Cytogenet Genome Res 102:100–108 (2003)
DOI: 10.1159/000075733

Cytogenetic and Genome Research

A new contribution to the integration of human and porcine genome maps: 623 new points of homology

A. Robic,[a] T. Faraut,[a] N. Iannuccelli,[a] Y. Lahbib-Mansais,[a] V. Cantegrel,[a] L. Alexander,[b] and D. Milan[a]

[a] Institut National de la Recherche Agronomique (INRA), Laboratoire de Génétique Cellulaire, BP27, Castanet Tolosan (France);
[b] Department of Veterinary PathoBiology, College of Veterinary Medicine, University of Minnesota, St Paul MN (USA)

Abstract. In this study we examined homologies between 1,735 porcine microsatellites and human sequence. For 1,710 microsatellites we directly used the sequence flanking the repeat available in GenBank. For a set of 305 microsatellites, a BAC library was screened and end-sequencing provided 461 additional sequences. Altogether 2,171 porcine sequences were tentatively aligned with the sequence of the human genome using the fasta program. Human homologies were observed for 652 microsatellite loci and porcine chromosome assignments available for 623 microsatellites provide useful links in the human and pig comparative map. Moreover for 92 STS, a significant sequence similarity was detected using at least two sequences and in all cases corresponding human locations were consistent. The present study allowed the integration of anonymous markers and the porcine linkage map into the framework of the comparative data between human and porcine genomes (http://w3.toulouse.inra.fr/lgc/pig/msat/). Moreover all conserved syntenic segments were defined on human chromosomes.

In the last few years, construction of mapping tools, such as radiation hybrid panels (Yerle et al., 1998, 2002), and the development of cDNA libraries in pigs, has been important for the identification and mapping of genes in this species. However the number of mapped loci needs to be increased in order to improve the resolution of the comparative mapping between human and porcine species, which is a prerequisite for positional cloning strategies of quantitative trait loci (QTL). Bidirectional painting has provided a framework of conserved synteny groups conserved between the pig and human (Goureau et al., 1996). Moreover gene mapping has confirmed the large-scale correspondences identified between human and pig (Lahbib-Mansais et al., 1999, 2000, 2003; Pinton et al., 2000; Rink et al., 2002).

Currently more than 1,200 markers have been mapped on porcine genetic linkage maps (Rohrer et al., 1994, 1996; Archibald et al., 1995; Marklund et al., 1996). Moreover mapping on the radiation hybrid panel IMpRH (Yerle et al., 2002) is in full swing (Hawken et al., 1999; Karnuah et al., 2001; Korwin-Kossakowska et al., 2002; Krause et al., 2002; Rink et al., 2002; Lahbib-Mansais et al., 2003) and 6000 markers are already mapped. A first generation RH comparative map of the porcine and human genome has already been published (Rink et al., 2002; Lahbib-Mansais et al., 2003) but this map includes only EST markers. Since the human genome project is nearly finished, comparative mapping approaches using the human information greatly facilitate the construction of physical maps in other mammalian species. Moreover the incorporation of noncoding sequence into the comparative mapping framework could be now considered even if this strategy seems laborious.

Farber and Medrano (2003) developed an approach to identify homologies existing between livestock microsatellite flanking sequences (STS sequence) and the human genome sequence. In this study we present two systematic approaches to identify genome-wide relationships between microsatellite loci and human sequence for 1,735 porcine microsatellite loci. This study allowed characterization of 623 new anchoring loci on the human and pig comparative map.

Supported by grants from the French program "AGENAE".

Received 6 May 2003; manuscript accepted 28 July 2003.

Request reprints from Annie Robic, INRA
Laboratoire de Génétique Cellulaire, BP27
31326 Castanet Tolosan Cedex (France)
telephone: (33) 05 61 28 51 15; fax: (33) 05 61 28 53 08
e-mail: arobic@toulouse.inra.fr

KARGER
Fax + 41 61 306 12 34
E-mail karger@karger.ch
www.karger.com

0301–0171/03/1024–0100$19.50/0

Accessible online at:
www.karger.com/cgr

Table 1. Evaluation grid to select hit after a fasta analysis. Selections were based on the percentage of identity. The minimal rate is a function of the length of the alignment detected by the fasta program. The second lines (italics) allowed taking into account gaps which were accepted only for large alignments. This last possibility was not used for BAC-end sequences. Indeed they did not contain microsatellites which were masked and which were suitable for inducing a major gap inside the alignment. The examples suggested are fictitious.

Description of hits obtained by fasta analysis		
Alignment (bp)	% Identity minimal	% Ungapped minimal
40	88	88
50	84	84
60	80	80
	76	*82*
70	78	78
	74	*82*
80	76	76
	72	*78*
90	74	74
	66	*76*
100	72	72
	64	*74*
110	70	70
	60	*72*
120	68	68
	58	*70*
120	66	66
	56	*68*
140	65	65
	54	*66*
>150	65	65
	50	*65*

Examples	
90% identity (90% ungapped) in 72 nt overlap	hit retained
74% identity (83% ungapped) in 72 nt overlap	hit retained
70% identity (83% ungapped) in 72 nt overlap	hit not retained
73% identity (83% ungapped) in 78 nt overlap	hit retained

Materials and methods

Microsatellites

Porcine microsatellites flanking sequences were downloaded from the GenBank sequence database using the Entrez nucleotide query website (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide).

Comparative sequence analysis

RepeatMasker with option "other mammals" was used to mask simple repeat and interspersed repetitive elements from each sequence set (Smit and Green at http://ftp.genome.washington.edu/RM/repeatMasker.html). Each masked sequence set was queried against the human genome sequence (primate division of GenBank, January 2003) using fasta program (Pearson and Lipman, 1988; software version 3.3t09, May 2001). A first script was written to automate the masking and fasta analysis. A second script was written to extract results from the fasta output. The table was sorted and all matches with an *expected value* (e-value) $>10^{-3}$ were discarded. This table contained all hits found for each sequence. When the number of hits was low (<8) only the first was retained. When the number of hits was very high (between 50 and 3,000) and/or when the probability of the first was similar to the probability of others we retained no hits. We developed an evaluation grid to retain or discard hits and four examples are given to illustrate this selection (Table 1).

Match annotation

Whereas to speak about the contents of the *fasta* analysis output file we employ the term "hit" and prefer to reserve the term of "match" to indicate the selected and annoted hit.

Porcine microsatellite map locations were identified from GenBank annotation, or from available information on ARKdb genome databases (Hu et al., 2001, http://www.thearkdb.org), or on the USDA database (http://sol.marc.usda.gov/), or on the IMpRH database (Milan et al., 2000; http://imprh.toulouse.inra.fr/).

