Measuring conservation of contiguous sets of autosomal markers on bovine and porcine genomes in relation to the map of the human genome

Zhihua Jiang, Jenna S. Melville, Honghe Cao, Sudhir Kumar, Alan Filipski, and Ann M. Verrinder Gibbins

Abstract: Based on published information, we have identified 991 genes and gene-family clusters for cattle and 764 for pigs that have orthologues in the human genome. The relative linear locations of these genes on human sequence maps were used as "rulers" to annotate bovine and porcine genomes based on a CSAM (contiguous sets of autosomal markers) approach. A CSAM is an uninterrupted set of markers in one genome (primary genome; the human genome in this study) that is syntenic in the other genome (secondary genome; the bovine and porcine genomes in this study). The analysis revealed 81 conserved syntenies and 161 CSAMs between human and bovine autosomes and 50 conserved syntenies and 95 CSAMs between human and porcine autosomes. Using the human sequence map as a reference, these 991 and 764 markers could correlate 72 and 74% of the human genome with the bovine and porcine genomes, respectively. Based on the number of contiguous markers in each CSAM, we classified these CSAMs into five size groups as follows: singletons (one marker only), small (2–4 markers), medium (5–10 markers), large (11–20 markers), and very large (>20 markers). Several bovine and porcine chromosomes appear to be represented as di-CSAM repeats in a tandem or dispersed way on human chromosomes. The number of potential CSAMs for which no markers are currently available were estimated to be 63 between human and bovine genomes and 18 between human and porcine genomes. These results provide basic guidelines for further gene and QTL mapping of the bovine and porcine genomes, as well as insight into the evolution of mammalian genomes.

Key words: Human, cattle, pig, orthologous genes, CSAM, comparative mapping.

Résumé: À partir d'information déjà publiées, les auteurs ont identifié 991 gènes ou familles de gènes bovins et 764 gènes ou familles de gènes porcins ayant des homologues au sein du génome humain. Les positions linéaires relatives de ces gènes sur les cartes de séquences humaines ont été employées pour annoter les génomes bovin et porcin à l'aide d'une approche CSAM (« contiguous set of autosomal markers »). Un CSAM représente une série de marqueurs ininterrompue chez un génome (génome primaire ; le génome humain dans le cas présent) et qui est organisée de façon identique (syntène) chez un autre génome (génome secondaire ; les génomes bovin ou porcin dans ce travail). Cette analyse a révélé 81 régions de synténie et 161 CSAM entre les autosomes humain et bovin. Pareillement, 50 régions de synténie et 95 CSAM ont été observés entre les autosomes humain et porcin. En utilisant la séquence du génome humain comme point de référence, ces 991 ou 764 marqueurs ont permis de corréler 72 % ou 74 % du génome humain avec les génomes bovin ou porcin, respectivement. En fonction du nombre de marqueurs contigus au sein de chaque CSAM, cinq classes de CSAM ont été définies : à marqueur unique (un seul marqueur), petit (2 à 4 marqueurs), moyen (5 à 10 marqueurs), grand (11 à 20 marqueurs) et très grand (>20 marqueurs). Le nombre de CSAM appartenant à chacune de ces classes diminue tandis que leur taille physique moyenne augmente tant lors des comparaisons entre génomes humain et bovin que lors des comparaisons entre génomes humain et porcin. Plusieurs chromosomes bovins ou porcins semblent avoir été dupliqués sous forme de CSAM répétés en tandem ou dispersés chez les chromosomes humains. Le nombre de CSAM potentiels pour lesquels aucun marqueur n'est présentement disponible a été estimé à 63 pour les génomes humain et bovin et à 18 pour les génomes humain et porcin. Ces résultats fournissent des points de repère pour de futurs travaux de cartographie de gènes ou de QTL chez les génomes bovin ou

Received 14 January 2002. Accepted 15 April 2002. Published on the NRC Research Press Web site at http://genome.nrc.ca on 29 June 2002.

Corresponding Editor: S. Varmuza.

Z. Jiang,[1] J.S. Melville, H. Cao, and A.M. Verrinder Gibbins. Department of Animal and Poultry Science, University of Guelph, Guelph, ON N1G 2W1, Canada.
S. Kumar and A. Filipski. Department of Biology, Arizona State University, Tempe, AZ 85287-1501, U.S.A.

[1]Corresponding author (e-mail: zjiang@lsd.uoguelph.ca).

 DOI: 10.1139/G02-038 © 2002 NRC Canada

porcin de même qu'un aperçu de l'évolution des génomes au sein des mammifères.

Mots clés : humain, bœuf, porc, gènes orthologues, CSAM, cartographie comparée.

[Traduit par la Rédaction]

Introduction

The release of a draft sequence of the human genome was considered to be a landmark event in biology. The current draft sequence covers about 94% of the human genome (The International Human Genome Sequencing Consortium 2001). In addition, a 2.91-billion base pair (bp) consensus sequence of the euchromatic portion of the human genome has been generated by whole-genome shotgun sequencing (Venter et al. 2001). Although divergence of the various mammalian lineages occurred 60–110 million years ago (Kumar and Hedges 1998), all mammals have a highly conserved genome size and presumably share most of their orthologous genes (Andersson et al. 1996). The draft sequence of the human genome will, therefore, be a basic and ultimate reference for comparative mapping of genomes among mammals. The applications of comparative maps are manifold and include linkage prediction, the candidate gene approach to disease gene identification, characterization of the genetic basis for complex traits, and studies of genome organization and evolution (Eppig and Nadeau 1995; Eppig 1996; Nadeau and Sankoff 1998; O'Brien et al. 1999).

