Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 86 (2005) 739 – 752

GENOMICS

www.elsevier.com/locate/ygeno

Piggy-BACing the human genome II. A high-resolution, physically anchored, comparative map of the porcine autosomes

Stacey N. Meyers[a], Margarita B. Rogatcheva[a], Dennis M. Larkin[a], Martine Yerle[b], Denis Milan[b], Rachel J. Hawken[c], Lawrence B. Schook[a], Jonathan E. Beever[a,*]

[a]*University of Illinois at Urbana–Champaign, 220 Edward R. Madigan Laboratory, MC-051, 1201 West Gregory Drive, Urbana, IL 61801, USA*
[b]*Laboratoire de Génétique Cellulaire, INRA, Toulouse, France*
[c]*CSIRO, University of Queensland, St. Lucia, QLD, Australia*

Received 18 February 2005; accepted 19 April 2005
Available online 24 October 2005

Abstract

Using the INRA–Minnesota porcine radiation hybrid panel, we have constructed a human–pig comparative map composed of 2274 loci, including 206 ESTs and 2068 BAC-end sequences, assigned to 34 linkage groups. The average spacing between comparative anchor loci is 1.15 Mb based on human genome sequence coordinates. A total of 51 conserved synteny groups that include 173 conserved segments were identified. This radiation hybrid map has the highest resolution of any porcine map to date and its integration with the porcine linkage map (reported here) will greatly facilitate the positional cloning of genes influencing complex traits of both agricultural and biomedical interest. Additionally, this map will provide a framework for anchoring contigs generated through BAC fingerprinting efforts and assist in the selection of a BAC minimal tiling path and assembly of the first sequence-ready map of the porcine genome.

Keywords: Comparative mapping; Radiation hybrid; BAC; Porcine

A primary focus of the animal genetics field is the elucidation of genes influencing diverse phenotypes of both agricultural and biomedical relevance. Most of these phenotypes are genetically complex, i.e., controlled by multiple genes occupying chromosomal positions referred to as quantitative trait loci (QTL). Mapping of QTL has become a common first step toward understanding the molecular basis of complex genetic traits and has provided the impetus for developing detailed genome maps in agricultural species. With this aim, more than a decade of research has been devoted to mapping and characterizing the genomes of domestic livestock species, including the pig (*Sus scrofa domestica*). Moderate-resolution genetic linkage maps of the porcine chromosomes have been produced using different mapping populations [1–4] (see also http://www.marc.usda.gov/genome/genome.html) and have facilitated the identification of chromosomal regions harboring QTL of interest to the swine genomics community [5]. To refine the map positions of these QTL and make use of valuable genome sequence information from extensively studied species such as the human [6,7] and mouse [8], a number of physical and comparative mapping techniques, including somatic cell hybrid analysis, *in situ* hybridization, and ZOO-FISH, have been employed. However, as with linkage mapping, these techniques lack the resolution necessary for positional candidate cloning and detailed molecular analysis of QTL.

Whole-genome radiation hybrid (WG-RH) mapping [9] appears to be the most effective physical and comparative mapping method used to date. Unlike other mapping techniques, the WG-RH method can be used to construct ordered maps containing both polymorphic markers, such as

* Corresponding author. Fax: +1 217 244 5617.
E-mail address: jbeever@uiuc.edu (J.E. Beever).

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.04.010

microsatellites used for linkage map construction, and nonpolymorphic markers, such as expressed sequence tags (ESTs). This unique feature enables the ordering of loci in chromosomal regions of low recombination that are unable to be resolved by meiotic mapping, aids in the estimation of physical distances between loci [10], allows for efficient integration of the physical and linkage maps, and facilitates the construction of detailed comparative maps.

The construction [11] and application [12,13] of WG-RH panels for mapping has expedited the development of physical and comparative maps of the porcine genome. The availability of these panels as well as the recent accumulation of porcine sequence information in public databases (~560,000 porcine sequences as of January 2005; http://www.ncbi.nlm.nih.gov/) provides a tremendous resource for the rapid generation of a high-resolution porcine RH map. In 2002, Rink and colleagues [13] took advantage of this resource and constructed an RH map composed of 1058 porcine ESTs. Following the notion that nearly all of these coding sequences are likely to be conserved across mammalian species, this map was then used to form the basis of comparative analysis between the human and the pig genomes. This human–pig comparative map improved the overall resolution of the porcine genome map and identified at least 60 breakpoints and 90 microrearrangements between the two genomes.

Although significant progress has been made in mapping the porcine genome, the individual RH maps for most chromosomes remain noncontiguous as a result of experimental design limitations. The strategy used to create the current RH maps involved mapping a set of markers representing a random sample of nonredundant porcine EST sequences [13]. The selection of only EST sequences for mapping limits the amount of usable sequence information as ESTs represent only coding sequences. Therefore, "gene deserts" or gene-poor regions of the porcine genome may be neglected using this approach. Moreover, the selection of random loci from this subset of usable sequences further impedes the mapping of markers with the spatial distribution needed to produce contiguous maps. Thus, a random EST mapping approach often requires the mapping of excessive numbers of loci to minimize the number of linkage groups. Finally, many EST sequences, e.g., those representing members of gene families, share sequence similarity with multiple genome locations, i.e., paralogous sequences. Rink et al. [13] reported that only ~10% of all mapped pig ESTs had a unique sequence similarity within the human genome, with an average of three similar sequences per EST. Consequently, it is possible for these nonunique sequences to be positioned ectopically on the comparative map. This may explain many of the singletons and single-locus synteny groups that have been provisionally assigned to the current comparative maps.

The recent availability of bacterial artificial chromosome (BAC) libraries [14–16] (see also http://bacpac.chori.org/porcine242.htm) representing approximately 31× coverage of the porcine genome offers relief from these experimental limitations and allows for a more targeted approach to physical and comparative mapping. End-sequencing of approximately 53,000 genomic inserts from these libraries (M. Rogatcheva et al., submitted for publication) has provided a previously untapped source of both coding and noncoding porcine sequence information. Exploitation of this resource as well as the complete human sequence and bioinformatics tools allows for the establishment of an ordered list of unique sequences from which to select evenly spaced markers prior to mapping. This preselection of evenly spaced, unique sequences facilitates the construction of contiguous maps as well as preventing the inconclusive mapping of paralogous loci. Additionally, this strategy utilizes physically anchored sequences, i.e., derived from BAC clones, and thus permits the rapid integration of data from BAC fingerprinting efforts and provides the necessary reagents for fine-mapping and sequencing efforts. Here we report the application of this targeted approach toward the construction of the first high-resolution, physically anchored, contiguous WG-RH comparative maps of the porcine autosomes.