Locations on the human genome were extracted from the UCSC human genome assembly of November 2002 (http://genome.cse.ucsc.edu/) or from the OMIM database (http://www.ncbi.nlm.nih.gov/Omim/).

To validate matches, known conserved syntenic relationships were examined using information available at http://www.toulouse.inra.fr/lgc/pig/compare/SSC.htm.

Results

A total of 1,735 microsatellites were analyzed in this study. For 1,710 microsatellites, STS sequences (including flanking sequences) were downloaded from GenBank. BAC end sequences were determined for 323 clones screened from the INRA library (Rogel-Gaillard et al., 1999) with 305 microsatellites providing 461 additional sequences (GenBank BX465382–BX465833 and BX511324–BX511332).

In total 2,171 sequences (1,710 + 461) were then compared to the human genome sequence with a fasta process with a threshold for the e-value of 10^{-3}. Approximately 25,000 hits were obtained and after the selection of the best hit for each comparison, we obtained approximately 1,000 hits (description in Materials and methods). Usually hits with e-values $>10^{-5}$ are considered as non-significant and are discarded. We empirically developed an evaluation grid to select hits on the identity percentage (Table 1). After this manual selection we retained 830 hits, and among them 25% had an e-value $>10^{-5}$. The human location was determined for all 830 human sequences

Table 2. All scores obtained in this study.

Hits		Microsatellites		
33[a]	hits with an e- value $>10^{-5}$ + incompatiblity with current knowledge of comparative map		33	they were discarded
36	hits are incompatible with current knowledge of comparative map		36	they were not considered
503	hits found using STS sequence	503	652[b]	623 with porcine localization **= 623 anchor loci on the comparative map**
258	hits found using BAC extremities (in 52 cases, the homology was obtained with two or three BACs sequences characterizing the same microsatellite)	203		29 without porcine localization
830	Total			

[a] These hits were not reported in Tables.
[b] We find the number of homologies detected using the two strategies (503 + 203 – 652 = 54).

implicated in these 830 hits as described in Materials and methods.

Knowing conserved syntenic relationships between human and porcine genomes, a first examination of results made it possible to discard 33 matches which were not compatible with current comparative data and whose e-values were not low enough to be considered significant ($>10^{-5}$). On the other hand, only 36 homologies with significant e-values ($<10^{-5}$) were not considered. These results were not discarded but only not considered to define conserved syntenic fragments: they appear in all Table results. These 36 matches were not compatible with current knowledge of the comparative map and 30 of 36 were only mapped on the IMpRH panel (Korwin-Kossakowska et al., 2002; Krause et al., 2002). Moreover if these 36 matches were considered, they would characterize 36 new segments of homology. After the final elimination of 33 matches and the not-considered 36 others, there remain 761 matches which relate to 652 microsatellites (Table 2). All were validated but porcine localization was not available for 29 microsatellites and these corresponding matches were not informative. Therefore 623 new anchors on the comparative map were characterized.

Using STS sequences, 503 points of orthology were characterized (29% of cases). With the second strategy 323 BACs were sequenced for one or two extremities with 513 bp available on average for each clone (Iannuccelli et al., in preparation). We obtained a corresponding human location for 209 BACs and 203 microsatellites (score = 65%). When more than one BAC-end sequence was available for one microsatellite, corresponding deduced human locations were always in accordance (52 cases). Moreover for 54 microsatellites, orthologous positions were detected using the two strategies (STS/BAC-end) and they were always consistent. Among the 92 cases (52+54) where we found a significant homology with at least two porcine sequences characterizing the same microsatellite, 30/33 hits obtained with an e-value $>10^{-5}$ were confirmed by a second (obtained with a second porcine sequence) with a very significant probability ($<10^{-10}$). Therefore human homology was obtained for 652 porcine microsatellites but 29 STS were not assigned on the porcine genome and only 623 homology points were characterized and exploitable.

These 623 new points of homology between porcine and human genomes were compatible with current knowledge of the comparative map available at http://www.toulouse.inra.fr/lgc/pig/compare/SSC.htm. In particular our results allowed a confirmation of correspondence between SSC3 and HSA7 (three matches), between SSC10 and HSA9 (four matches), between SSC14 and HSA9 (one match), between SSC15 and HSA4 (one match), between SSC15 and HSA8 (four matches), and between SSC17 and HSA4 (two matches).

To illustrate different categories of results obtained, we present hits observed for porcine microsatellite mapped from SSC1 (Table 3). Seventy seven hits were observed for 68 microsatellites. When several sequences were examined, similar results were obtained. For example, for S0320, three BAC-end sequences and the STS sequence matched with two human sequences on HSA9 with approximate positions of 113.26 and 113.36 Mb. These match results are consistent each other and this homology point is also compatible with present knowledge of the comparative map and so this result has been validated. As seen in Table 3 we obtained several hits with e-values $>10^{-5}$. When these hits were compatible with current knowledge of the comparative map (SW1997, SW1020, SW373–) they were retained. On the other hand when we have no additional arguments to retain these hits, they were discarded (SW1824, SW1957). Four microsatellites (UMNp431, UMNp373, UMNp528, UMNp616) harbored an e-value sufficient to be considered as sure but these four matches were not compatible with current comparative genome data. These results were not discarded but only not considered. Therefore 65 microsatellites originating from SSC1 had at least one significant match to human genomic sequence and were grouped in five chromosomal segments on HSA6, HSA9, HSA14, HSA15 and HSA18 (Table 3). Among these 65 matches only 39 have been previously mapped on the USDA linkage map (http://sol.marc.usda.gov/). The comparative map including the porcine linkage map and the human physical map (Fig. 1a) was compatible with information available at http://www.toulouse.inra.fr/lgc/pig/compare/table.htm. Nevertheless the conserved syntenic fragment between SSC1 and HSA18 which had already been detected was confirmed and enlarged on the q-arm of SSC1.

Table 3. 77 hits involving 68 porcine microsatellites localized on SSC1. When at least two porcine sequences characterizing a same microsatellite were used, hits are indicated by a vertical connecting line. Hits were classified by the position of the homology point on a human chromosome.

