Animal Biotechnology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597228

Piggy-BACing the Human Genome I: Constructing a Porcine BAC Physical Map Through Comparative Genomics

Margarita B. Rogatcheva [a]; Kefei Chen [a]; Denis M. Larkin [a]; Stacey N. Meyers [a]; Brandy M. Marron [a]; Weisong He [a]; Lawrence B. Schook [ab]; Jonathan E. Beever [a]
[a] Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[b] Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Online Publication Date: 01 January 2008
To cite this Article: Rogatcheva, Margarita B., Chen, Kefei, Larkin, Denis M., Meyers, Stacey N., Marron, Brandy M., He, Weisong, Schook, Lawrence B. and Beever, Jonathan E. (2008) 'Piggy-BACing the Human Genome I: Constructing a Porcine BAC Physical Map Through Comparative Genomics', Animal Biotechnology, 19:1, 28 - 42
To link to this article: DOI: 10.1080/10495390701807634
URL: http://dx.doi.org/10.1080/10495390701807634

Animal Biotechnology, 19: 28–42, 2008

ISSN: 1049-5398 print/1532-2378 online
DOI: 10.1080/10495390701807634

Taylor & Francis
Taylor & Francis Group

PIGGY-BACING THE HUMAN GENOME I: CONSTRUCTING A PORCINE BAC PHYSICAL MAP THROUGH COMPARATIVE GENOMICS

Margarita B. Rogatcheva[1], Kefei Chen[1], Denis M. Larkin[1], Stacey N. Meyers[1], Brandy M. Marron[1], Weisong He[1], Lawrence B. Schook[1,2], and Jonathan E. Beever[1]

[1]*Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA*

[2]*Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA*

Availability of the human genome sequence and high similarity between humans and pigs at the molecular level provides an opportunity to use a comparative mapping approach to piggy-BAC the human genome. In order to advance the pig genome sequencing initiative, sequence similarity between large-scale porcine BAC-end sequences (BESs) and human genome sequence was used to construct a comparatively-anchored porcine physical map that is a first step towards sequencing the pig genome. A total of 50,300 porcine BAC clones were end-sequenced, yielding 76,906 BESs after trimming with an average read length of 538 bp. To anchor the porcine BACs on the human genome, these BESs were subjected to BLAST analysis using the human draft sequence, revealing 31.5% significant hits ($E < e^{-5}$). Both genic and non-genic regions of homology contributed to the alignments between the human and porcine genomes. Porcine BESs with unique homology matches within the human genome provided a source of markers spaced approximately 70 to 300 kb along each human chromosome. In order to evaluate the utility of piggy-BACing human genome sequences, and confirm predictions of orthology, 193 evenly spaced BESs with similarity to HSA3 and HSA21 were selected and then utilized for developing a high-resolution (1.22 Mb) comparative radiation hybrid map of SSC13 that represents a fusion of HSA3 and HSA21. Resulting RH mapping of SSC13 covers 99% and 97% of HSA3 and HSA21, respectively. Seven evolutionary conserved blocks were identified including six on HSA3 and a single syntenic block corresponding to HSA21. The strategy of piggy-BACing the human genome described in this study demonstrates that through a directed, targeted comparative genomics approach construction of a high-resolution anchored physical map of the pig genome can be achieved. This map supports the selection

This work was partially supported by grants from the USDA-National Research Initiative (2001-35205-09965 and 2002-35205-12712), the USDA Cooperative State Research, Education, and Extension Service (AG2002-34480-11828), and the USDA Agricultural Research Service (Agreement No. 58-5438-2-313). We would like to thank the W. M. Keck Center for Comparative and Functional Genomics for their assistance in BAC-end sequencing and bioinformatics support.

Address correspondence to Lawrence B. Schook, Department of Animal Sciences, University of Illinois at Urbana-Champaign, 1201 West Gregory Drive, Urbana, IL 61801, USA. E-mail: schook@uiuc.edu

of BACs to construct a minimal tiling path for genome sequencing and targeted gap filling. Moreover, this approach is highly relevant to other genome sequencing projects.

Keywords: BAC-end sequence; Pig; Radiation hybrid map; SSCI3

INTRODUCTION

Completion of the human genome sequence provides a foundation for the understanding of genetic complexity and how genetic variation contributes to diverse phenotypes. It is clear that model organisms will continue to play a valuable role in the synthesis of this understanding. The pig (*Sus scrofa domestica*), a representative of the artiodactyl clade, has become an important agricultural animal as one of the major human nutritional sources of animal based protein. The pig is one of the first eutherian mammals domesticated that has co-evolved with humans (1) and is phylogenetically closer to primates than rodentia (2). Additionally, the pig has played a central role in the scientific and medical communities, providing scientific justification for sequencing of the porcine genome (3). Moreover, the recent ability to genetically modify the porcine genome, genetically manipulate embryonic fibroblasts, and "clone" genetically modified somatic cells through nuclear transfer demonstrates how the pig can provide relevant genetic models of appropriate phenotypes (4–6).

Comparative maps have indicated that the porcine and human genomes have higher conservation than either when compared with mouse genome. The mean length of conserved syntenic segments between human and pig is approximately twice as long as the average length of conserved segments between the human and mouse (7,8). Furthermore, the organizational similarities between the human and porcine genomes are reflected in similarities at the nucleotide level. Comparison of more than 600 non-coding DNAs aligned by orthologous exonic sequences on human chromosome 7 (HSA7) revealed that pig sequences consistently grouped closer to human and non-human primate sequences than did rodent sequences indicating that rodent genomes appear to be evolving at a faster rate than primate and artiodactyl genomes (9).