Results

We have constructed a high-resolution radiation hybrid map of the porcine autosomes composed of 2274 markers, including BAC-end sequences (BESs) and ESTs (Fig. 1, Supplementary Table A). Our original intention was to improve the resolution of the current EST RH map [13] by mapping additional EST loci. Therefore, 206 (~9%) of the newly mapped markers are porcine ESTs and are predominantly found on chromosome maps that were constructed first. The remaining 2068 (~91%) new markers are porcine BESs that were primarily chosen based on unique similarity to the human genome sequence as well as the relative genomic position of each orthologous human sequence. Using the human genome sequence coordinates as a guide, BESs were typically selected at 1- to 1.5-Mb intervals. Consequently, the average spacing between comparative anchor loci across all chromosomes is 1.15 Mb relative to the human genome sequence, ranging from 0.89 Mb for HSA13 to 1.35 Mb for HSA2 (Table 1). Greater than 95% (2137/2235) of all marker intervals were $\leq$2 Mb in size, also demonstrating uniformity of anchor loci spacing (Table S3). Additionally, 647 microsatellite markers from the first-generation porcine WG-RH map [12] were incorporated into our map. However, due to systematic differences in data generation between labs, integration of these markers significantly expanded the RH map; therefore, these data are not presented here, but are included in Supplementary Table 1.

Using a lod score threshold of 6 and a distance threshold of 50 cR, all 2274 loci were assigned to 34 linkage groups (Table 2). Each arm of the metacentric autosomes (SSC1–

SSC12) and the acrocentric autosomes (SSC13–SSC18) is represented by a single linkage group, with four exceptions; SSC1q, SSC2q, SSC3p, and SSC16 are each represented by two linkage groups. Linkage groups were oriented based on the previously determined order of the integrated microsatellite markers [3,4,12] (see also http://www.marc.usda.gov/genome/genome.html). It was assumed that the existence of centromeres was responsible for breaks in linkage, and therefore centromeres were assumed to be located between the two linkage groups of each metacentric autosome. In the cases of metacentric autosomes with three linkage groups (SSC1–SSC3), centromere position assumptions were based on in situ hybridization data [17–19] as well as the lengths of the linkage groups compared to the relative length of each chromosome arm.

The length (cR_{7000}) of each chromosome map is reported in Table 2. These values were determined by summing the lengths of the individual linkage groups. No additional length was included to account for gaps or centromeres, and thus the reported lengths are underestimated. Chromosome map lengths were generally consistent with relative chromosome sizes [20]. As would be expected, the longest chromosome maps are those of the largest porcine chromosomes, SSC1 and SSC13, and the shortest chromosome map is that of the smallest porcine autosome, SSC18. Additionally, as the selection of evenly spaced markers determines the number of markers mapped for each chromosome, the SSC1 and SSC13 maps comprise the greatest numbers of mapped markers (Table 2). Likewise, the SSC18 map reflects the fewest markers mapped. The sum of the individual map lengths, or total map length, is 52,053.7 cR_{7000}.

We have identified a total of 51 conserved synteny groups (Tables 1 and 2). Of these, 38 groups were previously identified in each of three whole-genome comparative mapping studies using the ZOO-FISH, in situ hybridization, and EST RH mapping techniques [13,20,21]. We were able to confirm 7 groups identified by at least one other study: SSC3/HSA7 [13,21,22], SSC10/HSA1 [13,21], SSC10/HSA9 [13], SSC14/HSA1 [13,21], SSC14/HSA9 [21,23], SSC15/HSA4 [13,24,25], and SSC15/HSA8 [13,26]. Additionally, we report 6 new conserved synteny groups. Four of these groups, SSC2/HSA1 [13], SSC3/HSA9 [13], SSC17/HSA4 [27], and SSC17/HSA8 [13,26], had been suggested previously by the provisional assignments of single loci, but have now been demonstrated to include at least two markers. Two groups, SSC14/HSA4 and SSC15/HSA15, are newly reported. Provisional assignments of markers from HSA17 to SSC2 [28], HSA19 to SSC5 [29], HSA4 (http://www.toulouse.inra.fr/lgc/pig/compare/SSC.htm) and HSA21 [29] to SSC6, HSA19 to SSC7 [22], HSA21 to SSC9 [30], HSA10 to SSC12 [26], and HSA5 to SSC17 [31], however, were not supported by our data. SSC1 and SSC14 represent evolutionarily complex chromosomes as they contain 5 and 7 conserved synteny groups, respectively, whereas five porcine autosomes (SSC8, SSC11, SSC12, SSC16, and SSC18) contain only 1 synteny group (Fig. 1, Table 2). HSA1 shares orthology with six different porcine chromosomes, whereas each of five human chromosomes (HSA3, HSA13, HSA17, HSA20, and HSA21) is orthologous to only one porcine autosome (Table 1).

Within the 51 conserved synteny groups are 173 conserved segments as well as one singleton that may represent an additional segment (Tables 1 and 2). Segment sizes are reported, unless noted otherwise, as the human distance spanned (excluding, when appropriate, a fixed human centromere size of 3 Mb and any heterochromatic region ≥5 Mb in size; http://genome.ucsc.edu/cgi-bin/hgGateway) by the first and last marker of a given segment, i.e., the difference in human megabase position between the boundary loci. Thus, the mean and median comparative segment sizes are 14.04 and 6.75 Mb, respectively. The 12 smallest segments, representing approximately 7% of the segments, are less than or equal to 0.5 Mb in size. The smallest segment, located on SSC8, is 0.02 Mb, or 20 kb. However, if size is adjusted to include the distance between each boundary marker and its flanking marker from a different segment, this segment may be as large as 0.51 Mb. Another segment from the group of 12, located on SSC15, is reported as 0.37 Mb, but may be as large as 7.25 Mb due to suboptimal spacing of markers flanking this segment. Eleven segments (~6%) are larger than 50 Mb. The largest segment, located on SSC3, spans 108.9 Mb of orthologous sequence from HSA2. The most evolutionarily complex chromosomes, SSC1 and SSC14, contain the most segments; they are made up of 20 and 18 conserved segments, respectively (Table 2). SSC4 comprises only 3 segments and represents the fewest segments per chromosome. Three human autosomes, HSA1, HSA7, and HSA15, have the greatest number of segments; each comprises 16 segments (Table 1). For its relative size, HSA15 appears to be the most fragmented. An entire human chromosome, HSA21, is represented by only 1 segment, located at the telomeric end of SSC13 (Fig. 1). Segment orientation was determined by the maximum likelihood marker order and is represented by arrows in Fig. 1. Ten (~6%) of the 173 segments, plus the one included singleton, could not be oriented due to ambiguous ordering of the markers (Table S2) and are represented by blocks in Fig. 1.

