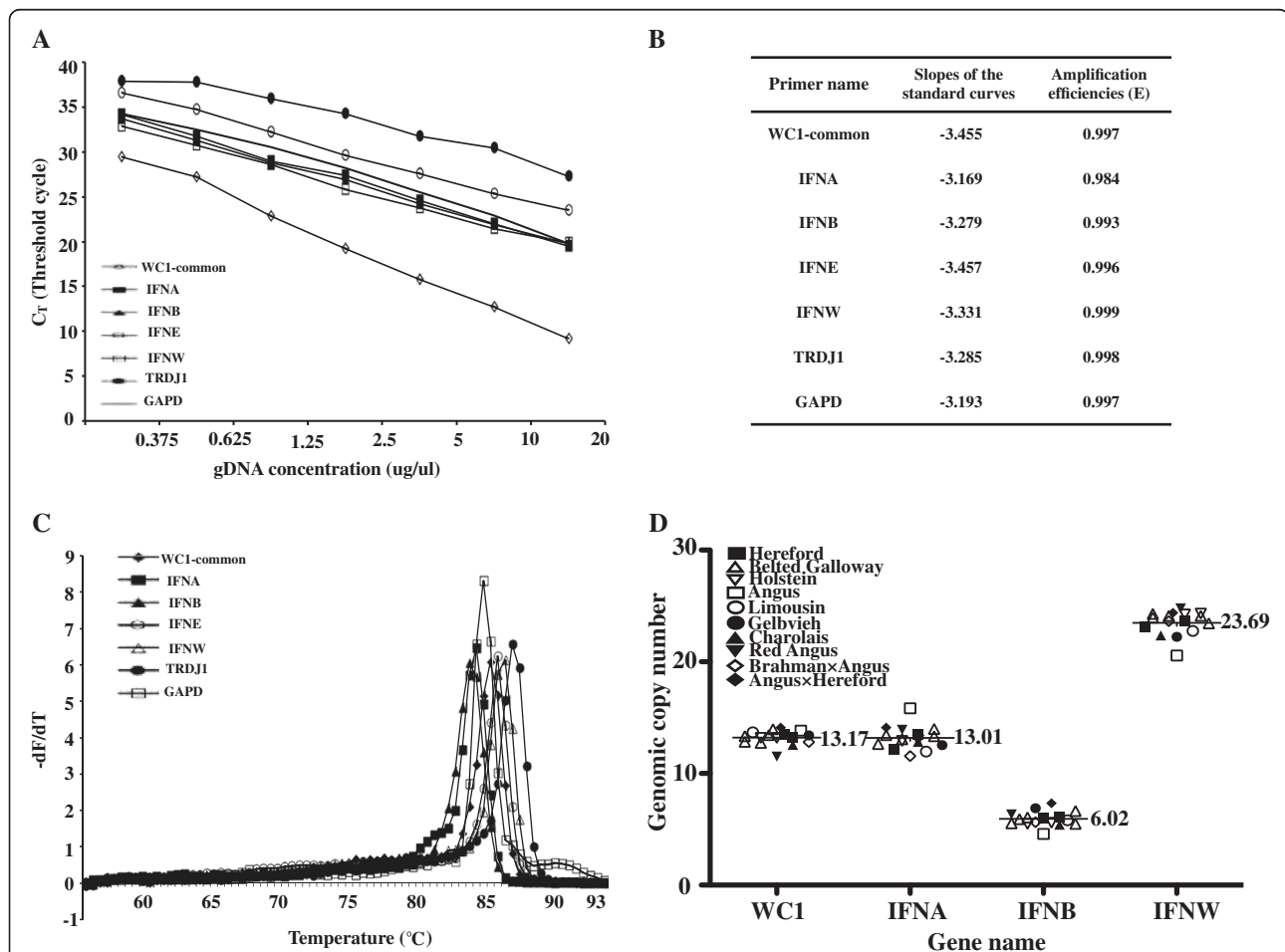


Figure 2 Q-PCR for WC1 gene number. (A) The standard curves for primer sets of WC1s (WC1-com) and other control genes. The standard curves were constructed with a series of 10-fold dilution of cDNA derived from ex vivo PBMC. Each standard dilution was amplified by real-time QPCR in duplicate. For each primer set, C_T values determined from real-time QPCR were plotted against the logarithm of their known initial gene numbers. A standard curve was generated by linear regression through these points. (B) The slopes of the standard curves for primers of WC1s (WC1-com) and other control genes. From the slopes, amplification efficiencies were also determined. (C) Primer specificities of WC1s (WC1-com) and other control genes' primer sets. Confirmation of PCR amplification specificities by melting curve for primer sets of WC1 common (WC1-com), bovine *TRDJ1*, bovine *GAPD*, bovine *IFNA*, bovine *IFNB*, bovine *IFNE*, and bovine *IFNW*. Melting peaks were examined for WC1 common (WC1-com, solid line with empty diamond marker), bovine *TRDJ1* (solid line with solid circle marker), bovine *GAPD* (solid line with solid square marker), bovine *IFNA* (solid line with solid triangle marker), bovine *IFNB* (solid line with empty circle marker), and bovine *IFNW* (solid line with empty square marker). (D) Gene numbers of bovine WC1 genes and other control genes. The breeds of tested cattle are indicated in the figure. The $\Delta\Delta C_T$ method was applied for relative quantification. Some breeds contained more than one animal, and each evaluation was performed at least twice, yielding similar results.

donor animal used for the Bovine Genome Sequencing and Annotation project [26], as well as in additional breeds of cattle.

Although considerable repetition of sequence occurs among repeating SRCR domains of WC1 molecules (i.e., b,c,d,e,d'), the most distal SRCR domain (domain 1 which has an "a" pattern [1]) of each known WC1 molecule is unique in terms of structure and sequence relative to all other WC1 domains [3] and coded for by a single exon. Thus, we reasoned that the number of SRCR domain 1 gene exons would be proportional to the WC1 gene number. As controls, bovine *IFNA*, bovine *IFNB*, and bovine *IFNW* genes were evaluated in our system since they are multigene families with known gene numbers [34]. Bovine T cell receptor δ J1 gene (*TRDJ1*) and *IFNE* were both used as calibrators since they are present as single gene copies in the bovine genome [28,34]. Bovine glyceraldehyde-3-phosphate dehydrogenase (*GAPD*) was used as a reference gene for DNA quality [27].

Primer sets were designed to amplify all members within a family (*IFNA*, *IFNB* and *IFNW*) and for all known WC1 domain 1 sequences identified to date (see common primers for WC1 in Figure 1). Standard curves for primer sets that amplify all WC1 known genes and for bovine *TRDJ1*, *GAPD*, *IFNA*, *IFNB*, *IFNE*, and *IFNW* were constructed with a range from 20 ug/ul to 0.375 ug/ μ l (Figure 2A). All curves were linear in the range tested ($R^2 > 0.95$) in duplicate reactions. The slopes of the standard curves and amplification efficiencies (E) were determined to be in the tested range (Figure 2B) and thus the primers were used for further relative quantification. The specificity for primers was determined by the melting curve analysis to rule out amplification of non-specific PCR products [35]. Figure 2C shows sharp peaks in the fluorescence signal around the melting temperature (T_M) of the PCR products. In addition, the $2^{-\Delta\Delta CT}$ relative quantification method requires amplification efficiencies of the target and reference to be approximately equal to be valid [36]. It was observed that the difference between amplification efficiencies of the targets and the reference (bovine *GAPD*) were less than 0.1 which indicates that the amplification efficiencies of the target and reference were similar enough to perform the $2^{-\Delta\Delta CT}$ relative quantification method.

The results from the relative quantification and the calculated gene numbers are shown in Figure 2D, which were based on amplification efficiencies calculated as described above and the equation (relative amount of target = $(1 + E)^{-\Delta\Delta CT}$) described previously [21]. The results for sixteen animals of ten different breeds of cattle (Herford, Belted Galloway, Holstein, Red Angus, Angus, Charolais, Gelbvieh, Limousin, Brahman cross Angus, and Angus cross Hereford) showed a mean gene number of 13.01, 6.02, and 23.69 for bovine *IFNA* (13

expected), bovine *IFNB* (6 expected), and bovine *IFNW* (24 expected), respectively, which are consistent with results in previous studies [26,34]. For bovine WC1 genes we obtained a mean gene number of 13.17 (Figure 2D). According to the obtained Q-PCR results, the number of WC1 genes for some tested cattle (one Holstein and one Red Angus) was less than thirteen. It is possible that those cattle have fewer than 13 WC1 genes, but statistical analysis indicated that the mean number of WC1 genes was thirteen without variation among all the tested individuals and breeds. Thus we conclude that the bovine genome contains thirteen WC1 genes and that this number is consistent among ten breeds of cattle.

