A high-resolution whole-genome cattle–human comparative map reveals details of mammalian chromosome evolution

Annelie Everts-van der Wind*†, Denis M. Larkin*†, Cheryl A. Green*, Janice S. Elliott‡, Colleen A. Olmstead*, Readman Chiu§, Jacqueline E. Schein§, Marco A. Marra§, James E. Womack‡¶, and Harris A. Lewin*¶‖

*Department of Animal Sciences and ‖Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; ‡Department of Veterinary Pathobiology, Texas A & M University, College Station, TX 77843; and §Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada V5Z 4S6

Contributed by James E. Womack, October 26, 2005

Approximately 3,000 cattle bacterial artificial chromosome (BAC)-end sequences were added to the Illinois–Texas 5,000-rad RH (RH, radiation hybrid) map. The BAC-end sequences selected for mapping are ≈1 Mbp apart on the human chromosomes as determined by BLASTN analysis. The map has 3,484 ordered markers, of which 3,204 are anchored in the human genome. Two hundred-and-one homologous synteny blocks (HSBs) were identified, of which 27 are previously undiscovered, 79 are extended, 26 were formed by previously unrecognized breakpoints in 18 previously defined HSBs, and 23 are the result of fusions. The comparative coverage relative to the human genome is ≈91%, or 97% of the theoretical maximum. The positions of 64% of all cattle centromeres and telomeres were reassigned relative to their positions on the previous map, thus facilitating a more detailed comparative analysis of centromere and telomere evolution. As an example of the utility of the high-resolution map, 22 cattle BAC fingerprint contigs were directly anchored to cattle chromosome 19 [*Bos taurus*, (BTA) 19]. The order of markers on the cattle RH and fingerprint maps of BTA19 and the sequence-based map of human chromosome 17 [*Homo sapiens*, (HSA) 17] were found to be highly consistent, with only two minor ordering discrepancies between the RH map and fingerprint contigs. The high-resolution Illinois-Texas 5,000-rad RH and comparative maps will facilitate identification of candidate genes for economically important traits, the phylogenomic analysis of mammalian chromosomes, proofing of the BAC fingerprint map and, ultimately, aid the assembly of cattle whole-genome sequence.

comparative genomics | radiation hybrid | map integration

Detailed whole-genome maps are a major currency for comparative and functional genomics (1). With the human, rat, and mouse genomes completely sequenced (2–6), eight mammalian genomes to follow with >6-fold coverage, and eight others to be sequenced in draft form (www.genome.gov/10002154), greater insights into mammalian genome evolution and function can be obtained because of the improved accuracy and resolution of comparative maps (1). We previously reported two expressed sequence tag (EST)-based 5,000-rad cattle radiation hybrid (RH) maps containing 768 and 1,463 anchor points in the human genome, corresponding to ≈45% and 65% of comparative human genome sequence coverage, respectively (7, 8). These maps, collectively known as the Illinois-Texas 5,000-rad radiation hybrid panel (IL-TX $RH_{5,000}$) maps, facilitated the identification of genes responsible for double-muscling (9, 10), chondrodysplasia (11), and two major quantitative trait loci for milk production (12, 13).

Recently, Itoh *et al.* (14) constructed a cattle–human comparative RH map consisting of 5,593 EST and microsatellite markers. This map contains 1,716 anchor points between the human and cattle genomes, of which most are ESTs, and was reported to have ≈72% comparative coverage of the human genome. A significant drawback of EST-based comparative maps is the uneven marker distribution that often results in relatively large gaps in comparative coverage. These gaps significantly complicate comparative genome analysis, especially around the evolutionary breakpoint regions. To overcome this limitation, we initiated the development of high-resolution RH maps that use the sequenced ends of bacterial artificial chromosome (BAC) inserts as comparatively anchored markers (15). The BAC-end sequences (BESs) selected for mapping were linked by sequence similarity to the human and mouse genome sequences and were evenly spaced in the reference genomes at ≈1-Mbp intervals. The approach resulted in the construction of detailed comparative maps of BTA15 and BTA29, with ≈86% comparative coverage of HSA11 (15). Furthermore, the study revealed the presence of chromosome breakpoints that were "reused" in mammalian evolution, results that were recently confirmed in a more detailed multispecies analysis (1). Herein, we report extension of our methodology to mapping the whole cattle genome. The new map has ≈91% comparative coverage of the human genome sequence and contains single-linkage groups for all cattle autosomes. In addition, we used the new high-resolution cattle–human comparative map to study the evolution of centromeres and telomeres, conduct an analysis of gene content within evolutionary breakpoints, and demonstrate the power of the map for proofing the order of contigs in the physical map.