Over the past decade, tremendous progress has been made in mapping and characterizing the swine genome. Currently, moderate to high-resolution genetic linkage maps containing highly polymorphic loci have been produced using independent mapping populations (10,11). Somatic cell hybrid analysis, *in situ* hybridization and ZOO-FISH have been employed to enrich the gene map and whole-genome radiation hybrid (WG-RH) panels (7,000 and 12,000 rad) (12–15) have rapidly increased the number of comparative mapped loci. To date over 10,000 loci have been mapped on the $IMpRH_{7,000}$ and $IMNpRH_{12,000}$ panels with more than 5,000 anchoring points on the human genome providing a framework for sequencing the pig genome (16). Cross-species chromosome painting has also been employed to define the conserved blocks of human-pig-dog synteny (17). Access to resources such as bacterial artificial chromosome (BAC) libraries (18–23) have facilitated the production of high-resolution physical maps in specific chromosomal regions (18,24,25) and supported the construction of sequence-ready mapping resources for the porcine genome.

The development of these resources as well as establishment of detailed comparative genome information between humans and pigs, uniquely positions the porcine genome for sequencing. The present study provides the first results utilizing the human genome sequence to support the construction of a porcine physical map. We report here the ability to "piggy BAC" on the human sequence to construct a high-resolution (1.22 Mb) comparative map of SSC13 that represents a fusion of HSA3 and HSA21. Our approach links the sequence to the BAC clone map that supports subsequent targeted closure of genomic regions of particular interest. This strategy is justified through the outcomes associated with the human, mouse, and rat sequencing efforts that were done in parallel with the BAC map development.

METHODS

BAC DNA Isolation and End-Sequencing

BAC library RPCI-44 constructed from four crossbred male pigs (breed composition: 37.5% Yorkshire, 37.5% Landrace, and 25% Meishan) (21) and CHORI-242 constructed from a single Duroc female pig (23) were selected for BAC-end sequencing. BAC clones were cultured in 2 ml 96-well culturing blocks containing 1.5 ml of 2X Luria Broth and 12.5 μg/ml chloramphenicol for 24 h at 320 rpm. DNA extractions were performed using the Montage BAC_{96} Miniprep Kit (Millipore, Billerica, MA) according to the manufacturer's protocol, yielding 40 μl of a BAC DNA solution. BAC DNA was transferred to V-bottom plates and stored at 4°C.

Dye terminator sequencing reactions were conducted using 10 μl of BAC DNA solution and 10 μl master mix, containing 1.5 μl BigDye Terminator v 3.1 Cycle Sequencing Mix (Applied Biosystems, Foster, CA), 6.5 μl buffer (200 mM Tris-HCl, pH 9.0 and 5 mM $MgCl_2$), 2 μl standard T7 sequencing primer (20 μM solution) or custom-designed SP6 (GGC CGT CGA CAT TTA GGT GAC A) primer (15 μM). Sequencing reactions were performed using PTC-100 thermocyclers (MJ Research, Waltham, MA) with standard profile as follows: denature at 96°C for 4 min followed by 60 cycles of 96°C for 30 sec, 56°C for 10 sec, 60°C for 4 min. Reaction products were precipitated in 75 μl of (0.3 mM $MgSO_4$ in 70% ethanol), and washed in 70% ethanol. Samples were analyzed using ABI 3730 automated capillary sequencers (Applied Biosystems, Foster, CA).

Sequence Processing and Bioinformatics

DNA sequences were analyzed for quality assessment using the Genome Project Management System (GPMS), a local laboratory information management system for large-scale DNA sequencing projects (40). The BAC-end sequences were trimmed of vector sequences, and a total of 76,906 trimmed BESs have been deposited in a local Oracle database and also GenBank database under the accession numbers CL320164-CL390360, CL408399-CL414467, and CL439616-CL440255. All sequences were analyzed for quality by the Phred software (41,42) using $Q \geq 20$ as the cutoff. All ambiguous positions having $Q < 20$ were masked with "N". Sequences were trimmed for vector and mitochondrial DNA sequences and edited

read lengths with ≥ 200 bp were used in the final analysis. Repeats in DNA sequences were masked using Repeat Masker software (43) and BESs were analyzed for similarity with human genome using NCBI-BLASTn (44) running on an SGI/Gray 2000 16-processor supercomputer.

Comparative Mapping

Build 33 of the human genome pre-release sequence (April 2003) was used as the target database (45). An expectation value (E) of e^{-5} was set as the significance threshold for comparison of porcine BESs with the human sequence. Up to ten BLAST hits exceeding the threshold were stored for each BES. A Perl script was used to parse the BLAST output and the resulting data was placed into a spreadsheet to reveal the hit position within the human contigs and to assign a position within the assembled human genome sequence (build 33). Porcine BESs with unique hits were selected. Additionally, a second BLAST comparison using a lower significance threshold of e^{-2} was performed and the output, named as "e^{-2} database" was used for anchoring paired ends of those BACs that had only end uniquely anchored to human genome. For each selected porcine BAC Perl scripts analyzed all hits in the "e^{-2} database," including multiple hits and hits with E-value $\geq e^{-5}$ and assigned positions for paired end when one of the matches was found to be on the same human chromosome within 400 Kb from the position of a selected unique hit BES. Human genome coverage by porcine BACs was calculated by summing: 1) the estimated length of porcine BACs with mate-paired BES hits; and 2) the number of BESs having only one high confidence hit ($E < e^{-5}$) in the human genome more then 180 Kb apart and was multiplied by 180 Kb (average length of human sequence covered by porcine BAC).

