Comparative organization of cattle chromosome 5 revealed by comparative mapping by annotation and sequence similarity and radiation hybrid mapping

Akihito Ozawa*, Mark R. Band*, Joshua H. Larson*, Jena Donovan*, Cheryl A. Green*, James E. Womack†, and Harris A. Lewin*‡

*Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801; and †Department of Veterinary Pathobiology, Texas A&M University, College Station, TX 77843

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 27, 1999.

Contributed by James E. Womack, January 7, 2000

A whole genome cattle-hamster radiation hybrid cell panel was used to construct a map of 54 markers located on bovine chromosome 5 (BTA5). Of the 54 markers, 34 are microsatellites selected from the cattle linkage map and 20 are genes. Among the 20 mapped genes, 10 are new assignments that were made by using the comparative mapping by annotation and sequence similarity strategy. A LOD-3 radiation hybrid framework map consisting of 21 markers was constructed. The relatively low retention frequency of markers on this chromosome (19%) prevented unambiguous ordering of the other 33 markers. The length of the map is 398.7 cR, corresponding to a ratio of ≈2.8 $cR_{5,000}$/cM. Type I genes were binned for comparison of gene order among cattle, humans, and mice. Multiple internal rearrangements within conserved syntenic groups were apparent upon comparison of gene order on BTA5 and HSA12 and HSA22. A similarly high number of rearrangements were observed between BTA5 and MMU6, MMU10, and MMU15. The detailed comparative map of BTA5 should facilitate identification of genes affecting economically important traits that have been mapped to this chromosome and should contribute to our understanding of mammalian chromosome evolution.

radiation hybrid | expressed sequence tag | BTA5 | HSA12

Comparative genomics has been a useful tool for deducing the rearrangements of chromosomes that accompany plant and animal evolution (1, 2). A dominant feature of comparative maps is the apparent conservation of chromosome segments among classes and orders within the crown group of eukaryotes. The available data suggests that the more closely related two species are, the greater the conservation of synteny (3). Although the discovery of conserved syntenies enables estimation of the minimum number of chromosome rearrangements separating any two species, the determination of gene order within conserved segments is necessary for a detailed understanding of chromosomal evolution (4). The identification of internal rearrangements within conserved syntenies is important for both fundamental and applied aspects of comparative mapping (5). For example, comparative maps are used to identify candidate genes for economically important traits in crops and livestock (6, 7). Without knowledge of gene order, the selection of candidate genes using comparative mapping data is prone to error.

A major obstacle to identifying candidate genes is the poor quality of comparative maps among mammals, particularly for farm animals. For example, only about 200 genes are on the ordered genetic map of cattle (8–10). Thus, the boundaries of conserved chromosomal segments are rather poorly defined, and very few genes are ordered reliably. The recent development of a whole genome cattle-hamster radiation hybrid (RH) panel (11) has fueled a new thrust in cattle genomics and comparative gene mapping (12–15). However, mapping on the cattle RH panel has not progressed rapidly because of limited homologous DNA sequence information. Recently, a novel strategy for high throughput comparative mapping was proposed that couples the power of RH mapping with expressed sequence tags (ESTs) and bioinformatics (16). This strategy was termed "comparative mapping by annotation and sequence similarity" (COMPASS). COMPASS permits the prediction of map location of a randomly chosen DNA sequence, provided there is detailed sequence and mapping data for a reference species and adequate existing comparative mapping information for the target species. The first step in the COMPASS strategy is the production of ESTs or sequence tagged sites. This is followed by the production of a comparative map *in silico*. The *in silico* map is produced by identifying a human ortholog in the public domain DNA databases using the BLAST algorithm (21), finding the map location of the human ortholog in UniGene (http://www.ncbi.nlm.nih.gov/UniGene/), and then predicting a chromosome assignment in the target species on the basis of human mapping data and available comparative mapping information. The *in silico* assignments can then be used on a chromosome-by-chromosome basis to augment the production of ordered RH maps.

Quantitative trait loci (QTL) affecting somatic cell score (17) and yield of milk fat (18) have recently been mapped to bovine chromosome 5 (BTA5). Although 46 Type I genes have been assigned to BTA5, the relative order of only five of these genes has been determined with a high degree of confidence. Thus, the low density of ordered Type I markers on BTA5 makes it difficult to identify a set of candidate genes for QTL located on this chromosome. The paucity of detailed mapping information for BTA5 and the physiologically important QTL mapped to this chromosome led us to select BTA5 for detailed comparative mapping using the combined COMPASS-RH mapping approach. Synteny on BTA5 is conserved with genes on human chromosome (HSA) 12 and HSA22 (19). In the present study, we used a dual strategy to construct a framework RH map of BTA5. First, a set of 34 microsatellite markers and 10 known genes was used to construct a framework RH map. Then, from among cattle EST and extant database sequences, COMPASS was used to select for RH mapping eight orthologs predicted to be located on BTA5. In total, 20 Type I loci were mapped by using the RH panel, including 10 new assignments, thus greatly improving the resolution of the BTA5 gene map. This map revealed previously unknown details of the cattle-human and mouse-human comparative maps.

Abbreviations: RH, radiation hybrid; EST, expressed sequence tag; COMPASS, comparative mapping by annotation and sequence similarity; QTL, quantitative trait loci.

