

Radiation hybrid map of the porcine genome comprising 2035 EST loci

Anette Rink,^{1,*} Katie Eyer,¹ Benjamin Roelofs,¹ Kimberly J. Priest,^{1,*} Kathleen J. Sharkey-Brockmeier,¹ Sudkamon Lekhong,¹ Elif K. Karajusuf,¹ Jason Bang,¹ Martine Yerle,² Denis Milan,² Wan-Sheng Liu,¹ Craig W. Beattie¹

¹Department of Animal Biotechnology, College of Agriculture, Biotechnology and Natural Resources, University of Nevada, Reno, Nevada 89557, USA

²Laboratoire Genetique et Cytogenetique INRA, Tolosan-Castenet, France

Received: 6 September 2005 / Accepted: 16 March 2006

Abstract

The IMpRH_{7000-rad} radiation hybrid panel was used to map 2035 expressed sequence tags (ESTs) at a minimum LOD score of 4.0. A total of 134 linkage groups covers 57,192 cR or 78% of the predicted size of the porcine and 71% of the human genome, respectively. Approximately 81% (1649) of the porcine ESTs were annotated against the NCBI nonredundant database; 1422 mapped *in silico* to a location in build 35.1 of the human genome sequence (HGS) and 1185 to a gene and location in build 35.1 HGS. The map revealed 40 major breaks in synteny ($1.00e^{-25}$ and lower) with the human genome, 37 of which fall within a single chromosome. At this improved level of resolution and coverage, porcine chromosomes (SSC) 2, 5, 6, 7, 12, and 14 remain "gene-rich" and homologous to human chromosomes (HSA) 17, 19, and 22, while SSC 1, 8, 11, and X have been confirmed to correspond to the "gene-deserts" on HSA 18, 4, 13, and X.

Introduction

Robust, integrated, comprehensive maps were crucial to assembling the sequence of the human

(Venter et al. 2001; Lander et al. 2001, Bentley et al. 2001), mouse (Gregory et al. 2002), and rat genomes (Kwitek et al. 2001). In addition, radiation hybrid (RH) maps form the basis of comparative genomics by identifying regions of orthology between genomes. The power of comparative mapping increases with marker density in the species of interest (Marra et al. 1998) and integration with its genetic map (Breen et al. 2001) because genes can be assigned to a genetic interval (Milan et al. 2002). Successful integration of information between well- and lesser-characterized genomes requires that major breaks in synteny and internal rearrangements be identified, because chromosomal breaks result in fragment inversions, inter- and intrachromosomal translocations, or even loss of genome fragments (Rink et al. 2002a). We reported a first-generation expressed sequence tag (EST) map (Rink et al. 2001a, 2002a) of the porcine genome that integrated 1058 ESTs into our earlier microsatellite (MS)-based, whole-genome radiation hybrid (WG-RH) map for swine (Hawken et al. 1999). This map refined corresponding regions between the human and porcine genomes, identified 60 potential breakpoints in synteny, and improved resolution over the entire genome (Rink et al. 2002a). It also, for the first time, confirmed that synteny of "gene-rich" and "gene-desert" regions of the human genome were conserved in a closely related genome other than the mouse (Hudson et al. 2001). The IMpRH_{7000-rad} panel and an abundant supply of porcine ESTs primarily from immune tissues (Rink et al. 2002b) provide a continuing opportunity to identify genomic intervals that contain gene sequences in a species of interest, improve the comparative map, and hasten the assembly of the porcine genome sequence. We have hastened that

*Present address: Animal Disease and Food Safety Laboratory, Nevada Department of Agriculture, Reno, Nevada

Correspondence to: Craig W. Beattie, Department of Animal Biotechnology, College of Agriculture, Biotechnology and Natural Resources (CABNR), University of Nevada, Mail Stop 202, 1664 N. Virginia Street, Reno, NV 89557, USA; E-mail: cbeattie@cabnr.unr.edu

assembly by constructing a second-generation EST map of the porcine genome and used it to improve the resolution around putative synteny breaks.