Radiation Hybrid Mapping

A total of 193 comparatively anchored BESs were selected to provide an evenly spaced distribution along the length of HSA3 (166) and HSA21 (27). Primers were designed using available tools including Primer 3 (46), and Vector NTI v.7.0 software (InforMax) to have an optimal length of 20–22 bp, a GC content of 45–60%, a melting temperature of 60°C, and amplicons of 100–600 bp. RH mapping was performed using the IMpRH panel containing 90 hybrids and six (four positive and two negative) controls (15). Multipoint maximum likelihood RH maps were constructed using carthagene software v0.99 (47,48).

RESULTS AND DISCUSSION

BAC-End Sequencing Statistics

As a first step towards sequencing the pig genome, BAC-end sequencing of 50,300 large genomic inserts from the RPCI-44 (21) and CHORI-242 (23) porcine BAC libraries yielded 80,077 (80% success rate) high quality sequences (>200 bp in length and Phred score >20) with average read length of 538 bp (Table 1). Filtering of vector and mitochondrial DNA sequences reduced this number by 4%, to

Table 1 Pig BAC-end sequencing statistics

BAC-end sequencing reads	100,600
BAC-end sequences (BESs)[a] (success rate)	80,077 (80%)
BESs after trimming[b]	76,906 (96%)
Average read length, bp	538
BESs with mate-pairs[c]	66,620 (87%)
Total bases sequenced	40.3 Mb
Total unmasked sequence	27.9 Mb

[a]Number of sequences with edited read length $\geq$200 bp and Phred score $>$20.
[b]Number of BESs free of vector contamination with edited read length $\geq$200 bp.
[c]Number of BESs with paired ends successfully sequenced.

76,906 BESs. Among these, 86.8% of the BESs were mate-pairs. Assuming an average insert length of 165 kb for each BAC insert (21) and an estimated porcine genome size of approximately 2.7 gigabases (Gb), this corresponds to approximately 2.0X coverage of the porcine genome using only BAC inserts with paired BESs (n = 33,310). The cumulative length of sequences generated was 40.3 Mb or 0.013 genome equivalents.

Although the majority of the BESs were generated from Segment 1 of the RPCI-44 library (40,621 clones), the inclusion of 2,862 clones from Segment 2 of the CHORI-242 library permitted a comparison between libraries with respect to relative repetitive element content (Table 2). Repetitive sequences were present in 63% and 72% of the BESs derived from the RPCI-44 and CHORI-242 libraries, respectively. Further comparison between the libraries showed that 32.2% of the cumulative length of the BESs from the RPCI-44 library and 34.5% of those from the CHORI-242 library represented repeats. The most frequently identified repeats between the RPCI-44 BESs were LINEs (16.5%), SINEs (9.9%), or long terminal repeats (2.6%). The CHORI-242 BESs were slightly richer in SINEs (11.5%) and satellite DNA (1.4% versus 0.4%). Interestingly, differences in both repeat number and composition between porcine RPCI-44 and CHORI-242 libraries were similar to those observed for the mouse RPCI-23 and RPCI-24 libraries (26) that were constructed using the same restriction enzymes, *Eco*RI and *Mbo*I, respectively. This observation is best explained by differences in the distribution of restriction sites for *Eco*RI and *Mbo*I within the porcine genome. However, assuming this pool

Table 2 Repeat composition of RPCI-44 and CHORI-242 libraries

Characteristic	RPCI-44	CHORI-242
Total BESs	71,701	5,052
Total length	37.7 Mb	2.7 Mb
Bases masked	12.1 Mb (32%)	0.9 Mb (34.5%)
Interspersed repeats	30%	31.4%
SINEs	9.9%	11.5%
LINEs	16.5%	16.3%
LINE1	14.9%	14.3%
LTR elements	2.6%	2.2%
Satellite DNA	0.4%	1.4%

of BESs is representative of the porcine genome, it can be concluded that known repetitive elements occupy less total sequence within the porcine genome than reported for human (9) and cow (27) yet are comparable to the number of repetitive elements found within the dog and mouse genomes (9). These findings suggest that the pig genome may be physically smaller than the human genome due to less expansion of repetitive sequences.

Anchoring of Porcine BESs to the Human Genome Sequence and Comparative Mapping

After masking repetitive elements, 27.9 Mb porcine sequence was available for BLAST similarity searches against the human genome, thus providing a significant resource for anchoring the porcine and human genomes (Table 3). A total of 67,402 BESs (>100 bp of continuous unmasked sequence) were used for comparative analyses. Approximately one third (31.5%) of the BESs (21,195 BESs) had at least one match with a human genome sequence (E-value $< e^{-5}$). In comparison, a similar analysis of the cattle genome, another artiodactyla revealed a lower percentage (29.4%) of BESs with hits to the human genome (27). Limiting the analysis to only BESs with unique matches (19,550 unique hits, see Supplementary Table 1S; http://www.swinegenomics.com/publication_detail.php?id=111) we found the frequency of BESs with unique BLAST hits in the human genome (29%) among repeat masked porcine BESs was also higher compared to the reported frequency for cattle (23.0%) (27). The higher frequency of unique hits in human genome for porcine BESs could be due to either lower repeats number in the porcine genome, and/or higher overall human-porcine genome similarity.

These unique matches allowed comparative anchoring of 16,066 of the 43,483 sequenced BAC inserts with 2,408 anchored on both ends (i.e., BES mate-pairs, both with unique matches) (Supplementary Table S2a; and S2b; http://www.swinegenomics.com/publication_detail.php?id=111). To anchor additional BAC inserts at both ends, a database containing 57,765 similarities generated at an E-value $< e^{-2}$ was searched for mate-pairs of unique matches identified at the higher threshold.