Comparative coverage was calculated for all human autosomes (Table 1). First, the total human genomic distance spanned by the comparative segments of a given chromosome was determined by summing the individual segment sizes (as calculated above) of that chromosome. For example, the sizes of all 16 segments of HSA1 (located on SSC2, SSC4, SSC6, SSC9, SSC10, and SSC14) were totaled, resulting in a covered distance of 204.0 Mb. Next, the length of each human chromosome that can be used for comparison, referred to here as "comparative length," was calculated. This length represents the total length of the chromosome, minus a 3-Mb centromere and any region of heterochromatin ≥5 Mb in size (http://genome.ucsc.edu/cgi-bin/hgGateway). Such regions include the pericentromeric

regions of HSA1, HSA9, HSA16, and HSA19 as well as the short heterochromatic arms of HSA13, HSA14, HSA15, HSA21, and HSA22. For HSA1, a distance of 21.1 Mb, representing a 3-Mb centromere and 17.1 Mb of pericentromeric heterochromatin, was subtracted from the total length of HSA1, 245.2 Mb, resulting in a comparative length of 224.1 Mb. Finally, to estimate comparative coverage, the total human distance spanned by the comparative segments of a given chromosome was divided by the comparative length of that chromosome. Thus, for HSA1, a covered distance of 204.0 Mb divided by a comparative length of 224.1 Mb yields an estimated comparative coverage of 91.0%. Comparative coverage ranged from 66.7% for HSA22 to 97.1% for HSA3. Overall comparative coverage, or the total human distance spanned by all segments divided by the total comparative length of all 22 human autosomes, is estimated to be 90.3% (Table 1). All comparative coverage calculations do not account for distances between segments (i.e., breakpoints). Therefore, these values are underestimated, especially for those chromosomes with many segments. However, if the overall comparative coverage is calculated simply by summing the differences in megabase position between the first and the last markers mapped per chromosome arm for all human autosomes, without differentiating between the marker intervals within segments and those between segments (i.e., spanning breakpoints), and dividing by the total comparative length of all human autosomes, coverage is estimated to be 95.7%.

Discussion

In an effort to exploit the full potential of available genome resources, such as porcine radiation hybrid panels and BAC libraries, the complete human genome sequence, and bioinformatics tools, we have employed a targeted comparative mapping strategy to produce high-resolution, contiguous maps of the porcine autosomes. This powerful approach utilized physically anchored porcine BESs, representing both coding and noncoding sequences of the porcine genome, as well as the comparative sequence information obtained from similarity searches (M. Rogatcheva et al., submitted for publication). By analyzing the sequences prior to selecting loci for mapping, we were able to select a large number of porcine BESs with unique similarity to the human genome, thus improving map resolution while avoiding inconclusive mapping of paralogous sequences. Additionally, this analysis allowed us to select loci at evenly spaced intervals, based on human genome sequence coordinates, as is essential for producing contiguous maps.

Our targeted strategy has proven successful in producing porcine RH maps with a minimal number of linkage groups. Our goal of one linkage group per arm of the metacentric autosomes and per acrocentric autosome was nearly met, as there are only four additional, unresolved gaps in linkage. Two of these gaps, located on SSC2q and SSC16, are likely due to unusually low (~22%) and high (~90%) retention patterns, respectively, of the markers adjacent to these breaks. Interestingly, the sets of markers with low and high retention patterns are orthologous to adjacent chromosomal regions near the telomeric end of HSA5q. The remaining two breaks in linkage, located on SSC1q and SSC3p, separate adjacent loci from HSA15 and HSA16, respectively. The distance between the two markers from HSA15 is 1.37 Mb. Because this distance may approach the limit of linkage (LOD $\geq$6) for the INRA–Minnesota porcine radiation hybrid (IMpRH) panel used [11], it is possible that mapping of an additional marker between these two loci would resolve the gap on SSC1q, and the two comparative segments would join to reveal a single conserved segment. Unfortunately, only one BES with multiple sequence similarities was available for mapping in this region, and this locus did not map to SSC1. It is also possible, however, that this gap could not be resolved for other reasons; technical or mapping errors may have occurred, this region of the chromosome may represent a fragile site, or perhaps an as-of-yet unidentified comparative segment lies within this gap. Such a segment may be so small that it went undetected or was dismissed as an “unconfirmed singleton” or may even represent pig-specific sequence that is impossible to map using our comparative approach. These possibilities may also explain the gap on SSC3p, since the adjacent markers from HSA16 remain unlinked despite a distance of only 0.26 Mb.

Using our targeted approach to comparative mapping, we have significantly improved the resolution of the existing porcine genome maps. By exploiting the unique ability of the RH mapping method to map both polymorphic and nonpolymorphic markers, we were able to integrate microsatellite markers from the first-generation porcine WG-RH map [12] and compare our data with those of the existing moderate-resolution genetic linkage maps [3,4] (see also http://www.marc.usda.gov/genome/genome.html). Using a lod score of 6 and a distance threshold of 50 cR, the same map parameters used to construct the maps presented in Fig. 1, we were able to incorporate 647 (~96%) of 677 microsatellites into our map. Many markers unresolved by meiotic mapping could be ordered, and the order of integrated markers was generally consistent with the order determined by genetic linkage analysis (Table S1). Slight deviations from this order generally involved markers mapped within 5 cM of each other. As the linkage map has a reported resolution of approximately 5 cM [3], these deviations may represent the true marker order and improved map resolution. Marker order deviations of more than 5 cM may reflect incorrect assignment of the markers on the linkage map or may have been incorrectly integrated here due to the possible incompatibility of data generated in different laboratories. Indeed, the integration of the microsatellite data with our data resulted in significant (~25%) expansion of the RH map, suggesting that compatibility issues may exist.