Complete SRCR domain 1 sequences of the thirteen WC1 genes in the donor/reference animal Dominette

We previously annotated thirteen WC1 genes distributed between two regions on bovine chromosome 5 in the bovine genome Btau_3.1 assembly [3]. However, only

Table 2 Available sequence information from the 13 WC1s in previous study^a

Gene Name	Annotated genomic sequences from Dominette ^b	RNA transcripts representing the expressed gene sequences obtained from other animals
WC1-1	+	+
WC1-2	Db7-Dd'11 ^c	-
WC1-3	D1 ^c	+
WC1-4	+	+
WC1-5	+	+
WC1-6	D1-Dd'11 ^c	-
WC1-7	+	-
WC1-8	Dc8-Dd'11 ^c	-
WC1-9	+	+
WC1-10	+	+
WC1-11	+ ^d	+
WC1-12	+	-
WC1-13	+	+
WC1-nd1	-	+ ^{d, e}
WC1-nd2	-	D1 only ^e

a. Herzig et al., 2009, *BMC Genomics*, [3].

b. Dominette was the reference animal whose DNA was used by The Bovine Genome Sequencing and Analysis Consortium, [26]; D = domain, and they are described by their pattern type (a-d') as well as presumed position within the molecule (Domain 1 to 11). + = all 11 domains present in assembled genome or corresponding cDNA obtained.

c. Only partial sequences were obtained due to gaps in the genome sequences.

d. A transcript corresponding to SRCR Domain 1 through the intracytoplasmic region were obtained but it coded for only six extracellular domains, thus it was smaller than the majority of other WC1 molecules [3,14].

e. While transcripts for the SRCR Domain 1 of the proposed genes were found, they were named *WC1-nd1* and *WC1-nd2* because their placement was "not determined" in the genome of Dominette.

partial sequences for four of the genes (*WC1-2*, *WC1-3*, *WC1-6*, *WC1-8*) were annotated due to gaps in the genome sequences while some lacked complete transcript sequences [3] (see Table 2 for a summary). Coincident with the annotation, cDNA analysis of material derived from a different animal provided evidence for two additional WC1 domain 1 sequences (designated *WC1-nd1* and *WC1-nd2*), that were not placed in the genome [3] (Table 2). We reasoned that these represented sequences that were not identified as a result of gaps in the genome assembly of Dominette or breed-dependent polymorphisms in WC1 gene sequences since above we report there are only thirteen WC1 genes in all animals evaluated. To distinguish between these two possibilities, Dominette's genomic DNA as well as cDNA from her PBMC were amplified by primers specific for each of the thirteen WC1 domain 1 sequences available to us including sequences for *WC1-nd2* and *WC1-nd1* which were not present in the genome assembly. The WC1

gene-specific forward primers were based on one of the most variable regions of domain 1 sequences (denoted by double asterisk in Figure 1) while the reverse primer was designed within a conserved region.

To verify the amplification specificity of those primer sets, PCR reactions were performed on plasmids containing each of thirteen WC1 domain 1 sequences. The primer pair for bovine *IFNB* was used as the negative control since bovine *IFNB* is not related to bovine WC1. Each set of thirteen WC1 PCR primers generated a single prominent band with expected size (200bp) when the templates contained its corresponding WC1 gene (Figure 3A), indicating that the thirteen specific WC1 forward primers in conjunction with the common reverse primer amplified only their corresponding domain 1 sequence. This was validated by sequencing the amplicons. In addition, we verified the absence of nonspecific amplification of carrier sequences by PCR-amplification of negative controls (vector without inserts and with unrelated inserts), and the

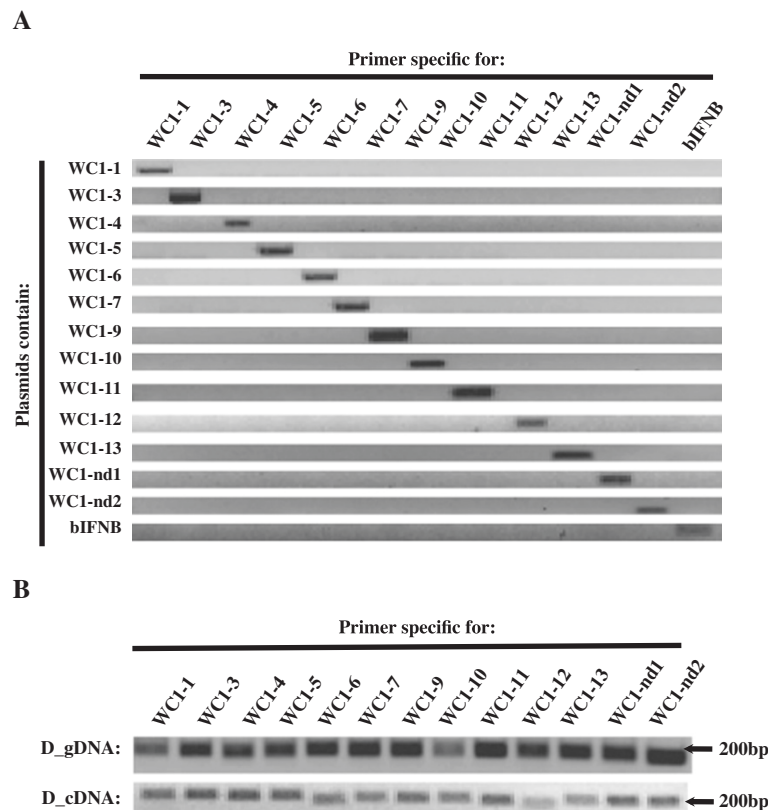


Figure 3 Specific amplification of each WC1 Domain 1. (A) Primer specificities of thirteen WC1 specific primer sets. Confirmation of PCR amplification specificities by gel electrophoresis for thirteen WC1 specific primer sets (*WC1-1*, *WC1-3*, *WC1-4*, *WC1-5*, *WC1-6*, *WC1-7*, *WC1-9*, *WC1-10*, *WC1-11*, *WC1-12*, *WC1-13*, *WC1-nd1*, and *WC1-nd2*) and the bovine *IFNB* primer pair. Plasmids containing thirteen WC1 domain 1 gene sequences were used as templates in PCR reactions. For each primer set, the identities of the amplified products were confirmed by DNA sequencing analysis. **(B)** Genomic DNA and cDNA evidence for Dominette. PCR analysis was conducted by using genomic DNA and cDNA derived from the reference animal Dominette of the Hereford breed (designed with prefixes of "D_gDNA and D_cDNA", respectively). Primer pairs for amplification of thirteen WC1 SRCR domain 1 sequences with the specific primer sets tested in (A). For each primer set, the identities of the amplified products were confirmed by DNA sequencing analysis.

absence of contamination by the PCR-amplification of negative 'no template' control (data not shown).

Analysis of PCR products obtained showed that all thirteen known WC1 domain 1 sequences, including that for *WC1-nd1* and *WC1-nd2*, were present in both Dominette's genomic DNA and cDNA (Figure 3B). Thus, we conclude that *WC1-nd1* and *WC1-nd2* correspond to gaps in the assembled genome and reasoned that they might represent the missing *WC1-2* and *WC1-8* domain 1 sequences.

Generating templates to obtain complete coding sequences for all thirteen WC1 genes

To attempt to obtain the complete coding sequences, intracytoplasmic tail sequences were aligned and a common intracytoplasmic tail sequence primer was designed along with a forward primer in the 5' signal sequence (Figure 4A). These primers amplified material of approximately 4.4kb (as described previously [3]) using cDNA from Dominette's ex vivo PBMC (Figure 4A). In

addition there was a smaller major band of 2.7kb (Figure 4B). Because the intracytoplasmic tail sequence of the gene we previously designated *WC1-11* was so different from the others [3], a separate reverse primers had to be designed (Figure 4A); when used it amplified two bands of 2.9kb and 2.2kb (Figure 4B). All four bands were extracted from the gels and evaluated as templates from which to amplify each of the thirteen WC1 domain 1's using gene-specific primer sets. PCR products for ten of the WC1 genes were obtained using the 4.4 kb material with *WC1-nd1*, *WC1-nd2* and *WC1-11* being the exceptions (see Figure 4C for those amplified). The 2.7kb bands proved to contain templates for twelve WC1 genes with *WC1-11* being the one exception as expected (Figure 4C). The *WC1-11* gene sequences were associated with both the 2.9kb and 2.2kb bands, also as expected. Because we were not able to amplify *WC1-nd1* and *WC1-nd2* from the 4.4 kb band, we enriched for WC1 transcripts by sorting for WC1.1⁺ γδ T cells. The