Table 3 BLASTn comparison of porcine BAC-end sequences with human genome sequence

BESs used for comparative analysis[a]	67,402
BESs with significant BLAST hits[b]	21,195 (31.5%)
Significant BLAST hits total[c]	28,957
BESs with unique match in human genome (% among BESs for analysis)	19,550 (29%)
Hits in genes (% among unique BLAST hits)[d]	12,537 (64%)
BACs with at least one end sequenced	43,483
BACs comparatively anchored on draft human genome	16,066 (37%)
BACs with paired-end BLAST hits (% among comparatively anchored)[e]	3,726 (11%)

[a]Number of BESs with $\geq$100 bp contiguous non-repetitive porcine sequences.
[b]Number of BES with significant (E $< e^{-5}$) BLAST match within human draft sequence (build 33).
[c]Total number of BLAST matches (E $< e^{-5}$) for 21,189 BESs.
[d]Number of BLAST hits annotated as genes in the human genome draft sequence.
[e]Number of BAC with both ends anchored on human genome graft sequence.

If the mate-pair for each BES with a unique similarity at E-value $\geq e^{-5}$ had a match to the same human chromosome within 400 Kb, at the lower significance threshold, the insert was considered anchored at both ends. This approach yielded an additional 1,318 comparatively anchored BESs (highlighted in blue in Supplementary Tables S2a and S2b; http://www.swinegenomics.com/publication_detail.php?id=111). The majority of these anchored BESs would have been missed using the conventional threshold ($E < e^{-5}$), since 1,087 of them had hits with E-value $\geq e^{-5}$.

Thus, among the 33,310 BAC clones with BES mate-pairs, 3,726 inserts had both ends anchored to human sequence (Table 3). The median distance between anchored paired-ends for the RPCI-44 library (3,483 clones) was 185.9 Kb, approximately 12.7% longer than the average insert size of 165 Kb previously estimated for this library (21). Similarly, the CHORI-242 library (243 clones) had an estimated median distance of 202 Kb between anchored mate-pairs. The predicted average size is longer than the average insert size of 173 Kb previously estimated for the CHORI-242 library (23). Such differences between estimates most likely represent non-linear correspondence between the porcine and human genome sequences due to expansion or contraction during genome evolution. Assuming the human genome size of 3 Gb, these data suggest the porcine genome size may be proportionally smaller in the range of 2.62 to 2.75 Gb.

As estimated by BESs with overlapping hits, only 5.8% of all BESs were redundant, demonstrating that there was no selection bias among the sequenced BAC clones. Unique BES hits were distributed evenly along human chromosomes with an average density of one anchor per 73 Kb of completed human genome sequence (Table 4). Overall the distribution of similarities varied between human chromosomes. The highest density of unique matches for human autosomes was observed on HSA3 and the lowest density on HSA19. The frequency of hits within known genes varied from 54.5% on HSA6 to 77.5% on HSA22. Estimated genome coverage was highest for HSA14 (69.4%) and lowest for HSA19 (21.6%). The total number of hits and hit density for human X chromosome were lower than would be expected relative to its size. The calculated average ratio of single hits for each chromosome was approximately 65%. In contrast, only 50% of all hits with X chromosome were unique and a lower ratio of 33.9% was observed for HSA19. Of the 54 BESs assigned to the Y-chromosome ($E < e^{-5}$), only two had unique BLAST hits, the remainder having additional hits to autosomes and the pseudoautosomal region of the X chromosome.

In general, the densities of hits were in agreement with the estimated gene densities since 64.2% of the assigned BESs were observed in orthologous sequences of known human genes. Similarly, telomeric, centromeric and repeat-rich regions of human chromosomes had relatively few unique matches with porcine BESs. One controversial comparison is represented by HSA19, a chromosome characterized as gene rich (1,761 genes per 56 Mb) (28) but which demonstrated the lowest comparative single hit density (1 per 139.33 kb) and correspondingly lowest coverage by mapped porcine BACs (21.6%). The same low coverage was reported for mouse (29) and cattle (27) BESs. Human chromosome 19 is known to contain an extraordinary number of clustered gene families (28), members of which encode the Kruppel-type zinc finger-containing proteins, olfactory receptors, serine proteases,