Materials and methods

Marker development and amplification. All ESTs were cloned from a panel of normalized cDNA libraries (Rink et al. 2002b). Sequences were annotated using gapped Advanced BLAST (Altschul et al. 1997) against dbEST, the nonredundant part of NCBI and international protein databases. Primers for the original 1058 ESTs and 977 new ESTs analyzed were designed to amplify an optimal 140-bp product size (range = 100–200 bp), 20-bp optimal primer length, 45%–60% GC content with PRIMER3 (http://www.frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). All primers were optimized by determining the highest annealing temperature at which successful amplification of porcine genomic DNA took place. Primers were then tested with porcine and Chinese hamster genomic DNA at that temperature for species specificity. During the course of this study, we avoided designing redundant primer pairs for a given gene or EST.

Annotation of ESTs. Sequences of the 2035 porcine ESTs were subjected to a BLASTN search against build 35.1 (May 2004) of the human genome sequence (HGS) and all other sequences deposited in GenBank by the University of Minnesota, Center for Computational Genomics and Bioinformatics (CCGB) (<http://www.cccb.umn.edu/>). All sequences of the porcine ESTs were also BLAT-searched against the HGS (build 35.1) (<http://www.genome.ucsc.edu/cgi-bin/hgBlat?command=start>), which allowed us to quickly identify sequences of 95% and greater similarity of length 40 bases or more. The most significant match to a human sequence in the results was documented with the human chromosome number, gene (ortholog) symbol, score, *e* value, and start position (bp) in the human sequence. An *e* value $\leq 1 \times 10^{-5}$ was set as a threshold based on the significant number of sequences in the data set and the need to return nearly exact matches of greater than 50 bp. An annotation threshold was set at a score of greater than 125 and $e \leq 10^{-25}$ to ensure the integrity of the final comparative maps. Data can be accessed online at http://www.cabnr.unr.edu/beattie/docs/complete_info.pdf.

Genotyping. The IMpRH_{7000-rad} panel (Yerle et al. 1998; Hawken et al. 1999; Rink et al. 2002a) was used in this study. All PCR typing reactions were performed in 96-well Techne® Touchgene

thermocyclers (Techne Inc., Princeton, NJ). Each PCR reaction contained 25 ng of hybrid DNA, 0.4 μ M of each primer, 1.5 mM MgCl₂, 66 μ M of each dNTP, 1 \times cresol loading dye (Hodges et al. 1997), and 0.3 U Immolase™ DNA Polymerase and 1 \times PCR buffer (Bioline USA Inc., Randolph, MA) in a total volume of 15 μ l. The amplification cycle included an initial 95°C, 7-min denaturation step, followed by cycling between 94°C for 15 sec, 57–69°C for 30 sec, 72°C for 30 sec, and a final extension step at 72°C for 5 min. Controls consisted of porcine and hamster genomic DNA and a reaction containing no DNA. The first PCR was routinely run for 40 cycles; the second PCR was run for between 33 and 40 cycles depending on the level of background amplification and intensity of pig-specific bands. If more than six discrepancies were observed between the first and second PCR, a third PCR was performed. The PCR products were electrophoresed on 2% agarose gels, visualized, and photographed with an AlphaImager™ 2200 (Alpha Innotech Corp, San Leandro, CA). At least two gel images per marker were independently scored by two individuals using GelScore (<http://www.wesbarris.com/GelScore/>). Consensus vectors were established by two individuals using GelScore data as a baseline. Markers were scored as present (1), absent (0), and ambiguous (2), and those with unusually high or low retention frequencies or with more than six discrepancies were eliminated.

Construction of the IMpRH_{7000-rad} map. All EST vectors were initially subjected to a two-point analysis against framework microsatellites using the online IMpRH mapping tool (<http://www.imprh.toulouse.inra.fr/Action=Menu?Do=Map+one+marker+on+IMpRH&USER=&PASS=>) that assigned each EST to 1 of 19 individual input files (one for each autosome, plus X) based on the closest-linked MS framework marker. Input files contained all the EST and MS vectors for a given chromosome. The CarthaGene software (Schiex and Gaspin 1997; deGivery et al. 2005) was then used to build and analyze the RH map in the following manner: Markers on each chromosome were grouped into distinct linkage groups using a two-point threshold set at LOD 4 and a distance threshold of less than 100 cR between any two markers. Any marker not assigned to a linkage group was labeled as a singleton. Linkage groups were analyzed individually using simulated annealing to improve CarthaGene's default marker order. Any improvements in order were put into resident memory as the new best map. After the simulated annealing was completed, the best map for that linkage group was subjected to a