Data deposition: The sequences reported in this article have been deposited in the GenBank database (accession nos. AF093736, AF093737, AF093738, AF093739, AF093740, AF093741, and AF093742).

‡To whom reprint requests should be addressed at: Department of Animal Sciences, University of Illinois at Urbana-Champaign, 1201 West Gregory Drive, Urbana, IL 61801. E-mail: h-lewin@ux1.cso.uiuc.edu.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.050007097. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.050007097

Table 1. Markers, marker retention frequencies, and placements

Marker (Alias)	Type	RF	Placement, cR			Reference
			Primary†	LOD‡	Secondary	
D5S64 (BM1095)	MS	0.26	0.0	F		30
D5S51 (ILSTS042)	MS	0.21	16.0	F		31
D5S11 (BM6026)	MS	0.22	29.4	F		32
MYF-5	Gene	0.26	37.9	**	55.1	Table 2
D5S12 (BP1)	MS	0.22	45.2	F		33
D5S72 (BMS610)	MS	0.26	54.2		39.2	30
OV0182	EST	0.21	65.6	**	72.9	Table 2
D5S14 (BL23)	MS	0.21	70.3	F		33
D5S66 (BMS1315)	MS	0.18	84.0	F		30
D5S79 (OARFCB5)	MS	0.21	92.1	F		34
PDE1B	Gene	0.18	96.0	***		Table 2
D5S000 (ILSTS022)	MS	0.23	102.5	F		31
D5S15 (BMC1009)	MS	0.24	105.0		99.9	33
KRT1B	Gene	0.24	105.0		99.9	Table 2
OV0251	EST	0.23	107.6	**	99.5	Table 2
D5S6 (CSSM34)	MS	0.21	112.8	F		32
COL2A1	Gene	0.19	120.5	*	108.2	Table 2
LALBA	Gene	0.17	121.7	***		Table 2
OV0190	EST	0.18	122.7	***		Table 2
LYZ1	Gene	0.21	133.0	*	157.7	Table 2
D5S80 (BL4)	MS	0.17	149.3	F		10
D5S17 (BL37)	MS	0.17	157.7		138.9	33
IFNG	Gene	0.13	157.8		137.9	Table 2
OV0373	EST	0.18	158.4		174.6	Table 2
D5S4 (RM500)	MS	0.17	166.2	F		32
D5S19 (BR2936)	MS	0.24	176.5	**	133.4	33
D5S20 (AGLA254)	MS	0.18	178.3		189.9	33
D5S3 (ETH10)	MS	0.20	181.2		188.0	32
OV0268	EST	0.22	185.0	F		Table 2
D5S52 (ILSTS066)	MS	0.20	195.1	F		31
IGF1	MS, Gene	0.14	215.2	F		33
D5S21 (BM1819)	MS	0.18	229.3	F		33
MG1	Gene	0.10	241.6	***		35
RHOD	Gene	0.11	242.3	**	265.7	Table 2
D5S57 (BMS1248)	MS	0.12	256.2	F		30
D5S75 (BM8230)	MS	0.13	261.5	**	246.3	30
M6PR	Gene	0.18	270.6		277.2	Table 2
D5S68 (BMS1658)	MS	0.09	271.9		276.2	30
D5S22 (BM315)	MS	0.09	271.9		276.2	33
D5S41 (URB52)	MS	0.10	274.1	F		8
D5S2 (ETH2)	MS	0.13	288.6	F		32
D5S74 (BMS772)	MS	0.13	292.8		284.4	30
CD9	Gene	0.13	292.8		284.4	Table 2
OV0097	EST	0.14	295.2	**	283.9	Table 2
D5S26 (BM49)	MS	0.18	301.3		310.0	33
D5S25 (BM2830)	MS	0.18	301.3		310.0	33
D5S23 (BM733)	MS	0.20	305.5	F		33
D5S27 (IDVGA9)	MS	0.24	327.5	F		36
ACO2	Gene	0.24	327.5			Table 2
D5S1 (ETH152)	MS	0.28	367.5	*	384.0	33
D5S71 (BMS597)	MS	0.28	373.2		378.7	30
D5S42 (URB60)	MS	0.27	376.1			8
D5S76 (BM8126)	MS	0.27	376.1	F		30
ACR	Gene	0.34	398.7	**	360.7	Table 2

MS, microsatellite marker; EST, expressed sequence tag; RF, retention frequeny.

†Distance in $cR_{5,000}$ from the most centromeric marker, *D5S64*. For markers placed with a LOD score <3, the next best placement is given as "secondary."

‡LOD score compared with secondary placement. (F, framework; three asterisks, LOD > 3; two asterisks, 3 > LOD > 2; one asterisk, 2 > LOD > 1; no asterisk, 1 > LOD.)

