ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

SHORT COMMUNICATION doi:10.1111/j.1365-2052.2010.02140.x

A high-resolution comparative map of porcine chromosome 4 (SSC4)

J.-G. Ma*†, T.-C. Chang†, H. Yasue‡, A. D. Farmer§, J. A. Crow§, K. Eyer¶, H. Hiraiwa‡, T. Shimogiri, S. N. Meyers††, J. E. Beever††, L. B. Schook††, E. F. Retzel§, C. W. Beattie‡‡ and W.-S. Liu†**

*Department of Biological Science and Engineering, Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China. †Department of Dairy and Animal Science, College of Agricultural Sciences, Pennsylvania State University, 305 Henning Building, University Park, PA 16802, USA. ‡Genome Research Department, National Institute of Agrobiological Sciences, Ikenodai, Tsukuba, Ibaraki 305-0901, Japan. §National Center for Genome Resources, Santa Fe, NM 87505 USA. ¶Department of Biology, College of Sciences, University of Nevada, Reno, NV 89557, USA. **Faculty of Agriculture, Kagoshima University, Korimoto, Kagoshima 890-0065, Japan. ††Department of Animal Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ‡‡Department of Surgical Oncology, Room 618 820 CSB, University of Illinois COM, 840 South Wood St., Chicago, IL, USA

Summary

We used the IMNpRH2$_{12\,000\text{-rad}}$ RH and IMpRH$_{7\,000\text{-rad}}$ panels to integrate 2019 transcriptome (RNA-seq)-generated contigs with markers from the porcine genetic and radiation hybrid (RH) maps and bacterial artificial chromosome finger-printed contigs, into 1) parallel framework maps (LOD≥10) on both panels for swine chromosome (SSC) 4, and 2) a high-resolution comparative map of SSC4, thus and human chromosomes (HSA) 1 and 8. A total of 573 loci were anchored and ordered on SSC4 closing gaps identified in the porcine sequence assembly Sscrofa9. Alignment of the SSC4 RH with the genetic map identified five microsatellites incorrectly mapped around the centromeric region in the genetic map. Further alignment of the RH and comparative maps with the genome sequence identified four additional regions of discrepancy that are also suggestive of errors in assembly, three of which were resolved through conserved synteny with blocks on HSA1 and HSA8.

Keywords comparative map, radiation hybrid map, RNA-seq, sequence assembly, SSC4, swine.

High-resolution radiation hybrid (RH) maps provide a 'blueprint' for genome sequence assembly and comparative mapping across genomes (Lewin *et al.* 2009). A RH-based bacterial artificial chromosome (BAC) contig map serves as the template for sequencing and assembly of the swine genome, but does not provide a robust comparative synteny map that identifies breakpoints within and between minimally sequenced regions of the genome. Because closure of contig gaps would improve overall sequence assembly, we selected SSC4, one of the most deeply sequenced chromosomes in the draft sequence assembly, to test whether we could significantly and rapidly improve the pathway to final assembly by integrating RNA sequence (RNA-seq) data into a high-resolution comparative RH map of SSC4 and human HSA1 and HSA8.

Mapping vectors from the IMNpRH$_{12\,000\text{-rad}}$ panel and IMpRH$_{7\,000\text{-rad}}$ panel were first merged using Carthagene (Schiex & Gaspin 1997), and the merged data set (Table S1) was analysed using a maximum multipoint likelihood linkage strategy. Three linkage groups with 89, 63 and 421 markers, respectively (Table S2), were initially mapped to SSC4 at a 2pt LOD≥10, and an FW map for each linkage group was built simultaneously at a likelihood ratio of 1000:1 (Fig. S1c, d). A total of 328 FW markers were ordered on the SSC4 IMNpRH$_{12\,000\text{-rad}}$ FW map, over an accumulated map distance of 11 207.3 cR$_{12\,000}$ and 5507.8 cR$_{7\,000}$ (Tables S1 & S2). This yielded a Kb/cR$_{12\,000}$ ratio of ~13.0, a twofold increase in resolution over the IMpRH$_{7\,000\text{-rad}}$ FW map (Table S2) and slightly lower than the 2.2- to 3.0-fold increase reported for several other porcine chromosomes (Yerle *et al.* 2002; Liu *et al.* 2005, 2008; Ma *et al.* 2009).