second improving method where marker orders were flipped internally. A window size of at least four markers was applied, with the largest linkage groups using a window size of either six or eight markers. The MS locations on the final best map were compared to the genetic map (<http://www.marc.usda.gov/genome/swine/swine.html>) to verify MS order. If a discrepancy occurred between MS order within the linkage groups and MS order on the genetic map, chromosomes were recalculated at LOD 6 and 8. Any EST or MS vectors potentially causing the rearrangements were removed from the associated linkage group and the order of the entire linkage group recalculated. If the removal of a marker corrected a discrepancy in the linkage group, it was reclassified as a singleton and not analyzed further. One hundred ninety-one ESTs were removed from a linkage group because they were reclassified as a singleton. All putative breaks in synteny fell within linkage groups in the RH maps. At these LOD levels we are confident that any marker identified as a break in synteny is in the correct relative position within the swine genome. Final maps were drawn using MapCreator (<http://www.wesbarris.com/mapcreator/>).

Comparative mapping. Chromosomal homologies between the swine and human genomes were determined essentially following an idea described by Ehrlich et al. (1997) and Bourque et al. (2004). First, chromosomal homologies between the two species were determined based on previous gene comparative mapping and bidirectional chromosome painting results (<http://www.toulouse.inra.fr/lgc/pig/compare/>). Bidirectional painting indicates that individual swine chromosomes exhibit synteny with between 1 and 5 human chromosomes with a total of 39 identifiable chromosomal breaks (Goureau et al. 1996). Second, synteny breaks refer to adjacent porcine EST markers that do not match their predicted human correspondence. In the present study, any markers that did not match the predicted human chromosome region as its top search result were initially labeled as putative synteny breaks and were subject to further analysis. Third, each marker identifying a putative synteny break was subjected to a manual BLAT search against HGS (build 35.1) on the UCSC website (<http://www.genome.ucsc.edu/>) with the same parameters used for CCGB searching, and the search results were checked for any annotation anywhere in the list that matched the predicted human segment. If any annotation to the predicted human segment was found, the putative synteny break marker was then checked with its adjacent markers in the pig RH linkage group to see if that

human annotation was in the corresponding location (bp). If this marker and the adjacent markers were in synteny in the human genome and in the correct location, this marker was no longer considered a synteny break. Finally, the number of synteny blocks between the pig and the human genome were counted based on the method of Bourque et al. (2004) and the data was integrated with the above analysis and our RH mapping results. If all porcine markers in a given RH linkage group were in agreement with established human synteny, it was called one synteny block (conserved segment between the two species).

Conservation of synteny and genome coverage. The resolution a RH map is able to provide is dependent on the resolution of the RH panel and on the number and distribution of markers in the data set. The swine genome is estimated to span approximately 2.7 Gb (Schmitz et al. 1992), with varying marker densities across the individual swine chromosomes due to various known "gene-rich" regions and "gene-deserts" (Rink et al. 2002b). Published predicted averages for RH resolution of the IMpRH_{7000-rad} panel range from 37 kb/cR₇₀₀₀ (Yerle et al. 2002) to 75 kb/cR₇₀₀₀ (Hawken et al. 1999). A comparative analysis between the IMpRH_{7000-rad} and IMNpRH_{212000-rad} panels found an average resolution of 37 kb/cR₇₀₀₀ (Yerle et al. 2002). We calculated the coverage of the swine genome as:

$$C = R \times L/S \text{ or } (37 \times 57191.8/2.7 \times 106) = 78\%$$

where C = coverage, R = RH map resolution (kb/cR), L = length of RH maps (cR), S = size of swine genome (kb).

Chromosomal locations and start positions of porcine orthologs of genes in the human genome were also established for all porcine genes analyzed using the NCBI human Map Viewer (build 35.1) (<http://www.ncbi.nlm.nih.gov/mapview/>). Map distances (cR) were the accumulated sum of all linkage groups. Coverage of the human genome, currently thought to span approximately 3.011 Gb (NCBI), was estimated using the start position of ESTs that returned a BLAST search result with $e \leq 10^{-5}$. All markers with a human genome position annotation were ordered by their human chromosome number and then by their position. The distance between each marker was calculated by subtracting the position of every marker from the next-closest marker. The resulting value is the size of the gap between the orthologs of those two markers. This gap value was held in memory. If the resulting gap value was greater than or equal to the gap threshold, the actual