Genome mapping in livestock species has also advanced significantly in recent years, especially by taking advantage of and building on the information available through the human genome program. O'Brien and colleagues (1993) proposed a list of 321 reference anchor loci to improve comparative mammalian maps and standardize their outputs. More recently, attempts have also been made to systematically develop reagents that can be used for comparative gene mapping. For example, genetic markers, such as universal mammalian primer sets (UMPs) (Venta et al. 1996), comparative anchor tagged sequences (CATs) (Lyons et al. 1997), and traced orthologous amplified sequence tags (TOASTs) (Jiang et al. 1998), have been developed that are easy to map and for which homology in different species is readily established. On the other hand, random cDNA sequencing has proven to be a most efficient way of gaining sequence data for hitherto unknown genes (Adams et al. 1991) and has led to assignment of novel genes in livestock genomes, especially for cattle and pigs (Fridolfsson et al. 1997; Ma et al. 1998; Band et al. 2000; Ozawa et al. 2000; Tosser-Klopp et al. 2001; Karnuah et al. 2001; and Jiang et al. 2002). Progress in livestock genome mapping has enabled not only alignment of livestock genomes with the human genome, but also identification of genes of interest in domestic animal genomes that underlie quantitative traits (Womack and Kata 1995).

Comparative mapping of mammals enables the identification of chromosomal segments conserved in mammalian genomes since divergence from a common ancestor. The conserved segments are usually identified by examining the relative order of contiguous orthologous gene landmarks in the chromosomes of the species being compared (Sankoff and Nadeau 1996; Nadeau and Sankoff 1998). However, there are few pairs of mammalian species for which many orthologous genes have been well ordered in their genomes. Kumar and colleagues (2001) recently proposed an approach intermediate to the conserved synteny and conserved segment approaches that uses contiguous sets of autosomal markers (CSAMs). Conserved synteny is defined as at least two pairs of homologous genes on the same chromosome regardless of order, whereas a conserved segment is a maximally contiguous chromosomal region with identical gene content and order in the two species being compared (Nadeau and Sankoff 1998). A CSAM is an uninterrupted set of markers in one genome (primary genome) that is syntenic in the other genome (secondary genome). The CSAM approach requires relative marker order information in one genome and only the chromosome number of those markers in the other genome. Based on published linkage maps, cytogenetic maps, and RH (radiation hybrid) maps, we have defined 991 autosomal genes and gene-family clusters for cattle and 764 for pigs as being homologous to those found in the human genome. The relative linear orders of these genes on human sequence maps (see Human Genome Resources at http://www.ncbi.nlm.nih.gov) were used as a means of annotating bovine and porcine genomes. Here we report on the identification, characterization, and distribution of CSAMs between human and bovine and between human and porcine genomes.

Materials and methods

Data resources and orthologous gene identification

Gene-mapping data were gathered for cattle mainly from the Bovine Genome Database at the Texas A&M University, College Station, Tex., and from the BOVMAP database at the Centre Institut National de la Recherche Agronomique (INRA) de Jouy-en-Josas, France. The ARKdb database developed by the Roslin Institute, Edinburgh, U.K., and the "Cytogenetic map of the pig" database provided by Centre INRA de Toulouse, France, were major resources for the collection of porcine gene mapping data for this study. Recently published gene-mapping data (late 2000 and 2001) were also collected for cattle and pigs, mainly from the journals *Animal Genetics*, *Journal of Animal Breeding and Genetics*, *Journal of Animal Science*, *Selection, Genetics and Evolution*, and *Mammalian Genome*. Human genome resources (as of September 5, 2001) at the National Centre for Biotechnology Information (NCBI), Bethesda, Md., were used as a means of identifying orthologous genes and their cytogenetic and sequence locations in the human genome for those genes that have been mapped in bovine and porcine genomes. The orthologous status was checked and corrected when a marker that mapped in either the bovine or porcine genome (*i*) indicated no match with an orthologous gene in the human genome, (*ii*) failed to match the chromosome locations determined by chromosome painting between human and porcine genomes (Goureau et al. 1996) or between hu-

© 2002 NRC Canada

man and bovine genomes (Chowdhary et al. 1996), or (*iii*) originated from an EST sequence. The original sequence of such a marker was used in a BLAST search against the human genome database and an orthologous relationship was established if the sequence identity was more than 82% when matched with a coding sequence or was more than 72% when matched with an untranslated region (UTR) of a given human gene within the length of a query. Checked or corrected genes for which the bovine or porcine symbols are different from the human symbols are listed at http://cgil.uoguelph.ca/pub/molecularpage.html.

Comparative characterization of bovine and porcine genomes

All of the orthologous genes and gene-family clusters defined as above were arranged in order based on their locations (in Mb) on the sequence map of each of the 22 human autosomes. The location of each gene was represented by its start base number on the human sequence map, as defined by NCBI. Because the relative orders of these genes in the human genome are known, the human genome was treated as primary, whereas bovine and porcine genomes were treated as secondary for CSAM mapping. The gene or gene-family cluster symbol and sequence location in the human genome, as well as the orthologous gene or marker symbol and chromosomal location in bovine and porcine genomes, are listed at http://cgil.uoguelph.ca/pub/molecularpage.html. In total, 991 genes and gene-family clusters were defined as orthologous between human and cattle, and 764 were defined as such between human and pigs. As defined by Kumar and colleagues (2001), a CSAM is an uninterrupted set of markers in the primary genome (human genome in our study) that are syntenic in the secondary genome (bovine or porcine genome in our study). The length (in Mb) of each CSAM was estimated as the distance between the "start" base of the first marker and the "stop" base of the last marker in a CSAM based on human sequence map information in our study. We considered a singleton to be a special case of CSAM in which a conserved region between two species was established with mapping evidence from only one gene. A singleton could potentially become an actual CSAM once more markers are mapped in both species.