Table 4 Distribution of BES BLAST hits in the human genome, and genome coverage

HSA	Chr. size (Mb)	[1]No. contigs in HSA	[2]Effective size (Mb) and %	[3]No. unique hits	Hit length (bp)	Distance btw hits median (Kb)	Coverage by BACs (Mb) and %	Hits in genes and %
1	246	94	218.71 (88.9)	1,593	182	76.77	123.76 (56.59)	1088 (68.32)
2	243	22	237.05 (97.55)	1,833	210	67.92	141.52 (59.70)	1165 (63.60)
3	199	15	193.61 (97.29)	1,756	202	64.82	131.24 (67.79)	1134 (64.57)
4	192	18	186.58 (97.18)	1,352	195	73.33	109.52 (58.70)	782 (57.83)
5	181	12	177.53 (98.08)	1,424	195	67.14	114.64 (64.58)	893 (62.71)
6	170	11	166.88 (98.16)	1,168	200	76.54	93.09 (55.78)	636 (54.45)
7	158	13	154.55 (97.82)	1,099	196	66.21	86.72 (56.11)	742 (67.51)
8	146	17	141.69 (97.05)	903	191	80.90	70.85 (50.00)	591 (65.45)
9	135	44	115.89 (85.84)	827	182	71.42	61.41 (52.99)	509 (61.54)
10	136	26	130.71 (96.11)	866	181	73.74	66.96 (51.23)	631 (72.86)
11	135	10	131.80 (97.63)	881	173	77.44	71.37 (54.15)	570 (64.70)
12	134	11	129.33 (96.51)	1048	188	63.64	81.32 (62.88)	721 (68.79)
13	114	6	95.51 (83.78)	624	186	80.29	50.09 (52.44)	344 (55.13)
14	106	1	87.19 (82.25)	766	205	66.71	60.49 (69.38)	499 (65.14)
15	100	11	81.12 (81.12)	610	219	65.37	49.33 (60.81)	412 (67.54)
16	90	13	79.90 (88.78)	424	172	92.23	33.16 (41.50)	286 (67.45)
17	82	13	77.48 (94.49)	440	178	91.97	37.49 (48.39)	313 (71.13)
18	78	6	74.54 (95.56)	568	181	71.99	47.82 (64.15)	347 (61.09)
19	64	4	55.78 (87.16)	153	142	139.33	12.02 (21.55)	116 (75.82)
20	64	7	59.43 (92.86)	330	165	87.06	26.82 (45.13)	237 (71.82)
21	47	5	33.92 (72.17)	183	196	105.43	10.91 (32.16)	109 (59.29)
22	50	11	34.35 (68.70)	152	163	106.56	12.8 (37.26)	117 (77.48)
X	153	29	147.69 (96.53)	537	213	130.57	47.46 (32.13)	295 (54.94)
Total genome	3,023	399	2811.24 (92.98)	19,537	192	72.64	1540.8 (54.30)	12537 (64.17)
Y				2				
Unassembled contigs				25				

[1]Number of contigs associated with chromosomal assembly using human draft sequence (build 33).
[2]Cumulative length of human genomic DNA available for comparative analysis and a percentage of total chromosome length.
[3]Number of porcine BESs found to have a unique match within human genome on a specific human chromosome.

cytochrome P_{450} and many others. Since most of the HSA19 genes are clustered in gene families, and our analysis only considered unique hits, this would suggest that most of genes families members were excluded from our analysis due to multiple hits and thus, possibly resulting in fewer hits calculated than should be expected.

Utility of Piggy-BACing Human Genome Sequences

Using only porcine BESs with putative orthology to 1.4% of the human genome sequence, we were able to anchor 6,336 orthologous human genes and 7,032 BESs matched conserved intergenic regions. Selected BESs were then utilized for constructing a high-resolution porcine RH map. To maximize the efficiency of RH mapping, only unique BESs with a single BLAST hit in the human genome were used as potential markers for RH mapping. The evolution of gene families in different species involves either species-specific expansion or deletion of their members (30). Hence porcine BESs with multiple homology matches within the human genome could either represent gene segments with similar functional domains or members of various gene families present in both pigs and humans or expanded in human gene families. Thus, we avoided using such markers for RH mapping to avoid multiple PCR products and inconclusive mapping. Considering the optimal resolution of the $IMpRH_{7,000}$ rad panel to be about 1 marker per 150 Kb, markers were chosen approximately 1 Mb apart in the human genome sequence.

To evaluate the utility of this approach, we targeted SSC13 (Figure 1). This porcine chromosome was selected since it represents the remnant of a putative mammalian ancestral chromosome and is the largest evolutionarily conserved block known in mammals consisting entirely of two human syntenic groups, HSA3 and HSA21 (31). A total of 193 comparatively anchored BESs were selected to provide an evenly spaced distribution along the length of HSA3 (166) and HSA21 (27). Of these, 186 (96%) BESs were successfully mapped, corresponding to a comparative anchor every 1.22 Mb relative to the human sequence. Mapped markers were orthologous to regions spanning from 0.14 Mb to 197.53 Mb on HSA3 and 14.53 Mb to 46.76 Mb on HSA21 providing approximately 99% and 97% coverage of each human chromosome, respectively. The estimated RH_{7000} map length was 4,905 cR, equivalent to a ratio of 21.9 cR per Mb when excluding the centromeric region on HSA3 (15).

An absence of single hits around the HSA3 centromere restricted our choice of markers to two closest from the p- and q-arms, which were still 5.5 Mb apart (89.8–95.3 Mb). The SSC13 RH map demonstrates these two markers closely linked together at a distance of 36 cR, that corresponds to the distance between two adjacent markers being 1.3 Mb apart on the human map and suggests an excess of genomic DNA around the HSA3 centromere. A similar result was obtained for the most proximal marker of the HSA21 linkage group. It was closely linked with the marker from HSA3 linkage group, suggesting that both the p-arm and centromeric regions of HSA21 were gained later in evolution by expansion or transposition of genomic material (31). Further studies of porcine centromere regions should shed light on chromosome evolution, composition of human centromere and provide a tool for annotation of human genome.