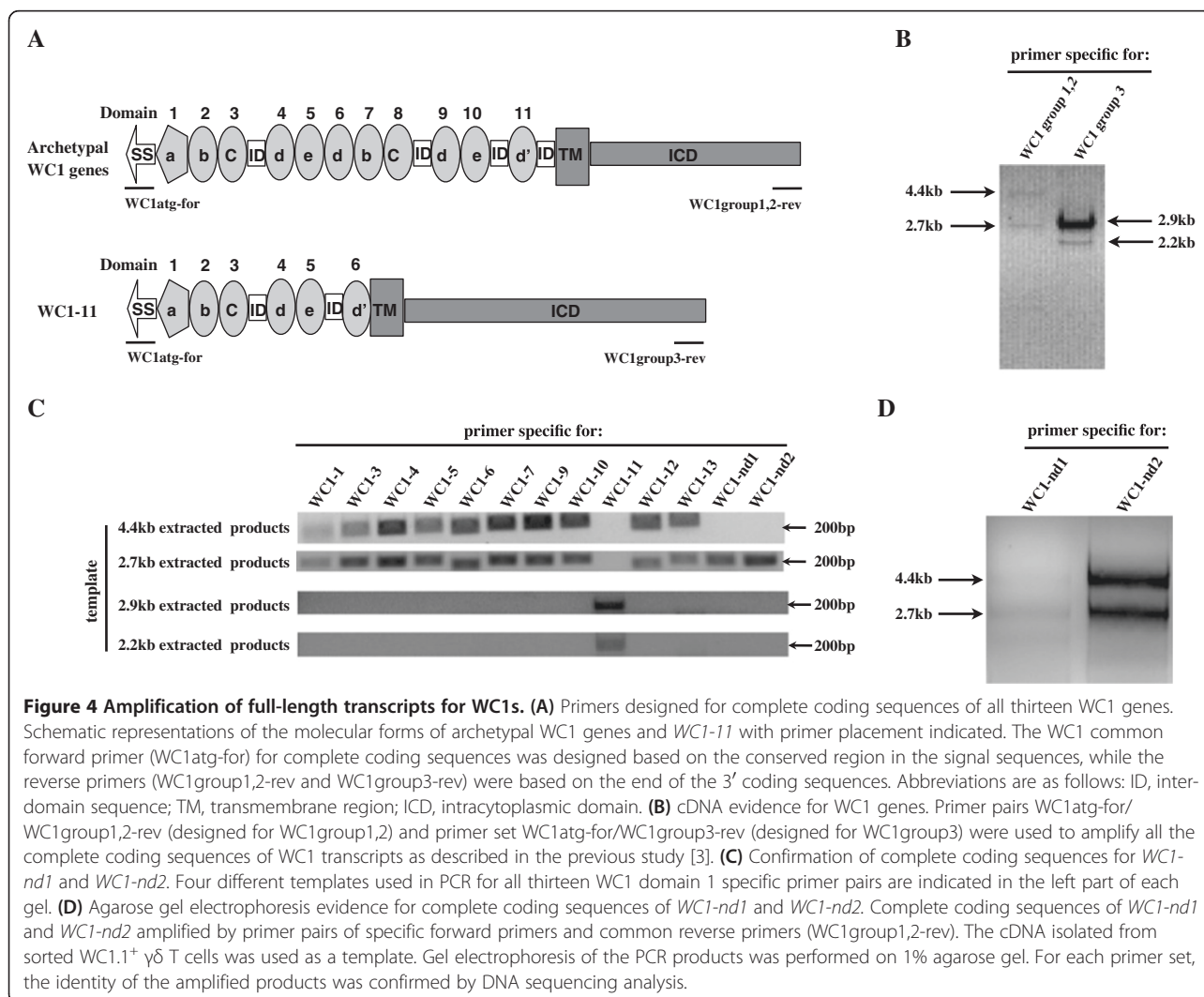


Figure 4 Amplification of full-length transcripts for WC1s. (A) Primers designed for complete coding sequences of all thirteen WC1 genes. Schematic representations of the molecular forms of archetypal WC1 genes and *WC1-11* with primer placement indicated. The WC1 common forward primer (WC1atg-for) for complete coding sequences was designed based on the conserved region in the signal sequences, while the reverse primers (WC1group1,2-rev and WC1group3-rev) were based on the end of the 3' coding sequences. Abbreviations are as follows: ID, inter-domain sequence; TM, transmembrane region; ICD, intracytoplasmic domain. **(B)** cDNA evidence for WC1 genes. Primer pairs WC1atg-for/WC1group1,2-rev (designed for WC1group1,2) and primer set WC1atg-for/WC1group3-rev (designed for WC1group3) were used to amplify all the complete coding sequences of WC1 transcripts as described in the previous study [3]. **(C)** Confirmation of complete coding sequences for *WC1-nd1* and *WC1-nd2*. Four different templates used in PCR for all thirteen WC1 domain 1 specific primer pairs are indicated in the left part of each gel. **(D)** Agarose gel electrophoresis evidence for complete coding sequences of *WC1-nd1* and *WC1-nd2*. Complete coding sequences of *WC1-nd1* and *WC1-nd2* amplified by primer pairs of specific forward primers and common reverse primers (WC1group1,2-rev). The cDNA isolated from sorted WC1.1⁺ γδ T cells was used as a template. Gel electrophoresis of the PCR products was performed on 1% agarose gel. For each primer set, the identity of the amplified products was confirmed by DNA sequencing analysis.

cDNA derived from this was amplified with common primers as described above and the 4.4 kb PCR products was found to contain transcripts corresponding to *WC1-nd1* and *WC1-nd2* domain 1s (Figure 4D, Figure 5, and Figure 6).

The results also confirmed and extended our previous observations that most WC1 transcripts display alternative splicing of coding exons given that domain 1 sequence of all 13 WC1 genes could be amplified from the smaller (as well as the larger) bands [3].

Complete coding sequences for the annotated *WC1-6*, *WC1-7* and *WC1-12* genes

To confirm complete sequences for transcripts (i.e. cDNA clones) representing the complete coding sequences of *WC1-6*, *WC1-7* and *WC1-12*, the template material described in the previous section was used. Forward primers specific for domain 1's of

WC1-6, *WC1-7* and *WC1-12* (Figure 1) were combined with the common reverse primer in the intracytoplasmic tail (Figure 4) [3] and used to successfully amplify the material extracted from the 4.4 kb band described above. The PCR products were sequenced and the deduced amino acid sequences were aligned with the archetypal WC1.1 sequence (now known as *WC1-3*) (Figure 5 and Figure 6). The percent identities were calculated based on those alignments (Table 3) and showed that indeed these represent the complete coding sequence for *WC1-6*, *WC1-7* and *WC1-12*.

WC1-nd2* is *WC1-8

Using the templates generated above and the forward primer specific for *WC1-nd2* domain 1 sequence (Figure 1), coding sequence for *WC1-nd2* was obtained from the 4.4 kb amplified material. This sequence was compared to the partially annotated *WC1-8* sequence