gap distance was recorded. All gap values greater than the threshold for individual chromosomes were then subtracted from the individual human chromosome length. The remaining value was the approximate comparative coverage of the chromosome that the annotated data represented, discounting gaps smaller than the threshold. As more annotated markers are added to the data set, the actual coverage and comparative coverage continuously improve. Two thresholds were used for our computations: 10 Mb (<0.332%) and 5 Mb (<0.166%). This threshold is referred to as the gap threshold. To avoid having to manually recompute the coverage whenever new data became available, a Perl script was developed and will be available for download at <http://www.ag.unr.edu/beattie/research.htm>. The script requires a tab-delimited input file containing three fields: marker name (or any unique identifier), human chromosome number (1–22 or X), and human genome position (bp). The gap value threshold is chosen by the user at runtime. The output file is a tab-delimited text file.

Results and Discussion

Genome coverage. Table 1 illustrates the number and characteristics of markers mapped on the IMpRH_{7000-rad} panel. Genome-wide marker distribution was not stochastic in the first-generation RH map (Rink et al. 2001a, b, c) or in this second, more detailed version. The DNA/marker ratio ranged from 0.53 for SSC 11 to 2.02 for SSC 12, respectively. Average marker retention frequency (RF) was 30.3%. We mapped a total of 977 new ESTs onto the first-generation map of 1058 ESTs and 743 MS (Rink et al. 2002a). The genome-wide RF for the new markers was 30.2% with a minimum of 9.3% (EST-AR078F01 on SSC 13 and EST-UNR6191H10 on SSC 14) and a maximum of 87.3% (EST-AR095G02 on SSC X). High RF in the region directly surrounding the centromere, called “centromeric effect,” was observed in several species including human (Gyapay et al. 1996; Stewart et al. 1997), cattle (Kurar et al. 2003), and chicken (Pitel et al. 2004; Rabie et al. 2004). It was also observed in pig on the centromere of SSC 12 when the 12,000-rad IMNpRH2 panel was used (Liu et al. 2005). To analyze whether this “centromeric effect” is present in other porcine chromosomes, the RFs of all 2035 ESTs were plotted in histograms for each respective chromosome. Because we did not have a single complete linkage group for each chromosome, the ESTs were plotted in the order in which they were placed on the RH maps. By examining which MS (on both the genetic and the RH maps) was nearest the centromere of any given

Table 1. Number and characteristics of markers on the IMpRH_{7000-rad} panel by chromosome

Porcine chromosome (SSC) number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	X
Total number of markers mapped	219	227	185	173	181	225	199	124	153	78	57	134	202	187	135	73	67	59	100
Microsatellite (MS)	59	46	47	49	39	60	42	55	39	27	20	22	62	59	37	25	16	11	28
Expressed Sequence Tag (EST)	160	181	138	124	142	165	157	69	114	51	37	112	140	128	98	48	51	48	72
2-pt LOD score ^a	6	6	6	6	6	6	6	6	6	6	4	8	6	8	6	4	4	6	6
Number of linkage groups	18	6	6	7	7	10	10	10	10	7	4	5	3	10	11	2	6	2	8
Number of EST markers not in linkage groups ^b	6	26	13	9	4	24	9	8	11	8	2	9	13	15	9	3	5	5	12
Number of syntenic human chromosomes ^b	5	3	3	2	2	4	3	1	3	3	1	2	4	2	1	1	1	1	1
ESTs that find matches outside syntenic chromosomes	4	2	2	0	2	0	2	3	3	0	0	1	1	2	8	6	4	1	0
Average retention frequency (RF)	0.32	0.29	0.33	0.32	0.37	0.27	0.32	0.31	0.32	0.32	0.33	0.28	0.24	0.26	0.32	0.4	0.27	0.34	0.42
Average RF of ESTs flanking the centromere ^c	0.44	0.41	0.44	0.4	0.43	0.38	0.55	0.37	0.44	0.34	0.46	0.45	0.28	0.28	0.31	0.4	0.25	0.29	0.44

^aThe 2-pt LOD score criterion for creating the second-generation porcine EST radiation hybrid map was set at 4 or higher. Good statistical support for local gene order was considered to be more useful than creating large linkage groups. However, with increasing marker density, complete chromosomal coverage through a minimum number of linkage groups is rapidly expanding.

^bBased on existing chromosomal paintings (Goureau et al., 1996).