A total of 245 non-FW markers (Table S2) were then added to the RH$_{12\,000\text{-rad}}$ FW map of SSC4 using CarthaGene (Schiex & Gaspin 1997), which improved the resolution of the SSC4

Address for correspondence

W.-S. Liu, Department of Dairy and Animal Science, College of Agricultural Sciences, The Pennsylvania State University, 305 Henning Building, University Park, PA 16802, USA.
E-mail: wul12@psu.edu

Accepted for publication 3 August 2010

map to ~254.8 Kb/marker (146 Mb/573), a >3-fold increase in marker density on SSC4 over the 0.8–1.0 Mb/marker interval on the current IMpRH$_{7\,000\text{-rad}}$ maps (Meyers *et al.* 2005; Rink *et al.* 2006; Humphray *et al.* 2007).

Seven FPC contigs (ctg) (4001–4005, 4007 and 4009), (http://www.sanger.ac.uk/Projects/S_scrofa/WebFPC/porcine/small.shtml) were identified, where 134 BACs/BESs were shared between the RH and FPC contig maps (Fig. S1c–e) with identical order of BAC end sequence (BES) in both the RH map and BACs, except for a small region of inconsistency that was observed for four BESs (362B20G08, 306B10F11, 274A10B02 and 310B20A04) bridging FPC ctg4004 and ctg4005 (Fig. S1c–e). The first two BESs, 362B20G08 and 306B10F11, mapped at the end of cgt4004, while BES 274A10B02-310B20A04 were at the beginning of ctg4005. In contrast, 274A10B02_306B10F11_362B20G08_310B20A04 were ordered within the same linkage group in the RH map (Fig. S1c–e), suggesting how contigs ctg4004 and ctg4005 became separated in the FPC maps.

Microsatellites (MSs) binned in the genetic map of SSC4 (http://www.marc.usda.gov/genome/swine/swine.htm) were all ordered on the RH$_{12\,000}$ map, providing a significant increase in map resolution. A discrepancy of the centromeric region (Fig. S1b–d) on the genetic map resulted from an inversion of five MSs (*SWR362*, *SW1998*, *SW1003*, *SW1513* and *SW1520*) (Fig. S1b, c). *SWR362* and *SW1998* mapped to SSC4p, while *SW1003*, *SW1513* and *SW1520* mapped to SSC4q in the RH map (Fig. S1c, d). We validated MS order by blasting the sequences of the five MSs against the NCBI pig high-throughput genomic sequence database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify BACs containing the MS sequence and to check the map positions of the BACs in the FPC maps. BAC CH242-417I24 contains SWR362, while three overlapping BACs, CH242-512E4, CH242-18C1 and Ch242-55N4, all contain SW1998, indicating that their order in FPC cgt4003 is 422B20A10_CH242-417I24_418A20F02_CH242-512E4_408A10F07 (http://www.sanger.ac.uk/Projects/S_scrofa/WebFPC/porcine/large.shtml), identical to their order in the RH map (Fig. S1c–e). Similar analyses identified that BACs CH242-321A10, CH242-387O5 and CH242-62L8 contain SW1520, SW1513 and SW1003, respectively. The order of these BACs in FPC ctg4004 is 285A10D09_CH242-62L8_286A10B03_CH242-321A10_422B20A12, again identical to their order in the RH map (Fig. S1c–e) (http://www.sanger.ac.uk/Projects/S_scrofa/WebFPC/porcine/large.shtml). The results suggest that the limited number of meioses (104 animals) in the two-generation backcross population used to generate the swine genetic map (Rohrer *et al.* 1994) and low rate of recombination in this chromosomal region in the reference population (Yu *et al.* 2001; DeWan *et al.* 2002) have confounded the linkage ordering in the centromeric region of SSC4 from *SW1520* to *SW362* (Fig. S1b) (http://www.marc.usda.gov/genome/swine/swine.htm).