Estimation of number of CSAMs between two species

To make a comprehensive estimate of the total number of CSAMs, the number of unobserved ("zero-size," Kumar et al. 2001) CSAMs was estimated. This was done by fitting a probability distribution to the observed data, reading the number of zero-size CSAMs from this distribution, and adding this estimated value to the observed total. For this purpose, the expected number of genes per CSAM (x) was assumed to be γ distributed with shape parameter α and trivial-scale parameter β as follows:

$$[1] \qquad \varphi(x) = [\beta^{\alpha} / \Gamma(\alpha)]x^{\alpha-1} e^{-\beta x}$$

This γ distribution is a generalization of the exponential distribution and accounts for variation in both length and gene density of the segments. The α shape parameter of this distribution describes the steepness of the distribution for small values of x, with values <1 indicating an excess of segments with few markers. This gamma model was chosen because it was found to provide a better fit to some observed mammalian CSAM characteristics than the more restricted exponential model (Kumar et al. 2001). For any conserved unit with mean size x, the actual size, k (number of markers), is described by a Poisson distribution as follows:

$$[2] \qquad q(k) = (x^k / k!)e^{-x}, \text{ where } k = 0,1,2,\ldots, n$$

Combining equations 1 and 2, we obtain a negative binomial distribution for k

$$[3] \qquad p(k) = \frac{[\Gamma(\alpha + k)\ \beta^{\alpha}]}{[k!\Gamma(\alpha)\ (1+\beta)^{k+\alpha}]}$$

In this distribution, $p(0)$ is the category of conserved units containing 0 markers. In the actual data, this category is unobserved and we need to estimate N, the total number of unobserved conserved segments. As previously described (Kumar et al. 2001), the observed portion ($k > 0$) of the distribution was smoothed and truncated before α and N (and hence the total number of markers) were estimated using the method of moments (Johnson and Kotz 1969).

For human–cattle, the α parameter was estimated to be 0.62, reflecting that there are more unobserved CSAMs than would be inferred using a simple exponential model (assuming $\alpha = 1$); for human–pig, the value was 0.96. The fact that the α value for pig is close to 1 indicates that the estimate using the Kumar method is very close to that derived using the Nadeau–Taylor method (Nadeau and Taylor 1984) in this case.