	I	II
WC1.1	MALGRHLSLRGLCVLLGTVMVGGQALELRLKDGVHRCEGRVEVKHQGEWGTVDGYRWTLKDA SVVCRQLGCGAAIGFPGGAYFPGPLGPIWLLYTSCEG-----TESTVSDCEHSNIKDYRNDGYNHGRDAGVVC SGFVRLAGGDGPCSGR	
CC6V.....IV.....T.H.....DME.....V.....K.E.....SLS.DC.....H.....E.L.....	
CC7V.....IV.....V.....A.....M.YK.....YTEY.LIESIE.....R.E.V.DA.R.S.....R.....F.V.....SREVSTLSL.E.V.IDF.N.....TFS.DQ.I.A.....F.S.....F.....R.....	
CC12W.I.....V.S.....AS.....N.GR.....K.....NDHK.S.EE.....R.....DA.Q.S.....AF.....FS.I.R.....ALKQ.RS.TL.....CPQ.S.AH.....A.....M.....E.....	
WC1-8K.....V.....A.D.....D.NLIME.....K.....Y.....S.S.DD.....H.....V.....	
CCnd2K.....V.....A.D.....D.NLIME.....K.....Y.....S.S.DD.....H.....V.....	
WC1-2Q.....A.H.....NDVK.S.....I.....D.....R.....I.....S.S.DW.....A.....	
WC1-1Q.....V.....A.....DLE.....S.....V.....R.....S.L.DW.....E.....Q.....	
WC1-13Q.....V.....A.....D.....SDY.S.....D.....R.....T.N.S.DD.....S.....	
CCnd1V.....A.D.....SDY.S.....D.....R.....T.N.S.DD.....S.....	
	III	IV
WC1.1	VEVHSGEAWIPVSDGNFTLATAQIICAEELGCGKAVSVLGHGELFRSSAQVWABEFCRCEGPEELWCPVPCPGGTCHHSGSAQVVC SAYSEVRLMTNGSSQCEGQVEMNISGQWRALCASHWSLANANVICRQLGCGVAISTPFGGPHLV	
CC6Y.....P.....V.....R.....A.S.....R.....A.....A.....IA.....L.....G.....R.....V.....H.....A.....	
CC7K.....D.TF.....R.....PF.....V.....K.MP.....DG.....G.....R.....L.....M.....A.....L.....L.....G.....R.....V.....R.....	
CC12T.....P.....V.....V.....VP.....G.....GG.....K.....R.....V.....K.....T.....H.....V.....F.....R.....F.....	
WC1-8T.....P.....V.....V.....VP.....DRW.....S.....A.....V.T.....K.T.....H.V.....V.....F.....N.....A.....	
CCnd2T.....P.....V.....V.....VP.....DRW.....S.....A.....V.T.....K.T.....H.V.....V.....F.....N.....A.....	
WC1-2Y.....P.....V.....R.....VK.....R.....V.....	
WC1-1Y.....P.....V.....Q.....R.....S.L.....A.....V.....H.....	
WC1-13Y.....P.....V.....Q.....R.....S.L.....A.....V.....H.....	
CCnd1P.....V.....Q.....R.....S.L.....A.....V.....H.....	
	V	VI
WC1.1	EEGDQILARFHCSGAESFLWSCPVTALGGPDCSHGNTASVICSGNQIQVLPQCNDVSVQPTGSAASEDSAPYCSDSRQLRLVDDGGGFCAGRVEILDQGSWGTICDDGWLDDARVCRQLGCGEALNATGSAHFAGSGPIWLDNLNCT	
CC6G.....T.....N.L.E.A.M.....ERV.....L.....N.....I.....D.....A.....	
CC7G.....S.V.....L.....SG.....T.....S.L.E.....N.....H.....Q.....R.....D.....	
CC12G.....S.....F.....Y.....T.....R.E.A.....DE.....E.....T.....P.....GD.....	
WC1-8GR.....LWK.W.....K.....V.....R.....HT.M.....N.....EQA.....E.....Y.....H.....R.....D.....	
CCnd2GS.....LWK.W.....K.....V.....R.....HT.M.....N.....EQA.....E.....Y.....H.....R.....D.....	
WC1-1G.....T.....N.MAE.A.....VE.....V.G.A.S.....F.....H.....D.....	
WC1-2G.....T.....N.MAE.A.....VE.....V.G.A.S.....F.....H.....D.....	
WC1-13G.....T.....N.MAE.A.....VE.....V.G.A.S.....F.....H.....D.....	
CCnd1G.....T.....N.MAE.A.....VE.....V.G.A.S.....F.....H.....D.....	
	VII	VI
WC1.1	GRVEILDQGSWGTICDDRDLDDARVCRQLGCGEALDVTVSFFGTGSGPIWLDVENVCRGEEQVRCPSWGRQHNHCNQEDAGVICSGFVRLAGGDGPCSGRVEVHSGEAWIPVSDGNFTLPTAQVCAELGCGKAVSVLGHMPPRE	
CC6H.....S.....R.....V.S.ME.AP.....A.....I.....H.Y.ERM.....H.....Q.....Q.RGR.....G.E.....H.V.E.....Y.....D.I.....R.....E.....EL.....	
CC7A.....NG.....R.....D.S.MG.AH.....A.....R.L.NTLD.....K.....H.K.....Q.....V.....S.EN.....V.....V.D.....RP.....D.....E.....EL.....	
CC12A.....G.....V.N.MQ.AE.....Q.....I.....Q.....G.M.H.....Q.....Q.RGR.D.G.E.....V.....E.....Q.....D.....	
WC1-8Y.S.....K.....S.....K.T.N.H.....R.....GK.Y.D.SR.....H.....I.....V.....A.....	
CCnd2Y.S.....K.....S.....K.T.N.H.....R.....GK.Y.D.SR.....H.....I.....V.....A.....	
WC1-2N.....S.....R.....K.....L.S.A.....N.R.T.K.H.....RD.GK.Y.D.S.....L.....I.....V.....	
WC1-1N.....S.....R.....K.....L.S.A.....N.R.T.K.H.....RD.GK.Y.D.S.....L.....I.....V.....	
WC1-13N.....S.....R.....K.....L.S.A.....N.R.T.K.H.....RD.GK.Y.D.S.....L.....I.....V.....	
CCnd1H.....R.....M.....N.R.T.K.H.....RD.GK.Y.D.S.....L.....I.....V.....	

Figure 5 Complete sequences of cDNA clones CC6 (*WC1-6*), CC7 (*WC1-7*), CC12 (*WC1-12*), CCnd1 (*WC1-nd1*) and CCnd2 (*WC1-nd2*). Deduced amino acids sequence from the coding sequences of *WC1-1*, *WC1-2*, *WC1-6*, and *WC1-13*, and the archetypal WC1 (Wc1.1) sequence were aligned using ClustalW2 and the default parameters and refined manually. GenBank accession numbers for amino acid sequences used for comparison are described in Materials and Methods. Identities are indicated by dots (.), gaps resulting from the alignment are indicated by tildes (~), gaps resulting from lack of genomic sequence (when the gaps were found adjacent and not within a coding region) are indicated by dashes (-). SRCR domains are indicated in Roman numerals and the transmembrane region is shown underlined for archetypal WC1 sequence. Continued in Figure 6.

Table 3 cDNA evidence for transcription of WC1-6, WC1-7, WC1-12, WC1-nd1 and WC1-nd2^a

Gene Name	cDNA clone ^b				
	CC6	CC7	CC12	CCnd1	CCnd2
WC1-1	94	62	71	96	90
WC1-2 ^c	88	63	65	97	91
WC1-3	93	63	69	93	91
WC1-4	64	96	61	65	62
WC1-5	93	66	70	95	93
WC1-6	99	66	68	92	90
WC1-7	66	98	62	65	62
WC1-8 ^c	90	62	66	92	97
WC1-9	64	96	62	63	61
WC1-10	65	58	74	66	66
WC1-11	70	62	72	67	67
WC1-12	68	62	98	70	66
WC1-13	97	65	70	94	91

a. Percent identity is shown for the deduced amino acid sequences of WC1 genes and cDNA clones.

b. Bold values indicate the highest level of identity and thus classification of the cDNA clone as a particular WC1 gene.

c. The annotation of these genes (WC1-2 and WC1-8) was found to be partial due to gaps in the genome assembly but shown here to be represented by WC1-nd1 and WC1-nd2, respectively.

and intracytoplasmic tail sequences corresponded to the annotated WC1-2 sequence and our evidence that only thirteen WC1 genes exist.

WC1 genes in other breeds of cattle

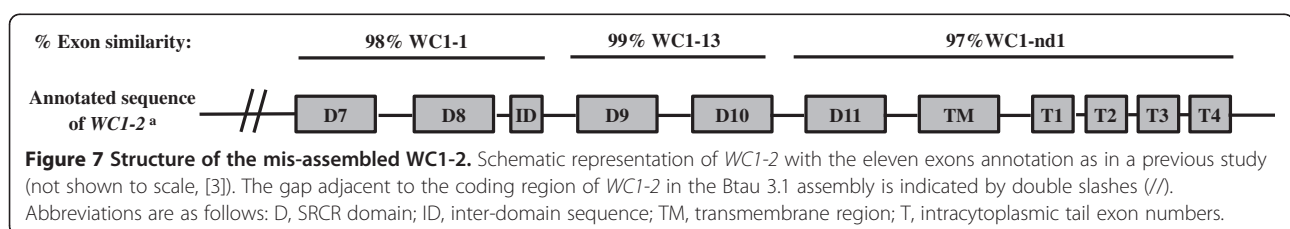
To determine whether the same thirteen known WC1 domain 1 gene sequences are conserved in other breeds, genomic DNA and cDNA from two other breeds of cattle (Belted Galloway and Holstein) were amplified using the thirteen WC1 gene-specific primers. Each animal showed products corresponding to all thirteen WC1 domain 1 sequences (Figure 8). To determine sequence similarity among individuals and breeds for the thirteen domain 1 sequences we used domain 1 common forward and reverse primers (Figure 1) to amplify genomic DNA and cDNA from three different breeds (Hereford, Belted Galloway, and Holstein). A total of 160 clones were sequenced until a complete set of thirteen different domain 1 sequences for each breed was obtained. The deduced amino acid sequences were aligned and showed

complete identity in the majority of cases and a maximum of three amino acid differences in the least similar (Figure 9). Those amino acid differences among animals of different breeds (such as the WC1-13 domain 1 genomic sequences from Hereford and Belted Galloway) suggest that a modest number of individual or breed-related polymorphisms are present (Figure 9). Allelic polymorphisms were also found since an individual animal had two different transcript sequences in some instances (such as the WC1-1 derived from a Belted Galloway) (Figure 9).

A phylogram generated based on deduced amino acid domain 1 sequences (Figure 10) further confirmed that each of the domain 1 sequences clustered with its corresponding annotated sequence from the reference animal. Of 160 sequences from three breeds of cattle, no additional sequences beyond the thirteen described WC1 genes were found. Thus we conclude that polymorphisms of WC1 genes are rare in *Bos taurus*. Moreover, those gene sequences derived previously from *Bos indicus* cattle (i.e. WC1-3, previously WC1.1, and WC1-4, previously WC1.2) were also conserved [7].