The comparative map of SSC4 and HSA1 and HSA8 (Fig. 1) revealed that a large block of sequence (0–93 Mb) from the distal region of SSC4p to the proximal region of SSC4q1.3 (Fig. S1a) is highly conserved within the long arm (48.4–146 Mb) of HSA8. The remainder of SSC4q (q1.3–2.5, Fig. S1a) from 93 to 136 Mb is conserved with HSA1 segments 1p22.3–12 (86–120 Mb) and 1q21.1–1q24.3 (145–171 Mb) (Fig. 1). However, this conservation in synteny was not complete. Firstly, the centromeric region (120–145 Mb) of HSA1 is not conserved with SSC4q (Fig. 1), indicating that a centromere origination and/or a genome-rearrangement event occurred between the two species during evolution (Liu *et al.* 2005). Second, two genes (*PTK2* and *TADA1L*) in SSC4-HSA8 and two genes (*TBP* and *PLA2G4A*) in SSC4-HSA1 are not syntenic (Fig. 1), suggesting micro-rearrangements between the two species. The length ratio of the corresponding sequence blocks for SSC4:HSA8 is ~1:1 (93/97 Mb) and SSC4:HSA1 ~1:1.4 (43/60 Mb), which may explain the higher gene density (4X) observed in the SSC4q1.3-2.5 region (Figs 1 and S1).

Approximately 42 million Illumina RNA-seq reads (46-bp pair-end) generated from porcine macrophage and lymph node RNA were assembled *de novo* into contigs using a hybrid protocol incorporating ABySS-P (Simpson *et al.* 2009) and SSAKE (Warren *et al.* 2007). The resulting contigs were extended using PCAP (Huang *et al.* 2003). Reads were aligned to the pig genome sequence assembly (Sscrofa9) using GSNAP (Wu & Watanabe 2005). All high-throughput sequencing data were managed using the Alpheus pipeline and database resource (Miller *et al.* 2008). *De novo* assembled contigs, porcine sequences currently assembled in the NCBI unigene database, and previously assembled ESTs were aligned to the porcine genome using GMAP (Wu & Nacu 2010). Porcine transcripts derived from this combined resource were aligned to the human RefSeq protein data set and the strongest hit was taken as the value for a specific gene in that sample. All data from Alpheus alignments were loaded into a modified version of the Comparative Map and Trait Viewer (CMTV) (Sawkins *et al.* 2004). Approximately 22.5 million reads aligned to the NCBI Unigene; 15.7 million of the reads aligned uniquely; 17.5 million of the reads aligned to Sscrofa9; and 12.8 million reads aligned uniquely. Aligned vs. uniquely aligned reads in RNA-seq data provide a measure of the uniqueness of genes or of their components (e.g. a motif may be repeated in different genes). Hybrid assembly (methods) yielded 44 356 contigs longer than 100 bp. The contig N50 was 257, and the B2 000 was 257 785 bp, or 2.4% of the total assembly. Overall, we identified 17 328 *de novo* assembled contigs that aligned to the pig genome; 4863 contigs aligning only to the human genome, with 221 contigs aligning better to the human genome than porcine, suggesting that the sequence in those regions of the porcine genome is incomplete; and 15 737 contigs aligned to both