Methods

Marker Selection and Mapping Strategy. BESs from the cattle male CHORI-240 (http://bacpac.chori.org/bovine240.htm) BAC library were downloaded from GenBank and repeat-masked by using REPEATMASKER software (ftp://ftp.genome.washington.edu). National Center for Biotechnology Information (NCBI) BLASTN was used to identify BESs with sequence similarity to the human (NCBI build 33) and mouse (NCBI build 30) genome sequences. Default BLASTN search parameters and an expectation value threshold of 0.00001 were used for comparison with the human genome sequence. More sensitive BLASTN search parameters (-W 7 –r 17 –q –21 –f 280 –G 29 –E 22 –X 240 –e 0.01) were used for comparison with the mouse genome sequence. Only those BESs with a single significant match in the human genome were extracted into a database (http://genome.ucsc.edu/cgi-bin/hgTrack?position=chr1&hgsid=63602597&bacendsCow=squish). In addition, the top BLASTN hit in the mouse genome was added to each BES annotation in the database to reduce the number of comparative singleton

Conflict of interest statement: No conflicts declared.

Abbreviations: BAC, bacterial artificial chromosome; BES, BAC-end sequence; BTA, *Bos taurus*; HSA, *Homo sapiens*; HSB, homologous synteny block; IL-TX $RH_{5,000}$, Illinois-Texas 5,000-rad radiation hybrid panel; NCBI, National Center for Biotechnology Information; RH, radiation hybrid; TSP, traveling salesman problem.

†A.E.-v.d.W. and D.M.L. contributed equally to this work.

¶To whom correspondence may be addressed. E-mail: jwomack@cvm.tamu.edu or h-lewin@uiuc.edu.

Table 1. Summary statistics for the high-resolution cattle $RH_{5,000}$ map

	ESTs		Microsatellites		BAC ends		Total no.					
BTA	Ordered	Binned	Ordered	Binned	Ordered	Binned	Ordered	Binned	Retention frequency	Length, TSP	BLASTN hits	No. of HSBs
1	26	32	9	4	140	22	175	58	0.28	993	164	7
2	25	30	8	1	122	15	155	46	0.18	1,190	147	9
3	43	16	9	3	102	12	154	31	0.18	878	143	3
4	28	14	5	5	119	10	152	29	0.20	900	146	10
5	43	24	24	8	110	16	177	48	0.18	826	148	9
6	33	15	14	2	114	6	161	23	0.18	1,263	146	6
7	40	16	11	4	109	5	160	25	0.17	1,410	147	9
8	32	19	12	3	109	4	153	26	0.19	1,307	139	13
9	22	15	9	3	103	7	134	25	0.14	764	124	3
10	38	11	4	4	95	7	137	22	0.23	694	132	9
11	48	13	8	0	107	2	163	15	0.27	961	152	9
12	15	8	5	3	92	6	112	17	0.18	702	105	4
13	25	25	5	3	95	8	125	36	0.20	856	120	12
14	17	17	8	11	91	18	116	46	0.23	708	107	3
15	32	5	8	0	71	5	111	10	0.18	709	101	6
16	35	7	6	0	65	2	106	9	0.29	688	99	12
17	24	22	10	3	67	8	101	33	0.23	748	90	6
18	30	45	7	2	60	9	97	56	0.23	900	84	5
19	21	33	0	0	71	19	92	52	0.42	1,156	92	9
20	10	5	7	3	73	1	90	9	0.20	814	81	2
21	24	8	9	3	66	6	99	17	0.18	673	85	9
22	19	9	5	0	60	6	84	15	0.26	819	79	5
23	8	31	5	6	58	4	71	41	0.26	544	66	3
24	18	13	5	1	67	8	90	22	0.22	468	84	4
25	15	19	5	1	44	8	64	28	0.24	678	59	6
26	13	15	4	0	47	6	64	21	0.23	589	58	3
27	12	8	9	5	48	6	69	19	0.14	628	58	7
28	8	7	1	5	42	2	51	14	0.30	409	50	4
29	15	11	2	1	41	6	58	18	0.33	516	54	5
X	17	25	18	2	128	9	163	36	0.16	1,640	144	7
Total	736	518	232	86	2,516	243	3,484	847	0.22	25,431	3,204	201

markers in the final map. One cattle BES with a BLASTN hit every ≈1 human-Mbp was selected from the database, giving preference to cattle BESs also having an orthologous hit in the mouse genome (http://genome-mm3.cse.ucsc.edu/cgi-bin/hgTracks?org=mouse&db=mm3&position=chr1&hgt.customText=http://cagst.animal.uiuc.edu/BACcontig/New_tracks/BAC_ends_track_build_30_mouse). Oligonucleotide primers for amplification of cattle BESs were designed by using VECTOR NTI 7.0 software (InforMax). Polymerase chain reaction was performed at annealing temperatures of 55–65°C for 35 cycles according to a procedure described in refs. 8 and 15. The BESs were typed in duplicate on a 5,000-rad cattle-hamster RH panel (16) and combined with markers scored in refs. 7, 8, and 15. After an initial round of mapping (see below) 16 BESs were selected to confirm comparative singletons (by synteny extensions) and decrease the marker spacing both inside homologous synteny blocks (HSBs) and between HSBs when the distance between the comparative anchors exceeded 1.6 Mbp in the human genome sequence.