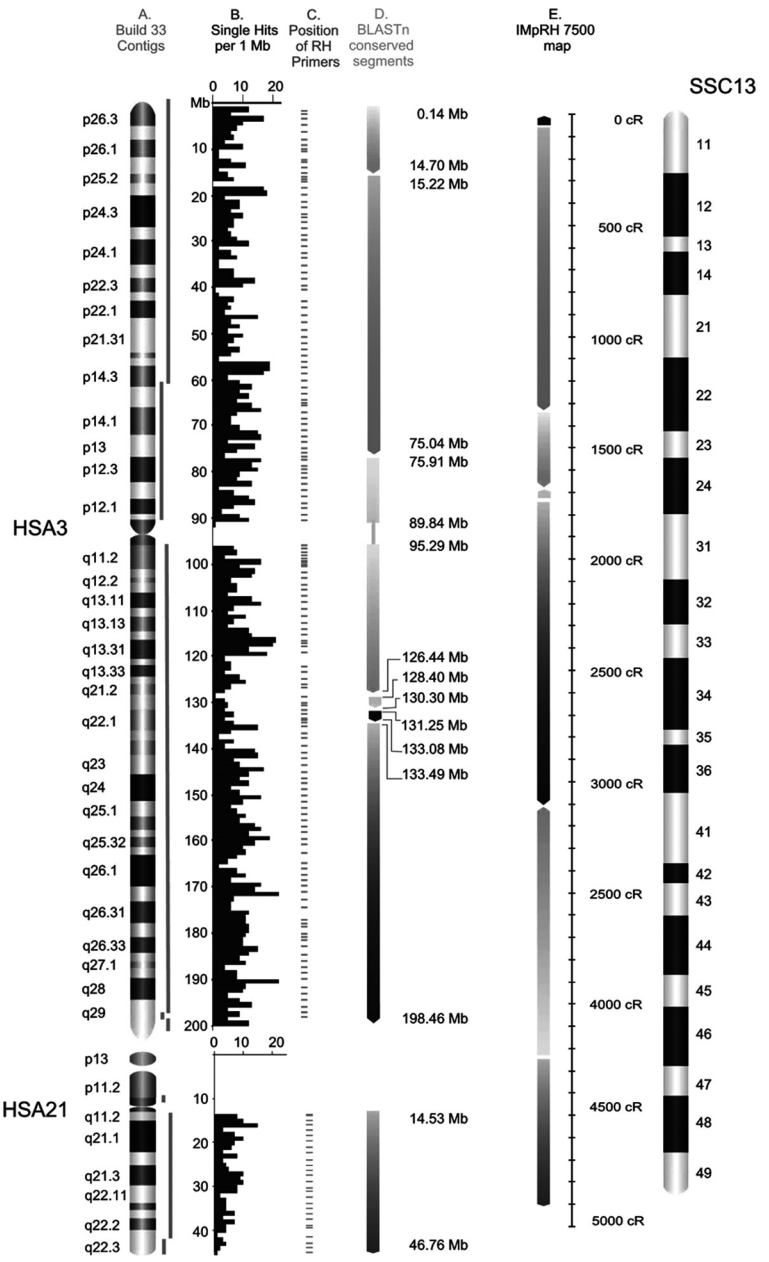

Figure 1 Assignment of porcine BAC-end sequences on human homologous of SSC13. **[A]**. The build 33 contigs for HSA3 and HSA21 that were used for BLAST analysis. **[B]**. The distribution of porcine BESs on the human chromosomes and the number of single (unique) BLAST hits/Mb of human sequence. **[C]**. Distribution of porcine BESs selected for RH mapping. **[E]**. The RH mapping of the selected markers **[C]** resulted in the identification of 7 fragments of synteny between pig and human. The position of syntenic blocks is shown on porcine RH map **[E]** and on the human build 33 contigs **[D]**. The beginning and end points of segments in Mb are from results of BLASTn comparison with build 33 contigs. Arrowheads provide orientation with respect to the human gene order.

Conserved Syntenies between SSC13 and HSA3/HSA21

Seven evolutionarily conserved blocks were identified with six corresponding to segments from HSA3 and a single syntenic block corresponding to HSA21. The mapping allows confirmation of syntenic segments previously detected by FISH mapping (17,32,33) and closure of the gaps by redefining the positions of the evolutionary breakpoints to less then 1 Mb between adjacent segments while detecting two novel small fragments of syntenic homology. Though below the resolution of the current $IMpRH_{7,000}$ rad panel, shorter rearrangements could not be detected, however, gene order was conserved within the homologous segments (15,34). The number of shared syntenic groups between humans and other mammals has recently been calculated. A total of 217 conserved syntenic blocks were observed between human and mouse (35). A lesser number (159) was estimated when comparing the dog-human genomes (36). The human genome is highly conserved relative to the genomes of all placental mammals, whereas the mouse and dog genomes are highly rearranged through large proportion of translocations (37). The HSA3/HSA21 synteny group was shown to have 28 fragments of synteny with mouse genome (29) and 13 with dog genome (36). Only 7 syntenic groups with porcine genome were detected, suggesting fewer and longer orthologous segments between human and pig genomes, making the pig genome more promising for resolving ambiguities in sequencing and annotations of the human genome.

CONCLUSIONS

The strategy of piggy-BACing the human genome described in this study demonstrates that through a directed, targeted comparative genomics approach construction of a high-resolution anchored physical map of the pig genome can be achieved (15). This map supports the selection of BACs to construct a minimal tiling path for genome sequencing and targeted gap filling. Moreover, this approach is highly relevant to other genome sequencing projects (38,39).

LIST OF ABBREVIATIONS

BAC: bacterial artificial chromosome; BES: BAC-end sequencing; cM: centimorgan; cR: centi-ray; FISH: fluorescence *in situ* hybridization; Gb: gigabase pairs; HSA: human (*homo sapiens*) chromosome; IMpRH: INRA-Minnesota porcine Radiation Hybrid panel; Kb: kilobase pairs; LINE: long interspersed nuclear element; LTR: long terminal repeat; Mb: megabase pairs; RH: radiation hybrid; SINE: short interspersed nuclear element; SSC: *Sus scrofa* chromosome; WG-RH: whole genome radiation hybrid.

REFERENCES

1. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science 2005; 307:1618–1621.