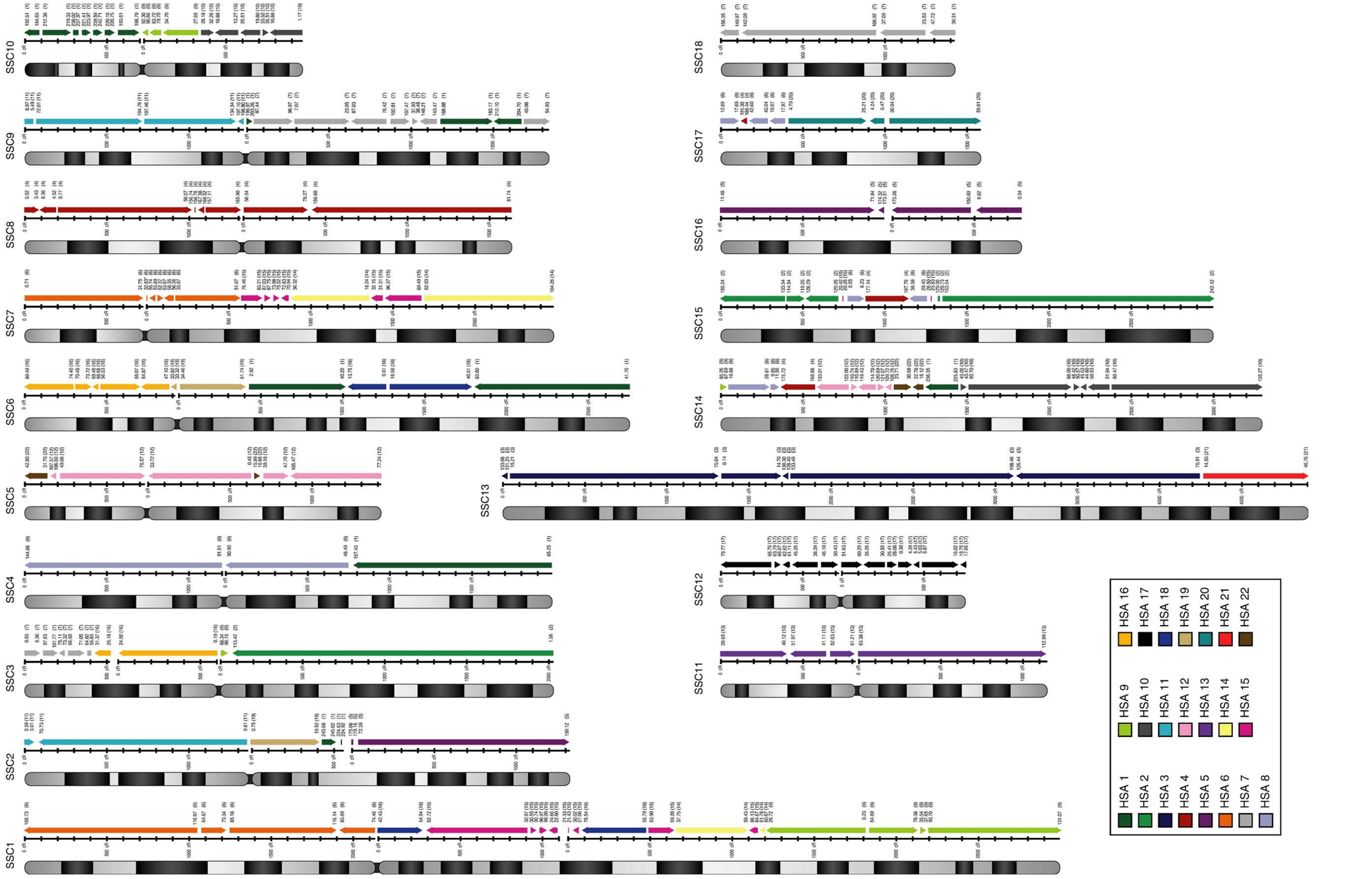

Fig. 1. Comparative RH maps of all 18 porcine autosomes. An idiogram of each porcine autosome is shown with corresponding RH and comparative mapping information. To the right of each autosome, and depicted as ticked black bars, are the RH maps. Each tick mark represents 100 cR_{7000}. Gaps between ticked black bars represent breaks in linkage. No direct positional relationship should be assumed between the idiograms and RH_{7000} maps. Conserved comparative segments are displayed adjacent to each RH_{7000} map. Human genome sequence coordinates, based on NCBI build 33, are provided for each of the two boundary loci per segment. Corresponding human chromosomes are provided in parentheses after each coordinate. Segments have also been color-coded based on orthology to each of the human autosomes, as shown in the color key. Lengths of the segments are approximately to scale. Spaces have been inserted between segments for ease of viewing and do not reflect gaps in linkage. Arrowheads illustrate orientation of segments. Those segments that could not be oriented due to ambiguous ordering of the markers are represented by blocks lacking arrowheads. The megabase position of the one singleton included in the map, located on the proximal end of SSC7q, is denoted by an asterisk.

Table 1
Summary statistics of the human-porcine comparative RH map sorted by human chromosome (HSA)

HSA	Comparative length in Mb[a] (total length in Mb)	Spanned human distance in Mb[b]	% Comparative coverage[c]	No. of loci mapped	Average spacing in Mb	No. of conserved synteny groups	SSC correspondence	No. of segments
1	224.1 (245.2)[d]	204.0	91.0	192	1.15	6	2, 4, 6, 9, 10, 14	16
2	240.3 (243.3)	225.3	93.7	174	1.35	2	3, 15	6
3	196.4 (199.4)	190.6	97.1	159	1.23	1	13	6
4	188.6 (191.6)	176.6	93.7	166	1.14	4	8, 14, 15, 17	11
5	178.0 (181.0)	171.9	96.5	143	1.24	2	2, 16	6
6	167.7 (170.7)	154.8	92.3	143	1.16	2	1, 7	10[e]
7	155.4 (158.4)	135.6	87.3	138	1.08	3	3, 9, 18	16
8	142.9 (145.9)	129.7	90.8	133	1.06	4	4, 14, 15, 17	9
9	117.5 (134.5)[d]	96.4	82.1	95	1.11	4	1, 3, 10, 14	9
10	132.5 (135.5)	113.8	85.9	108	1.20	2	10, 14	11
11	132.0 (135.0)	123.8	93.8	119	1.09	2	2, 9	6
12	130.5 (133.5)	116.3	89.1	108	1.20	2	5, 14	10
13	96.2 (114.2)[f]	88.5	92.0	105	0.89	1	11	4
14	87.2 (105.3)[f]	82.2	94.3	69	1.27	2	1, 7	4
15	81.8 (100.1)[f]	63.3	77.4	79	1.01	3	1, 7, 15	16
16	80.2 (90.0)[d]	69.4	86.5	66	1.15	2	3, 6	7
17	78.7 (81.7)	62.8	79.8	75	1.01	1	12	12
18	74.8 (77.8)	68.3	91.4	55	1.34	2	1, 6	4
19	55.8 (63.8)[d]	46.6	83.4	45	1.10	2	2, 6	3
20	60.6 (63.6)	53.1	87.6	49	1.14	1	17	3
21	34.0 (47.0)[f]	32.2	94.8	27	1.24	1	13	1
22	35.1 (49.5)[f]	23.4	66.7	26	1.07	2	5, 14	4
Total	2690.3 (2866.9)	2428.6	90.3	2274	1.15	51	—	174[e]