Quality Control of RH Vectors. The selected cattle BESs were assigned to the cattle chromosomes on the basis of two-point linkage analysis by using logarithm of odds >8.0 as a threshold (RHMAPPER 1.22). The order of the cattle sequence anchor points in the human genome sequence was then used to perform a quality control on the RH vectors. First, the cattle markers were ordered according to their orthologous human chromosome sequence position. Microsatellites and cattle ESTs without a BLASTN match in the human genome sequence were placed between the ordered markers on the basis of two-point linkage analysis. Next, RH vectors were compared for adjacent markers. Every position of the vector that was not consistent with the scoring on that same hybrid cell line for two adjacent markers was flagged by using a PERL script. Finally, all flagged reaction scores in each vector were checked for accuracy by using the original gel pictures. Each RH vector with more than three ambiguous scores after quality control, or with more than two inconsistencies with its adjacent vectors, were excluded from the initial build of the chromosome maps. Some of these markers were placed on the final build of chromosome maps if they allowed for closing gaps in spacing and were consistent with positions of surrounding markers.

Map Construction. The cattle BESs were assigned to cattle chromosomes on the basis two-point linkage analysis with previously mapped markers (8, 15). Multipoint linkage analysis was done for all markers that passed the quality control by using the packages RH_TSP_MAP version 2.0 (17) and CONCORDE (18). Map distances are expressed in arbitrary "traveling salesman problem" (TSP) units. Five maps were produced for each chromosome or linkage group, of which two maps were based on the minimum number of obligate chromosome breaks and three maps were variants of the maximum likelihood estimate approach. Next, the five maps were evaluated, and the frequency of the positions of a given marker in the five resulting maps was determined. A consensus map was generated with marker order that was consistent with the majority of the five maps. A PERL script was written to extract intermarker distances from any of the five maps for each particular marker combination found on the consensus map and in any of the five individual maps. The microsatellite order was checked for consistency with the order in a reference cattle genetic linkage map (19).

Comparative Analysis. Human gene sequence coordinates were retrieved from the RefSeq database (RefSeq Release 9). Homologous synteny blocks were defined on the basis of the anchor points in the human genome (NCBI, build 33 and 35), according to a set of rules described by Murphy *et al.* (1). Briefly, (*i*) an HSB was defined as two or more markers on the same chromosome in the

Table 2. Comparative coverage of the high-resolution cattle–human comparative map

HSA	Human genome length, Mbp	Modified length,* Mbp	Comparative coverage (%)	No. of BLASTN hits†	Average marker spacing, Mbp	No. of HSBs
1	246	224	91	272	0.82	18
2	243	238	93	264	0.90	16
3	200	195	93	213	0.92	9
4	191	188	93	213	0.88	11
5	181	178	94	199	0.89	10
6	171	167	96	190	0.88	6
7	159	155	90	171	0.91	15
8	146	143	91	165	0.87	12
9	138	116	86	132	0.88	9
10	135	132	90	135	0.98	11
11	134	131	87	157	0.84	12
12	132	131	93	149	0.88	8
13	114	96	92	105	0.92	4
14	106	88	88	111	0.80	5
15	100	82	81	99	0.83	12
16	89	79	89	81	0.98	6
17	79	78	87	92	0.85	9
18	76	75	93	84	0.89	4
19	64	56	85	68	0.82	3
20	62	59	92	79	0.75	6
21	47	34	88	34	0.99	3
22	50	35	89	47	0.75	5
X	155	152	91	144	1.05	7
Total	3,019	2,833	91	3,204	0.88	201

*Modified length of chromosome is the total length minus the length of telomere, centromere, and heterochromatin regions.

†Mapped cattle markers with BLASTN hits ($E < e^{-5}$) in the human genome.

human and cattle genomes not interrupted by an HSB from a different region of the same chromosome or a different chromosome; (*ii*) an internal rearrangement (inversion) was defined by a minimum of three consecutive markers on the cattle RH map in the same order as in the human genome, with adjacent markers separated by a span of >1 human-Mbp; (*iii*) a comparative singleton was a single marker placed on the cattle RH map, but out-of-place with respect to its expected position on the basis of the human comparative mapping information; and (*iv*) to minimize disruption of HSBs, a comparative singleton could be "jumped" to its expected position in a HSB on the same chromosome, provided that the distance for the jump was <2 human-Mbp.

Integration with the Cattle BAC Fingerprint Map. The cattle BAC fingerprint map was downloaded from the Genome Sciences Centre, January 2004, freeze (available upon request). Fingerprint contigs were integrated with the BTA19 RH map by using BESs from BAC clones placed on both fingerprint and RH map as connection points. Each BES from a BAC clone placed on both the fingerprint and RH map was assigned an additional number after the fingerprint contig identifier on the basis of the order of the BAC clones (represented by >1 BES on the RH map) within the fingerprint contig.

Results

A High-Resolution Cattle $RH_{5,000}$ Map. In total, RH vectors for 2,888 cattle BESs were added to the IL-TX $RH_{5,000}$ database, which contained 2,030 vectors for ESTs, genes, BESs, and microsatellites from earlier studies (7, 8, 15). Two-point linkage analysis was performed for 4,918 markers; 4,893 markers were linked at logarithm of odds ≥8. Multipoint analysis was performed for each chromosome (excluding BTAY), resulting in contiguous linkage groups for all 29 cattle autosomes and three linkage groups for BTAX.