Discussion

The complexity of the WC1 multi-gene family in cattle has been resolved in this study: thirteen functional genes were found associated with ten different breeds of animals. The question of gene number variation among cattle was addressed by adapting Q-PCR for quantification. While the result was consistent with our previous WC1 gene annotation undertaken as part of the Bovine Genome Sequencing and Annotation Consortium [3], errors existed in the annotation due to incomplete or mis-assembly of the genome and those were corrected herein. The confirmation of thirteen WC1 genes corresponds reasonably well to the estimate derived by Southern blot, which suggested nineteen genes [14], and another study from our group suggesting thirteen genes based on the number of unique intracytoplasmic tail transcripts obtained [15]. However, it is fewer than the fifty WC1 genes predicted for sheep by Southern blotting [12,13]. Recently, we obtained evidence that sheep have twice the number of WC1 genes as cattle (Kim, Chen and Baldwin, unpublished data). Sequences of SRCR domain 1, the most divergent among the WC1 domains, showed that the domain 1 sequence for an



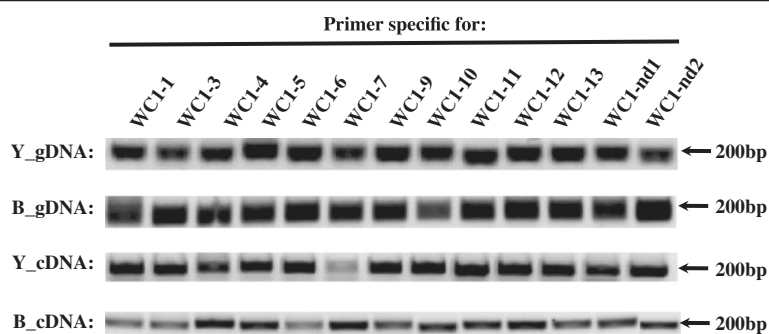


Figure 8 Genomic DNA and cDNA evidence for the 13 WC1 genes in two bovine breeds. PCR analysis was conducted by using genomic DNA and cDNA derived from cattle of two different breeds (Belted Galloway, designed with prefixes of "Y_gDNA and Y_cDNA", respectively; Holstein, designed with prefixes of "B_gDNA and B_cDNA", respectively). Primer pairs for distinguishing thirteen WC1 domain 1s were thirteen WC1 specific primer sets. For each primer set, the identities of the amplified products were confirmed by DNA sequencing analysis.

individual gene is highly conserved among breeds, with zero to three amino acids differences found per gene. Despite these differences, phylograms confirmed that the evolutionary divergence between individual WC1 genes was still greater than the divergence among animals for a particular gene. This suggests that the array of WC1 genes has been conserved for diverse functions. Also, we now conclude that there are three distinct WC1 molecular forms based on variation in the number of extracellular domains and intracytoplasmic tail sequences including their signaling motifs (Figure 11). These differences in the molecular structure of members of this multi-gene family have implications regarding ligand binding capacity and its signaling outcomes, which would be consistent among animals.

The conservation of WC1 gene sequences among animals and the number of family members is similar to those characteristics of other pattern recognition receptor (PRR) families. It has been proposed that under natural selection pressure, closely related non-rearranging immunoreceptors found on lymphocytes and antigen-presenting cells diversify in response to multiple ligands, such as bacterial and viral pathogen-associated molecular patterns (PAMPs) [38,39]. PRR's that recognize PAMPs include Toll-like receptors (TLRs) and the functionally similar but structurally distinct NOD-like receptors (NLRs). Individual TLRs and NLRs specifically recognize individual PAMPs, but also act together to recognize diverse microorganisms, initiating a range of host defense mechanisms [40,41]. The TLR family consists of 10 functional genes in humans [42], 12 in mice and ten in cattle [38] while NLRs [39] have 22 genes in humans and 34 in mice [43]. Two other multi-gene families expressed on NK and $\gamma\delta$ T cells are the C-type lectin-like Ly49 family [44-46], which is encoded by 15 functional genes in mice [47,48] but only a single related gene in humans and cattle [49,50], and the killer-IG-like receptor (KIR) family [50-52] which underwent rapid

repeated gene duplication in humans and cattle and has 4-14 genes depending upon the individual [47,50,51,53,54]. The ligands for Ly49 and KIR are comprised of a large family [50,55], including MHC class I-related molecules, that are rapidly evolving to evade the immune system. For example, infection of mice with murine cytomegalovirus (MCMV) caused the outgrowth of MCMV mutants which allowed the virus to escape recognition by the activating NK-cell receptor Ly49H [56].

Thus, we hypothesize that the WC1 family also expanded to keep pace with immune challenges from multiple pathogenic microorganisms and may be particularly important to $\gamma\delta$ T cells given that the TCR γ gene usage of WC1⁺ cells is restricted [11]. Evidence to support this comes from our and other's studies showing that the expression of particular WC1 molecules defines subpopulations of bovine WC1⁺ $\gamma\delta$ T cells that differ in their response to pathogens [9,37] and irradiated/stressed autologous monocytes [9]. In addition, shRNA-mediated selective reduction of WC1 expression by $\gamma\delta$ T cells decreases $\gamma\delta$ T cell response to *Leptospira*, supporting the hypothesis that WC1 proteins function as PRRs [29]. Moreover, some members of the SRCR superfamily have been shown to bind PAMPs via interactions with one or multiple SRCR domains. That is, the group B SRCR molecules CRP-ductin, Sp α and CD6 specifically bind to the bacterial products lipoteichoic acid (LTA) and lipopolysaccharide (LPS) [57-59] and DMBT1 binds to selected bacteria through a RVEVLxxxW motif in most of its SRCR domains [60]. Recently, we have localized *Leptospira*-binding activity to five of the eleven individual SRCR domains of specific WC1 molecules (Hsu and Telfer, unpublished data).

With regard to correcting errors in the previous assembly and annotation, here we found that *WC1-nd1* and *WC1-nd2*, the two WC1 transcripts that did not correspond to sequences in the Btau_3.1 genome assembly in our previous study [3], are indeed present in the