[a] Comparative length is defined as the length of each chromosome used for comparison of coverage and represents the total chromosome length (indicated in parantheses) minus a 3-Mb centromere and any region of heterochromatin ≥5 Mb in size.
[b] Represents the sum of all segment sizes for a given chromosome. Segment sizes are defined as the human distance spanned (excluding centromere and other heterochromatic regions) by the first and last marker of a given segment, i.e., the difference in human megabase position between the boundary loci.
[c] Percentage comparative coverage is calculated as the spanned human distance divided by the comparative length for each chromosome as well as the total.
[d] Denotes human chromosomes with large (≥5 Mb) regions of heterochromatin.
[e] Reported number of segments includes one singleton.
[f] Denotes human chromosomes with short heterochromatic arms.

Use of the RH mapping technique also allowed for refinement of the whole-genome comparative maps generated by other lower resolution mapping methods, such as FISH [20,21]. Goureau and colleagues [20] used the ZOO-FISH technique to produce a comparative map of 38 conserved synteny groups. However, the resolution of this mapping technique was insufficient, particularly at the borders of synteny, and offered little information on the number of segments within each of the conserved synteny groups. In an effort to enhance this mapping information, Pinton et al. [21] localized 113 comparative loci in pigs, focusing primarily on synteny breakpoint regions, and attempted to estimate the number of conserved segments within synteny groups. This study confirmed the 38 conserved synteny groups identified by Goureau, as well as 4 additional conserved synteny groups, and estimated at least 82 conserved segments. As expected, by using a higher resolution mapping method, we were able to identify more conserved synteny groups and conserved segments, a total of 51 and 173, respectively. We were unable, however, to confirm the existence of 2 different segments of orthology with HSA9 on SSC14 [21], nor were we able to demonstrate that the segments of SSC14 orthologous to HSA12 and HSA22 are separated by other comparative segments [21]. Additionally, our data did not support the presence of a comparative segment from HSA8 positioned between segments from HSA1 on SSC4 [21]; instead, we observe only 3 comparative segments, 2 with orthology to HSA8, located on SSC4p and the proximal end of SSC4q, and 1 with orthology to HSA1, located at the distal end of SSC4q. No overlap of these segments is observed.

Much of the improvement in map resolution was previously made by Rink et al. [13]. By also using the WG-RH mapping method, Rink and colleagues were able to add at least three new conserved synteny groups to the comparative map and suggest the provisional assignments of others, as well as identifying at least 60 breakpoints and 90 micro-rearrangements between the human and the porcine genomes. However, by mapping a greater number of markers, and selecting markers with a relatively even spatial distribution prior to mapping, we were able to achieve even higher resolution of the human–pig comparative map. Utilizing a large number of evenly spaced markers not only enabled the construction of contiguous maps, but also permitted the confirmation of putative segments suggested by singletons, aided in the identification of new conserved synteny groups,

Table 2
Summary statistics of the human–porcine comparative RH map sorted by porcine chromosome (SSC)

SSC	No. of loci mapped	No. of linkage groups	Map length in cR (% total)	No. of conserved synteny groups	HAS correspondence	No. of segments
1	255	3	6245.1 (12.0)	5	6, 9, 14, 15, 18	20
2	147	3	3245.7 (6.2)	4	1, 5, 11, 19	7
3	138	3	3147.2 (6.0)	4	2, 7, 9, 16	9
4	140	2	3189.4 (6.1)	2	1, 8	3
5	100	2	2149.8 (4.1)	2	12, 22	7
6	144	2	3661.9 (7.0)	4	1, 16, 18, 19	11
7	127	2	3200.8 (6.1)	3	6, 14, 15	14[a]
8	143	2	2946.0 (5.7)	1	4	8
9	150	2	3171.9 (6.1)	3	1, 7, 11	14
10	81	2	1668.9 (3.2)	3	1, 9, 10	15
11	105	2	1966.7 (3.8)	1	13	4
12	75	2	1471.6 (2.8)	1	17	12
13	186	1	4905.4 (9.4)	2	3, 21	7
14	154	1	3296.9 (6.3)	7	1, 4, 8, 9, 10, 12, 22	18
15	128	1	2998.2 (5.8)	4	2, 4, 8, 15	10
16	79	2	1781.4 (3.4)	1	5	4
17	64	1	1582.6 (3.0)	3	4, 8, 20	7
18	58	1	1424.2 (2.7)	1	7	4
Total	2274	34	52,053.7 (100.0)	51	—	174[a]

[a] Reported number of segments includes one singleton.

and refined the boundaries of conserved segments. Additionally, our selection of markers having unique sequence similarities to the human genome avoided the ectopic placement of paralogous loci that can result in the provisional assignment of singletons and single-locus conserved synteny groups.

All four whole-genome human–pig comparative mapping studies [13,20,21] appear to be in full agreement with respect to the number of conserved synteny groups for 12 of 18 porcine autosomes (SSC1, SSC4, SSC5, SSC6, SSC7, SSC8, SSC9, SSC11, SSC12, SSC13, SSC16, and SSC18). Therefore, FISH mapping, although insufficient for detailed map construction, has provided a strong basis for comparative analysis. The number of conserved synteny groups detected for the remaining 6 autosomes (SSC2, SSC3, SSC10, SSC14, SSC15, and SSC17) has increased over the past 5 years, reflecting improvements in mapping strategies and methods. Use of a strategy focusing on breakpoint regions [21] allowed for the identification of one additional conserved synteny group on both SSC3 and SSC10 and two additional groups on SSC14. These groups were subsequently confirmed through the use of the RH method [13], with one exception; Rink et al. was unable to confirm orthology between HSA9 and SSC14. Use of the RH method also allowed the detection of new conserved synteny groups on SSC10 and SSC15. Finally, application of our targeted strategy confirmed putative conserved synteny groups suggested by single loci assigned to SSC2 [13], SSC3 [13], and SSC17 [13,26,27], as well as detected new groups on SSC14 and SSC15.