Five-hundred-and-seventy-eight markers that did not pass the stringent quality controls, and an additional 25 unlinked markers were excluded from the map (Table 3, which is published as supporting information on the PNAS web site). Because the RH mapping procedure works optimally with markers that are evenly distributed (20), another 847 markers (243 BESs, 518 ESTs, and 86 microsatellites) with identical or nearly identical map positions were flagged in the data set and, ultimately, were mapped by binning to the ordered markers on the basis of two-point linkage analysis (all binned markers are listed in Table 4, which is published as supporting information on the PNAS web site). A backbone map was constructed after these quality steps were implemented. Subsequently, an additional 16 BESs were added to the map either to confirm singletons on the comparative map or to reduce gaps of >1.6 human-Mbp. On the final map, 3,484 markers were ordered, of which 2,516 are cattle BESs, 736 are ESTs, and 232 are microsatellites (Table 1; see also Table 5, which is published as supporting information on the PNAS web site). The RH maps of all of the cattle chromosomes can be found in Fig. 4, which is published as supporting information on the PNAS web site.

The number of markers ordered varies from 51 on BTA28 to 177 on BTA5. Cattle chromosome 28 has the shortest map length (409 TSP units), whereas BTAX is the longest (1,640 TSP units). The $RH_{5,000}$ whole-genome map is 25,431 TSP units (Tables 1 and 5). Assuming a cattle genome size of ≈3 Gbp, one TSP unit corresponds to ≈120 cattle-Kbp. This ratio yields a predicted marker density of 1.2 markers per cattle-Mbp, with the highest predicted density found on BTA5 (≈1.8 markers per Mbp) and the lowest density observed on BTA19 (≈0.7 marker per Mbp). The average retention frequency of the ordered markers is 22%, ranging from 14% on BTA27 and BTA9 to 42% on BTA19, which contains the selectable marker thymidine kinase 1 (TK1; Table 1).

A Whole-Genome Cattle–Human Comparative Map. Among the 3,484 markers ordered on the map, 3,204 (92%) have putative orthologs (or significant sequence similarity) in the human genome (NCBI build 35; Table 2). In total, the 201 defined HSBs (≥2 markers)

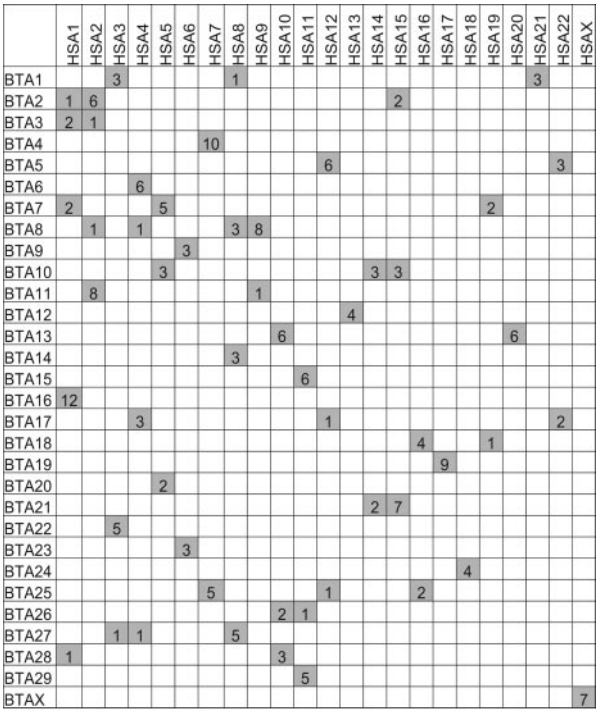

Fig. 1. A modified Oxford grid of HSBs shared among cattle and human chromosomes. The HSBs are illustrated by gray squares. The numbers within the gray squares represent the number of HSBs found shared on the cattle and human chromosomes.

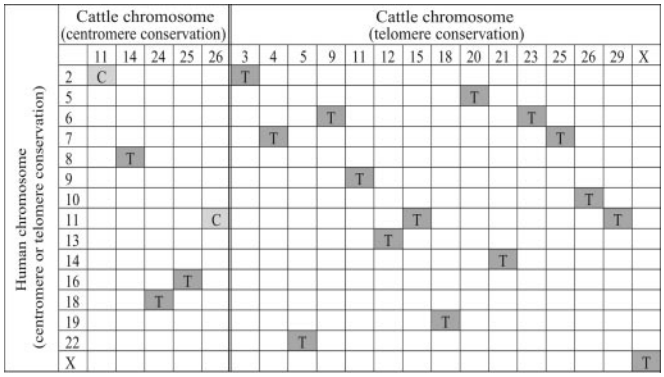

Fig. 2. Conservation of human centromere (C) and telomere (T) positions with positions of centromeres or telomeres in the cattle genome. The left side of the figure shows cattle chromosomes that have the position of their centromere conserved with a human centromere or telomere. The right side of the figure shows cattle chromosomes that have telomere positions conserved with a human centromere or telomere.

resulted from 178 evolutionary breakpoints between the human and cattle genomes (Fig. 4). The map positions of 14 of 21 markers that were not part of an HSB (comparative singletons) on the "backbone" map were confirmed by adding 16 targeted BESs. The final map contained only seven comparative singletons, of which three are on BTA13 (CC515837, BZ900194, BZ908049), one on BTA15 (BZ863012), one on BTA18 (CC767636), one on BTA19 (BZ896574), and one on BTA27 (BZ919176).