^cThe average retention frequency for chromosomes 1–12 and X was determined by selecting ESTs surrounding the predicted centromere. The number of ESTs used for each chromosome varied from 5 (SSC 6) to 16 (SSC 10). For chromosomes 13–18, the top 10 ESTs were used.

Table 2. Chromosomal coverage by the current ImpRH_{7000-rad} map

<i>Chromosome</i>	<i>Size (cR)</i>	<i>Size (Mb)^a</i>	<i>Resolution (kb/cR)</i>
1	5649.3	295	52.2
2	3399	168	49.4
3	3736.4	149	39.9
4	3974.1	146	36.7
5	3407.3	118	34.6
6	4240.3	177	41.7
7	3666.4	141	38.5
8	2790.5	158	56.6
9	3069.6	145	47.2
10	1441.9	103	71.4
11	1721.2	94	54.6
12	2287.2	74	32.4
13	4319.3	230	53.2
14	3290	168	51.1
15	3142	161	51.2
16	1727.9	101	58.5
17	1698.8	77	45.3
18	1273.1	68	53.4
X	2357.5	145	61.5
Total	57191.8	2718	47.5

^aSchmitz et al. (1992).

chromosome, the ESTs flanking that marker were easily identified and visualized on the histogram. After all chromosomes were analyzed, we found that all submetacentric (SSC 1-7), and metacentric (SSC 8-12, X) chromosomes show higher RF around the centromere (Table 1).

Markers were assigned at LOD 4 on SSC 11, 16, and 17; LOD 6 on SSC 1-10, 13, 15, 18, and X; and LOD 8 on SSC 12 and 14. The 191 ESTs not assigned to a linkage group (singletons) were not included in the graphical representation of the final maps (http://www.ag.unr.edu/beattie/research/second_generation.htm) to conserve space. Two-point LOD scores to the closest framework MS for these singletons, as well as complete information on all ESTs, including annotation, can be found online at http://www.cabnr.unr.edu/beattie/docs/complete_info.pdf. Map resolution ranged from 32.4 kb/cR₇₀₀₀ on SSC 12 to 71.4 kb/cR₇₀₀₀ on SSC 10 (Table 2). The average resolution across the entire genome was calculated at 47.5 kb/cR₇₀₀₀, midway between earlier estimates of 37 kb/cR₇₀₀₀ (Yerle et al. 2002) to 75 kb/cR₇₀₀₀ (Hawken et al. 1999).

A total of 134 linkage groups covered 57,192 cR or 78% of the predicted size of the porcine genome (Table 2) and 71% of the human genome (Table 3), respectively. At least 1649 (81%) ESTs had an annotation in GenBank with 1422 ESTs (70%) also matching the human genome data set housed at the University of California, Santa Cruz (<http://www.genome.ucsc.edu/>). Of those 1422 ESTs, 1158 had an

annotation to a specific human gene, while the remaining 264 ESTs returned a definite location but did not match a specific gene in build 35.1 HGS. With a 10-Mb threshold, our coverage ranged from 41% on HSA 21 to 100% on HSA 17 and 20. The average coverage across the entire human genome is 71% (Table 3). At a threshold of 5 Mb, our coverage ranged from 12% on HSA 21 to 90% on HSA 17 (Table 3). A scatterplot representation of coverage of individual human chromosomes, including X, is presented in Fig. 1. Individual chromosomal scatterplots positioned next to their human ideograms are available at http://www.cabnr.unr.edu/beattie/research/hsa_coverage.htm.