Materials and Methods

Selection of BTA5 Markers. Thirty-four microsatellites and 10 bovine genes that map to BTA5 were chosen from published linkage or physical maps as a scaffold to build the RH map (Table 1). These markers were selected for their relative spacing and to cover the entire length of the chromosome. Oligonucleotide primers were

Table 2. Primer sequences and product sizes of mapped markers

Symbol	Bovine gene, accession number	Primer, 5′-3′	Annealing T, °C	Ext. time, s	Product, bp
ACO2	Aconitate hydratase, mitochondrial precursor, Z49931	GCAGTTCCGTGTCACCAGA TGCATGAAGCTGCTCAAGAT	54	30	198
ACR	Acrosin, X68212	GCTCATAGAGGTCTTGAAGG GATCAGGAGGAGGCATGTTG	60	30	99
CD9	CD9 antigen, M81720	CGTGAAGACCGCTGGTTATT TTCATTGCAGGATTTCTGCTT	54	30	256
COL2A1	Collagen 2A1, L10347	CTCTGCGACGACATAATCTG TCTCCAGGTTCTCCTTTCTG	56	150	≈2,500
IFNG	Interferon gamma, Z54144	ATCATAACACAGGAGCTACC ATCCATGCTCCTTTGAATGACC	57	120	1,563
IGF1	Insulin-like growth factor 1, X64400	GCTTGGATGGACCATGTTG CACTTGAGGGGCAAATGATT	58	30	229
KRT1B	Cytokeratin (type II) component 1b/c, K03534	GCT CCTCGGTTCTCACCTCT CTGGATGCCTGAAGGACAAG	53	30	130
LALBA	Lactalbumin, U63110	GGGTTGAGTGGGCCATGACA GAGGATGACAAGAGAGGCCAG	58	30	367
LYZ1	Lysozyme 1, M95097	CTTGGGGCCAATTCTAGTGA CAAACAGCAGAAATCAGCCA	52	30	185
M6PR	Cation-dependent mannose-6-phosphate receptor, M17025	TTCCTCACTTGGCCTTCTGG GGTGAGATGAGAGACTGCTT	56	30	249
MG1	Myoglobin, D00409	ATCCTTCAGACTTCGGTGC TAAACCAGGTGTCACTCCCC	56	30	634
MYF5	Myogenic factor-5, X52526	GGATCGGATCACCAACTCAG TTTGGTACCTCCTTCCTCCT	53	30	197
OV0097	Similar to human B-cell receptor associated protein (BAP) mRNA, AF0993739	ACGAGCGAGGAATCAACACC CTCTGCATCGCTCCTCCCAAC	54	30	141
OV0182	Similar to human decorin (DCN) gene AF093737	TGTAGTTTCCAAGCTGAACGGCA TATCTCGGCAATCGGCTCTAACG	58	30	167
OV0190	Similar to human alpha-tubulin (TUBA) mRNA, AF093736	ATCTTTACTTTTGAAACAGG GCAGCATGTCATACTCAACT	52	30	139
OV0251	Similar to human TEGT mRNA, AF093741	TTGATTCTCTGCTAGGCCTG GTCATTCTTGGCACTGTTTC	56	30	116
OV0268	Similar to human CDK2 mRNA, AF093740	CATCACTCCAGAGTGGATTG CCTTCGACTCTGATCGCCTT	56	30	159
OV0373	Similar to human glucosamine-6-sulphatase (GNS) mRNA, AF093738	AAAGTGCAGCAGTGAACAGG CCACTGACTTTGTGAATTCC	58	30	129
PDE1B	63-kDa calcium/calmodulin-dependent 3′,5-cyclic nucleotide phosphodiesterase, M94867	GCCGAAGACGAGCACAACCA CATGTCCCTCGGAGCAGATG	64	60	~600
RHOD	Rhodanese (thiosulfate sulfurtransferase), M58561	TTTGAAAAGAGCCCAGAGGA GCTGGATCACTCTACCAGCA	54	30	284

synthesized for microsatellite markers according to published sequences (Table 1). Cattle-specific primers were designed for all genes by using the program PRIMER DESIGNER 2 (Scientific and Educational Software, Durham, NC; Table 2).

COMPASS. Cattle ESTs originating from an ovary cDNA library were mapped *in silico* by COMPASS as described previously (16) with modification. In brief, BLAST homology searches for cattle ESTs using default search parameters were performed against the nonredundant and dbEST databases of GenBank to define human orthologs. Map locations of the human orthologs were then retrieved from gene assignments made using the GB4 RH panel, as summarized in GeneMap ′98 (20). Assignments to cattle chromosomes were carried out by comparison with cattle-on-human synteny maps presented in the U.S. ArkDB database (http://bos.cvm.tamu.edu/bovgbase.html). Oligonucleotides used for PCR amplification of ESTs were designed from 3′ sequences (Table 2).

A new variant of COMPASS was used for identification of genes on BTA5 that would be useful for developing a cattle-human comparative map. This approach involved using UniGene to select genes mapping to HSA12 or HSA22 that are predicted to map to BTA5 on the basis of comparative mapping information. Then, cattle orthologs existing in GenBank were identified by using BLAST (21). In this way, four cattle genes predicted to be on BTA5 were found, essentially representing "free" sequences for RH mapping (Table 2). Oligonucleotides used for PCR amplification of these genes were designed from the cattle sequences whereas primers for amplification of collagen 2A1 (*COL2A1*) were designed from the human sequence (Table 2).