The largest new conserved synteny group was first suggested by the placement of a singleton on the EST RH map of SSC17 [13]. This singleton, originally dismissed as an error in the human genome assembly [13], may actually represent one of three conserved segments detected on SSC17 spanning nearly 8.8 Mb of orthologous HSA8 sequence (Fig. 1). Two segments are orthologous to a contiguous HSA8 region of approximately 6.2 Mb (breakpoint included), and the third segment spans approximately 2.6 Mb of a separate region. The second largest new conserved synteny group, located on SSC14, covers nearly 6.9 Mb of a contiguous region of HSA4. Due to its relatively large size, it is somewhat remarkable that this conserved synteny has gone undetected in other studies. Other smaller groups were identified as well. Singletons assigned to the current RH maps suggest the possibility of conserved synteny groups on SSC3 and SSC2 with orthology to HSA9 and HSA1, respectively [13]. We have identified one segment, made up of four different markers spanning 1.8 Mb of HSA9, on SSC3 and have detected two small segments (0.3 and 1.3 Mb in size) on SSC2, with orthology to two different regions of HSA1. Finally, two small regions of orthology were detected between HSA15 and SSC15 that have not previously been reported; these segments are approximately 0.2 and 0.7 Mb in size.

Our targeted approach also enabled the refinement of conserved segment boundaries within synteny groups. For example, the SSC10/HSA1 synteny group is currently represented by four markers assigned to two putative segments [13]. One segment is represented by a singleton and the other segment is represented by the three remaining markers. Two of the three markers have similar human genome sequence coordinates, differing by only 0.4 Mb, whereas the position of the third marker is 26 Mb from the first two. These markers were used to designate a single comparative segment. In contrast, our selection of evenly spaced markers allowed us to characterize this region further

and demonstrate that this conserved synteny group consists of up to seven small segments (Fig. 1).

Within the 51 conserved synteny groups, we have identified a total of 173 conserved segments as well as one singleton that may represent an additional segment. Segments were designated as such based on several criteria, including number of loci, assigned linkage group, likelihood of marker order compared to alternative orders (as determined using the flips algorithm of Carthagene), and relative position of sequences in the human genome. Our first criterion was that segments must include at least two loci. In an effort to eliminate markers with spurious placement resulting from technical errors or the occasional use of markers with nonunique sequence similarity, any mapped locus that could not be confirmed by the addition of a second marker with approximately the same human genome sequence coordinates was designated an "unconfirmed singleton" and was not assigned to the map. Only one singleton was assigned, despite our inability to confirm it with a second marker, to the proximal end of SSC7q (Fig. 1). This assignment was permitted for essentially two reasons. First, the comparative sequence from HSA6 contains the human major histocompatibility complex genes, and the characterization and sequencing of the orthologous region of the swine genome has revealed that these genes are located on both sides of the centromere on SSC7 [32,33]. Therefore, we believed that this locus was positioned correctly and could reflect an extension of the segment located across the centromere on SSC7p. However, this singleton could not be included as part of the segment from SSC7p because it failed to meet our second criterion—that all markers of a segment must be assigned to the same linkage group. Due to the presence of a centromere, this singleton could not be linked to the markers with adjacent orthologous sequence in the human genome. Based on this same criterion, other regions of orthology with apparent centromere insertions were designated as 2 segments instead of 1. This was the case for 3 or 4 other metacentric autosomes, including SSC4, SSC11, SSC12, and perhaps SSC6 (Fig. 1). Therefore, the actual number of conserved segments may be less than the total of 173 reported here. Analysis using the flips algorithm of Carthagene, however, supports opposite orientations of the segments on either side of the SSC6 centromere, suggesting that these 2 segments may truly be separate segments. Interestingly, these apparent centromere insertions, occurring in 4 or 5 of the 12 metacentric autosomes, may support the hypothesis of neocentromere emergence [34], whereas the other 7 or 8 metacentric autosomes may support the hypothesis of centromere repositioning from an acrocentric ancestor [34].

Our determination of segments was also based on a conservative interpretation of our data with respect to gene order. Following the formation of linkage groups, we applied the annealing and taboo algorithms of Carthagene to determine the multipoint maximum likelihood order of markers within each group (Supplementary Table A). Initial assessment of these data revealed several blocks of markers having essentially the same order as found in the human genome. Surprisingly, only 67 deviations from the human gene order could be detected within these blocks, suggesting a high level of gene order conservation between the human and the porcine genomes. To assess the strength of the maximum likelihood orders, we used the Carthagene flips algorithm to explore other possible permutations with a log-likelihood difference of <3 and within a sliding window of six markers. These alternative permutations and corresponding log-likelihood differences are provided in Supplementary Table 2. As nearly 79% (53/67) of the deviations detected could not be supported with a difference $\geq$3, we concluded there was conserved gene order and determined segments accordingly. If two markers mapped immediately adjacent to one other (either within the same segment or at the boundary of adjacent segments), but were inverted with respect to the human gene order, flipping of the two markers was allowed to define a conserved segment, regardless of the difference between log-likelihoods. However, if two nonadjacent loci, i.e., separated by at least one other mapped locus, were inverted with respect to the human gene order, flipping of these markers was allowed only if the difference in log-likelihood was <3. For example, six groups of three markers with inverted gene order were observed. If flipping of the first and third markers of a given group was possible with a difference <3, as was the case for three of the six groups, inversion of the segment was allowed and one conserved segment was maintained; however, inversion of the other three groups was not allowed, and each group of three was designated as an individual conserved segment.

The strength of order, as assessed using the flips algorithm, was also used to determine the orientation of segments, particularly the smaller segments, as indicated by arrowheads in Fig. 1. Segments of two or more markers ordered with a likelihood difference >3 were assigned an orientation based on the human genome coordinates. Orientation was not determined, however, for those small segments containing only a few markers ordered with a likelihood difference <3. Segments for which orientation could not be determined are depicted as blocks lacking arrowheads in Fig. 1.

Due to the strength of our maximum likelihood orders and our establishment of the above criteria for segment determination, we do not report as many microrearrangements as have been suggested [13]. Conservative interpretation of our data indicates that gene order within segments is largely conserved, with the exception of relatively few minor rearrangements that may be due to technical errors. Clear evidence exists for major rearrangements within conserved synteny groups, as demonstrated in Fig. 1, and these rearrangements can range from the simple inversion of a single segment to complex inversions of several segments. For example, SSC11 consists of a single synteny group, and gene order is conserved throughout the chromosome with the exception of a single inverted segment located on the p

arm. SSC12, on the other hand, also consists of a single synteny group, but reflects several rearrangements involving many segments. It is also interesting to observe the differences in the degree of segmentation for each autosome. There are large chromosomes with relatively few, yet large, segments, such as SSC13, or many small segments, such as SSC14. The same is true for the smaller chromosomes as well; SSC16 and SSC18 have relatively few, somewhat large, segments, and SSC10 and SSC12 both consist of many small segments.