Porcine microsatellite	Porcine linkage position (USDA, cM)	Porcine sequence[a]	Human sequence	HSA	Approximate position (Mb)	*Expected Value*	Description of alignment
SW1957[b]		AF253752	AC108153	**4**	**53.9**	**0.0083**	69.369% identity (72.642% ungapped) in 111 nt overlap
UMNp431[c]		AF511275	AC079768	**4**	**171.21**	2.2e-11	70.701% identity (72.549% ungapped) in 157 nt overlap
microsatellite		AJ277795	AL590668	6	49.41	3.1e-23	67.708% identity (79.592% ungapped) in 288 nt overlap
SW1653	49.4	AF253682	AL109922	6	65.67	1.3e-05	59.574% identity (71.795% ungapped) in 141 nt overlap
S0396		U78022	AL357507	6	74.86	2.2e-09	72.165% identity (76.503% ungapped) in 194 nt overlap
SW2185	67.6	*BX465721*	AL049696	6	80.95	0.00057	67.568% identity (67.568% ungapped) in 111 nt overlap
SW1430	58.5	AF253621	AL359715	6	81.09	2.1e-05	75.000% identity (80.488% ungapped) in 88 nt overlap
SO317		X77282	AL590143	6	85.36	2.8e-15	74.390% identity (75.776% ungapped) in 164 nt overlap
SO318		X77281	AL590143	6	85.36	2.8e-15	74.390% identity (75.776% ungapped) in 164 nt overlap
SW1123	52.9	AF225089	AL590392	6	88.58	5.7e-09	80.851% identity (81.720% ungapped) in 94 nt overlap
SW1997	53.6	AF253768	AL592428	6	92.58	0.0015	52.217% identity (72.109% ungapped) in 203 nt overlap
UMNp192		AF511132	AL512490	6	108.16	2.9e-13	70.335% identity (73.869% ungapped) in 209 nt overlap
SW952	56.5	*BX465473*	AL500524	6	114.56	2.2e-27	75.325% identity (77.852% ungapped) in 308 nt overlap
SW952	56.5	*BX465474*	AL021327	6	114.74	3.3e-17	72.000% identity (73.303% ungapped) in 225 nt overlap
M. triadin		AJ224992	AL603902	6	123.37	3.6e-06	75.439% identity (78.182% ungapped) in 114 nt overlap
SW1851	44.6	AF225128	AL357274	6	130.79	5.2e-05	78.462% identity (78.462% ungapped) in 65 nt overlap
S0008	43.5	M97235	AC005587	6	131.90	1.0e-08	69.444% identity (73.529% ungapped) in 144 nt overlap
S0008	43.5	*BX465652*	AC005587	6	131.90	2.9e-09	79.798% identity (81.443% ungapped) in 99 nt overlap
UMNp467		AF511298	AL589674	6	142.22	1.3e-34	77.936% identity (79.348% ungapped) in 281 nt overlap
sZ002		AF279701	AL049844	6	144.05	6.4e-05	71.560% identity (74.286% ungapped) in 109 nt overlap
UMNp160		AF375756	AL109755	6	144.17	3.1e-26	73.294% identity (78.165% ungapped) in 337 nt overlap
SW1332	29.2	AF253594	AL023283	6	145.41	0.00032	80.328% identity (85.965% ungapped) in 61 nt overlap
SW137	23.5	AF235212	AL359252	6	149.01	3.8e-08	66.364% identity (68.545% ungapped) in 220 nt overlap
SW64	23.5	AF225100	AL359252	6	149.01	3.5e-14	73.410% identity (76.506% ungapped) in 173 nt overlap
UMNp380		AF511243	AL078582	6	152.32	9.3e-38	70.757% identity (72.267% ungapped) in 383 nt overlap
SWR485	16.4	AF235454	AL589963	6	152.74	0.00015	78.571% identity (79.710% ungapped) in 70 nt overlap
SW2184		AF253814	AL078583	6	161.98	4.5e-08	90.278% identity (90.278% ungapped) in 72 nt overlap
UMNp373[c]		AF511238	AC005686	**7**	**33.94**	9.4e-07	57.143% identity (71.429% ungapped) in 175 nt overlap
HY-N13		AB050040	AL591644	9	1.94	6.7e-15	69.430% identity (72.043% ungapped) in 193 nt overlap
HY-N26		AB050046	AL353741	9	6.15	4.9e-13	69.608% identity (72.821% ungapped) in 204 nt overlap
S0020	83.2	*BX465393*	AL160053	9	16.33	3.7e-44	75.245% identity (78.920% ungapped) in 408 nt overlap
S0020	83.2	*BX465555*	AL160053	9	16.33	7.3e-05	73.118% identity (75.556% ungapped) in 93 nt overlap
S0142	83.2	*BX465393*	AL160053	9	16.33	3.7e-44	75.245% identity (78.920% ungapped) in 408 nt overlap
SW1970	83.2	AF253756	AL353895	9	18.86	7.3e-08	55.729% identity (79.851% ungapped) in 192 nt overlap
SW780		*BX465635*	AL133281	9	19.86	3.5e-22	73.874% identity (74.886% ungapped) in 222 nt overlap
UMNp55		AF375685	AL512635	9	20.46	6.4e-10	71.429% identity (77.957% ungapped) in 203 nt overlap
SW1020	83.7	AF253566	AL451137	9	26.94	0.0066	76.543% identity (78.481% ungapped) in 81 nt overlap
SW2551	95.8	AF225182	AL161781	9	37.15	2.1e-09	87.013% identity (89.333% ungapped) in 77 nt overlap
UMNp484		AF511312	AL162412	9	64.21	0.0093	65.432% identity (71.622% ungapped) in 162 nt overlap
HY-N15		AB050042	AL158154	9	76.35	6.6e-06	67.722% identity (74.306% ungapped) in 158 nt overlap
SW1462	93.9	*BX465776*	AL357032	9	76.75	1.4e-07	85.057% identity (87.059% ungapped) in 87 nt overlap
SW1462	93.9	AF253631	AL162726	9	76.91	7.5e-09	83.871% identity (84.783% ungapped) in 93 nt overlap
HY-N19		AB050043	AC068050	9	102.53	2.4e-23	76.232% identity (79.697% ungapped) in 345 nt overlap
SW974	102.9	AF225111	AL358779	9	103.25	5.6e-07	81.481% identity (88.000% ungapped) in 81 nt overlap
S0302	102.9	*BX465686*	AL162733	9	104.01	9.6e-05	75.269% identity (76.923% ungapped) in 93 nt overlap
S0302	102.9	U10321	AL162733	9	104.01	1.8e-07	73.451% identity (74.107% ungapped) in 113 nt overlap
S0354	104.5	L29228	AL139041	9	107.41	1.4e-22	80.588% identity (83.537% ungapped) in 170 nt overlap
HY-N6		AB050036	AL359455	9	108.00	3.7e-12	64.122% identity (68.852% ungapped) in 262 nt overlap
UMNp367		AF511235	AL161630	9	112.00	5.1e-09	73.950% identity (75.862% ungapped) in 119 nt overlap
S0320	112.5	*BX465446*	AL157780	9	113.26	4.5e-20	67.333% identity (68.942% ungapped) in 300 nt overlap
S0320	112.5	*BX465438*	AL512602	9	113.36	3.5e-29	72.107% identity (75.000% ungapped) in 337 nt overlap
S0320	112.5	*BX465447*	AL512602	9	113.36	3.7e-19	72.107% identity (75.000% ungapped) in 337 nt overlap
S0320		X77284	AL512602	9	113.36	5.2e-12	60.891% identity (75.926% ungapped) in 202 nt overlap
SW705	122.6	AF235342	AL137846	9	118.82	1.3e-06	70.732% identity (74.359% ungapped) in 123 nt overlap
SW1301	140.5	*BX465487*	AL354855	9	125.76	3.4e-11	69.744% identity (71.958% ungapped) in 195 nt overlap
SW1301	140.5	*BX465488*	AL358781	9	125.89	6.6e-22	80.503% identity (81.013% ungapped) in 159 nt overlap
HYN1		*BX465502*	AL513102	9	126.48	3.9e-13	67.511% identity (71.749% ungapped) in 237 nt overlap
SW1824[b]		AF225127	AC117502	**12**	**69.7**	**0.0025**	80.000% identity (80.000% ungapped) in 65 nt overlap
UMNp528[c]		AF511339	AC084881	**12**	**119.34**	8.6e-09	71.324% identity (73.485% ungapped) in 136 nt overlap
UMNp616[c]		AF511394	AL162455	**13**	**88.91**	7.8e-34	66.728% identity (69.157% ungapped) in 541 nt overlap
UMNp41		AF375673	AL138498	14	36.15	0.00002	74.790% identity (77.391% ungapped) in 119 nt overlap
SO313		X76937	AL161664	14	40.41	2.3e-17	67.949% identity (75.177% ungapped) in 312 nt overlap
SW962	80.5	AF235495	AL049874	14	54.87	0.0057	73.171% identity (77.922% ungapped) in 82 nt overlap
SW216	82.4	*BX465710*	AF215937	14	58.50	1.9e-55	78.555% identity (81.882% ungapped) in 443 nt overlap
sZ003[d]		AF279702	AL132641	**14**	**79.95**	**0.0021**	63.298% identity (67.614% ungapped) in 188 nt overlap
UMNp29[d]		AF511181	AL358292	**14**	**81.33**	**6.6e-05**	76.923% identity (76.923% ungapped) in 78 nt overlap
UMNp486		AF511314	AC091074	15	39.71	6.8e-12	73.288% identity (74.306% ungapped) in 146 nt overlap
UMNp256		AF511157	AC024061	15	46.92	1.0e-10	73.418% identity (73.418% ungapped) in 158 nt overlap
UMNp117		AF375732	AC018618	15	55.06	9.8e-05	72.951% identity (79.464% ungapped) in 122 nt overlap