2. Jorgensen FG, Hobolth A, Hornshoj H, Bendixen C, Fredholm M, Schierup MH. Comparative analysis of protein coding sequences from human, mouse and the domesticated pig. BMC Biol 2005; 32: doi: 10.1186/1741-7007-3-2.
3. Tumbleson ME, Schook LB. Advances in Swine in Biomedical Research; Plenum Press, New York, 1996.
4. Lai L, Kolber-Simonds D, Park KW, Cheong HT, Greenstein JL, Im GS, Samuel M, Bonk A, Rieke A, Day BN, Murphy CN, Carter DB, Hawley RJ, Prather RS. Production of alpha-1,3-galactosyltransferase knockout pigs by nuclear transfer cloning. Science 2002; 295:1089–1092.
5. Rogatcheva MB, Rund LA, Beever JE, Schook LB. Harvesting the genomic promise: Recombineering sequences for phenotypes. Anim Biotechnol 2003; 14:103–118.
6. Schook L, Beattie C, Beever J, Donovan S, Jamison R, Zuckermann F, Niemi S, Rothschild M, Rutherford M, Smith D. Swine in biomedical research: Creating the building blocks of animal models. Anim Biotechnol 2005; 16:183–190.
7. Ellegren H, Chowdhary BP, Johansson M, Marklund L, Fredholm M, Gustavsson I, Andersson L. A primary linkage map of the porcine genome reveals a low rate of genetic recombination. Genetics 1994; 137:1089–1100.
8. Rettenberger G, Klett C, Zechner U, Kunz J, Vogel W, Hameister H. Visualization of the conservation of synteny between humans and pigs by heterologous chromosomal painting. Genomics 1995; 26:372–378.
9. Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, Maskeri B, Hansen NF, Schwartz MS, Weber RJ, Kent WJ, Karolchik D, Bruen TC, Bevan R, Cutler DJ, Schwartz S, Elnitski L, Idol JR, Prasad AB, Lee-Lin SQ, Maduro VV, Summers TJ, Portnoy ME, Dietrich NL, Akhter N, Ayele K, Benjamin B, Cariaga K, Brinkley CP, Brooks SY, Granite S, Guan X, Gupta J, Haghighi P, Ho SL, Huang MC, Karlins E, Laric PL, Legaspi R, Lim MJ, Maduro QL, Masiello CA, Mastrian SD, McCloskey JC, Pearson R, Stantripop S, Tiongson EE, Tran JT, Tsurgeon C, Vogt JL, Walker MA, Wetherby KD, Wiggins LS, Young AC, Zhang LH, Osoegawa K, Zhu B, Zhao B, Shu CL, De Jong PJ, Lawrence CE, Smit AF, Chakravarti A, Haussler D, Green P, Miller W, Green ED. Comparative analyses of multi-species sequences from targeted genomic regions. Nature 2003; 424:788–793.
10. Archibald AL, Haley CS, Brown JF, Couperwhite S, McQueen HA, Nicholson D, Coppieters W, Van de Weghe A, Stratil A, Wintero AK. The PiGMaP consortium linkage map of the pig (Sus scrofa). Mamm Genome 1995; 6:157–175.
11. Rohrer GA, Alexander LJ, Hu Z, Smith TP, Keele JW, Beattie CW. A comprehensive map of the porcine genome. Genome Res 1996; 6:371–391.
12. Chowdhary BP, Raudsepp T, Fronicke L, Scherthan H. Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. Genome Res 1998; 8:577–589.
13. Hawken RJ, Murtaugh J, Flickinger GH, Yerle M, Robic A, Milan D, Gellin J, Beattie CW, Schook LB, Alexander LJ. A first-generation porcine whole-genome radiation hybrid map. Mamm Genome 1999; 10:824–830.
14. 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 2002; 97:219–228.
15. Meyers SN, Rogatcheva MB, Larkin DM, Yerle M, Milan D, Hawken RJ, Schook LB, Beever JE. Piggy-BACing the human genome II. A high-resolution, physically anchored, comparative map of the porcine autosomes. Genomics 2005; 86:739–752.
16. Milan D, Beever J, Lahbib Y, Schook L, Beattie C, Yerle M. An integrated RH map of the porcine genome with more than 5000 anchoring points on the human genome provides