Conservation of synteny. Porcine Chromosomes 2, 5, 6, 12, and 14 were confirmed to be "gene-rich" relative to HSA 17, 19, and 22 (Venter et al. 2001, and extrapolated from NCBI build 35.1 HGS) suggesting significant conservation of synteny on these chromosomes. We refer to "rich" using an average of 10.0 genes/Mb in the human genome (Venter et al. 2001, and NCBI build 35.1 HGS). Autosomal chromosomes in the human genome have an average of 9.7 genes/Mb, 9.4 when HSA X and HSA Y are included. We also observed that SSC 3 (HSA 2, 7, 16), SSC 4 (HSA 1 and 8), SSC 7 (HSA 4, 6, 14-16), and SSC 14 (HSA 1, 8-10, 12, 22), are also "gene-rich" using this convention, while SSC 1, 8, 11, and X continue to correspond to the "gene-deserts" on HSA 18, 4, 13, and X (Fig. 2). The known "gene-deserts" in the swine genome, SSC 1, 8, 11, and X, that correspond to HSA 18, 4, 13, and X had 72%, 58%, 66%, 53%, and 66% of the markers they statistically should have received, respectively (Fig. 2). In contrast to the first-generation porcine EST RH map (Rink et al. 2002a), SSC 10 received less than the number of "genes" it should have based on the amount of DNA present in the chromosome (Schmitz et al. 1992). Interestingly, SSC 10 corresponds to regions on HSA 1, 9, and 10, which are not listed as chromosome deserts in the HGS. SSC 2, 5, 6, 7, and 12 clearly correspond to the gene-rich HSA 17, 19, and 22. SSC 6, a new addition to the current set of chromosomes that appears "gene-rich," received 124% of the genes it statistically should have. SSC 6 exhibits synteny with a total of four human chromosomes: HSA 1, 16, 18, and 19 including the "gene-desert" associated with HSA 18. Swine autosomes 3, 4, and 14 show an increase in gene density compared with the first-generation map (Fig. 2) and present the porcine equivalent of "gene-rich" and "gene-desert" human chromosomes, avg. 9.3, 8.5, 9.3/Mb, respectively. However, the current porcine RH maps should be considered preliminary

Table 3. Coverage of human genome by porcine linkage groups

HSA Chromosome	HSA length (Mb)	Gaps >10 Mb (Mb)	Gaps >5 Mb (Mb)	Coverage - 10 Mb (%)	Coverage - 5 Mb (%)
1	245.0	37.2	104.4	84.8	57.4
2	242.0	44.7	140.5	81.5	41.9
3	200.0	41.5	110.1	79.3	44.9
4	191.0	72.1	148.4	62.3	22.3
5	180.0	60.0	123.4	66.7	31.5
6	171.0	29.6	111.2	82.7	35.0
7	158.0	51.6	97.8	67.3	38.1
8	146.0	34.4	80.3	76.4	45.0
9	138.0	72.0	98.2	47.8	28.8
10	135.0	29.6	65.9	78.1	51.2
11	134.0	34.6	62.7	74.2	53.2
12	132.0	15.3	41.1	88.4	68.8
13	114.0	53.8	72.9	52.8	36.1
14	106.0	42.2	57.4	60.2	45.9
15	100.0	20.6	58.0	79.4	42.0
16	88.0	26.4	54.5	70.0	38.0
17	80.0	0.0	8.0	100.0	90.0
18	76.0	25.1	56.3	67.0	26.0
19	64.0	15.8	15.8	75.3	75.3
20	62.0	0.0	29.7	100.0	52.1
21	46.0	27.1	40.1	41.0	12.8
22	49.0	16.1	16.1	67.2	67.2
X	154.0	56.4	89.2	63.4	42.1
			Total	72.4	45.5

until a higher resolution is reached and full integration of linkage, cytogenetic, and RH maps has been achieved.

We identified 95 ESTs that mark putative breakpoints in synteny after the initial BLASTN search at CCGB (<http://www.ccg.umn.edu/>). However, 21 of the 95 putative breaks in synteny

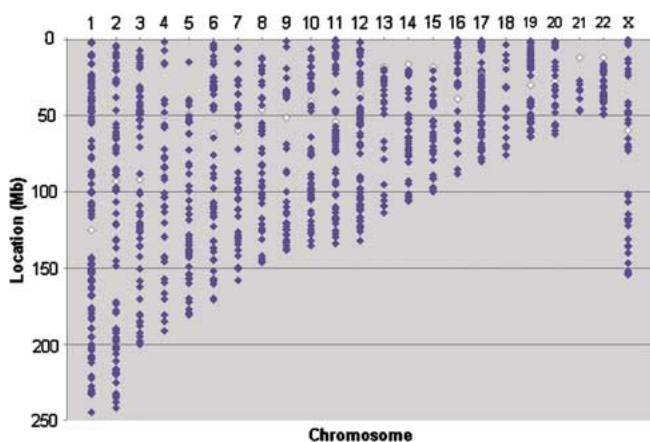


Fig. 1. Linear scatterplot of human chromosomes 1-22 and X. Each data point represents the physical base pair location of each of the 1422 porcine ESTs with strong annotation in build 35.1 (NCBI). Centromeres are indicated by white diamonds. This set of data represents approximately 70% of the total number of ESTs in the study. The remaining 30% at this time do not have any annotation to a position within the human genome.