In conclusion, by using a targeted approach to comparative mapping exploiting recent advances in genomics, we have generated the first contiguous maps of the porcine autosomes composed of evenly spaced markers arranged in a minimal number of linkage groups, providing unprecedented resolution of the human–pig comparative map. These maps define conserved segments of synteny and gene order, precisely localize evolutionary breakpoints, and can be integrated with the porcine linkage map. Because our approach utilized physically anchored markers, i.e., derived from BAC clones, the necessary tools are now primed for the targeted isolation of additional, informative markers required for maximal refinement of genetic intervals containing QTL and should greatly facilitate future positional candidate cloning efforts. Additionally, our BAC-based map can provide a framework for anchoring contigs generated through BAC fingerprinting efforts and may prove invaluable for doing so. Preliminary BAC fingerprinting contig data for SSC17 appear to be in full agreement with our map (S. Humphray, personal communication). Finally, our BAC RH map can also contribute to the selection of a BAC minimal tiling path and assembly of the first sequence-ready map of the porcine genome.

Materials and methods

Marker selection and primer design

BESs

Approxiately 91% of markers included in the map were porcine BESs. The majority of BESs were selected, at 1- to 1.5-Mb intervals, from an ordered list of unique, i.e., nonparalogous, sequences established by Rogatcheva et al. (submitted for publication). Briefly, repetitive elements in BESs were masked, and masked sequences were analyzed for similarity with build 33 of the human genome draft sequence (April 2003 release; http://genome.ucsc.edu/cgi-bin/hgGateway) using the NCBI BLASTn program. BESs with a single BLAST hit below a significance threshold E value of e^{-5} were compiled and sorted by matching position within the human genome, thus allowing for selection of evenly spaced markers. When given a choice among several markers with the same relative genomic position, markers with an E value less than e^{-10} and a match length of at least 100 bp were selected. When no markers were available for a given human genome position, BESs were selected from an alternative set of sequences provided by the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/) that were also analyzed for similarity with NCBI build 33 and ordered by human sequence coordinates. When possible, similar selection criteria (unique match, $E < e^{-10}$, match length $\geq$100 bp) were used to select markers. Occasionally, however, nonunique sequences with a significant difference between the best and the next best matches were selected, or the criteria for E value and/or match length were relaxed.

ESTs

Approximately 9% of markers included in the map were porcine ESTs, as our original intention was to improve the resolution of the current EST RH map by mapping additional EST loci. Initially, human sequences were selected by browsing the Ensembl ContigView of a particular chromosomal region of interest (http://www.ensembl.org/). These human sequences were then analyzed, using the NCBI BLASTn program [35] (see also http://www.ncbi.nlm.nih.gov/), for similarity with sequences in the NCBI "est_others" database containing nonhuman and nonmouse EST sequences. Porcine EST sequences with significant similarity, based on expectation (E) values of e^{-10} or lower, were then identified from the list of BLAST results and used to design primers for mapping. Several previously mapped ESTs had sequence similarity to more than one human sequence; therefore, these markers were eventually replaced by BESs having a unique similarity to the human genome at nearly the same genomic position as the most significant EST match. The number of replaced markers is reported in Supplementary Table 4.

Primer design

Primers were designed using available tools, including Primer Designer 2 (Scientific and Educational Software), Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), and Vector NTI v7.0 software (InforMax). Regardless of the primer design tool used, primers were typically designed to have a length of 20–22 bp, a GC content of 45–60%, and a melting temperature of 60°C, as well as amplifying products of 100–600 bp. For the EST sequences used, primers were designed within 3′ untranslated regions whenever possible.

PCR amplification of selected markers

Templates were prepared in 96-well PCR plates and included DNA from each of the 90 hybrids of the INRA–Minnesota porcine radiation hybrid panel [11] as well as six (four positive and two negative) controls. Positive controls contained porcine genomic DNA. One negative control contained Chinese hamster genomic DNA (Wg3H; Invitrogen), and the other contained no DNA. PCR was performed in a 10-μl reaction volume containing 20–25 ng of template DNA, 1× PCR buffer (containing 1.5 mM $MgCl_2$; Qiagen),

200 μM each dNTP (Fermentas), 0.5 μM each primer, and 0.25 U HotStarTaq DNA polymerase (Qiagen). Typical PCR cycling parameters included an initial denaturation step of 95°C for 15 min followed by 35 cycles of 94°C for 30 s, 55–70°C for 45 s, and 72°C for 45 s, plus a final extension step of 72°C for 5 min. PCR products were electrophoresed in 2% agarose, 0.5–1× TBE gels and visualized by ethidium bromide staining. Markers displaying only porcine-specific PCR products, or porcine products that could be easily distinguished from hamster products, were used to construct the RH maps. The number of markers not included in the map, due to nonspecific, weak, or no amplification, are reported in Table S4.

Construction of RH maps

Loci were scored as positive (1), negative (0), or ambiguous (2) for each hybrid. Vectors containing the 90 scored hybrids were then entered into the online IMpRH mapping tool [36] (see also http://imprh.toulouse.inra.fr/) to obtain maximum two-point lod scores and establish a chromosomal assignment for each marker. Marker vector data were then grouped by chromosome and used to construct multipoint maximum likelihood RH maps using Carthagene software v0.99 [37,38] (see also http://www.inra.fr/bia/T/CarthaGene/). Linkage groups were formed with a lod score threshold of 6 and a distance threshold of 50 cR. The annealing and taboo algorithms of Carthagene were then used to determine the multipoint maximum likelihood order of markers within each linkage group. Strength of these orders was assessed using a flips algorithm to identify possible alternative permutations within a window of six markers with a lod score <3. RH maps were initially constructed using only markers mapped in this study to alleviate concerns with the consistency of data generated by different individuals and from different laboratories. These maps were then used as a framework for integration of the microsatellite data from the first-generation porcine WG-RH map [12]. Linkage groups were oriented based on the previously determined order of these microsatellite markers on the genetic maps [3,4] (see also http://www.marc.usda.gov/genome/genome.html).