Radiation Hybrid Screening. All 56 markers were screened against a 5,000-rad whole genome RH panel in 20-μl PCR reactions as described by Band *et al.* (13). Annealing temperatures and extension times for PCRs are shown in Table 2. The PCR product for *COL2A1* was digested with *Rsa*I to distinguish between the cattle and hamster products. Digests of cattle *COL2A1* PCR products resulted in 2,100- and 400-bp fragments whereas those of hamster were 1,500 and 1,000 bp.

Map Construction. As a first step, 30 microsatellite markers were analyzed by using an equal retention model with the RHMAPPER software package (22). Six markers (*D5S64*, *D5S11*, *D5S79*, *D5S41*, *D5S14*, and *D5S42*) were chosen as an initial framework by finding

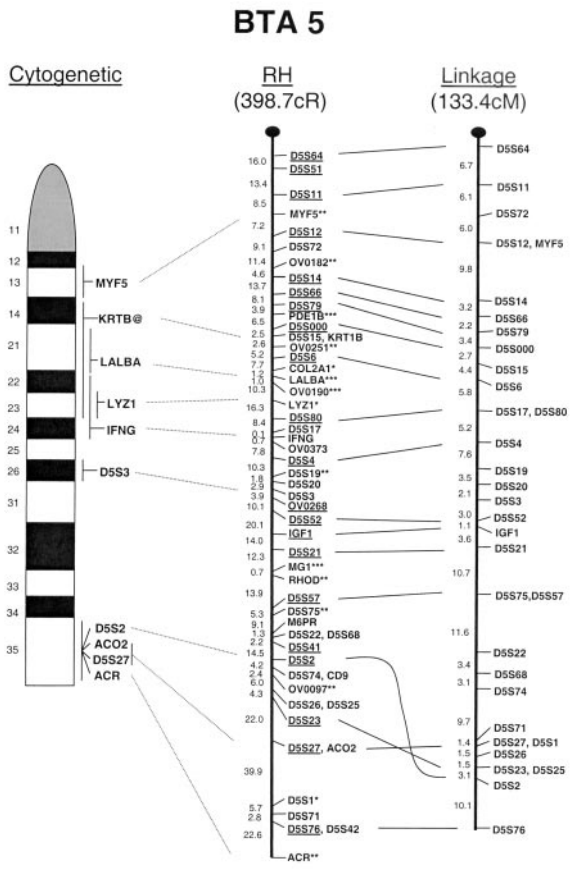

Fig. 1. Comparison of the radiation hybrid, cytogenetic, and linkage maps of BTA5. Framework markers (1,000:1) are underlined. Probabilities of placement within a given interval are denoted by asterisks as described in the legend of Table 1. Distances (cR or cM) between two markers are indicated in small numerals along the map. Framework markers are connected by solid lines to the corresponding locus on the linkage map. Cytogenetic and linkage assignments were taken from Barendse *et al.* (9), Comincini *et al.* (37), Friedl *et al.* (38), Fries *et al.* (25), Hayes *et al.* (39), Kappes *et al.* (10), and Ryan *et al.* (40).

the most likely order of marker triples at LOD-4, as described (23). This framework map was expanded by repeatedly using the grow_framework command of RHMAPPER with all remaining markers until no new markers were added. The final 1,000:1 framework map was used to create a placement map as described by Band *et al.* (13).

Results

A total of 56 markers, 34 microsatellites, 8 ESTs, and 14 known genes were genotyped on the 5,000-rad whole genome cattle-hamster RH panel. From these data, a LOD-3 (1,000:1) framework map was constructed, consisting of 20 microsatellite markers and one EST (Fig. 1). The remaining markers were used to produce a placement (comprehensive) map, which has a total length of 398.7 cR and an average interval size of 7 cR (Fig. 1). Two ESTs (OV0395 and OV0103) were not linked to any other markers and were therefore excluded from the map. Of the 90 hybrid lines that comprise the panel, 57 (63.3%) were informative for mapping BTA5 markers, 32 (35.6%) scored negative, and 1 (1.1%) scored positive for all markers. Retention frequency ranged from 34% (*ACR*) to 9% (*D5S68* and *D5S22*), with an average of 19% (Table 1). Ordering was not possible for the marker pairs *D5S15-KRT1B*, *D5S22-D5S68*, *D5S74-CD9*, *D5S26-D5S25*, and *D5S76-D5S42* because they had identical retention patterns. The order of markers on the RH map was essentially the same as the cattle cytogenetic and linkage maps (10), with the exception of the interval bound by *D5S2* and *D5S27* (Fig. 1).

Eight ESTs from a cattle ovary cDNA library were predicted to be on BTA5 by using the COMPASS strategy (Table 3). Six of these assignments were confirmed by RH mapping. The two ESTs not confirmed (OV0395 and OV0103) are classic examples of parology (Table 3). In addition to these ESTs, four cattle orthologs of human genes mapping to HSA12 or HSA22 were identified in GenBank: calmodulin-stimulated phosphodiesterase (*PDE1B*), rhodanese (*RHOD*), mannose-6-phosphate receptor (*M6PR*), and CD9 antigen (*CD9*). All were placed on the BTA5 RH map, confirming the predictions made from available comparative mapping information. Thus, a total of 10 new assignments were made by using COMPASS and were subsequently confirmed by RH mapping.