were eliminated because these markers were subsequently found to be adjacent to ESTs in the correct location (bp) in the human genome when rechecked using a manual BLAT search against build 35 on the UCSC website (<http://www.genome.ucsc.edu/>). The remaining 74 markers were subsequently considered possible breaks in synteny, without setting any significant *e*-value threshold. The score of all remaining putative breaks in synteny ranged widely from 21 to 650 (average of

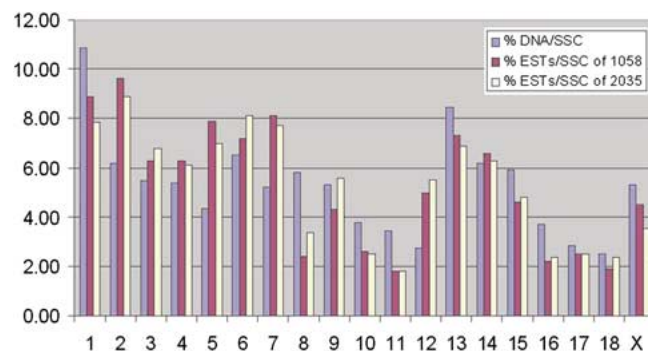


Fig. 2. Identification of "gene-deserts" and "gene-pools" in the porcine genome. SSC 1, 8, 11, and X correspond with HSA 18, 4, 13, and X, respectively. These represent "gene-deserts" both in the human and the porcine genomes. A total of six porcine chromosomes (SSC 2, 5, 6, 7, 12, and 14) are syntenic with the three most "gene-rich" human chromosomes (HSA 17, 19, and 22).

222.81). *E* values ranged from 4.16×10^{-6} to 0.00 (with 14 markers having *e* values $<10^{-100}$). The length, read quality, score, and *e* value of all putative breaks were then examined. Shorter ESTs had a strong tendency toward low scores and *e* values approaching the cutoff value when annotated. An annotation threshold was set at a score of greater than 125 and $e \leq 10^{-25}$ to ensure the integrity of the final comparative maps. At this level of stringency, 34 of the 74 breaks did not meet the threshold, leaving 40 markers considered truly significant breaks in synteny, where each marker represents one syntenic block. We also identified at least 90 synteny blocks (123 if all 74 possible breaks were considered) between the pig and human genomes. This number is significantly smaller than the reported 394 between human and mouse, and the 417 between human and rat (Bourque et al. 2004). The map revealed 40 breaks in synteny ($1.00e^{-25}$ and lower) with the human genome, extremely close to the 39 breaks identified by reverse chromosomal painting (Goureau et al. 1996). Marker numbers on this version of the physical map of the swine genome are still significantly lower than on the mouse or rat genome map, respectively. Therefore, the current number of syntenic breaks between the human and the porcine genome map should be considered preliminary.

Acknowledgments

The authors thank Dr. Ernie Retzel at the Center for Computational Genomics and Bioinformatics (CCGB), University of Minnesota, for database management and support. This project was supported by USDA NRI-CGP 2003-03686.