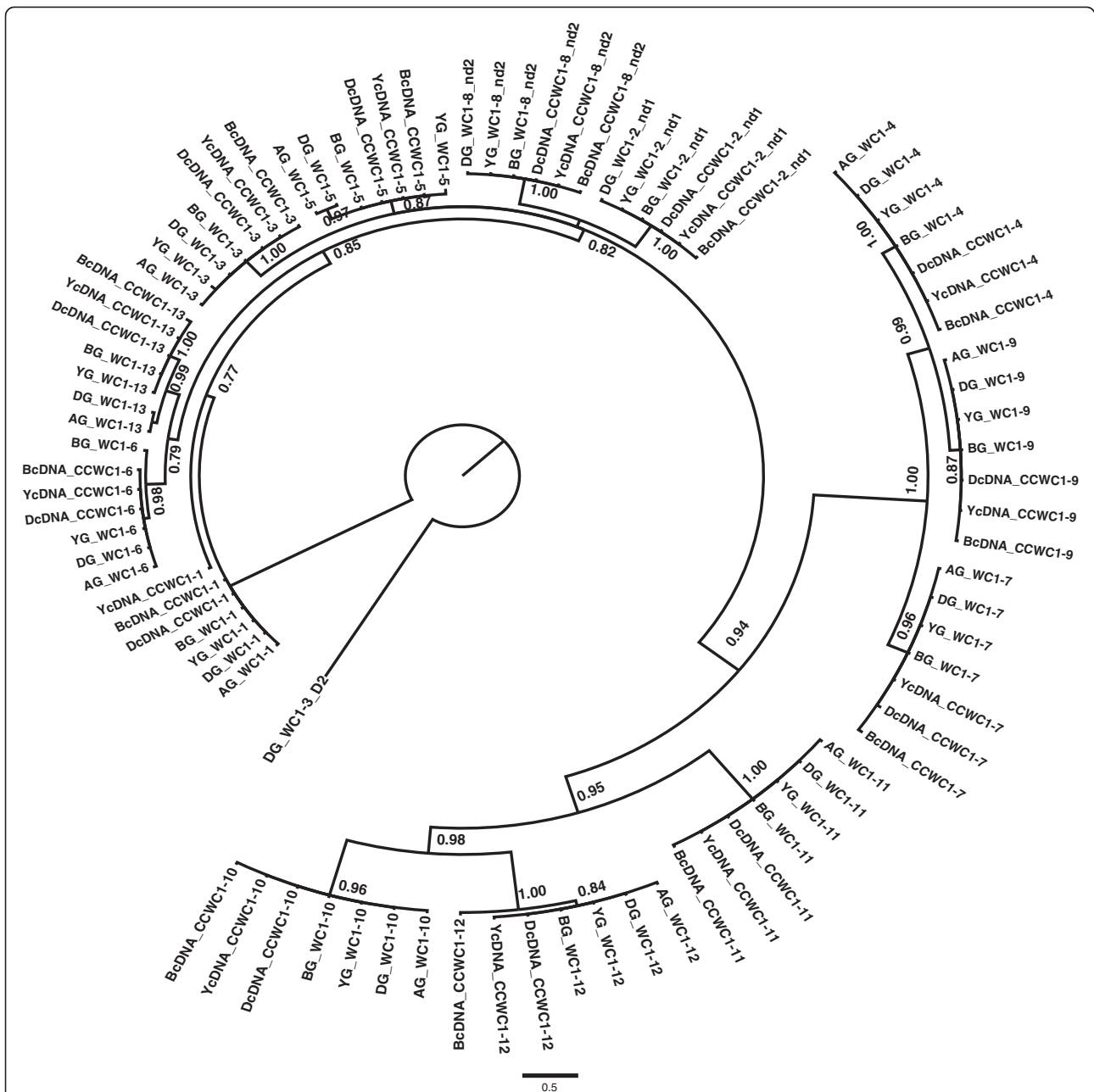


Figure 10 Phylogenetic tree of WC1 domain 1 sequences. The evolutionary history of 90 taxa was inferred using Bayesian analysis in MrBayes3.2 [33]. Annotated sequences were designed with a prefix of "A". Sequences obtained from individual animal of three breeds (Hereford, Belted Galloway, and Holstein) were designed with prefixes of "D", "Y", and "B", respectively. Genomic DNA and transcripts sequences were designed with prefixes of "G" and "cDNA", respectively. Markov chain Monte Carlo analysis was performed for 830,000 cycles, using 2 runs of 4 chains each, a temperature setting of 0.2, and an amino acid mixed model to approximate the posterior probabilities of trees, shown at branch nodes. The average standard deviation of split frequencies was 0.01, which was diagnostic of convergence at <0.05. Bootstrap values >70 are shown.

genome and are transcribed by Dominette. The inability to identify corresponding genomic sequence for *WC1-nd2* resulted from gaps in the assembly. That is, *WC1-8* was a partial sequence with no WC1 domain 1 sequence available, but we show here that it corresponds to *WC1-nd2* by analyzing the complete transcript sequence. The

second major error was regarding a gene previously annotated as *WC1-2*, which was found to be a concatenamer of mis-assembled SRCR domains corresponding to domains of *WC1-1*, *WC1-13*, and *WC1-nd1*. Using genomic DNA and cDNA from Dominette, we found that the most membrane-proximal SRCR domain, the

transmembrane region and the intracytoplasmic tail sequences of the previously annotated *WC1-2* corresponded to our unplaced to *WC1-nd1* sequence. Thus, *WC1-nd1* has been assigned as *WC1-2*, completing the panel of thirteen complete coding sequences for WC1 genes. In an attempt to further confirm our conclusions, we searched for WC1 sequences in the more recently released assemblies Btau_4.0 and UMD3 but found them to be less informative. WC1 coreceptors are unique to T cells of “ $\gamma\delta$ T cells high” species [2] including cattle [14] but not found for “ $\gamma\delta$ T cells low” species, such as human or mice [2]. Thus, the gaps regarding the WC1 coding region in assemblies Btau_3.1, Btau_4.0 and UMD3 may be a consequence of the absence in the human genome which was used for scaffolding the bovine genome.

WC1 gene products classified as Type I and II have eleven extracellular SRCR domains organized in the domain pattern of a-[b-c-d-e-d]-[b-c-d-e-d'] [3] (Figure 11) in comparison to the one Type III gene (*WC1-11*) which codes for six extracellular SRCR domains (a-[b-c-d-e]-[d']) and is closely related to the swine WC1 gene [3]. The complete sequences of the two new WC1 genes indicates that they can be classified as Type I WC1 genes based on their eleven extracellular domains and intracytoplasmic tail sequences coded for by four exons [3]. Thus, the Type I group is the largest with nine WC1 members. WC1 molecules that contain eleven extracellular domains with two repeating cassettes of highly related (domains b-c-d-e-d) may be advantageous for ligand binding. It is possible that the shorter molecule, *WC1-11*, represents an ancestral form although it is interesting that the alternatively spliced variants of the eleven-domain WC1 molecules [3] makes them similar in size to *WC1-11* and that these shorter splice variants are found associated with stimulated cells (Chen et al., unpublished data). This perhaps represents a regulatory mechanism for dampening the response to pathogens. Alternatively, the shorter forms being more similar in size as the TCR may function to co-bind pathogens more efficiently since $\gamma\delta$ TCRs do not see processed antigens but are able to interact with pathogens in a manner more akin to that of antibodies [61]. It will be important to develop an understanding of the functionality of those alternative splice forms and their affects on the immune response of $\gamma\delta$ T cells in future studies. Finally, it is notable that in all alternative spliced transcripts the extracellular SRCR domain apposed to the membrane is always d' as it is in full-length molecules, suggesting that the d' may be structural. It might facilitate co-clustering of WC1 molecules with the TCR since we have shown that both associate with lipid rafts following cell activation [15].

Differences in the intracytoplasmic tails likely play an important role in signal transduction. Type II WC1

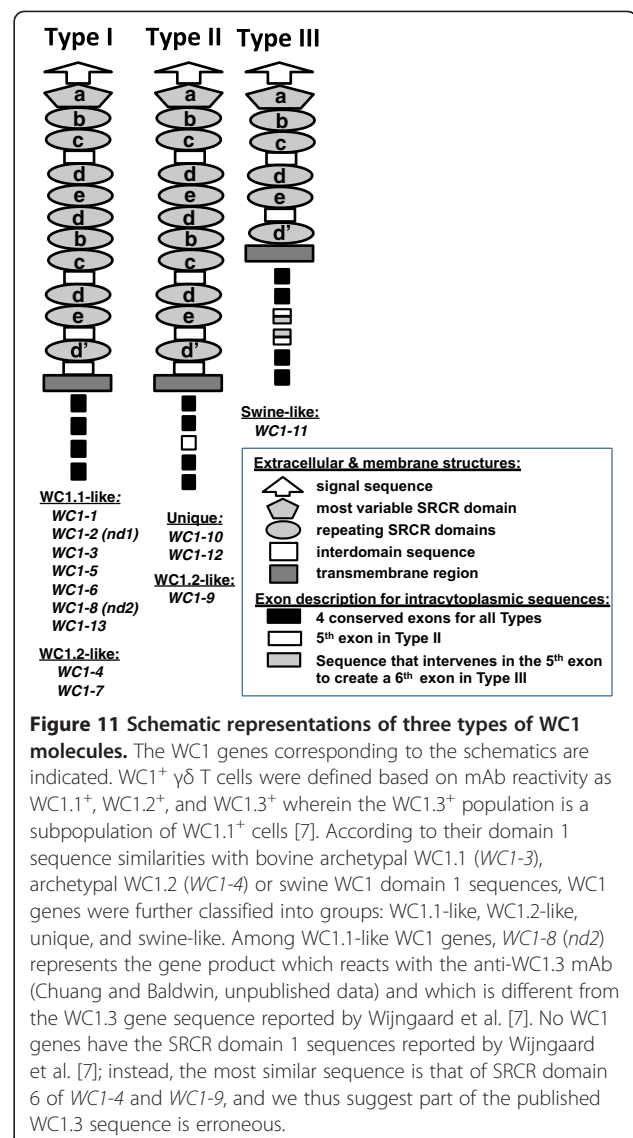


Figure 11 Schematic representations of three types of WC1 molecules. The WC1 genes corresponding to the schematics are indicated. WC1⁺ $\gamma\delta$ T cells were defined based on mAb reactivity as WC1.1⁺, WC1.2⁺, and WC1.3⁺ wherein the WC1.3⁺ population is a subpopulation of WC1.1⁺ cells [7]. According to their domain 1 sequence similarities with bovine archetypal WC1.1 (*WC1-3*), archetypal WC1.2 (*WC1-4*) or swine WC1 domain 1 sequences, WC1 genes were further classified into groups: WC1.1-like, WC1.2-like, unique, and swine-like. Among WC1.1-like WC1 genes, *WC1-8* (*nd2*) represents the gene product which reacts with the anti-WC1.3 mAb (Chuang and Baldwin, unpublished data) and which is different from the WC1.3 gene sequence reported by Wijngaard et al. [7]. No WC1 genes have the SRCR domain 1 sequences reported by Wijngaard et al. [7]; instead, the most similar sequence is that of SRCR domain 6 of *WC1-4* and *WC1-9*, and we thus suggest part of the published WC1.3 sequence is erroneous.

molecules have a “long tail” molecular form, with fifteen or more amino acids encoded by an additional (5th) exon [3] (Figure 11). Type III contains a very long intracytoplasmic domain resulting from a 6th exon coding for amino acids inserted into the sequence coded for by the middle exon (the 5th exon) of Type II WC1 genes [3] (Figure 11). Short and long tails are also found with other immunoreceptor families: KIR and NKG2D. Activating KIRs have short cytoplasmic tails with ITAMs that pair with DAP12/KARAP; inhibitory KIRs possess long cytoplasmic tails with ITIM motifs [62]. NKG2D long form tails associate with DAP10 [63], while the short form [64] associates with DAP10 or DAP12. The adaptor determines the outcome of signaling following ligand binding [63,64]. A signaling role for the most common WC1 tail sequence, which is the shortest, is shown by the requirement for phosphorylation of the

second tyrosine for transmission of signaling through the TCR [4]. It is notable that three gene products, *WC1-4*, *WC1-7*, and *WC1-9*, all have highly similar extracellular domains, possibly recognizing the same ligands, but the intracytoplasmic tails of *WC1-4* and *WC1-7* are archetypal (or short) while the tail of *WC1-9* is longer as illustrated in Figure 11. This may indicate that cells bearing *WC1-4* or *WC1-7* vs. *WC1-9* have different functional outcomes even if they bind the same ligands consistent with the paired receptor hypothesis for KIR molecules [62]. The signaling role for the other intracytoplasmic sequences of *WC1* molecules is under investigation.