The number of HSBs on each cattle chromosome ranges from 2 on BTA20 to 13 on BTA8 (Figs. 1 and 4 and Table 1). The HSB sizes ranged from 5.9 Kbp (HSA12 HSB on BTA25) to 86.2 Mbp (HSA1 HSB on BTA3), with a median HSB size of 7.6 Mbp (see Table 5). Coordinates of the 198 HSBs (NCBI build 33) were identified in the human genome and compared with 176 HSBs identified in our previous comparative map according to the same set of rules used to define the HSBs in this study (1). Twenty-seven HSBs were previously undiscovered regions of homology, 12 were previously comparative singletons, and the remaining 15 HSBs were located in regions of the comparative map for which no information previously was available. The 27 previously unidentified HSBs had a median size of 3.1 Mbp, ranging from 103 Kbp to 16.0 Mbp. Fifty-five previously identified HSBs were merged into 23 newly extended HSBs, whereas 18 previously identified HSBs were split into 44 HSBs (11 previously identified HSBs were extended and 7 previously identified HSBs were truncated). Of the HSBs that remained intact, 2 were of the same length (<1.0 Kbp difference), 79 were extended (>1.0 Kbp difference), and 17 were truncated (>1.0 Kbp difference). Only five previously defined breakpoint regions were repositioned as compared with the second-generation map.

On the basis of human genome sequence coordinates (NCBI build 35) the comparative coverage on the third-generation cattle RH map is ≈91%. Comparative coverage of each chromosome excludes the lengths of telomere, centromere, and heterochromatin regions. Excluding these regions, which often lack human map contig and sequence data, the comparative length of the human genome sequence covered in the cattle genome comprises ≈2.8 Gbp (Table 2). Comparative coverage is also lost when sequence gaps overlap with breakpoint regions on the comparative map. The resolution of the new comparative map, thus, is close to the predicted maximum comparative coverage of 94% (Table 2).

Four cattle chromosomes, BTA12, BTA19, BTA24, and BTAX, have complete homology to HSA13, HSA17, HSA18, and HSAX, respectively; however, internal rearrangements are observed for all comparisons (Figs. 1 and 4). In addition, eight cattle chromosomes are homologs to nine human chromosome arms (BTA18 is homologous to HSA16q and HSA19q, thus the difference of one cattle chromosome). A previously uncharacterized relatively small region of HSA2p (0.1–1.7 Mbp) was identified on BTA8, thus disrupting a chromosome arm previously thought to be a complete HSB on BTA11. Similarly, HSA12p, HSA10p, and HSA6q, previously believed to be homologous to large HSBs on BTA5, BTA13, and BTA9, were found to have small HSBs on BTA25, BTA28, and BTA23, respectively. Two of these newly assigned HSBs involve small regions located close to the position of the human centromere.

Conservation of Centromere and Telomere Positions. Among the 30 cattle centromere positions (excluding BTAY), 13 positions were confirmed and 4 were reassigned with respect to their location on HSBs on the second-generation comparative map (Table 6, which is published as supporting information on the PNAS web site). Four cattle centromeres were found associated with the same HSBs as described in ref. 8, but their relative positions are inverted on the new map due to the revised ordering of markers relative to those in adjacent HSBs (Table 6). The positions of nine cattle centromeres were assigned to HSBs for the first time, according to our data (Table 6). Among the 31 cattle telomere locations (excluding BTAY), the positions of 19 were confirmed and 5 were reassigned with respect to their location on HSBs on the second-generation comparative map (Table 6). Three cattle telomeres were found associated with the same HSBs as described in ref. 8, but their relative position was inverted because of the revised ordering of markers relative to those in adjacent HSBs (Table 6). The positions of four cattle telomeres were assigned to HSBs for the first time, according to our data (Table 6).

Comparative analysis of centromere and telomere positions in the cattle and human genomes allowed determination of whether the positions of centromeres and telomeres have been conserved in both species (Figs. 2 and 4). The positions of two human centromeres and three human telomeres are consistent with centromere