Table 3. Mapping of cattle ESTs by using COMPASS

			BLAST Results				
Clone	End	Identification/description	Accession no.	P(N)	%	Match length	Human interval (map location)
OV0097	3′	Human B-cell receptor associated protein mRNA	U72511	2e–30	90	124	*D12S328–D12S89* (12p13)
	5′	Human B-cell receptor associated protein mRNA	U72511	7e–54	83	336	
OV0182	3′	Human decorin gene, exon8	L01131	3e–47	84	251	*D12S322–D12S346* (12q23)
	5′	Human decorin gene, exon1	M98263	2e–69	93	177	
OV0190	3′	Human alpha-tubulin mRNA	K00558	e–101	92	291	*D12S85–D12S339* (12q13)
	5′	Human alpha-tubulin mRNA	X01703	0	96	407	
OV0251	3′	Human TEGT mRNA	X75861	2e–29	89	150	*D12S333–D12S325* (12q12–q13)
	5′	*R. norvegicus* TEGT mRNA	X75855	3e–37	94	105	
OV0268	3′	Human CDK2 mRNA	X62071	1e–67	91	199	*D12S325–D12S329* (12q13)
	5′	Human CDK2 mRNA	X62071	3e–41	87	201	
OV0373	3′	No significant hit					
	5′	Human glucosamine-6-sulfatase mRNA	Z12173	1e–55	86	277	*D12S83–D12S350* (12q14)
OV0103	3′	Human ribosomal protein L41	N30226	4e–16	90	70	No interval available (12q13.2–q13.3)
	5′	Human ribosomal protein L41	AF026844	1e–18	88	93	
OV0395	3′	EST highly similar to human decorin gene	AA099394	6e–88	89	297	*D12S322–D12S327* (12q15–q21.1) and *D6S1640–D6S422* (6p25–p22.2)
	5′	No significant hit					

GENETICS

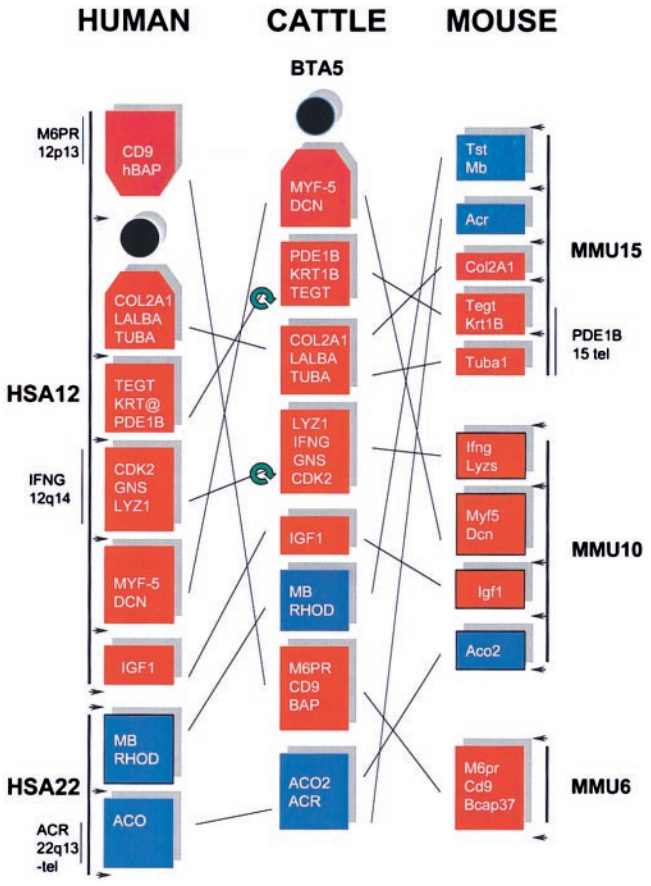

Fig. 2. Comparative maps of genes on BTA5. Bins are represented by shadowed boxes. The relative order of bins was defined on the basis of framework markers within the bins (Fig. 1). ''KRT@'' denotes the keratin gene cluster, including cytokeratin 2, with sequence similar to bovine *KRT1B*. The *TUBA*-like gene on HSA12 is also known as *K-ALPHA-1*. The order of markers within bins is confirmed for framework markers only. Circular arrows represent inversions within bins. Cattle ESTs are labeled with names of human orthologs. Arrowheads indicate evolutionary breakpoints relative to BTA5. Orders of human and mouse genes were taken from the human transcript map (20) and the Mouse Genome Database (http://www.informatics.jax.org/), respectively.

Type I genes were binned for comparison of gene order among cattle, humans, and mice (Fig. 2). Each bin was chosen to represent the smallest possible cluster of genes within a conserved synteny that is uninterrupted by a major rearrangement, using the cattle map as the reference. On the cattle map, order between bins is certain with odds of at least 1:1,000 compared with the next most likely order. However, local orders within the bins could not always be determined with odds >1:1,000 (Table 1). Therefore, the gene order within the shown bins may or may not be conserved among species. Within three of four bins with more than two genes the most likely order of the cattle RH map suggests that conservation of order has been maintained with the human map.