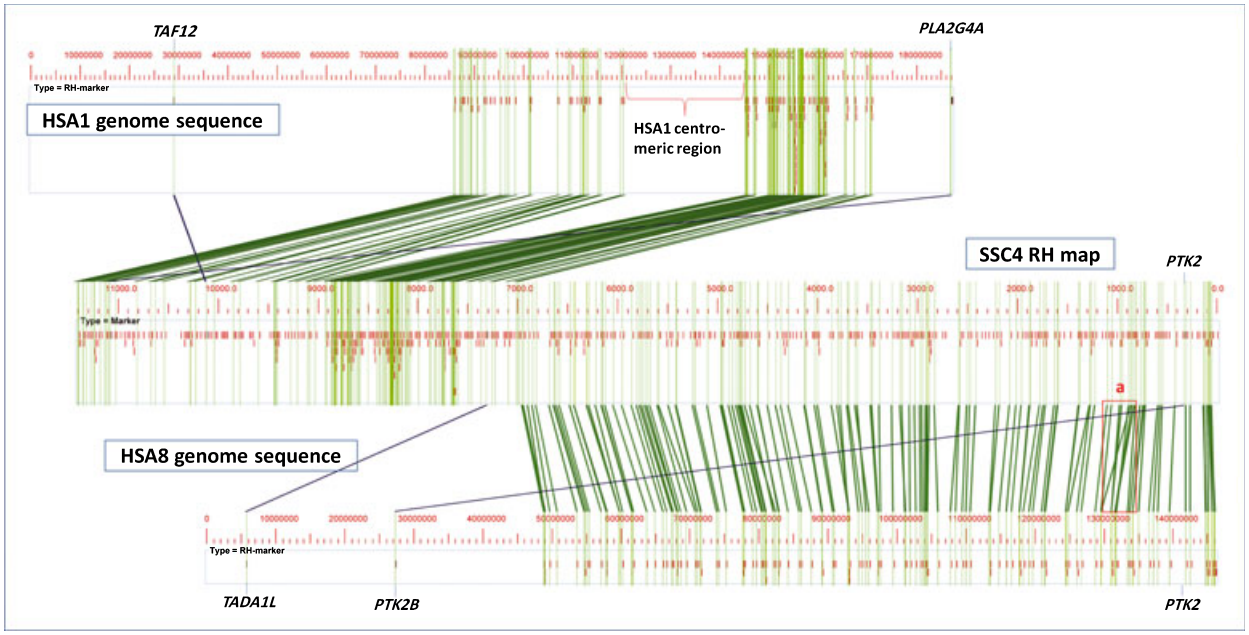

Figure 1 Comparison of STS order between the SSC4 RH map and HSA1 and HSA8 sequence maps. The corresponding SSC4 STS positions on HSA1 and HSA8 (Build 37) were aligned along the SSC4 $RH_{12\,000\text{-rad}}$ map (Fig. S1). Top: HSA1 genome sequence. Middle: SSC4 RH map with an accumulated map length of 11407 $cR_{12\,000}$. Bottom: HSA8 genome sequence. The red bracket indicates the centromeric region of 25 Mb (120–145 Mb) in HSA1 that is not conserved with SSC4, although the flanking regions are highly conserved between the two species. The map positions of *TAF12, PLA2G4A, PTK2, PTK2B* and *TADA1L* are marked by the gene symbol. Red box **a** refers to the region **a** in Fig. 1 where map discrepancy was observed. RH, radiation hybrid.

genomes. Contigs were then aligned to the SSC4 genome sequence (Sscrofa9) and the RH map (Fig. 2). A total of 2019 contigs mapped to SSC4. Two hundred and forty-eight of these contigs (Table S3) overlapped with sequences of genes/ESTs mapped on the $IMNpRH2_{12\,000\text{-rad}}$ RH panel, and the remaining 1171 contigs provided additional coding sequence for the RH map (Fig. 2). Several large blocks of SSC4 genomic sequence, i.e. 3–5 Mb, 22–27 Mb and 88–91 Mb (Fig. 2), did not match contig sequences, suggesting that a select group of genes are expressed in porcine macrophages and lymph nodes.