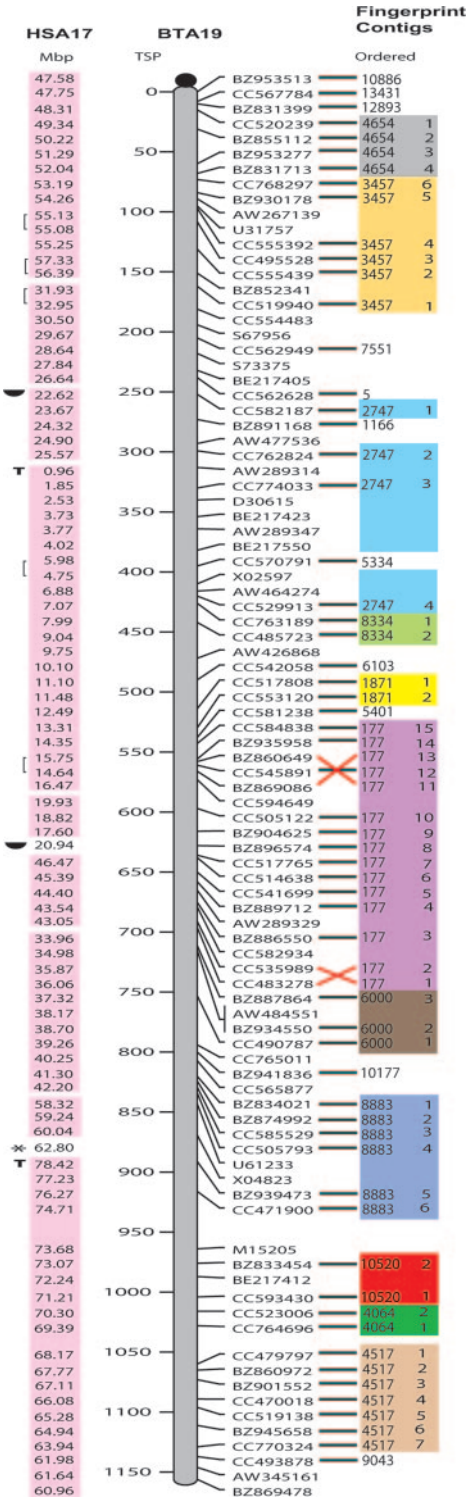

Fig. 3. The integrated RH map and BAC contig of BTA19. Cattle chromosome 19 is represented by a contiguous gray bar. Markers are placed to the right side of BTA19 with their corresponding GenBank accession numbers. Homologous synteny blocks on HSA17 are colored pink, with the homologous sequence coordinates in the human genome (NCBI build 35) inside the HSBs. Marker names (not map position) are aligned with the sequence coordinates on the human map (in Mbp, to the far left). On the left side of the RH map, the intermarker spacing is indicated in TSP units. Connectors (left brackets) on the left side of the HSBs indicate small inversions, which would make the order with respect to the human genome perfect. A filled black circle represents the cattle centromere. Half circles are used to indicate the location of the human centromere because of an evolutionary breakpoint at the human centromere location. An uppercase "T" positioned to the left of the HSBs represent the positions of HSA17 telomeres. An asterisk positioned on the left side of a marker sequence coordinate indicates an "out-of-place" marker. On the right side of the GenBank accession numbers are filled black horizontal lines connecting the cattle RH map to the cattle BAC fingerprint map. The numbers inside the colored boxes indicate fingerprint contig identification numbers, with coloring representing unique fingerprint contigs that contain multiple cattle BESs. The order of BESs on BAC clones within the fingerprint contigs is indicated with numbers on the right of the contig identifiers inside the colored boxes.

positions in cattle chromosomes. For example, on the basis of flanking HSBs, the centromere position of submetacentric HSA2 is consistent with the centromere position on acrocentric BTA11. The positions of 15 telomeres on human chromosomes also appear conserved as telomeres on 15 homologous cattle chromosomes. For example, a 10 human-Mbp HSB at the telomere of HSA2 is also found at the telomere of BTA3 in the same orientation (Figs. 2 and 4).

Gap Size Distribution and Gene Density. The 178 gaps between evolutionary breakpoints in the comparative map have an average size of 1.0 human-Mbp. There are 30 gaps in the comparative map >1.6 human-Mbp, but <5.0 human-Mbp. There are two gaps >5.0 Mbp, one of which is at 7.4 Mbp on HSA9p (proximal to the human centromere at 38.4 to 45.8 Mbp) and the other at 7.3 Mbp on HSA19 (4.9 Mbp and 2.4 Mbp on each side of the human centromere). These large gaps resulted from a paucity of cattle BESs with sequence similarity to sequences within the corresponding regions of human genome. For example, we were unable to find any cattle BESs having unique similarity to the human genome within the pericentric region of HSA19 (between 19.6 and 24.1 Mbp).

The density of RefSeq genes in the human genome is 6 per Mbp. However, the gene density in small gaps on the comparative map (<1.0 Mbp) is higher (10 genes per Mbp, $P < 0.10$), whereas the gene density in the larger gaps (>2.0 Mbp) is significantly lower (≈2 genes per Mbp; $P < 0.0001$) than the human genome on average. The average RefSeq gene size inside the large gaps (>2.0 Mbp) is greater (73.7 Kbp) than the average RefSeq gene size inside the small gaps (<1.0 Mbp; 37.4 Kbp).