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402
- Bentley DR, Deloukas P, Dunham A, French L, Gregory SG, et al. (2001) The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature* 409, 942–943
- Breen M, Jouquand S, Renier C, Mellersh CH, Hitte C, et al. (2001) Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Genome Res* 11, 1784–1795
- Bourque G, Pevzner PA, Tesler G (2004) Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. *Genome Res* 14, 507–516
- deGivery S, Bouchez M, Chabrier P, Milan D, Schiex T (2005) CARHTAGENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics* 21, 1703–1704
- Ehrlich J, Sankoff D, Nadeau JH (1997) Synteny conservation and chromosome rearrangements during mammalian evolution. *Genetics* 147, 289–296
- Goureau A, Yerle M, Schmitz A, Riquet J, Milan D, et al. (1996) Human and porcine correspondence of chromosome segments using bidirectional chromosome painting. *Genomics* 36, 252–262
- Gregory SG, Sekhon M, Schein J, Zhao J, Osoegawa K, et al. (2002) A physical map of the mouse genome. *Nature* 418, 743–750
- Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, et al. (1996) A radiation hybrid map of the human genome. *Hum Mol Genet* 5, 339–346
- Hawken RJ, Murtaugh J, Flickinger GH, Yerle M, Robic A, et al. (1999) A first-generation porcine whole-genome radiation hybrid map. *Mamm Genome* 10, 824–830
- Hodges E, Boddy SM, Thomas S, Smith J (1997) Modification of IgH PCR clonal analysis by the addition of sucrose and cresol red directly to PCR reaction mixes. *Mol Pathol* 50, 164–166
- Hudson TJ, Church DM, Greenaway S, Nguyen H, Cook A, et al. (2001) A radiation hybrid map of mouse genes. *Nat Genet* 29, 201–205
- Kurar E, Womack JE, Kirkpatrick BW (2003) A radiation hybrid map of bovine chromosome 24 and comparative mapping with human chromosome 18. *Anim Genet* 34, 198–204
- Kwitek AE, Tonellato PJ, Chen D, Gullings-Handley J, Cheng YS, et al. (2001) Automated construction of high-density comparative maps between rat, human, and mouse. *Genome Res* 11, 1935–1943
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
- Liu WS, Eyer K, Yasue H, Roelofs B, Hiraiwa H, et al. (2005) A 12,000-rad porcine radiation hybrid (IMNpRH2) panel refines the conserved synteny between SSC12 and HSA17. *Genomics* 86, 731–736
- Marra MA, Hillier L, Waterston RH (1998) Expressed sequence tags—ESTablishing bridges between genomes. *Trends Genet* 14, 4–7
- Milan D, Schiex T, Yerle M, Rink A, Beattie C, et al. (2002) Integration of genetic and radiation hybrid maps of the pig: The second generation IMpRH maps. *Plant, Animal and Microbe Genomes Conference X*, San Diego, CA, January 12–16, P229
- Pitel F, Abasht B, Morisson M, Crooijmans RP, Vignoles F, et al. (2004) A high-resolution radiation hybrid map of chicken chromosome 5 and comparison with human chromosomes. *BMC Genomics* 5, 66
- Rabie TS, Crooijmans RP, Morisson M, Andryszkiewicz J, van der Poel JJ, et al. (2004) A radiation hybrid map of chicken chromosome 4. *Mamm Genome* 15, 560–569
- Rink A, Santschi EM, Hess M, Eyer KE, Godfrey ML, et al. (2001a) A First Generation EST Radiation Hybrid

- Map of the Porcine Genome. Plant and Animal Genome IX, San Diego, CA, January 13–17: 2001
22. Rink A, Santschi EM, Eyer KE, Hess M, Karajusuf EK, et al. (2001b) Zooming in on the Porcine–Human Comparative Map. 12th North American Colloquium Animal Cytogenetics and Gene Mapping, Davis, CA, July 5–19, 2001
 23. Rink A, Santschi EM, Eyer KE, Hess M, Roelofs B, et al. (2001c) Divergent Molecular and Structural Resolution between the Porcine and Human Genomes. Vth Congress of the International Xenotransplantation Association, “Meeting the Challenges of the new millennium.” Chicago, IL, September 29–October 3, 2001
 24. Rink A, Santschi EM, Hess M, Eyer KE, Godfrey ML, et al. (2002a) A first generation est radiation hybrid comparative map of the porcine and human genome. *Mamm Genome* 13, 578–587
 25. Rink A, Santschi EM, Beattie CW (2002b) Normalized cDNA libraries from a porcine model of orthopedic implant associated infection. *Mamm Genome* 13, 198–205
 26. Schiex T, Gaspin C (1997) CARTHAGENE: constructing and joining maximum likelihood genetic maps. *Proc Int Conf Intell Syst Mol Biol* 5, 258–267
 27. Schmitz A, Chaput B, Fouchet P, Guilly MN, Frelat G, et al. (1992) Swine chromosomal DNA quantification by bivariate flow karyotyping and interpretation. *Cytometry* 13, 703–710
 28. Stewart EA, McKusick KB, Aggarwal A, Bajorek E, Brady S, et al. (1997) An STS-based radiation hybrid map of the human genome. *Genome Res* 7, 422–433
 29. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. *Science* 291, 1304–1351
 30. Yerle M, Pinton P, Delcros C, Arnal N, Milan D, et al. (1998) Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs. *Cytogenet Cell Genet* 82, 182–188
 31. Yerle M, Pinton P, Delcros C, Arnal N, Milan D, et al. (2002) Generation of a 12,000-rad radiation hybrid panel for fine mapping in pig. *Cytogenet Genome Res* 97, 219–228