➤

Table 3 (continued)

Porcine microsatellite	Porcine linkage position (USDA, cM)	Porcine sequence[a]	Human sequence	HSA	Approximate position (Mb)	*Expected Value*	Description of alignment
SW2073	79.4	AF253788	AC016355	15	60.89	2.7e-18	89.744% identity (92.920% ungapped) in 117 nt overlap
HY-N21		AB050044	AC087593	15	82.12	2.2e-20	71.318% identity (74.194% ungapped) in 258 nt overlap
UMNp84		AF375710	AC069029	15	93.27	3.4e-24	75.745% identity (79.111% ungapped) in 235 nt overlap
SW373	119.5	AF225095	AC090916	18	35.81	0.00049	81.667% identity (83.051% ungapped) in 60 nt overlap
SW1668	60.2	AF253686	AC023421	18	43.04	1.6e-08	73.276% identity (78.704% ungapped) in 116 nt overlap
UMNp345		AF511218	AC018994	18	53.10	4.4e-16	81.633% identity (83.333% ungapped) in 147 nt overlap
UMNp330		AF511204	AC107990	18	57.51	1.1e-09	55.450% identity (68.824% ungapped) in 211 nt overlap
S0331		L36911	AC011930	18	67.75	7.7e-06	80.000% identity (83.333% ungapped) in 75 nt overlap

[a] BAC-end sequences are indicated in italics.
[b] Two hits were discarded because they harbored too high *e-values* and were incompatible with current knowledge of comparative maps.
[c] Four hits were not compatible with current knowledge of comparative maps and were not retained.
[d] Hits considered as provisional.