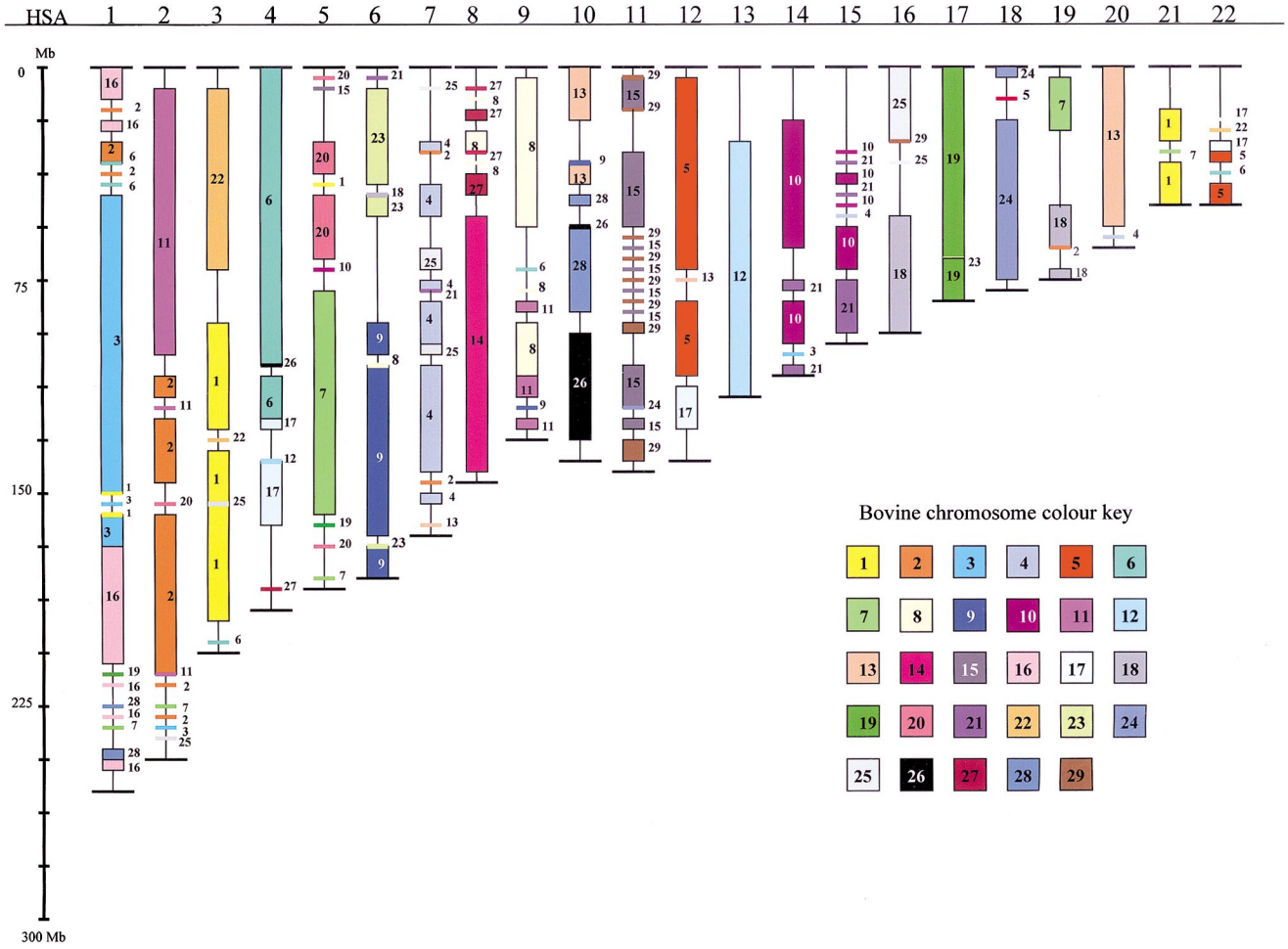
Results

A comparative view of bovine and porcine autosomes

Based on 991 genes or gene clusters mapped on bovine autosomes, 81 conserved syntenies and 161 CSAMs were identified between human and bovine autosomes (Fig. 1). Mapping of 764 genes or gene clusters in porcine autosomes enabled identification of 50 conserved syntenies and 95 CSAMs between human and porcine autosomes (Fig. 2). The numbers of conserved syntenies and CSAMs observed on each human chromosome varied from 1 to 8 and from 2 to 20, respectively, between human and bovine genomes (Fig. 1) and from 1 to 5 and 1 to 12, respectively, between human and porcine genomes (Fig. 2). The human autosomes are approximately 2906 Mb in total length, based on their sequence maps (NCBI data on Sept. 5, 2001). Using the human sequence map as a reference, 72 (2100/2906 Mb) and 74% (2148/2906 Mb) of the human genome was aligned with the bovine and porcine genomes, respectively. The aligned lengths (in Mb) to human genome sequences of each of the bovine autosomes were as follows: BTA1, 121; BTA2, 96; BTA3, 125; BTA4, 74; BTA5, 106; BTA6, 119; BTA7, 103; BTA8, 100; BTA9, 74; BTA10, 92; BTA11, 103; BTA12, 90; BTA13, 78; BTA14, 90; BTA15, 65; BTA16, 61; BTA17, 52; BTA18, 67; BTA19, 83; BTA20, 34; BTA21, 31; BTA22, 62; BTA23, 47; BTA24, 63; BTA25, 36; BTA26, 39; BTA27, 16; BTA28, 46; and BTA29, 27. The analogous figures for each of the porcine chromosomes were as follows: SSC1, 219; SSC2, 163; SSC3, 121; SSC4, 173; SSC5, 118; SSC6, 163; SSC7, 115; SSC8, 163; SSC9, 81; SSC10, 7;

© 2002 NRC Canada

Fig. 1. Comparative organization of the human and bovine genomes in a CSAM analysis using 991 gene markers. Bovine chromosomes are colour-coded, and the human chromosomes are painted with bovine chromosome-specific colours for homologous regions. The rectangles represent observed CSAMs and lines demonstrate unaligned regions.



SSC11, 95; SSC12, 70; SSC13, 217; SSC14, 186; SSC15, 116; SSC16, 50; SSC17, 43; and SSC18, 47.