Direct Anchoring of Cattle BAC Fingerprint Contigs to the RH Map of BTA19. Cattle BAC fingerprint contigs could be directly anchored to cattle chromosome 19 by using the RH map, because the selected cattle BESs were derived from the same BACs as those used for construction of the whole-genome BAC contig (Fig. 3). The mapped cattle BESs directly anchored 22 fingerprint contigs on the cattle RH map of BTA19 (Fig. 3). Four fingerprint contigs span six of eight evolutionary breakpoints on the comparative map. Fifteen fingerprint contigs are anchored with one or two cattle BESs, and seven fingerprint contigs are anchored with ≥3 cattle BESs. Of these seven fingerprint contigs, three have marker order in complete concordance on BTA19 and HSA17. Among the remaining four fingerprint contigs, only minor discrepancies between the order of markers in the RH map and order of markers/clones in the fingerprint contig or human map were identified. For example, in fingerprint contig ctg8883, one BES (CC505793) was found on the comparative map in a position conflicting with the other markers in the HSB. However, the order of markers surrounding CC505793 was in agreement in both the fingerprint contig and RH map. The RH and fingerprint contig data, thus, are in agreement, which suggests that a microrearrangement may exist between the human and cattle genomes at this location.

Discussion

Significant progress has been made in improving the IL-TX $RH_{5,000}$ map, with the number of markers increased 4-fold, from 1,087 to

4,331 (7, 8). The number of comparative anchor points in the human genome was increased 5-fold, from 638 to 3,204, by mapping BESs selected for spacing of ≈1 Mbp in the human genome. Although Itoh *et al.* (14) placed ≈2,000 more markers on their RH map, comparative coverage relative to the human genome is ≈20% less than our BES-based map, primarily as a result of the uneven distribution of ESTs used for mapping. Cattle BESs thus provide an ideal marker resource for enhancing RH maps because at least 50% can be anchored directly to the human or mouse genome, and precise marker spacing can be determined. Mapping BESs with only one significant BLASTN hit in the human genome circumvented additional problems created by mapping ESTs, such as distinguishing paralogs and pseudogenes. In addition, mapping BESs (*i*) increased the number of HSBs identified on the comparative map, (*ii*) lowered the number of comparative singletons, and (*iii*) dramatically increased the total comparative coverage with respect to the human genome.

A fundamental requirement for detailed comparative mapping is the definition of HSBs. The number of HSBs identified on the new cattle–human comparative map is 201, with a potential for 208 if the seven comparative singletons represent small HSBs. The six gaps in the comparative map that are >3.0 Mbp are potential regions in which new HSBs could be discovered. Twenty-seven previously undescribed HSBs (median size = 3.1 Mbp) were identified in the present study. The median size of the previously unidentified HSBs is smaller than the overall median HSB size of 7.6 Mbp, supporting the expectation based on the model of random chromosomal breakage proposed by Nadeau and Taylor (21). However, recent work has demonstrated that ≈20% of evolutionary breakpoints are "reused" (1), which tends to produce microsyntenies that are beyond the resolution of the current map. Thus, the remaining 172 gaps in the comparative map that are <3.0 Mbp are regions where only small HSBs may be discovered. These small HSBs await discovery by direct sequence-based comparisons.

Each gap between HSBs represents a gap in comparative coverage. In the mouse-human sequence-based comparison, the average gap size is 300 Kbp. The gaps in coverage are primarily because of the accumulation of unique DNA sequences (pairwise) within the breakpoint regions, as evidenced by a paucity of cattle BESs with sequence similarity to the human genome in map gaps. Although the actual gap sizes in the cattle genome are as yet unknown, the theoretical maximum comparative coverage of the cattle genome on the human genome can be estimated at 98.1% by multiplying the total number of gaps in the cattle–human comparative map (201 HSBs –23 human chromosomes = 178 gaps) by an assumed average gap size (300 Kbp), subtracting that amount from the total modified human genome length (human genome length excluding centromeres, telomeres, and heterochromatin), and dividing that amount by the modified total human genome length (Table 2). By contrast, a map with ≈1 human-Mbp spacing between the markers can be expected to cover a maximum of 93.6% of the modified human genome sequence length. Our new map covers ≈91% of the human genome or ≈97% of the theoretical maximum.

The relatively tight gaps in the new map facilitate study of the distinguishing features of DNA sequences within evolutionary breakpoint regions (1, 8). We observed a higher gene density in small gaps (<1.0 human-Mbp), whereas a lower gene density was found in the few relatively large gaps (>2.0 human-Mbp) than compared with the genome on average, consistent with results from the previous IL-TX $RH_{5,000}$ map (8). Murphy *et al.* (1) also found higher gene density (predicted genes plus RefSeq genes) in evolutionary breakpoint regions identified by comparison of genomes from seven different mammalian species. Evolutionary chromosomal breaks, thus, may be more likely to result in novel gene fusions and/or separation of genes under common regulatory control. Alternatively, segmental duplications may be promoted at the site of evolutionary chromosome breakages, as suggested from previous data (1), leading to the observed higher gene density. The adaptive evolutionary significance of evolutionary chromosome breakages remains to be conclusively determined, but there is accumulating evidence that not all breakages are evolutionarily neutral (22). Thus, lineage-specific phenotypes may be, in part, a result of chromosome rearrangements and the subsequent generation of novel genes within the breakpoint regions.