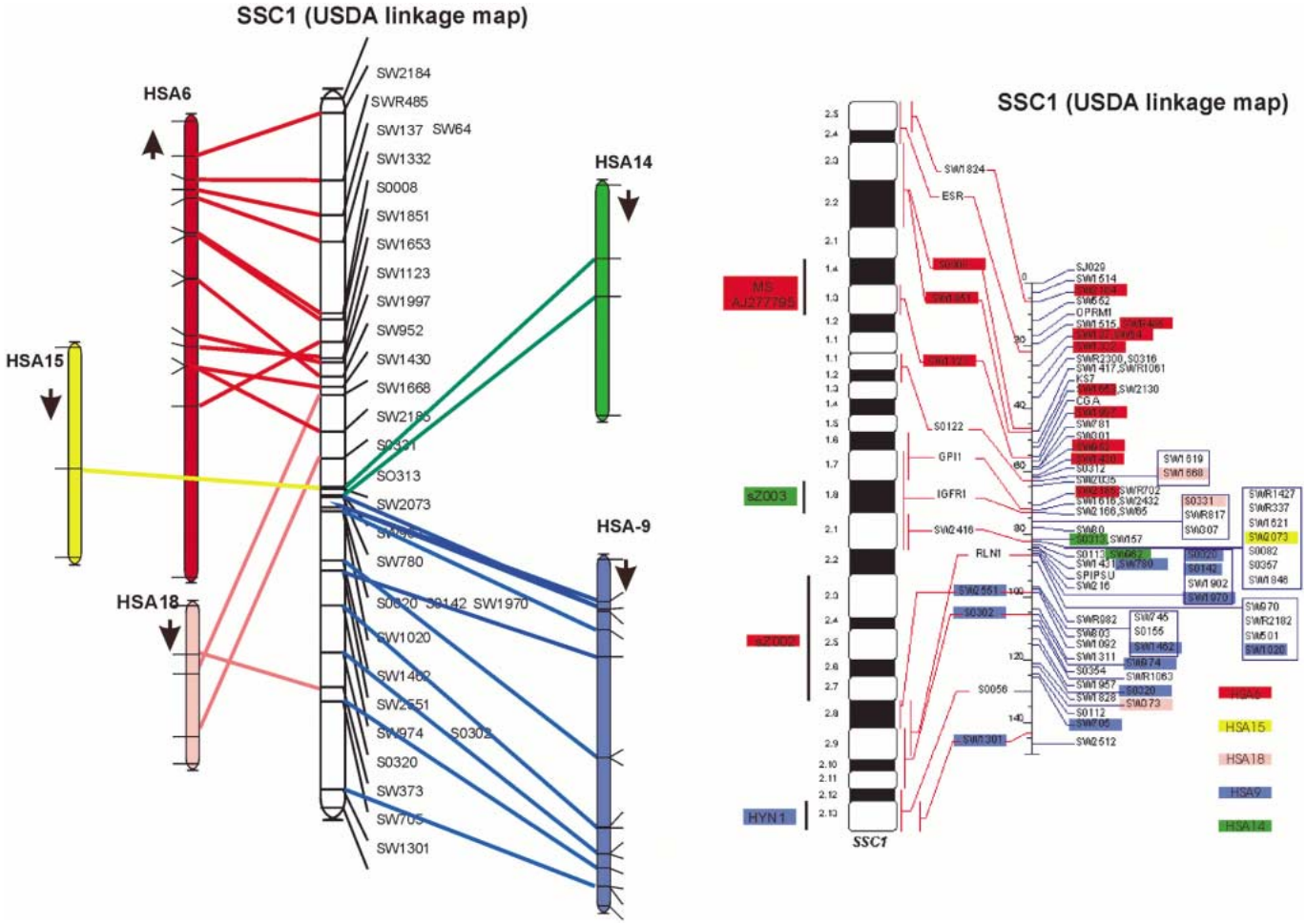


Fig. 1. Visualization of conserved syntenic fragments between human genome and SSC1. (**a**) The order of orthologous loci is compared between the USDA linkage map and human physical map. (**b**) Visualization of human homologies on USDA linkage map of SSC1. Some markers localized on the porcine cytogenetic map were added to the USDA linkage map (at the left of the chromosome drawing). The visualization of human homologies on the USDA linkage map is available for all porcine chromosomes on http://w3.toulouse.inra.fr/lgc/pig/msat/.

All homology data are reported in tables sorted by porcine or human chromosomes and are available at http://w3.toulouse.inra.fr/lgc/pig/msat/. When linkage data is available (60% of markers) it is possible to show conserved syntenic fragments on porcine linkage maps. All figures (e.g. Fig. 1b) are available at http://w3.toulouse.inra.fr/lgc/pig/msat/. When we examined all results sorted by human chromosomes, we were able to establish the precise location of the conserved syntenic breakpoints on human chromosomes. For example we are able to determine that the human chromosomal segment from HSA14 homologue to SSC1 was divided into two sub-segments. This observation is not concordant with Rink et al. (2002) and was based on two matches having a high e-value. Consequently, this segmentation in two sub-segments could be considered as provisional. Figure 2 allows visualization on human chromosomes of all conserved syntenic segments found here. The same figure is available in color at http://w3.toulouse.inra.fr/lgc/pig/msat/.

Discussion

Until now, results have been accumulating on the comparative map between the porcine and human genome due to the mapping of ESTs. Radiation hybrid mapping is included in this comparative map but it is very difficult to use the linkage map without the use of intermediate cytogenetic or RH maps. Here we propose a new strategy to include genetic markers in the comparative data.

We did not use a selection of hits based only on the e-value. We selected results including 25% with an e-value $>10^{-5}$. In sequence alignments those hits are generally discarded and here we tried to save results obtained with short sequences. We empirically developed an evaluation grid. Selections were made using the percentage of identity but the minimal rate was a function of the length of the alignment. Gaps inside the alignment were accepted only for large matches. Obtaining 92 consistent results by at least two different porcine sequences and especially 30 hits observed with a non-significant e-value ($>10^{-5}$) and confirmed by a very significant second hit, showed the interest of this evaluation grid. Nevertheless, this grid is empirical and 15% of hits having an e-value $>10^{-5}$ (33 discarded) and a maximum of 6% (36 not considered) of hits with an e-value $<10^{-5}$ were wrongly selected. Methodology employed must give a maximum of results entering within the framework of current knowledge of the comparative map to make a safe description of minimum innovations.

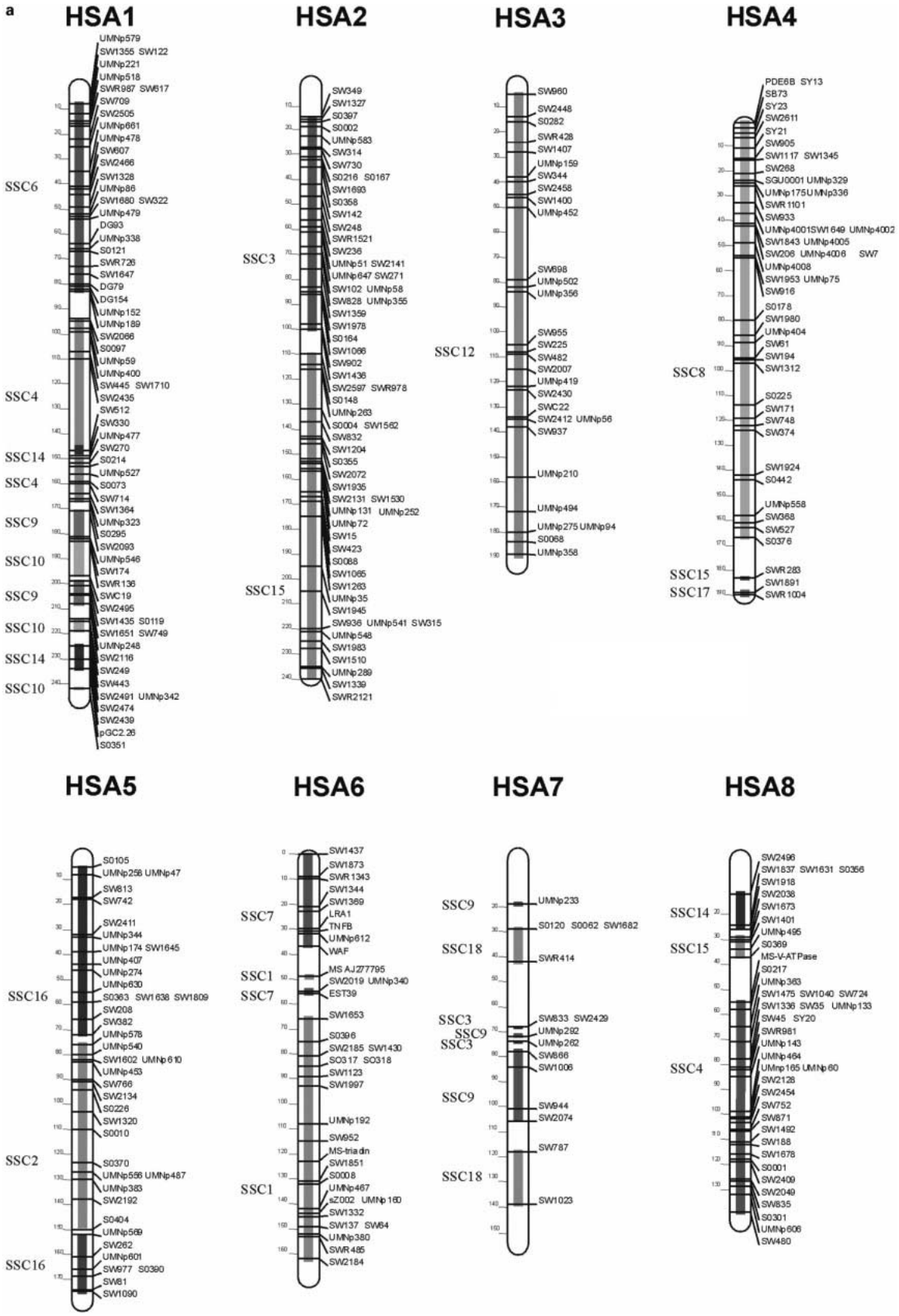
We did not retain 36 porcine markers with matches on the human genome with e-values sufficient to be considered as sure. These results were incompatible with present data accumulated on the comparative map between human and porcine genomes. Moreover none of these matches have been confirmed by another available match here or elsewhere. Therefore if these 36 matches were considered, they would characterize 36 new segments of homology. For most of them, the chromosomal assignment on IMpRH was obtained with a low LOD score. There is too much doubt about these 36 markers to say that 36 new conserved syntenic fragments were identified. Consequently they will not be considered until there is an eventual new chromosomal assignment or a confirmation by others. Consequently the grid used here to select hits is probably more effective than we are able to show (15 and 6% wrong selections with respectively e-values $>10^{-5}$ and $<10^{-5}$). Eventual novel conserved syntenic fragments are perhaps included among these 36 matches and they were included in all tables.