Conclusion

Using Q-PCR to quantitate gene number, we showed that the *WC1* immunoreceptor family comprises thirteen genes in the bovine genome, without variation in number among ten cattle breeds tested. Moreover, conservation of sequences for the thirteen *WC1* genes existed among breeds. We found that all thirteen *WC1* molecules fit into the three distinct molecular forms we previously described. While it has already been shown that functionally distinct subpopulations of bovine *WC1*⁺ $\gamma\delta$ T cells can be defined by the expression of particular *WC1* molecules, future studies need to address the significant questions of the signaling potential of each type of *WC1* molecule in $\gamma\delta$ T cell responses and the identification of ligand-binding domains in the various *WC1* molecules. *WC1* co-receptors on $\gamma\delta$ T cells may be a type of PRRs on nonconventional T cells that participate with the TCR for maximal cell activation. Understanding the mechanism of activation of nonconventional $\gamma\delta$ T cells that serve to bridge between innate and adaptive immune response might be exploited for efficacious vaccine design to improve human and domesticated animal health.

Authors' contributions

CC carried out the molecular studies, the sequence analyses and drafted the manuscript. CH helped to provide annotation data and participated in the design of the study. AL provided the genomic DNA and cDNA of the reference animal Dominette and revised the manuscript. KJ and MT provided genomic DNA from seven additional breeds of cattle and participated in data interpretation and manuscript editing. JT and CB participated in the design of the study, interpretation of data, securing funding for the study and helped draft the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was conducted as part of the Bovine Genome Annotation Sequencing Consortium and supported by CSREES USDA-NRI grant #2006-01691, 2005-01812 and CSREES-USDA Massachusetts Agricultural Experiment Station under project No.MA0209204, and by Agriculture and Food Research Initiative Competitive Grant no. 2011-67015-30736 from the National Institute of Food and Agriculture USDA and NIH (R01 HD070056-01) program titled Dual Purpose with Dual Benefit: Research in Biomedicine and Agriculture using Agriculturally Important Domestic Species. Mention of trade names or commercial products is solely for the purpose of providing information and does not imply recommendation, endorsement or exclusion of other suitable products by the U.S. Department of Agriculture.

Author details

¹Department of Veterinary and Animal Sciences, University of Massachusetts, Amherst, MA 01003, USA. ²Department of Agriculture, Fort Keogh Livestock & Range Research Laboratory, USDA-ARS Fort Keogh LARRL, Miles City, MT 59301, USA. ³USDA-ARS, U. S. Meat Animal Research Center, Clay Center, NE 68933, USA. ⁴Current Address: Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA.

Received: 28 December 2011 Accepted: 3 October 2012

Published: 16 October 2012

References

1. Sarrías MR, Gronlund J, Padilla O, Madsen J, Holmskov U, Lozano F: The Scavenger Receptor Cysteine-Rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. *Crit Rev Immunol* 2004, **24**(1):1–37.
2. Herzig CT, Waters RW, Baldwin CL, Telfer JC: Evolution of the CD163 family and its relationship to the bovine gamma delta T cell co-receptor *WC1*. *BMC Evol Biol* 2010, **10**:181.
3. Herzig CT, Baldwin CL: Genomic organization and classification of the bovine *WC1* genes and expression by peripheral blood gamma delta T cells. *BMC Genomics* 2009, **10**:191.
4. Wang F, Herzig C, Ozer D, Baldwin CL, Telfer JC: Tyrosine phosphorylation of scavenger receptor cysteine-rich *WC1* is required for the *WC1*-mediated potentiation of TCR-induced T-cell proliferation. *Eur J Immunol* 2009, **39**(1):254–266.
5. Holm D, Fink DR, Gronlund J, Hansen S, Holmskov U: Cloning and characterization of SCART1, a novel scavenger receptor cysteine-rich type I transmembrane molecule. *Mol Immunol* 2009, **46**(8–9):1663–1672.
6. Kisielow J, Kopf M, Karjalainen K: SCART scavenger receptors identify a novel subset of adult gamma delta T cells. *J Immunol* 2008, **181**(3):1710–1716.
7. Wijngaard PL, MacHugh ND, Metzelaar MJ, Romberg S, Bensaid A, Pepin L, Davis WC, Clevers HC: Members of the novel *WC1* gene family are differentially expressed on subsets of bovine CD4-CD8- gamma delta T lymphocytes. *J Immunol* 1994, **152**(7):3476–3482.
8. Rogers AN, VanBuren DG, Hedblom E, Tilahun ME, Telfer JC, Baldwin CL: Function of ruminant gamma delta T cells is defined by *WC1.1* or *WC1.2* isoform expression. *Vet Immunol Immunopathol* 2005, **108**(1–2):211–217.
9. Rogers AN, Vanburen DG, Hedblom EE, Tilahun ME, Telfer JC, Baldwin CL: Gamma delta T cell function varies with the expressed *WC1* coreceptor. *J Immunol* 2005, **174**(6):3386–3393.
10. Lahmers KK, Hedges JF, Jutila MA, Deng M, Abrahamsen MS, Brown WC: Comparative gene expression by *WC1*⁺ gamma delta and CD4⁺ alpha beta T lymphocytes, which respond to *Anaplasma marginale*, demonstrates higher expression of chemokines and other myeloid cell-associated genes by *WC1*⁺ gamma delta T cells. *J Leukoc Biol* 2006, **80**(4):939–952.
11. Blumerman SL, Herzig CT, Rogers AN, Telfer JC, Baldwin CL: Differential TCR gene usage between *WC1*⁻ and *WC1*⁺ ruminant gamma delta T cell subpopulations including those responding to bacterial antigen. *Immunogenetics* 2006, **58**(8):680–692.
12. O'Keefe MA, Metcalfe SA, Glew MD, Bowden T, McInnes S, Kimpton WG, Cahill RN, Hein WR, Walker ID: Lymph node homing cells biologically enriched for gamma delta T cells express multiple genes from the T19 repertoire. *Int Immunol* 1994, **6**(11):1687–1697.
13. Walker ID, Glew MD, O'Keefe MA, Metcalfe SA, Clevers HC, Wijngaard PL, Adams TE, Hein WR: A novel multi-gene family of sheep gamma delta T cells. *Immunology* 1994, **83**(4):517–523.
14. Wijngaard PL, Metzelaar MJ, MacHugh ND, Morrison WI, Clevers HC: Molecular characterization of the *WC1* antigen expressed specifically on bovine CD4-CD8- gamma delta T lymphocytes. *J Immunol* 1992, **149**(10):3273–3277.
15. Rogers AN, Vanburen DG, Zou B, Lahmers KK, Herzig CT, Brown WC, Telfer JC, Baldwin CL: Characterization of *WC1* co-receptors on functionally distinct subpopulations of ruminant gamma delta T cells. *Cell Immunol* 2006, **239**(2):151–161.
16. Heid CA, Stevens J, Livak KJ, Williams PM: Real time quantitative PCR. *Genome Res* 1996, **6**(10):986–994.