The long arm of HSA12, bound by *COL2A1* and *IGF1*, corresponds to a contiguous region of BTA5 bound by the centromeric marker *MYF5* and extending to the central region near *IGF1*. However, multiple disruptions in gene order have occurred within this region (Fig. 2). The human orthologs of *MYF5* and *DCN* are adjacent to *IGF1* on HSA12 whereas these genes are proximal to the centromere on BTA5, a large distance from *IGF1*. A break and a probable inversion are apparent between the bins containing *TEGT* and *COLA2A1*. An additional inversion and internal rearrangement involves the bin containing *CDK2*, *GNS*, *IFNG*, and *LYZ1*. The remainder of BTA5 distal to *IGF1* is composed of alternating sections that are homologous to HSA12 and HSA22, thus confirming observations from chromosome painting experiments (19). Given the comprehensive coverage of the framework markers, it is likely that the comparative organization shown in Fig. 2 represents all of the major rearrangements relative to the homologous human chromosomes.

Comparison of BTA5 with homologous mouse chromosomes shows a similarly high number of rearrangements. Of the 20 genes mapped, 17 have mapped mouse orthologs on three chromosomes, MMU6, MMU10, and MMU15. A larger number of evolutionary breakpoints ($n = 12$) can be identified in the cattle-mouse comparative map as compared with the cattle-human comparative map ($n = 9$).

Discussion

A RH map of BTA5 was constructed with 54 markers known or predicted to map to BTA5 using COMPASS. The map length is 398.7 cR, yielding a ratio of $\approx$2.8 $cR_{5,000}$/cM. Assuming 1 Mb/cM, this corresponds to 1 breakpoint every 354 kb, or an average fragment size of 35.4 Mb. This ratio is 70% of that found for BTA23 (13). The low retention frequencies (19% on average) and large fragment sizes result in an expected resolution of 2.1 Mb, considerably lower than that observed for other bovine chromosomes (13). This explains why only 21 of 48 unique loci could be ordered with 1,000:1 odds.

In general, the order of markers on the RH map was consistent with the order described on the linkage maps, excluding one region between the markers *D5S2* and *D5S27* that has large discrepancies for interval distances. On the RH map, these markers are framework markers. However, on the linkage map, the order of markers in this region is not reliable [LOD < 3.0 (10)]. Another type of discrepancy between the linkage and RH maps is the ratio of cM to cR. In the linkage map data, no recombinants were observed between *D5S27* and *D5S1*; however, 39.9 cR separates these loci on the RH map. In addition, the *D5S23-D5S27* interval is 3 cM as compared with 22 cR, a ratio greater than twice the average. The RF increased from 0.10 to 0.28 over the region spanning *D5S41–D5S1* (Table 1), thus implying variable breakage probabilities and possibly explaining the discrepancies between the linkage and RH maps.

Six of eight ESTs predicted to map to BTA5 were confirmed by RH mapping. Of the two that could not be assigned to BTA5, one (OV0395) having high similarity to the human *DCN* was assigned to BTA23 (discussed below). The second discrepancy (OV103) has high homology to the gene encoding ribosomal protein L41 but could not be localized to any specific interval with significant probability. Genes encoding ribosomal protein subunits are known to be associated with large numbers of pseudogenes (24). Thus, PCR amplification may have resulted in increased retention frequency, preventing conclusive assignment of this gene.

Although accurate assignments can be predicted for bovine genes by the COMPASS method, various sources of errors must be considered. Nadeau and Sankoff (4) describe several sources of error when identifying conserved segments, the most critical being incorrect identification of true orthologs. The existence of paralogs and pseudogenes creates difficulties in elucidating the true comparative chromosome organization among species. For example, at least nine keratin genes are clustered on human 12q13. In the present study, bovine epidermal cytokeratin type II 1B (*KRT1*; GenBank accession no. K03534), part of the cytokeratin class II gene cluster (25), was mapped to BTA5 as expected. This gene is highly similar to human cytokeratin 2 (*KRT2A*; GenBank accession no. M99063), although it also has high similarity with other keratin genes. Thus, it is not clear which is the true human ortholog when relying solely on sequence similarity. Furthermore, for recently duplicated genes, there may be no "true ortholog," as is the case for the cattle *DYA* MHC class II gene (26). Another example involves bovine ADP-ribosylation factor (*ARF*; GenBank accession no.

J03794), which is highly similar to human *ARF3* (94% of amino acids; GenBank accession no. M74491). This gene maps between *LALBA* and *TEGT* on HSA12; however, human *ARF1* (GenBank accession no. M84332), which has 100% protein identity with the bovine protein, maps to HSA1. Another very interesting example of confused orthology/parology involves ESTs OV0182 and OV0395, both of which have high similarity to human *DCN* and are classified in the same human UniGene cluster (Hs.76152). However, these two ESTs do not have significant similarity among themselves because they are derived from different ends of the cluster. EST OV0182 was mapped to BTA5, the predicted placement for *DCN*, whereas EST OV0395 maps to BTA23 (M.R.B., unpublished data). However, there are two alternative locations reported for sequence tagged sites within this UniGene cluster, one on HSA12 and the other on HSA6. These regions are predicted to be part of conserved syntenies on BTA5 and BTA23, respectively. Therefore, it is likely that these ESTs represent paralogs that are not clearly distinguished in UniGene.