All chromosome correspondences identified by chromosomal painting were confirmed. Among correspondences already characterized or suspected by other authors six were detected in our 623 results.

We identified human orthologous positions for 65 porcine microsatellite loci originated from SSC1 but only 39 could be used to produce a bi-directional comparative map (Fig. 1). These results are compatible with bi-directional painting (Goureau et al., 1996) and allow the definition of conserved syntenic segments on the linkage map of this porcine chromosome. The conserved syntenic fragments SSC1/HSA14 and SSC1/HSA15 appeared localized near SSC1q2.1, but the linkage map of this region did not allow a good exploration of the segment q2.2-q2.7. On the other hand the dispersion of the conserved synteny SSC1/HSA18 along the q-arm is confirmed. Moreover we were able to study the order of loci inside conserved syntenic chromosomal segments of HSA6 and HSA9. The loci order is conserved only inside small sub-regions and it appears very important to increase the density of loci on the comparative map to avoid concluding too quickly on the conservation of the order of loci.

This study allowed the integration of the porcine linkage map in the framework comparative map. This approach is very useful for QTL studies to avoid using RH mapping as an intermediate step. Figures including new results of comparative maps anchored on each porcine chromosome linkage maps are available on the web. It would be tedious to describe each contribution of the integration of the porcine linkage map in the framework of comparative mapping, therefore only one has been detailed here. Homologies have already been detected between loci from the p-arm of SSC5 and HSA22 (Rink et al., 2002; Lahbib-Mansais et al., 2003). We reported a match between SW152 and a human sequence originating from HSA22. The sub-region of SSC5 around SW152 was not "attributed" on the comparative map of Lahbib-Mansais et al. (2003) and it might be possible that a new conserved syntenic fragment has been characterized here.

It is not possible to integrate all results in comparative maps initialized on porcine chromosomes because markers have originated from several porcine maps. On the other hand all results are useful and are used to initialize a comparative map on human chromosomes. Moreover when we examined all results sorted by human chromosome (Table available on the web at http://w3.toulouse.inra.fr/lgc/pig/msat/), we were able to improve the precision of the localization of the conserved syntenic breakpoints on each human chromosome. Figure 2 allowed visualization on human chromosomes of all conserved syntenic segments found here. Contrary to Fig. 1, this representation is not a punctual drawing of found homologies: we connected the identical homologies for better visualizing conserved syntenic fragments. The risk is to include some sub-


a
HSA1
UMNp579
SW1355 SW122
UMNp221
UMNp518
SWR987 SW617
SW709
SW2505
UMNp661
UMNp478
SW607
SW2466
SW1328
UMNp86
SW1680 SW322
UMNp479
DG93
UMNp338
S0121
SWR726
SW1647
DG79
DG154
UMNp152
UMNp189
SW2066
S0097
UMNp59
UMNp400
SW445 SW1710
SW2435
SW512
SW330
UMNp477
SW270
S0214
UMNp527
S0073
SW714
SW1364
UMNp323
S0295
SW2093
UMNp546
SW174
SWR136
SWC19
SW2495
SW1435 S0119
SW1651 SW749
UMNp248
SW2116
SW249
SW443
SW2491 UMNp342
SW2474
SW2439
pGC2.26
S0351
SSC6
SSC4
SSC14
SSC4
SSC9
SSC10
SSC9
SSC10
SSC14
SSC10
HSA2
SW349
SW1327
S0397
S0002
UMNp583
SW314
SW730
S0216 S0167
SW1693
S0358
SW142
SW248
SWR1521
SW236
UMNp51 SW2141
UMNp647 SW271
SW102 UMNp58
SW828 UMNp355
SW1359
SW1978
S0164
SW1066
SW902
SW1436
SW2597 SWR978
S0148
UMNp263
S0004 SW1562
SW832
SW1204
S0355
SW2072
SW1935
SW2131 SW1530
UMNp131 UMNp252
UMNp72
SW15
SW423
S0088
SW1065
SW1263
UMNp35
SW1945
SW936 UMNp541 SW315
UMNp548
SW1983
SW1510
UMNp289
SW1339
SWR2121
SSC3
SSC15
HSA3
SW960
SW2448
S0282
SWR428
SW1407
UMNp159
SW344
SW2458
SW1400
UMNp452
SW698
UMNp502
UMNp356
SW955
SW225
SW482
SW2007
UMNp419
SW2430
SWC22
SW2412 UMNp56
SW937
UMNp210
UMNp494
UMNp275 UMNp94
S0068
UMNp358
SSC12
HSA4
PDE6B SY13
SB73
SY23
SW2611
SY21
SW905
SW1117 SW1345
SW268
SGU0001 UMNp329
UMNp175 UMNp336
SWR1101
SW933
UMNp4001 SW1649 UMNp4002
SW1843 UMNp4005
SW206 UMNp4006 SW7
UMNp4008
SW1953 UMNp75
SW916
S0178
SW1980
UMNp404
SW61
SW194
SW1312
S0225
SW171
SW748
SW374
SW1924
S0442
UMNp558
SW368
SW527
S0376
SWR283
SW1891
SWR1004
SSC8
SSC15
SSC17
HSA5
S0105
UMNp256 UMNp47
SW813
SW742
SW2411
UMNp344
UMNp174 SW1645
UMNp407
UMNp274
UMNp630
S0363 SW1638 SW1809
SW208
SW382
UMNp578
UMNp540
SW1602 UMNp610
UMNp453
SW766
SW2134
S0226
SW1320
S0010
S0370
UMNp566 UMNp487
UMNp383
SW2192
S0404
UMNp569
SW262
UMNp601
SW977 S0390
SW81
SW1090
SSC16
SSC2
SSC16
HSA6
SW1437
SW1873
SWR1343
SW1344
SW1369
LRA1
TNFB
UMNp612
WAF
MS AJ277795
SW2019 UMNp340
EST39
SW1653
S0396
SW2185 SW1430
SO317 SO318
SW1123
SW1997
UMNp192
SW952
MS-triadin
SW1851
S0008
UMNp467
sZ002 UMNp160
SW1332
SW137 SW64
UMNp380
SWR485
SW2184
SSC7
SSC1
SSC7
SSC1
HSA7
UMNp233
S0120 S0062 SW1682
SWR414
SW833 SW2429
UMNp292
UMNp262
SW866
SW1008
SW944
SW2074
SW787
SW1023
SSC9
SSC18
SSC3
SSC9
SSC3
SSC9
SSC18
HSA8
SW2496
SW1837 SW1631 S0356
SW1918
SW2038
SW1673
SW1401
UMNp495
S0369
MS-V-ATPase
S0217
UMNp363
SW1475 SW1040 SW724
SW1336 SW35 UMNp133
SW45 SY20
SWR981
UMNp143
UMNp464
UMnp165 UMNp60
SW2128
SW2454
SW752
SW871
SW1492
SW188
SW1678
S0001
SW2409
SW2049
SW835
S0301
UMNp606
SW480
SSC14
SSC15
SSC4