An important advance made possible by mapping cattle BESs was that the RH map could be anchored directly to the whole-genome BAC fingerprint contig (Fig. 1). This advance was demonstrated by the assignment of 22 BAC fingerprint contigs to their relative positions on the cattle chromosomes, with direct confirmation of BES order within the RH map and the fingerprint contigs for 93% (42 of 45) of BESs that are part of ordered HSBs. Comparisons of BES order on the RH maps and within the fingerprint contigs were used to identify inconsistencies in the maps and markers or clones that are presumably "out of place" on the basis of their cattle–human comparative map location (Fig. 3; e.g., CC505793). This comparison will ultimately be important in selecting the correct minimum tiling path for BAC-skim sequencing and correctly assembling the cattle genome sequence. Thus, the high-resolution RH map is a powerful tool for correctly assigning the fingerprint contigs and, ultimately, the cattle genome sequence scaffolds to the cattle chromosomes.

We thank Drs. R. Agarwale and A. A. Schäffer (NCBI) for advice on the use of TSP/CONCORDE code. This work was supported in part by U.S. Department of Agriculture (USDA) National Research Initiative Tools and Reagents Grant AG2004-3520-14196; USDA Cooperative State Research, Education, and Extension Service Livestock Genome Sequencing Initiative AG2004-34480-14417; and USDA Agricultural Research Service Agreement 58-5438-2-313.

1. Murphy, W. J., Larkin, D. M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L., Beever, J. E., Chowdhary, B. P., Galibert, F., Gatzke, L., *et al.* (2005) *Science* **309,** 613–617.
2. Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002) *Nature* **420,** 520–562.
3. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature* **409,** 860–921.
4. International Human Genome Sequencing Consortium (2004) *Nature* **431,** 931–945.
5. Gibbs, R. A., Weinstock, G. M., Metzker, D. M., Sodergren, E. J., Scherer, S., Scott, G., Steffen, D., Worley, K. C., Burch, P. E., Okwuonu, G., *et al.* (2004) *Nature* **428,** 493–521.
6. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A. *et al.* (2001) *Science* **291,** 1304–1351.
7. Band, M. R., Larson, J. H., Rebeiz, M., Green, C. A., Heyen, D. W., Donovan, J., Windish, R., Steining, C., Mahyuddin, P., Womack, J. E., *et al.* (2000) *Genome Res.* **10,** 1359–1368.
8. Everts-van der Wind, A., Kata, S. R., Band, M. R., Rebeiz, M., Larkin, D. M., Everts, R. E., Green, C. A., Liu, L., Natarajan, S., Goldammer, T., *et al.* (2004) *Genome Res.* **14,** 1424–1437.
9. Grobet, L., Martin, L. J., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Menissier, F., Massabanda, J., *et al.* (1997) *Nat. Genet.* **17,** 71–74.
10. Kambadur, R., Sharma, M., Smith, T. P. & Bass, J. J. (1997) *Genome Res.* **7,** 910–916.
11. Takeda, H., Takami, M., Oguni, T., Tsuji, T., Yoneda, K., Sato, H., Ihara, N., Itoh, T., Kata, S. R., Mishina, Y., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 10549–10554.
12. Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., *et al.* (2002) *Genome Res.* **12,** 222–231.
13. Cohen-Zinder, M., Seroussi, E., Larkin, D. M., Loor, J. J., Everts-van der Wind, A., Heon Lee, J., Drackley, J. K., Band, M. R., Hernandez, A. G., Shani, M., *et al.* (2005) *Genome Res.* **14,** 936–944.
14. Itoh, T., Watanabe, T., Ihara, N., Mariani, P., Beattie, C. W., Sugimoto, Y. & Takasuga, A. (2005) *Genomics* **85,** 413–424.
15. Larkin, D. M., Everts-van der Wind, A., Rebeiz, M., Schweitzer, P. A., Bachman, S., Green, C. A., Wright, C. L., Campos, E. J., Benson, L. D., Edwards, J., *et al.* (2003) *Genome Res.* **13,** 1966–1972.
16. Womack, J. E., Johnson, J. S., Owens, E. K., Rexroad, C. E., III, Schläpfer, J. & Yang, Y. P. (1997) *Mamm. Genome* **8,** 854–856.
17. Agarwala, R., Applegate, D. L., Maglott, D., Schuler, G. D. & Schäffer, A. A. (2000) *Genome Res.* **10,** 350–364.
18. Applegate, D., Bixby, R., Chvátal, V. & Cook, W. (1998) *Doc. Math. J. DMV Extra Volume ICM III*, 645–656.
19. Ihara, N., Takasuga, A., Mizoshita, K., Takeda, H., Sugimoto, M., Mizoguchi, Y., Hirano, T., Itoh, T., Watanabe, T., Reed, K. M., *et al.* (2004) *Genome Res.* **14,** 1987–1998.
20. Ben Dor, A. & Chor, B. (1997) *J. Comput. Biol.* **4,** 517–533.
21. Nadeau, J. H. & Taylor, B. A. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 814–818.
22. Fitzgerald, J. & Bateman, J. F. (2004) *Trends Genet.* **20,** 408–412.