Another limiting factor in the accuracy of COMPASS is the resolution and accuracy of gene maps in the target and reference species. As map resolution and accuracy increase by mapping additional ESTs and the use of higher resolution RH panels, the accuracy of *in silico* predictions should improve. When all chromosome breakpoints are defined between a target and a reference species, it should be possible to electronically bin any new EST or sequence tagged site that has a mapped ortholog in the reference species. For all mammals, the complete sequence of the human genome should be the gold standard for COMPASS, thus resolving many of the ambiguities encountered in RH mapping studies.

The assignment of 20 Type I markers to BTA5 supports previous studies that used chromosome painting, *in situ* hybridization, and somatic cell mapping. However, evidence of multiple rearrangements is apparent when comparing gene order between BTA5 and HSA12 and HSA22. Three markers, *PDE1B*, *CDK2*, and *CD63*, map within 6 cR on HSA12. By contrast, the bovine orthologs of *PDE1B* and *CDK2* are separated by 89 cR (Fig. 2). The large distance between these genes on the comparative map is attributable to multiple rearrangements within and between binned groups of markers. A comparison with the mouse gene assignments shows that *PDE1B* maps to MMU15 (27) whereas *CD63* maps to MMU10 (28). These observations point to common evolutionary breakpoint among the mouse and cattle chromosomes as compared with the human chromosome. However, when comparing this rearrangement among human and cattle chromosomes, synteny is conserved whereas a break in synteny is observed between human and mouse chromosomes.

Horvat and Medrano (29) identified a region on MMU10 that contains the high growth (*hg*) gene. A deletion in this region causes an increase in growth without resulting in an obese phenotype. This gene has obvious potential for elucidating mechanisms of growth and development and as a target for manipulation in livestock. The *hg* gene has been mapped to an interval flanked by mouse *Igf1* and *Dcn*, a region of conserved gene order on the human and mouse gene maps. However, in cattle, a major rearrangement has resulted in a large distance between these two loci. Unfortunately, it is not possible to predict at this time which region of the cattle chromosome contains the orthologous gene because the gene falls within a "gap" in the comparative map. Targeted closure of this gap on the RH map should be useful for obtaining the resolution necessary to determine whether the ortholog of *Hg* is a candidate gene for QTL in cattle.

An important goal of agricultural genomics is to create detailed comparative maps for economically important species of plants and animals. The COMPASS strategy is shown here to be effective in reliably predicting the position of genes *in silico*, exploiting the extensive knowledge of the human genome project. An important aspect of the COMPASS strategy is that the cost of mapping *in silico* is far less than the cost of RH mapping or linkage mapping. Thus, the expense and effort of mapping large numbers of genes to identify candidate genes for QTL can be minimized. In addition, COMPASS can be used in a highly selective and directed manner to fill in gaps in the RH and comparative maps. Sequencing and mapping of additional cattle ESTs and sequence tagged sites should allow the construction of high definition comparative maps, such as the one presented here for BTA5, creating a powerful tool to aid in identification of genes affecting economically important traits and contributing to our understanding of mammalian chromosome evolution.

This work was made possible in part by Grant 98-35205-6644 to H.A.L. and J.E.W. from the U.S. Department of Agriculture National Research Initiative Competitive Grants Program.