Finally, the high-resolution integrated map created through pairwise alignment of marker and sequence positions on the RH and SSC4 sequence maps, respectively, (Fig. S1c, d) identified four small regions (Fig. 2, regions a–d) of potential inconsistency. They were reanalysed using the conserved synteny between HSA8/HSA1 and SSC4 to determine whether the SSC4 RH or sequence map was correct based on agreement with the human sequence. Marker identity and order in the regions b, c and d on the SSC4 RH map (Fig. 2) are in agreement with the corresponding sequences on HSA8 and 1, eliminating the potential discrepancies identified in regions b-d (Fig. 2) and suggesting that a re-examination of the SSC4 sequence assembly in these three regions is necessary. The potential discrepancy identified as region a (Fig. 2) was not resolved because of limitations in the current level of resolution of the SSC4 RH map, sequence assembly and the HSA8 assembly.

We also analysed two genes (*PTK2 and TADA1L*) that were not in synteny between SSC4 and HSA8 (Fig. 1) and found that the human *PTK2* gene has two copies; *PTK2* at 141.7 Mb and *PTK2B* at 27.3 Mb on HSA8 (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=hum_chr.inf&query=PTK2). BLAT alignment (http://genome.ucsc.edu/) of the pig *PTK2* sequence (accession no. BW999672) against the human genome (Build 37) identified an orthologous copy at 27.3 Mb, but not the copy at 141.7 Mb on HSA8, suggesting that SSC4 may have only one copy of *PTK2*, similar to the human *PTK2B* gene, which is conserved between SSC4 and HSA8 (Fig. 1). A BLAT alignment also showed that the pig *TADA1L* gene (accession no. BI336804) has an orthologous sequence on HSA1 (score 360; similarity 94.2%) and a potentially orthologous sequence on HSA8 (score 305; similarity 86.1%). The human genome does, in fact, have two copies of *TADA1L,* one on HSA1 at 166.8 Mb and the other on HSA8 at 5.7 Mb (Fig. 1), suggesting that the assignment of the pig *TADA1L* to the SSC4 region in synteny with HSA8 is correct on the RH map, although we do not know if the pig has more than one copy of the *TADA1L* gene. Two additional genes, *TAF12* and *PLA2G4A,* are not in synteny between SSC4 and HSA1 (Fig. 1), but the multiple copies of the human *TAF12* may indicate a species-specific gene duplication and account for the lack of conservation in the position of this amplified gene family between human and pig.