Characterization of the observed CSAMs between two species

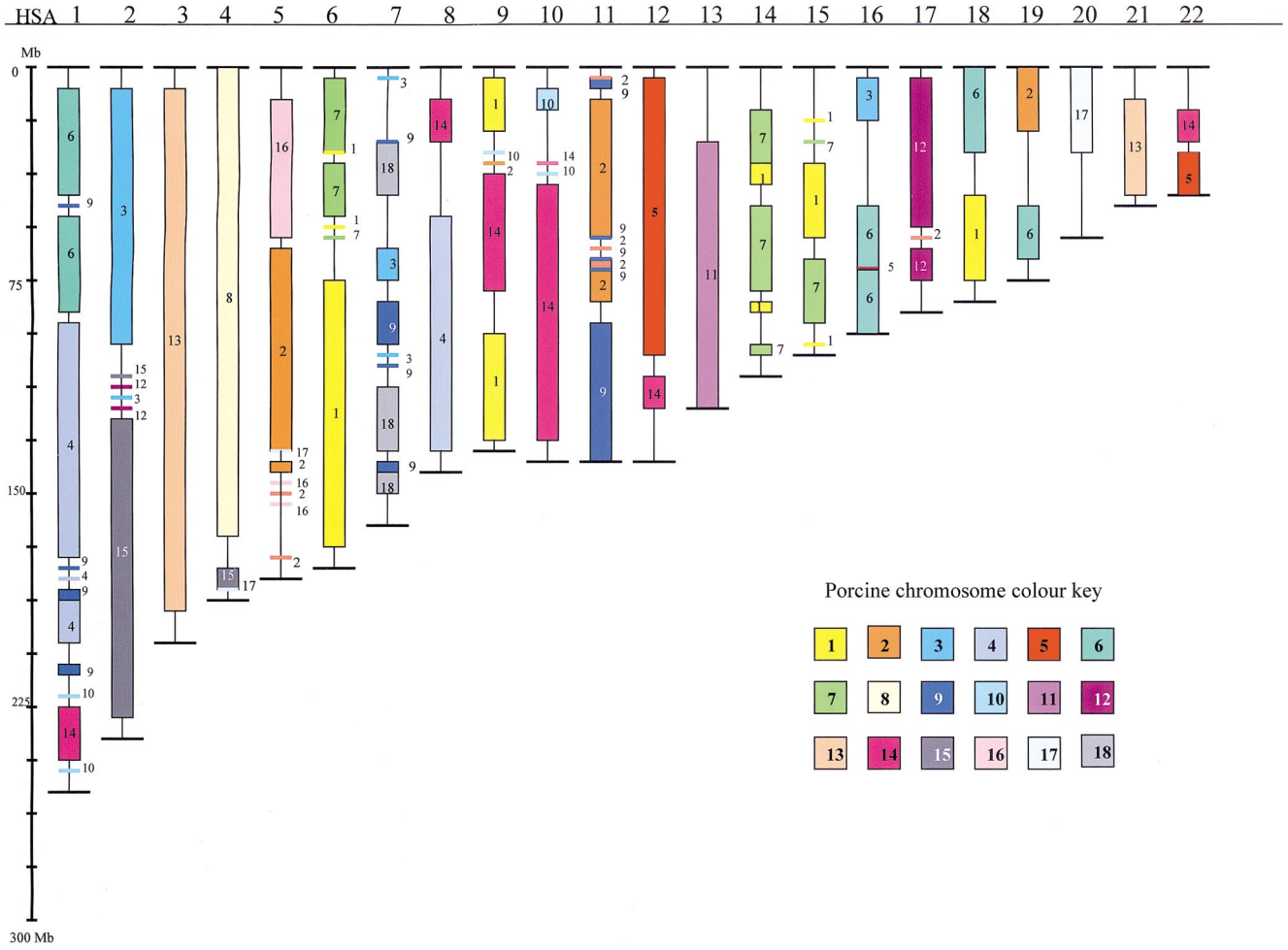
Based on the number of contiguous markers in each CSAM, we classified these CSAMs into five size groups: singletons (one marker only), small (2–4 markers), medium (5–10 markers), large (11–20 markers), and very large (>20 markers). The number of CSAMs detected for each of these five size groups, and their average sizes (Mb; in parentheses), were 59 (0.04), 43 (3.22), 31 (16.41), 15 (40.94), and 13 (64.25) for the human–bovine comparison and 32 (0.07), 21 (10.16), 20 (22.62), 16 (49.35), and 6 (115.34) for the human–porcine comparison. A single, large CSAM was observed for each of BTA12 (Fig. 1) and SSC11 (Fig. 2), corresponding to much of HSA13. Most of HSA3, 20, and 21 showed one CSAM syntenic to SSC13, 17, and 13, respectively (Fig. 2). In cattle (Fig. 1), BTA14 harbors a CSAM that is homologous to a large portion of HSA8 (93 out of a total of 143 Mb). In pigs (Fig. 2), SSC8 harbors a CSAM homologous to a large portion of HSA4 (163 out of a total of 189 Mb). A di-CSAM pattern, distributed as tandem or dispersed repeats (TR or DR, respectively), was found for some of the human chromosomes. The repeat number and type of either bovine or procine di-CSAM pattern (in parentheses) for HSA1 are 2 (BTA16–BTA2, TR), 2 (BTA3–BTA1, TR), 2 (BTA28–BTA16, TR), and 3 (SSC4–SSC9, TR); for HSA2, 3 (BTA11–BTA2, TR and DR); for HSA3, 2 (BTA22–BTA1, TR); for HSA5, 3 (SSC16–SSC2, TR and DR); for HSA6, 2 (BTA23–BTA9, DR) and 3 (SSC7–SSC1, TR); for HSA7, 3 (BTA25–BTA4, DR) and 3 (SSC9–SSC18, TR and DR); for HSA8, 3 (BTA27–BTA8, TR); for HSA10, 2 (BTA28–BTA26, TR) and 2 (SSC10–SSC14, TR); for HSA11, 8 (BTA29–BTA15, TR and DR) and 5 (SSC2–SSC9, TR); for HSA14, 2 (SSC7–SSC1, TR); and for HSA15, 3 (BTA10–BTA21, TR and DR) and 2 (SSC1–SSC7, TR).

Estimation of number of undetermined CSAMs in bovine and porcine genomes

183 segments between human and bovine genomes and 118 between human and porcine genomes remain unaligned with respect to potential CSAMs, because there are currently no orthologous gene markers available to correlate the genomes with each other (Figs. 1 and 2). These segments are

© 2002 NRC Canada

Fig. 2. Comparative organization of the human and porcine genomes in a CSAM analysis using 764 gene markers. Porcine chromosomes are colour-coded, and the human chromosomes are painted with porcine chromosome-specific colours for homologous regions. The rectangles represent observed CSAMs and lines demonstrate unaligned regions.