1. Gale, M. D. & Devos, K. M. (1998) *Science* **282,** 656–659.
2. Andersson, L., Archibald, A., Ashburner, M., Audun, S., Barendse, W., Bitgood, J., Bottema, C., Broad, T., Brown, S., Burt, D., Charlier, C., *et al.* (1996) *Mamm. Genome* **7,** 717–734.
3. Nadeau, J. H. & Sankoff, D. (1998) *Mamm. Genome* **9,** 491–495.
4. Nadeau, J. H. & Sankoff, D. (1998) *Trends Genet.* **14,** 495–501.
5. Womack, J. E. & Kata, S. R. (1995) *Curr. Opin. Genet. Dev.* **5,** 725–733.
6. Georges, M. & Andersson, L. (1996) *Genome Res.* **6,** 907–921.
7. Phillips, R. L. & Freeling, M. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 1969–1970.
8. Ma, R. Z., Beever, J. E., Da, Y., Green, C. A., Russ, I., Park, C., Heyen, D. W., Everts, R. E., Fisher, S. R., Overton, K. M., *et al.* (1996) *J. Hered.* **87,** 261–271.
9. Barendse, W., Vaiman, D., Kemp, S. J., Sugimoto, Y., Armitage, S. M., Williams, J. L., Sun, H. S., Eggen, A., Agaba, M., Aleyasin, S. A., *et al.* (1997) *Mamm. Genome* **8,** 21–28.
10. Kappes, S. M., Keele, J. W., Stone, R. T., McGraw, R. A., Sonstegard, T. S., Smith, T. P. L., Lopez-Corrales, N. L. & Beattie, C. W. (1997) *Genome Res.* **7,** 235–249.
11. Womack, J. E., Johnson, J. S., Owens, E. K., Rexroad, C. E., III, Schläpfer, J. & Yang, Y. (1997) *Mamm. Genome* **8,** 854–856.
12. Schläpfer, J., Yang, Y., Rexroad, C. E., III & Womack, J. E. (1997) *Chromosome Res.* **5,** 511–519.
13. Band, M., Larson, J. H., Womack, J. E. & Lewin, H. A. (1998) *Genomics* **53,** 269–275.
14. Yang, Y. & Womack, J. E. (1998) *Genome Res.* **8,** 731–736.
15. Yang, Y., Rexroad, C. E. III, Schläpfer J. & Womack, J. E. (1998) *Genomics* **48,** 93–99.
16. Ma, R. Z., van Eijk, M. J. T., Beever, J. E., Guérin, G., Mummery, C. L. & Lewin H. A. (1998) *Mamm. Genome* **9,** 545–549.
17. Heyen, D. W., Weller, J. I., Ron, M., Band, M., Beever, J. E., Feldmesser, E., Da, Y., Wiggans, G. R., Van Raden, P. M. & Lewin, H. A. (1999) *Physiol. Genomics* **1,** 165–175.
18. Zhang, Q., Boichard, D., Hoeschele, I., Ernst, C., Eggen, A., Murkve, B., Pfister-Genskow, M., Witte, L. A., Grignola, F. E., Uimari, P., *et al.* (1998) *Genetics* **149,** 1959–1973.
19. Solinas-Toldo, S., Lengauer, C. & Fries R. (1995) *Genomics* **27,** 489–496.
20. Deloukas, P., Schuler, G. D., Gyapay, G., Beasley, E. M., Soderlund, C., Rodriguez-Tome, P., Hui, L., Matise, T. C., McKusick, K. B., Beckmann, J. S., *et al.* (1998) *Science* **282,** 744–746.
21. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
22. Stein, L., Kruglyak, L., Slonim, D. & Lander, E. (1995) RHMAPPER (Whitehead Institute/Massachusetts Institute of Technology Center for Genome Research).
23. Slonim, D., Kruglyak, L., Stein, L. & Lander, E. (1997) *J. Comput. Biol.* **4,** 487–504.
24. Kenmochi, N., Kawaguchi, T., Rozen, S., Davis, E., Goodman, N., Hudson, T. J., Tanaka, T. & Page, D. C. (1998) *Genome Res.* **8,** 509–523.
25. Fries, R., Threadgill, D. W., Hediger, R., Gunawardana, A., Blessing, M., Joreano, J. L., Stranzinger, G. & Womack, J. E. (1991) *Cytogenet. Cell Genet.* **57,** 135–141.
26. van der Poel, J. J., Groenen, M. A., Dijkhof, R. J., Ruyter, D. & Giphart, M. J. (1990) *Immunogenetics* **31,** 29–36.
27. Reed, T. M., Browning, J. E., Blough, R. I., Vorhees, C. V. & Repaske, D. R. (1998) *Mamm. Genome* **9,** 571–576.
28. Gwynn, B., Eicher, E. M. & Peters, L. L. (1996) *Genomics* **35,** 389–391.
29. Horvat, S. & Medrano J. (1998) *Genomics* **54,** 159–164.
30. Stone, R. T., Pulido, J. C., Duyk, G. M., Kappes, S. M., Keele, J. W. & Beattie, C. W. (1995) *Mamm. Genome* **6,** 714–724.
31. Kemp, S. J., Hishida, O., Wambugu. J., Rink, A., Longeri, M. L., Ma, R. Z., Da, Y., Lewin, H. A., Barendse, W. & Teale, A. J. (1995) *Anim. Genet.* **26,** 299–306.
32. Barendse, W., Armitage, S. M., Kossarek, L. M., Shalom, A., Kirkpatrick, B. W., Ryan, A. M., Clayton, D., Li, L., Neibergs, H. L., Zhang, N., *et al.* (1994) *Nat. Genet.* **6,** 227–235.
33. Bishop, M. D., Kappes, S. M., Keele, J. W., Stone, R. T., Sunden, S. L. F., Hawkins, G. A., Solinas-Toldo, S., Fries, R., Grosz, M., Yoo, J. & Beattie C. W. (1994) *Genetics* **136,** 619–639.
34. Crawford, A. M., Dodds, K. G., Ede, A. J., Pierson, C. A., Montogomery, G. W., Garmonsway, H. G., Beattie, A. E., Davies, K., Maddox, J. F., Kappes, S. W., *et al.* (1995) *Genetics* **140,** 703–724.
35. Agaba, M. & Kemp, S. J. (1994) *Anim. Genet.* **25,** 187–189.
36. Mezzelani, A., Zhang, Y., Redaelli, L., Castiglioni, B., Leone, P., Williams, J. L., Solinas-Toldo, S., Wigger, G., Fries, R. & Ferretti L. (1995) *Mamm. Genome* **6,** 629–635.
37. Comincini, S., Drisaldi, B. & Ferretti, L. (1997) *Mamm. Genome* **8,** 486–831.
38. Friedl, R., Adham, I. M. & Rottmann, O. J. (1994) *Mamm. Genome* **5,** 830–831.
39. Hayes, H. C., Popescu, P. & Dutrillaux, B. (1997) *Mamm. Genome* **4,** 593–597.
40. Ryan, A. M., Schelling, C. P., Womack, J. E. & Gallagher, D. S. Jr. (1997) *Anim. Genet.* **28,** 84–87.