a framework for sequencing of the pig. Proceeding of the 30th International Conference on Animal Genetics 2006; P59:B520.
17. Biltueva LS, Yang F, Vorobieva NV, Graphodatsky AS: Comparative map between the domestic pig and dog. Mamm Genome 2004; 15:809–818.
18. 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 1999; 85:205–211.
19. Suzuki K, Asakawa S, Iida M, Shimanuki S, Fujishima N, Hiraiwa H, Murakami Y, Shimizu N, Yasue H. Construction and evaluation of a porcine bacterial artificial chromosome library. Anim Genet 2000; 31:8–12.
20. Anderson SI, Lopez-Corrales NL, Gorick B, Archibald AL. A large-fragment porcine genomic library resource in a BAC vector. Mamm Genome 2000; 11:811–814.
21. Fahrenkrug SC, Rohrer GA, Freking BA, Smith TP, Osoegawa K, Shu CL, Catanese JJ, de Jong PJ. A porcine BAC library with tenfold genome coverage: A resource for physical and genetic map integration. Mamm Genome 2001; 12:472–474.
22. Jeon JT, Park EW, Jeon HJ, Kim TH, Lee KT, Cheong IC. A large-insert porcine library with sevenfold genome coverage: A tool for positional cloning of candidate genes for major quantitative traits. Mol Cells 2003; 16:113–116.
23. BACPAC Resource Center. <http://bacpac.chori.org/porcine242.htm>; Date of Access, December 12, 2007.
24. Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, Rogel-Gaillard C, Paul S, Iannuccelli N, Rask L, Ronne H, Lundstrom K, Reinsch N, Gellin J, Kalm E, Roy PL, Chardon P, Andersson L. A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science 2000; 288:1248–1251.
25. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, Andersson L. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 2003; 425:832–836.
26. Zhao S, Shatsman S, Ayodeji B, Geer K, Tsegaye G, Krol M, Gebregeorgis E, Shvartsbeyn A, Russell D, Overton L, Jiang L, Dimitrov G, Tran K, Shetty J, Malek JA, Feldblyum T, Nierman WC, Fraser CM. Mouse BAC ends quality assessment and sequence analyses. Genome Res 2001; 11:1736–1745.
27. Larkin DM, Everts-van der Wind A, Rebeiz M, Schweitzer PA, Bachman S, Green C, Wright CL, Campos EJ, Benson LD, Edwards J, Liu L, Osoegawa K, Womack JE, de Jong PJ, Lewin HA. A cattle-human comparative map built with cattle BAC-ends and human genome sequence. Genome Res 2003; 13:1966–1972.
28. Kim J, Gordon L, Dehal P, Badri H, Christensen M, Groza M, Ha C, Hammond S, Vargas M, Wehri E, Wagner M, Olsen A, Stubbs L. Homology-driven assembly of a sequence-ready mouse BAC contig map spanning regions related to the 46-Mb gene-rich euchromatic segments of human chromosome 19. Genomics 2001; 74:129–141.
29. Gregory SG, Sekhon M, Schein J, Zhao S, Osoegawa K, Scott CE, Evans RS, Burridge PW, Cox TV, Fox CA, Hutton RD, Mullenger IR, Phillips KJ, Smith J, Stalker J, Threadgold GJ, Birney E, Wylie K, Chinwalla A, Wallis J, Hillier L, Carter J, Gaige T, Jaeger S, Kremitzki C, Layman D, Maas J, McGrane R, Mead K, Walker R, Jones S, Smith M, Asano J, Bosdet I, Chan S, Chittaranjan S, Chiu R, Fjell C, Fuhrmann D, Girn N, Gray C, Guin R, Hsiao L, Krzywinski M, Kutsche R, Lee SS, Mathewson C, McLeavy C, Messervier S, Ness S, Pandoh P, Prabhu AL, Saeedi P, Smailus D, Spence L, Stott J, Taylor S, Terpstra W, Tsai M, Vardy J, Wye N, Yang G, Shatsman S, Ayodeji B, Geer K, Tsegaye G, Shvartsbeyn A, Gebregeorgis E, Krol M, Russell D, Overton L, Malek JA, Holmes M, Heaney M, Shetty J, Feldblyum T, Nierman WC, Catanese JJ, Hubbard T,

Waterston RH, Rogers J, de Jong PJ, Fraser CM, Marra M, McPherson JD, Bentley DR. A physical map of the mouse genome. Nature 2002; 418:743–750.
30. O'Brien SJ, Murphy WJ. Genomics. A dog's breakfast? Science 2003; 301:1854–1855.
31. Muller S, Stanyon R, Finelli P, Archidiacono N, Wienberg J. Molecular cytogenetic dissection of human chromosomes 3 and 21 evolution. Proc Natl Acad Sci USA 2000; 97:206–211.
32. Van Poucke M, Yerle M, Tuggle C, Piumi F, Genet C, Van Zeveren A, Peelman LJ. Integration of porcine chromosome 13 maps. Cytogenet Cell Genet 2001; 93:297–303.
33. Van Poucke M, Yerle M, Chardon P, Jacobs K, Genet C, Mattheeuws M, Van Zeveren A, Peelman LJ. A refined comparative map between porcine chromosome 13 and human chromosome 3. Cytogenet Genome Res 2003; 102:133–138.
34. Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'brien SJ, Pevzner PA, Lewin HA. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 2005; 309:613–617.
35. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES. Mouse Genome Sequencing Consortium: Initial sequencing and comparative analysis of the mouse genome. Nature 2002; 420:520–562.
36. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC. The dog genome: Survey sequencing and comparative analysis. Science 2003; 301:1898–1903.

37. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. Molecular phylogenetics and the origins of placental mammals. Nature 2001; 409:614–618.
38. Porcine Genome Fingerprinting Project. <http://www.sanger.ac.uk/Projects/S_scrofa/WebFPC/porcine/>; Date of access, December 12, 2007.
39. Porcine Genome Sequencing Project. <http://www.sanger.ac.uk/Projects/S_scrofa/>; Date of access, December 12, 2007.
40. Liu L, Roinishvili L, Pan X, Liu Z, Kumar C. GPMS: A web based genome project management system. Proceeding of 4th World Multiconference on Systematics, Cybernectics, and Informatics 2000; SCI2000: 62–67.
41. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 1998; 8(3):86–194.
42. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 1998; 8(3):175–185.
43. RepeatMasker software. <ftp://ftp.genome.washington.edu> ; Date of access, December 12, 2007.
44. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 1997; 25(17):3389–3402.
45. Human Genome Sequence Build 33. <http://genome.ucsc.edu/cgi-bin/hgGateway>
46. Primer 3. <http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi>; Date of access, December 12, 2007.
47. Schiex T, Gaspin C. CARTHAGENE: Constructing and joining maximum likelihood genetic maps. Proc Int Conf Intell Syst Mol Biol 1997:5258–5267.
48. de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T. CARHTA GENE: Multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics 2005; 21(8): 1703–1704.