located either on telomeric regions or in gaps between observed CSAMs. Because the reported analysis was based on human autosomes, the total number of unaligned telomeric segments is 44. For the human–bovine comparison, 15 unaligned telomeric segments each measure less than 1 Mb in length, 8 are 1–3 Mb, 16 are 3–10 Mb, 2 are 10–20 Mb, and 3 are larger than 20 Mb; analogous numbers for the 44 unaligned telomeric segments in the human–porcine comparison are 5, 9, 16, 12, and 2, respectively. 139 CSAM gap segments remain between human and bovine genomes, and 74 between human and porcine genomes. Using the same length categories as described for the unaligned telomeric segments to classify these CSAM gap segments, 42, 38, 43, 13, and 3 were detected for the human–bovine comparison and 11, 21, 27, 9, and 6 for the human–porcine comparison. These unaligned segments span at least 800 Mb of the bovine genome and 760 Mb of the porcine genome. Using the statistical model developed by Kumar and colleagues (2001), an additional 63 CSAMs were estimated to exist between human and bovine genomes, and 18 between human and porcine genomes. The total numbers of CSAMs are expected to be about 224 between human and bovine genomes and about 113 between human and porcine genomes.

Discussion

In recent years, substantial effort has been made to compare the genomes of vertebrates, especially that of the human with various domestic animal species. Once reliable comparative maps are established, a comparative positional candidate-gene approach can be performed to search for candidate genes in a homologous region of the human transcript map to identify economically important trait loci in livestock species (Womack and Kata 1995). The success of this strategy depends, without doubt, on how well the human genome is aligned with that of the animal under study (Fridolfsson et al. 1997). Heterologous chromosomal painting (Zoo-FISH) has been widely used to reveal extensive, conserved, chromosomal segment homologies between chromosomes of the human and other species, and this approach has led to a global comparative map (Chowdhary et al. 1998). For example, Zoo-FISH analysis has identified and delineated about 50 segments of homology between human and bovine genomes (Solinas-Toldo et al. 1995) and about 40 such segments between human and porcine genomes (Rettenberger et al. 1995; Goureau et al. 1996). However, Zoo-FISH data alone can not provide the basis for comparative positional

© 2002 NRC Canada

candidate gene identification owing to its fairly crude resolution. The homologous segments generated by chromosomal painting still have to be defined more precisely both to evaluate how the synteny is conserved at the gene-order level and to detect genome rearrangement between species (Lahbib-Mansais et al. 2000). Linkage, somatic-cell hybrid, and FISH mapping of genes in bovine and porcine genomes (Eggen and Fries 1995; Womack and Kata 1995; Band et al. 2000; Archibald et al. 1995 and personal communication; Marklund et al. 1996; Rohrer et al. 1996; Hawken et al. 1999; Yerle et al. 1995, 1997; Pinton et al. 2000) provided us with the means to extend comparative mapping of bovine and porcine genomes at a whole-genome level with relatively high resolution and intensity. In this study, we have identified 991 genes and gene-family clusters for cattle and 764 for pigs as being orthologous to those found in the human genome. Based on the mapping information of these markers in human, bovine, and porcine genomes, 81 conserved syntenies and 161 CSAMs were identified between human and bovine autosomes and 50 conserved syntenies and 95 CSAMs were identified between human and porcine autosomes. The CSAM analysis revealed new homologies between human and bovine and between human and porcine genomes that were not observed in investigations using a Zoo-FISH approach (Solinas-Toldo et al. 1995; Rettenberger et al. 1995; and Goureau et al. 1996). As CSAM identification requires information on the relative locations of orthologous markers in one genome and only the chromosome number on which each marker resides in the other genome (Kumar et al. 2001). All markers can be combined in a single integrated map of a given genome, regardless of the strategy by which they were originally mapped. The maps constructed in this study using the CSAM approach can be regarded as a second generation of comparative maps between the human and bovine and between the human and porcine genomes (Figs. 1 and 2).

Orthologous genes are frequently used as landmarks in constructing comparative maps between human and other livestock species owing to their unambiguous similarity, large numbers, and precise locations (Nadeau and Sankoff 1998). The recent rapid expansion in numbers of ESTs, especially for cattle and pigs, provides us with an efficient way of gaining sequence data for identifying orthologous genes in livestock species. The numbers of EST entries in GenBank for cattle and pigs have reached 180 000 and 100 000, respectively (data accessed 1 October 2001). In fact, attempts have been made to systematically map ESTs in both the bovine (Band et al. 2000; Ma et al. 1998; Ozawa et al. 2000) and the porcine genome (Fridolfsson et al. 1997; Karnuah et al. 2001; Smith et al. 2001; and Tosser-Klopp et al. 2001). The use of ESTs has increased the density of markers in comparative maps between species, thus facilitating the candidate-gene approach for analyzing traits of interest. Placing ESTs on comparative maps has also facilitated their characterization as specific genes (Jiang et al. 2002). The orthologous status of ESTs should perhaps be determined before mapping to confirm their potential value in comparative mapping. During collection and verification of data for this study, we found that certain ESTs mapped in bovine and porcine genomes have no apparent human orthologs. Among the 768 genes on the cattle RH map, 687 were shown to have putative human orthologs, whereas the remaining 81 were ESTs or database sequences that have no significant human hits in UniGene (NCBI) (Band et al. 2000). Of 44 ESTs mapped in the porcine genome, only 22 orthologs were identified in human (Tosser-Klopp et al. 2001). In addition, some of the singletons identified in this study could also be an artifact of EST mapping because of the existence of paralogs and pseudogenes in the genome.