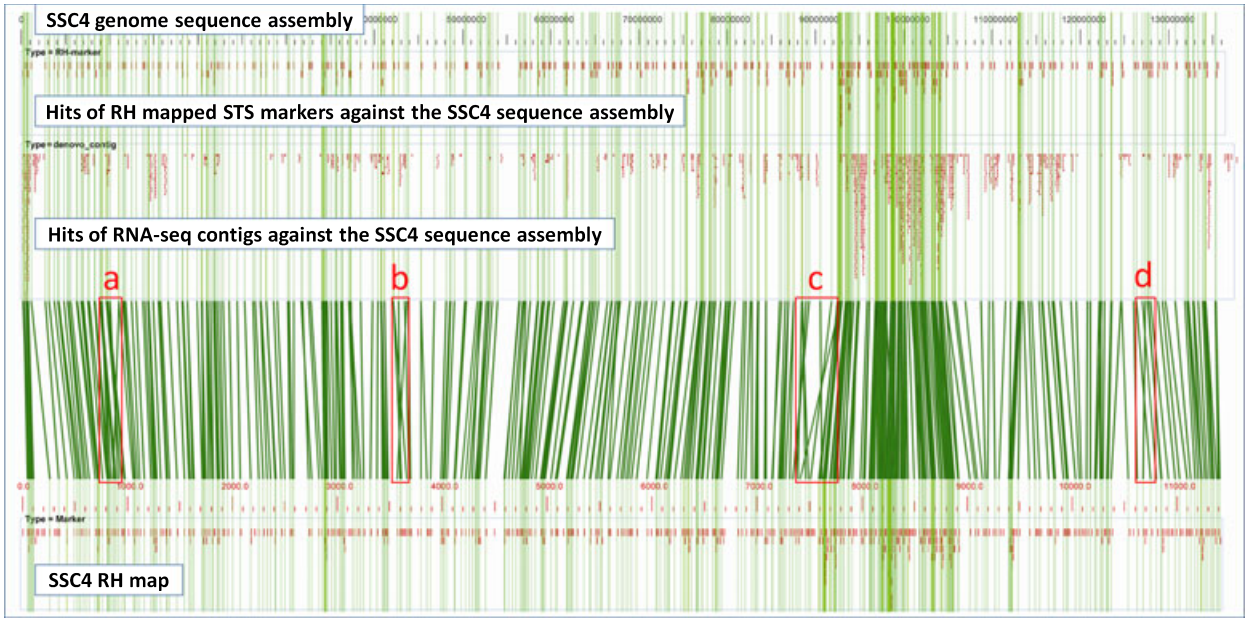

Figure 2 Comparison of SSC4 RH map, sequence assembly and RNA-seq data. The RNA-seq contigs and the RH marker sequences in the SSC4 RH map were aligned with the SSC4 genome assembly (Sscrofa9) and loaded into Comparative Map and Trait Viewer for visualization and exploration (Sawkins *et al*. 2004). Top: SSC4 draft assembly and the locations of the RH markers and RNA contigs. Bottom: marker order in the SSC4 RH map with an accumulated map length of 11407 $cR_{12\,000}$. The red bars under each scale represent the hits when conducting the BLAST search and alignment. The green vertical lines between the two maps indicate the corresponding map positions in the two maps. Red boxes (a–d) highlight the four regions where map discrepancies were observed. RH, radiation hybrid.

We then used the comparative RH map to estimate gap distance in the FPC (and sequence) map. Analysis of the genes and BACs adjacent to the gap between ctg4004 and ctg4005 (Fig. S1) identified two porcine genes, *FCGR2B* and *DUSP12*, and one MS, UMNp591, (Fig. S1) that mapped to BAC CH242-447O4, located at the terminus of ctg4004 in the FPC database (http://www.sanger.ac.uk/cgi-bin/Projects/S_scrofa/WebFPCdirect.cgi?contig=4004&clone=CH242-447O4). In contrast, the pig *NDUFS2* and *FCER1G* genes were located in BAC CH242-28G15, which maps to ctg4005 and overlaps with 310B20A04 (http://www.sanger.ac.uk/cgi-bin/Projects/S_scrofa/WebFPCdirect.cgi?contig=4005&clone=CH242-28G15). The cgt4004/ctg4005 boundary corresponds to a region of conserved synteny on HSA8. Human *FCGR2B* starts at 161647639 bp, while *FCER1G* ends at 161188748 bp, suggesting that the gap between ctg4004 and ctg4005 in the pig FPC map is ~328 Kb [(161647639–161188748)/1.4].

In conclusion, a comparative mapping approach allowed us to integrate results from the $IMNpRH2_{12\,000\text{-rad}}$ and $IMpRH_{7\,000\text{-rad}}$ panels, RNA-seq data, genetic and BAC FPC maps, to estimate the size of the remaining contigs on SSC4, to identify the gaps between porcine BAC contigs along the tiling path of SSC4, and to aid sequence assembly.

Acknowledgements

We thank Earl Landrito, Joseph Ekstrand, Michael Treat, Nicole Paes, Mark Lemos, Amy C Griffith and Mindy L Davis, at the University of Nevada at Reno, and Kelly Crawford, from the Pennsylvania State University, for their assistance in the RH typing. We are grateful to Thomas Wu of Genentech, Inc., for allowing us to have pre-publication access to the GMAP and GSNAP software. This work was supported by grants from USDA-CSREES-NRI Nos. 2004-35205-14244 and 2003-03686) and start-up funds from the Pennsylvania State University to Liu, W.-S.

References

DeWan A.T., Parrado A.R., Matise T.C. & Leal S.M. (2002) The map problem: a comparison of genetic and sequence-based physical maps. *American Journal of Human Genetics* **70**, 101–7.