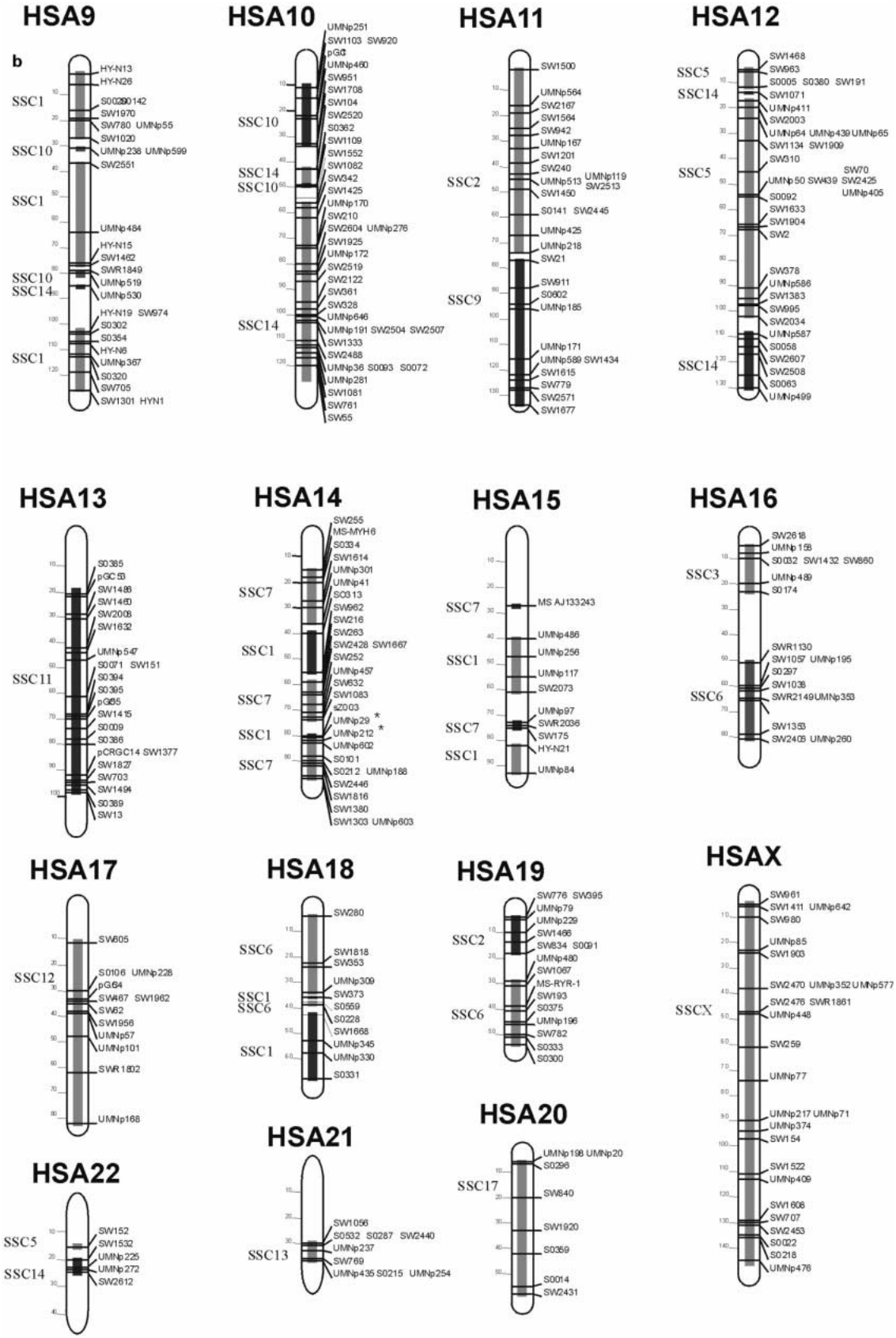


Fig. 2. Conserved syntenic breakpoints determined on human chromosomes. The existence of one segment is provisional (indicated by *). This same figure in color is available on the web at http://w3.toulouse.inra.fr/lgc/pig/msat/.