This paper provides the first report of a di-CSAM repeat pattern distributed tandemly or dispersed on human chromosomes when the human genome is aligned with either the bovine or the porcine genome (Figs. 1 and 2). This phenomenon should be further examined by fine mapping in these three mammalian species. The resolution of 59 singleton-CSAMs recorded in this study between bovine and human genomes, and 32 between porcine and human genomes, will be one of the factors determining the final status and number of these di-CSAM patterns. The release of the draft sequence of the human genome has left geneticists with "a mountain still to climb" (Butler and Smaglik 2000). Placing the contigs and genes in their correct order on each human chromosome remains difficult. For example, using the draft sequence data available about six months ago, we found that HSA1 consists of 13 CSAM segments equivalent to regions on SSC4, SSC6, SSC9, SSC10, and SSC14 in the order 6–14–4–14–4–9–4–9–4–14–10–14–10 (Melville et al. 2001). However, our current study, using updated human genome information, revealed 12 CSAM segments between HSA1 and the same five porcine chromosomes as described above, but the order is now changed to 6–9–6–4–9–4–9–4–9–10–14–10. As estimated, an additional 63 new CSAMs could exist between the human and porcine genomes and 18 between the human and porcine genomes. The placement of additional orthologous gene landmarks in bovine and porcine genomes to address those unaligned segments would also refine CSAM patterns among these three genomes. Eventually, the di-CSAM repeat patterns observed between human and bovine or porcine genomes could contribute to an understanding of the evolution of mammalian genomes.

Acknowledgements

We thank Ms. Heather L. Baltzer for assistance in data collection. Dr. Honghe Cao is a postdoctoral fellow supported by the Food System Biotechnology Centre at the University of Guelph. We are also grateful to anonymous reviewers for comments and advice on revising the manuscript. Financial assistance was provided by the Natural Sciences and Engineering Research Council of Canada; the GENEX Swine Group, Canada; the Dairy Cattle Genetics Research and Development Council (DairyGen), Canada; Ontario Swine Improvement, Canada; and the National Institutes of Health, U.S.A. (S.K.).

References

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dutonick, M., Polymeropoulos, M.H., et al. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science (Washington, D.C.), **252**: 1651–1656.

© 2002 NRC Canada

Andersson, L., Archibald, A., Ashburner, M., Audin, S., Barendse, W., et al. 1996. Comparative genome organization of vertebrates. The first international workshop on comparative genome organization. Mamm. Genome, **7**: 717–734.

Archibald, A.L., Haley, C.S., Brown, J.F., Copperwhite, S., McQueen, H.A., et al. 1995. The PiGMaP consortium linkage map of the pig (*Sus scrofa*). Mamm. Genome, **6**: 157–175.

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., and Lewin, H.A. 2000. An ordered comparative map of the cattle and human genomes. Genome Res. **10**: 1359–1368.

Butler, D., and Smaglik, P. 2000. Draft data leave geneticists with a mountain still to climb. Nature (London), **405**: 984–985.

Chowdhary, B.P., Fronicke, L., Gustavsson, I., and Scherthan, H. 1996. Comparative analysis of the cattle and human genomes: detection of ZOO-FISH and gene mapping-based chromosomal homologies. Mamm. Genome, **7**: 297–302.

Chowdhary, B.P., Raudsepp, T., Fronicke, L., and Scherthan, H. 1998. Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. Genome Res. **8**: 577–89.

Eggen, A., and Fries, R. 1995. An integrated cytogenetic and meiotic map of the bovine genome. Anim. Genet. **26**: 215–236.

Eppig, J.T. 1996. Comparative maps: adding pieces to the mammalian jigsaw puzzle. Curr. Opin. Genet. Dev. **6**: 723–730.

Eppig, J.T., and Nadeau, J.H. 1995. Comparative maps: the mammalian jigsaw puzzle. Curr. Opin. Genet. Dev. **5**: 709–716.

Fridolfsson, A.K., Hori, T., Wintero, A.K., Fredholm, M., Yerle, M., Robic, A., Andersson, L., and Ellegren, H. 1997. Expansion of the pig comparative map by expressed sequence tags (EST) mapping. Mamm. Genome, **8**: 907–912.

Goureau, A., Yerle, M., Schmitz, A., Riquet, J., Milan, D., Pinton, P., Frelat, G., and Gellin, J. 1996. Human and porcine correspondence of chromosome segments using bidirectional chromosome painting. Genomics, **36**: 252–262.

Hawken, R.J., Murtaugh, J., Flickinger, G.H., Yerle, M., Robic, A., Milan, D., Gellin, J., Beattie, C.W., Schook, L.B., and Alexander, L.J. 1999. A first-generation porcine whole-genome radiation hybrid map. Mamm. Genome, **10**: 824–30.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature (London), **409**: 860–921.

Jiang, Z.H., Priat, C., and Galibert, F. 1998. Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. Mamm. Genome, **9**: 577–587.

Jiang, Z.H., He, H., Hamasima, N., Suzuki, H., and Verrinder Gibbins, A.M. 2002. Comparative mapping of *Homo sapiens* chromosome 4 (HSA4) and *Sus scrofa* chromosome 8 (SSC8) using orthologous genes representing different cytogenetic bands as landmarks. Genome, **45**: 147–156.

Johnson, N.L., and Kotz, S. 1969. Distributions in statistics: discrete distributions. Houghton Mifflin, Boston, Mass.

Karnuah, A.B., Uenishi, H., Kiuchi, S., Kojima, M., Onishi, A., Yasue, H., and Mitsuhashi, T. 2001. Assignment of 64 genes expressed in 28-day-old pig embryo to radiation hybrid map. Mamm. Genome, **12**: 518–523.