17. Higuchi R, Fockler C, Dollinger G, Watson R: **Kinetic PCR analysis: real-time monitoring of DNA amplification reactions.** *Biotechnology (N Y)* 1993, **11**(9):1026–1030.
18. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic Acids Res* 2001, **29**(9):e45.
19. Hilscher C, Vahrson W, Dittmer DP: **Faster quantitative real-time PCR protocols may lose sensitivity and show increased variability.** *Nucleic Acids Res* 2005, **33**(21):e182.
20. Larionov A, Krause A, Miller W: **A standard curve based method for relative real time PCR data processing.** *BMC Bioinforma* 2005, **6**:62.
21. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 2001, **25**(4):402–408.
22. Ferreira ID, Rosario VE, Cravo PV: **Real-time quantitative PCR with SYBR Green I detection for estimating copy numbers of nine drug resistance candidate genes in Plasmodium falciparum.** *Malar J* 2006, **5**:1.
23. Bieche I, Olivi M, Champeme MH, Vidaud D, Lidereau R, Vidaud M: **Novel approach to quantitative polymerase chain reaction using real-time detection: application to the detection of gene amplification in breast cancer.** *Int J Cancer* 1998, **78**(5):661–666.
24. Klein D: **Quantification using real-time PCR technology: applications and limitations.** *Trends Mol Med* 2002, **8**(6):257–260.
25. Liu GE, Van Tassel CP, Sonstegard TS, Li RW, Alexander LJ, Keele JW, Matukumalli LK, Smith TP, Gasbarre LC: **Detection of germline and somatic copy number variations in cattle.** *Dev Biol (Basel)* 2008, **132**:231–237.
26. Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, Adelson DL, Eichler EE, Elnitski L, Guigo R, et al: **The genome sequence of taurine cattle: a window to ruminant biology and evolution.** *Science* 2009, **324**(5926):522–528.
27. Goossens K, Van Poucke M, Van Soom A, Vandesompele J, Van Zeveren A, Peelman LJ: **Selection of reference genes for quantitative real-time PCR in bovine preimplantation embryos.** *BMC Dev Biol* 2005, **5**:27.
28. Herzig CT, Lefranc MP, Baldwin CL: **Annotation and classification of the bovine T cell receptor delta genes.** *BMC Genomics* 2010, **11**:100.
29. Wang F, Herzig CT, Chen C, Hsu H, Baldwin CL, Telfer JC: **Scavenger receptor WC1 contributes to the gamma delta T cell response to Leptospira.** *Mol Immunol* 2011, **48**(6–7):801–809.
30. Yu Y, Lee C, Kim J, Hwang S: **Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction.** *Biotechnol Bioeng* 2005, **89**(6):670–679.
31. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999, **41**:95–98.
32. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD: **Multiple sequence alignment with the Clustal series of programs.** *Nucleic Acids Res* 2003, **31**(13):3497–3500.
33. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**(12):1572–1574.
34. Walker AM, Roberts RM: **Characterization of the bovine type I IFN locus: rearrangements, expansions, and novel subfamilies.** *BMC Genomics* 2009, **10**:187.
35. Ririe KM, Rasmussen RP, Wittwer CT: **Product differentiation by analysis of DNA melting curves during the polymerase chain reaction.** *Anal Biochem* 1997, **245**(2):154–160.
36. Nystrom K, Biller M, Grahm A, Lindh M, Larson G, Olofsson S: **Real time PCR for monitoring regulation of host gene expression in herpes simplex virus type 1-infected human diploid cells.** *J Virol Methods* 2004, **118**(2):83–94.
37. Lahmers KK, Norimine J, Abrahamsen MS, Palmer GH, Brown WC: **The CD4+ T cell immunodominant Anaplasma marginale major surface protein 2 stimulates gamma delta T cell clones that express unique T cell receptors.** *J Leukoc Biol* 2005, **77**(2):199–208.
38. Leulier F, Lemaître B: **Toll-like receptors—taking an evolutionary approach.** *Nat Rev Genet* 2008, **9**(3):165–178.
39. Chen G, Shaw MH, Kim YG, Nunez G: **NOD-like receptors: role in innate immunity and inflammatory disease.** *Annu Rev Pathol* 2009, **4**:365–398.
40. Medzhitov R, Janeway C Jr: **The Toll receptor family and microbial recognition.** *Trends Microbiol* 2000, **8**(10):452–456.
41. Aderem A: **Role of Toll-like receptors in inflammatory response in macrophages.** *Crit Care Med* 2001, **29**(7 Suppl):S16–S18.
42. Poulain-Godefroy O, Le Bacquer O, Plançq P, Lecoœur C, Pattou F, Fruhbeck G, Froguel P: **Inflammatory role of Toll-like receptors in human and murine adipose tissue.** *Mediators Inflamm* 2010, **2010**:823486.
43. Villaseñor-Cardoso MI, Ortega E: **Polymorphisms of innate immunity receptors in infection by parasites.** *Parasite Immunol* 2011, **33**(12):643–653.
44. Hara T, Nishimura H, Hasegawa Y, Yoshikai Y: **Thymus-dependent modulation of Ly49 inhibitory receptor expression on NK1.1+gamma/delta T cells.** *Immunology* 2001, **102**(1):24–30.
45. Held W, Roland J, Raulet DH: **Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells.** *Nature* 1995, **376**(6538):355–358.
46. Ortaldo JR, Winkler-Pickett R, Mason AT, Mason LH: **The Ly-49 family: regulation of cytotoxicity and cytokine production in murine CD3+ cells.** *J Immunol* 1998, **160**(3):1158–1165.
47. Hao L, Nei M: **Genomic organization and evolutionary analysis of Ly49 genes encoding the rodent natural killer cell receptors: rapid evolution by repeated gene duplication.** *Immunogenetics* 2004, **56**(5):343–354.
48. Wilhelm BT, Mager DL: **Rapid expansion of the Ly49 gene cluster in rat.** *Genomics* 2004, **84**(1):218–221.
49. Westgaard IH, Berg SF, Orstavik S, Fossum S, Dissen E: **Identification of a human member of the Ly-49 multigene family.** *Eur J Immunol* 1998, **28**(6):1839–1846.
50. McQueen KL, Wilhelm BT, Harden KD, Mager DL: **Evolution of NK receptors: a single Ly49 and multiple KIR genes in the cow.** *Eur J Immunol* 2002, **32**(3):810–817.
51. Martin AM, Freitas EM, Witt CS, Christiansen FT: **The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster.** *Immunogenetics* 2000, **51**(4–5):268–280.
52. Wilson MJ, Torkar M, Haude A, Milne S, Jones T, Sheer D, Beck S, Trowsdale J: **Plasticity in the organization and sequences of human KIR/ILT gene families.** *Proc Natl Acad Sci U S A* 2000, **97**(9):4778–4783.
53. Guethlein LA, Flodin LR, Adams EJ, Parham P: **NK cell receptors of the orangutan (Pongo pygmaeus): a pivotal species for tracking the coevolution of killer cell Ig-like receptors with MHC-C.** *J Immunol* 2002, **169**(1):220–229.
54. Hershberger KL, Shyam R, Miura A, Letvin NL: **Diversity of the killer cell Ig-like receptors of rhesus monkeys.** *J Immunol* 2001, **166**(7):4380–4390.
55. Gagnier L, Wilhelm BT, Mager DL: **Ly49 genes in non-rodent mammals.** *Immunogenetics* 2003, **55**(2):109–115.
56. French AR, Pingel JT, Wagner M, Bubic I, Yang L, Kim S, Koszinowski U, Jonjic S, Yokoyama WM: **Escape of mutant double-stranded DNA virus from innate immune control.** *Immunity* 2004, **20**(6):747–756.
57. Madsen J, Tornøe I, Nielsen O, Lausen M, Krebs I, Mollenhauer J, Kollender G, Poustka A, Skjodt K, Holmskov U: **CRP-ductin, the mouse homologue of gp-340/deleted in malignant brain tumors 1 (DMBT1), binds gram-positive and gram-negative bacteria and interacts with lung surfactant protein D.** *Eur J Immunol* 2003, **33**(8):2327–2336.
58. Sarrias MR, Rosello S, Sanchez-Barbero F, Sierra JM, Vila J, Yelamos J, Vives J, Casals C, Lozano F: **A role for human Sp alpha as a pattern recognition receptor.** *J Biol Chem* 2005, **280**(42):35391–35398.
59. Sarrias MR, Farnos M, Mota R, Sanchez-Barbero F, Ibanez A, Gimferrer I, Vera J, Fenutria R, Casals C, Yelamos J, et al: **CD6 binds to pathogen-associated molecular patterns and protects from LPS-induced septic shock.** *Proc Natl Acad Sci U S A* 2007, **104**(28):11724–11729.
60. Bikker FJ, Ligtenberg AJ, End C, Renner M, Blaich S, Lyer S, Wittig R, van't Hof W, Veerman EC, Nazmi K, et al: **Bacteria binding by DMBT1/SAG/gp-340 is confined to the VEVLXXXXW motif in its scavenger receptor cysteine-rich domains.** *J Biol Chem* 2004, **279**(46):47699–47703.
61. Adams EJ, Strop P, Shin S, Chien YH, Garcia KC: **An autonomous CDR3delta is sufficient for recognition of the nonclassical MHC class I molecules T10 and T22 by gammadelta T cells.** *Nat Immunol* 2008, **9**(7):777–784.
62. Parham P: **Killer cell immunoglobulin-like receptor diversity: balancing signals in the natural killer cell response.** *Immunol Lett* 2004, **92**(1–2):11–13.
63. Raulet DH: **Roles of the NKG2D immunoreceptor and its ligands.** *Nat Rev Immunol* 2003, **3**(10):781–790.
64. Eagle RA, Trowsdale J: **Promiscuity and the single receptor: NKG2D.** *Nat Rev Immunol* 2007, **7**(9):737–744.

doi:10.1186/1471-2156-13-86

Cite this article as: Chen et al: Gene number determination and genetic polymorphism of the gamma delta T cell co-receptor WC1 genes. *BMC Genetics* 2012 **13**:86.