Huang X., Wang J., Aluru S., Yang S.P. & Hillier L. (2003) PCAP: a whole-genome assembly program. *Genome Research* **13**, 2164–70.

Humphray S.J., Scott C.E., Clark R. *et al.* (2007) A high utility integrated map of the pig genome. *Genome Biology* **8**, R139.

Lewin H.A., Larkin D.M., Pontius J. & O'Brien S.J. (2009) Every genome sequence needs a good map. *Genome Research* **19**, 1925–28.

Liu W.S., Eyer K., Yasue H. *et al.* (2005) A 12,000-rad porcine radiation hybrid (IMNpRH2) panel refines the conserved synteny between SSC12 and HSA17. *Genomics* **86**, 731–8.

Liu W.S., Yasue H., Eyer K. *et al.* (2008) High-resolution comprehensive radiation hybrid maps of the porcine chromosomes 2p and 9p compared with the human chromosome 11. *Cytogenetic and Genome Research* **120**, 157–63.

Ma J.G., Yasue H., Eyer K.E., Hiraiwa H., Shimogiri T., Meyers S.N., Beever J.E., Schook L.B., Beattie C.W. & Liu W.S. (2009) An

integrated RH map of porcine chromosome 10. *BMC Genomics* **10,** 211.

Meyers S.N., Rogatcheva M.B., Larkin D.M., Yerle M., Milan D., Hawken R.J., Schook L.B. & Beever J.E. (2005) Piggy-BACing the human genome II. A high-resolution, physically anchored, comparative map of the porcine autosomes. *Genomics* **86,** 739–52.

Miller N.A., Kingsmore S.F., Farmer A. *et al.* (2008) Management of High-Throughput DNA Sequencing Projects: *Alpheus*. *Journal of Computer Science and Systems Biology* **1,** 132–48.

Rink A., Eyer K., Roelofs B. *et al.* (2006) Radiation hybrid map of the porcine genome comprising 2035 EST loci. *Mammalian Genome* **17,** 878–85.

Rohrer G.A., Alexander L.J., Keele J.W., Smith T.P. & Beattie C.W. (1994) A microsatellite linkage map of the porcine genome. *Genetics* **136,** 231–45.

Sawkins M.C., Farmer A.D., Hoisington D., Sullivan J., Tolopko A., Jiang Z. & Ribaut J.M. (2004) Comparative map and trait viewer (CMTV): an integrated bioinformatic tool to construct consensus maps and compare QTL and functional genomics data across genomes and experiments. *Plant Molecular Biology* **56,** 465–80.

Schiex T. & Gaspin C. (1997) CARTHAGENE: constructing and joining maximum likelihood genetic maps. *Proceedings/International Conference on Intelligent Systems for Molecular Biology* **5,** 258–67.

Simpson J.T., Wong K., Jackman S.D., Schein J.E., Jones S.J. & Birol I. (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research* **19,** 1117–23.

Warren R.L., Sutton G.G., Jones S.J. & Holt R.A. (2007) Assembling millions of short DNA sequences using SSAKE. *Bioinformatics* **23,** 500–1.

Wu T.D. & Nacu S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**(7), 873–81.

Wu T.D. & Watanabe C.K. (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21,** 1859–75.

Yerle M., Pinton P., Delcros C., Arnal N., Milan D. & Robic A. (2002) Generation and characterization of a 12,000-rad radiation hybrid panel for fine mapping in pig. *Cytogenetic and Genome Research* **97,** 219–28.

Yu A., Zhao C., Fan Y. *et al.* (2001) Comparison of human genetic and sequence-based physical maps. *Nature* **409,** 951–3.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1 High-resolution RH comprehensive and comparative maps of porcine chromosome (SSC) 4.

Table S1 Mapping vectors used in this study.

Table S2. Markers in the SSC4 12 000- and 7 000-rad RH maps.

Table S3. SSC4 marker information.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.