Kumar, S., and Hedges, S.B. 1998. A molecular timescale for vertebrate evolution. Nature (London), **392**: 917–920.

Kumar, S., Gadagkar, S.R., Filipski, A., and Gu, X. 2001. Determination of the number of conserved chromosomal segments between species. Genetics, **157**: 1387–1395.

Lahbib-Mansais, Y., Leroux, S., Milan, D., Yerle, M., Robic, A., Jiang, Z., Andre, C., and Gellin, J. 2000. Comparative mapping between humans and pigs: localization of 58 anchorage markers (TOASTs) by use of porcine somatic cell and radiation hybrid panels. Mamm. Genome, **11**: 1098–106.

Lyons, L.A., Laughlin, T.F., Copeland, N.G., Jenkins, N.A., Womack, J.E., and O'Brien, S.J. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat. Genet. **15**: 47–56.

Ma, R.Z., van Eijk, M.J., Beever, J.E., Guerin, G., Mummery, C.L., and Lewin, H.A. 1998. Comparative analysis of 82 expressed sequence tags from a cattle ovary cDNA library. Mamm. Genome, **9**: 545–549.

Marklund, L., Johansson Moller, M., Hoyheim, B., Davies, W., Fredholm, M., Juneja, R.K., Mariani, P., Coppieters, W., Ellegren, H., and Andersson, L. 1996. A comprehensive linkage map of the pig based on a wild pig – Large White intercross. Anim. Genet. **27**: 255–69.

Melville, J.S., Jiang, Z.H., and Verrinder Gibbins, A.M. 2001. Cytogenetic banding block counterparts between porcine and human genomes based on 639 orthologous genes mapped in both species. Proceedings of 12th North America Colloquium on Animal Cytogentics and Gene Mapping, Davis, Calif. p. 38.

Nadeau, J.H., and Sankoff, D. 1998. Counting on comparative maps. Trends Genet. **14**: 495–501.

Nadeau, J.H., and Taylor, B.A. 1984. Lengths of chromosomal segments conserved since divergence of man and mouse. Proc. Natl. Acad. Sci. U.S.A. **81**: 814–818.

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A., and Copeland, N.G. 1993. Anchored reference loci for comparative genome mapping in mammals. Nat. Genet. **3**: 103–12.

O'Brien, S.J., Eisenberg, J.F., Miyamoto, M., Hedges, S.B., Kumar, S., et al. 1999. Genome maps 10. Comparative genomics. Mammalian radiations. Wall chart. Science (Washington, D.C.), **286**: 463–478.

Ozawa, A., Band, M.R., Larson, J.H., Donovan, J., Green, C.A., Womack, J.E., and Lewin, H.A. 2000. Comparative organization of cattle chromosome 5 revealed by comparative mapping by annotation and sequence similarity and radiation hybrid mapping. Proc. Natl. Acad. Sci. U.S.A. **97**: 4150–4155.

Pinton, P., Schibler, L., Cribiu, E., Gellin, J., and Yerle, M. 2000. Localization of 113 anchor loci in pigs: improvement of the comparative map for humans, pigs, and goats. Mamm. Genome, **11**: 306–315.

Rettenberger, G., Klett, C., Zechner, U., Kunz, J., Vogel, W., and Hameister, H. 1995. Visualization of the conservation of synteny between humans and pigs by heterologous chromosomal painting. Genomics, **26**: 372–378.

Rohrer, G.A., et al. 1996. A comprehensive map of the porcine genome. Genome Res. **6**: 371–391.

Sankoff, D., and Nadeau, J.H. 1996. Conserved synteny as a measure of genomic distance. Discrete Appl. Math. **71**: 247–257.

Smith, T.P., Fahrenkrug, S.C., Rohrer, G.A., Simmen, F.A., Rexroad, C.E., and Keele, J.W. 2001. Mapping of expressed sequence tags from a porcine early embryonic cDNA library. Anim. Genet. **32**: 66–72.

Solinas-Toldo, S., Lengauer, C., and Fries, R. 1995. Comparative genome map of human and cattle. Genomics, **27**: 489–496.

Tosser-Klopp, G., Bonnet, A., Yerle, M., and Hatey, F. 2001. Functional study and regional mapping of 44 hormono-regulated genes isolated from a porcine granulosa cell library. Genet. Sel. Evol. **33**: 69–87.

Venta, P.J., Brouillette, J.A., Yuzbasiyan-Gurkan, V., and Brewer, G.J. 1996. Gene-specific universal mammalian sequence-tagged sites: application to the canine genome. Biochem. Genet. **34**: 321–341.

© 2002 NRC Canada

Venter, J.C., et al. 2001. The sequence of the human genome. Science (Washington, D.C.), **291**: 1304–1351.

Womack, J.E., and Kata, S.R. 1995. Bovine genome mapping: evolutionary inference and the power of comparative genomics. Curr. Opin. Genet. Dev. **5**: 725–733.

Yerle, M., Lahbid-Mansais, Y., Mellink, C., Goureau, A., Pinton, P., et al. 1995. The PiGMaP consortium cytogenetic map of the domestic pig (*Sus scrofa domestica*). Mamm. Genome, **6**: 176–86.

Yerle, M., Lahbib-Mansais, Y., Pinton, P., Robic, A., Goureau, A., Milan, D., and Gellin, J. 1997. The cytogenetic map of the domestic pig. Mamm. Genome, **8**: 592–607.

© 2002 NRC Canada