Acknowledgments

This work was supported in part by USDA/NRI-CSREES Grant AG2001-35205-11698 and USDA-ARS AG58-5438-2-313.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ygeno.2005.04.010.

References

[1] A.L. Archibald, et al., The PiGMaP Consortium linkage map of the pig (Sus scrofa), Mamm. Genome 6 (1995) 157–175.
[2] H. Ellegren, et al., A primary linkage map of the porcine genome reveals a low rate of genetic recombination, Genetics 137 (1994) 1089–1100.
[3] G.A. Rohrer, et al., A comprehensive map of the porcine genome, Genome Res. 6 (1996) 371–391.
[4] G.A. Rohrer, L.J. Alexander, J.W. Keele, T.P. Smith, C.W. Beattie, A microsatellite linkage map of the porcine genome, Genetics 136 (1994) 231–245.
[5] J.P. Bidanel, M.F. Rothschild, Current status of quantitative trait locus mapping in pigs, Pig News Inf. 23 (2002) 39N–53N.
[6] E.S. Lander, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.
[7] J.C. Venter, et al., The sequence of the human genome, Science 291 (2001) 1304–1351.
[8] R.H. Waterston, et al., Initial sequencing and comparative analysis of the mouse genome, Nature 420 (2002) 520–562.
[9] L.C. McCarthy, Whole genome radiation hybrid mapping, Trends Genet. 12 (1996) 491–493.
[10] R.J. Leach, P. O'Connell, Mapping of mammalian genomes with radiation (Goss and Harris) hybrids, Adv. Genet. 33 (1995) 63–99.
[11] M. Yerle, et al., Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs, Cytogenet. Cell Genet. 82 (1998) 182–188.
[12] R.J. Hawken, et al., A first-generation porcine whole-genome radiation hybrid map, Mamm. Genome 10 (1999) 824–830.
[13] A. Rink, et al., A first-generation EST RH comparative map of the porcine and human genome, Mamm. Genome 13 (2002) 578–587.
[14] S.I. Anderson, N.L. Lopez-Corrales, B. Gorick, A.L. Archibald, A large-fragment porcine genomic library resource in a BAC vector, Mamm. Genome 11 (2000) 811–814.
[15] S.C. Fahrenkrug, et al., A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration, Mamm. Genome 12 (2001) 472–474.
[16] C. Rogel-Gaillard, N. Bourgeaux, A. Billault, M. Vaiman, P. Chardon, Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements, Cytogenet. Cell Genet. 85 (1999) 205–211.
[17] L.J. Alexander, et al., Physical assignments of 68 porcine cosmid and lambda clones containing polymorphic microsatellites, Mamm. Genome 7 (1996) 368–372.
[18] H. Ellegren, B. Chowdhary, M. Johansson, L. Andersson, Integrating the porcine physical and linkage map using cosmid-derived markers, Anim. Genet. 25 (1994) 155–164.
[19] G. Rettenberger, et al., Chromosomal assignment of seventeen porcine microsatellites and genes by use of a somatic cell hybrid mapping panel, Anim. Genet. 26 (1995) 269–273.
[20] A. Goureau, et al., Human and porcine correspondence of chromosome segments using bidirectional chromosome painting, Genomics 36 (1996) 252–262.
[21] P. Pinton, L. Schibler, E. Cribiu, J. Gellin, M. Yerle, Localization of 113 anchor loci in pigs: improvement of the comparative map for humans, pigs, and goats, Mamm. Genome 11 (2000) 306–315.
[22] C.B. Jorgensen, A.K. Wintero, M. Yerle, M. Fredholm, Mapping of 22 expressed sequence tags isolated from a porcine small intestine cDNA library, Mamm. Genome 8 (1997) 423–427.
[23] M. Kojima, K. Ohata, H. Yasue, Assignment of the porcine spleen tyrosine kinase gene, SYK, to chromosome 14q14 by fluorescence in situ hybridization, Mamm. Genome 8 (1997) 379–380.
[24] N.J. Larsen, et al., New insights into porcine–human synteny conservation, Mamm. Genome 10 (1999) 488–491.
[25] A.K. Wintero, C.B. Jorgensen, A. Robic, M. Yerle, M. Fredholm, Improvement of the porcine transcription map: localization of 33 genes, of which 24 are orthologous, Mamm. Genome 9 (1998) 366–372.

[26] R. Davoli, et al., Isolation of porcine expressed sequence tags for the construction of a first genomic transcript map of the skeletal muscle in pig, Anim. Genet. 33 (2002) 3–18.
[27] L.A. Messer, et al., Mapping of the melatonin receptor 1a (MTNR1A) gene in pigs, sheep, and cattle, Mamm. Genome 8 (1997) 368–370.
[28] P.D. Thomsen, A.K. Wintero, M. Fredholm, Chromosomal assignments of 19 porcine cDNA sequences by FISH, Mamm. Genome 9 (1998) 394–396.
[29] A.K. Fridolfsson, et al., Expansion of the pig comparative map by expressed sequence tags (EST) mapping, Mamm. Genome 8 (1997) 907–912.
[30] M.M. Leong, C.C. Lin, R.F. Ruth, Assignment of superoxide dismutase (SOD-1) gene to chromosome No. 9 of domestic pig, Can. J. Genet. Cytol. 25 (1983) 233–238.
[31] Y. Lahbib-Mansais, et al., A successful strategy for comparative mapping with human ESTs: 65 new regional assignments in the pig, Mamm. Genome 10 (1999) 145–153.
[32] P. Chardon, et al., Sequence of the swine major histocompatibility complex region containing all non-classical class I genes, Tissue Antigens 57 (2001) 55–65.
[33] P. Chardon, et al., Physical organization of the pig major histocompatibility complex class II region, Immunogenetics 50 (1999) 344–348.
[34] G. Montefalcone, S. Tempesta, M. Rocchi, N. Archidiacono, Centromere repositioning, Genome Res. 9 (1999) 1184–1188.
[35] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[36] D. Milan, et al., IMpRH server: an RH mapping server available on the Web, Bioinformatics 16 (2000) 558–559.
[37] S. de Givry, M. Bouchez, P. Chabrier, D. Milan, T. Schiex, CARTHAGENE: multipopulation integrated genetic and radiation hybrid mapping, Bioinformatics 21 (2005) 1703–1704.
[38] T. Schiex, C. Gaspin, CARTHAGENE: constructing and joining maximum likelihood genetic maps, Proc. Int. Conf. Intell. Syst. Mol. Biol. 5 (1997) 258–267.