chromosomal regions in a segment of conserved synteny without any point of anchoring (50 Mb for example for HSA1). Nevertheless we did not detect important disagreements with bi-directional painting descriptions (Goureau et al., 1996). Our results were concordant with all published results for HSA3, HSA13, HSA17, HSA20, HSA21 and HSAX. On HSA2, HSA5, and HSA11 we describe a simpler situation than that reported by Rink et al. (2002): no overlaps were detected. On HSA16 and HSA19, we did not find overlaps but the covering of these chromosomes is not complete here (Fig. 2). For HSA4, HSA6, HSA10, HSA14, HSA15 and HSA18 our results allowed slightly different definitions of chromosomal fragments than results published by Rink et al. (2002). After a new analysis of results (positions on human sequence) reported by Nonneman and Rohrer (2003), MAPK8 confirmed the existence of the short sub-segment in the correspondence HSA10/SSC10 characterized here by only one result (near 48 Mb on HSA10). Comparative maps of HSA1, and HSA7 appeared more complex and it was very difficult to describe disagreement between our description and the one of Rink et al. (2002) or the one of Lahbib-Mansais et al. (2003). Nevertheless definition of conserved syntenic chromosomal fragments HSA1/SSC4 and HSA8/SSC4 were compatible with those reported by Fujishima-Kanaya et al. (2003). On HSA22, our results showed a possible segmentation of the correspondence with SSC5 in two sub-segments (upstream and downstream of the conserved syntenic fragment with SSC14). The first (SW152, already described here previously) would be new and we did not detect the second characterized by Lahbib-Mansais et al. (2003) and Rink et al. (2002) because the covering of HSA22 is not complete here. On HSA12 (near 12 Mb) and on HSA9 (near 80 Mb), the segmentation of correspondence respectively with SSC14 (only one marker) and SSC10 (two markers) detected here, have not been reported by Lahbib-Mansais et al. (2003) or by Rink et al. (2002). Lastly our results allowed a description of the correspondence between HSA8 and SSC15 (four markers) not detected by Lahbib-Mansais et al. (2003) or by Rink et al. (2002).

In summary these results increase the number of links between porcine maps and the human physical map. Microsatellites are usually considered as anonymous markers and in this study we demonstrate their possible integration in the comparative map in 29% of the cases. Flanking sequences of microsatellites are most of the time very short and comparison of BAC end sequences is more effective (65%). The score observed with the second strategy demonstrates the interest and the feasibility of the sequencing of the porcine genome by a shotgun method using the human sequence as support.

Acknowledgements

The authors express their appreciation to Dr Patrick Chardon and collaborators for providing the BAC library.

References

Archibald AL, Haley CS, Brown JF, Couperwhite S, McQueen HA, Nicholson D, Coppieters W, Van de Weghe A, Stratil A, Wintero AK, et al: The PiGMaP consortium linkage map of the pig *(Sus scrofa)*. Mammal Genome 6:157–175 (1995).

Farber CR, Medrano JF: Putative in silico mapping of DNA sequences to livestock genome maps using SSLP flanking sequences. Anim Genet 34:11–18 (2003).

Fujishima-Kanaya N, Toki D, Suzuki K, Sawazaki T, Hiraiwa H, Iida M, Hayashi T, Uenishi H, Wada Y, Ito Y, Awata T: Development of 50 gene-associated microsatellite markers using BAC clones and the construction of a linkage map of swine chromosome 4. Anim Genet 34:135–141 (2003).

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

Hu J, Mungall C, Law A, Papworth R, Nelson JP, Brown A, Simpson I, Leckie S, Burt DW, Hillyard AL, Archibald AL: The ARKdb: genome databases for farmed and other animals. Nucl Acids Res 29:106–110 (2001).

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

Korwin-Kossakowska A, Reed KM, Pelak C, Krause E, Morrison L, Alexander LJ: Radiation hybrid mapping of 118 new porcine microsatellites. Anim Genet 33:224–227 (2002).

Krause E, Morrison L, Reed KM, Alexander LJ: Radiation hybrid mapping of 273 previously unreported porcine microsatellites. Anim Genet 33:477–485 (2002).

Lahbib-Mansais Y, Dalias G, Milan D, Yerle M, Robic A, Gyapay G, Gellin J: A successful strategy for comparative mapping with human ESTs: 65 new regional assignments in the pig. Mammal Genome 10:145–153 (1999).

Lahbib-Mansais Y, Leroux S, Milan D, Yerle M, Robic A, Jiang Z, Andre C, Gellin J: Comparative mapping between humans and pigs: localization of 58 anchorage markers (TOASTs) by use of porcine somatic cell and radiation hybrid panels. Mammal Genome 11:1098–1106 (2000).

Lahbib-Mansais Y, Tosser-Klopp G, Leroux S, Cabau C, Karsenty E, Milan D, Barillot E, Yerle M, Hatey F, Gellin J: Contribution to high-resolution mapping with 101 type I markers and progress in comparative map between humans and pigs. Mammal Genome 14:275–288 (2003).

Marklund L, Johansson Moller M, Hoyheim B, Davies W, Fredholm M, Juneja RK, Mariani P, Coppieters W, Ellegren H, Andersson L: A comprehensive linkage map of the pig based on a wild pig-Large White intercross. Anim Genet 27:255–269 (1996).

Milan D, Hawken R, Cabau C, Leroux S, Genet C, Lahbib Y, Tosser G, Robic A, Hatey F, Alexander L, Beattie C, Schook L, Yerle M, Gellin J: IMpRH Server: an RH mapping server available on the Web. Bioinformatics 16:558–559 (2000).

Nonneman DJ, Rohrer GA: Comparative mapping of a region on chromosome 10 containing QTL for reproduction in swine. Anim Genet 34:42–46 (2003).

Pearson WR, Lipman DJ: Improved tools for biological sequence comparison. Proc natl Acad Sci, USA 85:2444–2448 (1988).

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

Rink A, Santschi EM, Eyer KM, Roelofs B, Hess M, Godfrey M, Karajusuf EK, Yerle M, Milan D, Beattie CW: A first-generation EST RH comparative map of the porcine and human genome. Mammal Genome 13:578–587 (2002).

Rogel-Gaillard C, Bourgeaux N, Billault A, Vaiman M, Chardon P: Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements. Cytogenet Cell Genet 85:205–211 (1999).

Rohrer GA, Alexander LJ, Keele JW, Smith TP, Beattie CW: A microsatellite linkage map of the porcine genome. Genetics 136:231–245 (1994).

Rohrer GA, Alexander LJ, Hu Z, Smith TP, Keele JW, Beattie CW: A comprehensive map of the porcine genome. Genome Res 6:371–391 (1996).

Yerle M, Pinton P, Robic A, Alfonso A, Palvadeau Y, Delcros C, Hawken R, Alexander L, Beattie C, Schook L, Milan D, Gellin J: Construction of a whole-genome radiation hybrid panel for high- resolution gene mapping in pigs. Cytogenet Cell Genet 82:182–188 (1998).

Yerle M, Pinton P, Delcros C, Arnal N, Milan D, Robic A: Generation and characterization of a 12,000-rad radiation hybrid panel for fine mapping in pig. Cytogenet Genome Res 97:219–228 (2002).