Original Article

Cytogenet Genome Res 115:134–137 (2006)
DOI: 10.1159/000095233

Cytogenetic and Genome Research

Sequence analysis of the porcine *IFNAR1* and *IFNGR2* genes

T. Leeb K. Dolle B. Haase

Institute of Genetics, Vetsuisse Faculty, University of Berne, Berne (Switzerland)

Manuscript received 3 March 2006; accepted in revised form for publication by M. Schmid, 10 April 2006.

Abstract. A porcine BAC clone harboring the tightly linked *IFNAR1* and *IFNGR2* genes was identified by comparative analysis of the publicly available porcine BAC end sequences. The complete 168,835 bp insert sequence of this clone was determined. Sequence comparisons of the genomic sequence with EST sequences from public databases were performed and allowed a detailed annotation of the *IFNAR1* and *IFNGR2* genes. The analyzed genes showed a conserved genomic organization with their known mammalian orthologs, however the sequence conservation of these genes across species was relatively low. In addition to the *IFNAR1* and *IFNGR2* genes, which were completely sequenced, the analyzed BAC clone also contained parts of an orphan gene encoding a putative transmembrane protein (TMEM50B). In contrast to the *IFNAR1* and *IFNGR2* genes the sequence conservation of the *TMEM50B* gene across different mammalian species was extremely high.

Genes of the interferon signaling pathway are of interest as candidate genes for disease resistance in livestock animals as mutations in these genes were found in human patients with greatly elevated susceptibility to specific infections (Ottenhoff et al., 2002). Interferons mediate their actions through two different interferon receptors (interferon α/β/ω receptor and interferon γ receptor). Further components of the interferon signaling cascade are the janus kinases and stat transcription factors (Briscoe et al., 1996). We had previously characterized the porcine *JAK1* and *TYK2* genes, which encode janus kinases mediating downstream effects of the interferon receptors (Kuiper et al., 2003; Leeb et al., 2004). The two interferon receptors themselves are composed of two subunits each that are encoded by separate genes. We report here the characterization of the clustered genes for the interferon α/β/ω receptor 1 subunit *(IFNAR1)* and for the interferon γ receptor 2 subunit *(IFNGR2)*.

Materials and methods

In silico identification of a porcine gene-specific BAC clone

A 400-kb segment from build 35.1 of the human genome containing the *IFNAR1* and *IFNGR2* genes was downloaded from NCBI (http://www.ncbi.nlm.nih.gov/mapview/). Repetitive sequences in the human sequence were masked with Repeatmasker (http://www.repeatmasker.org/). The repeat-masked human sequence was then used as query in a BLASTN search against the GSS section of GenBank limited by the Entrez query 'Sus scrofa [ORGN]' (Altschul et al., 1990). The resulting BLAST hits were visually inspected for porcine BAC end sequences that matched in a position and orientation on the human genome, which was compatible with a clone that contained the porcine orthologs of the desired human genes. BAC end sequences were publicly available from the CHORI-242, RPCI-44, PigE, PiGI and KNP pig BAC libraries (http://www.sanger.ac.uk/Projects/S_scrofa/). From the positive results the RPCI-44 BAC clone RP44-362L19 (http://bacpac.chori.org/) was chosen for sequencing.

Request reprints from Tosso Leeb
Institute of Genetics, Vetsuisse Faculty, University of Berne
Bremgartenstrasse 109a, CH–3001 Berne (Switzerland)
telephone: +41 31 631 2326; fax: +41 31 631 2640
e-mail: Tosso.Leeb@itz.unibe.ch

1424–8581/06/1152–0134$23.50/0

Accessible online at:
www.karger.com/cgr

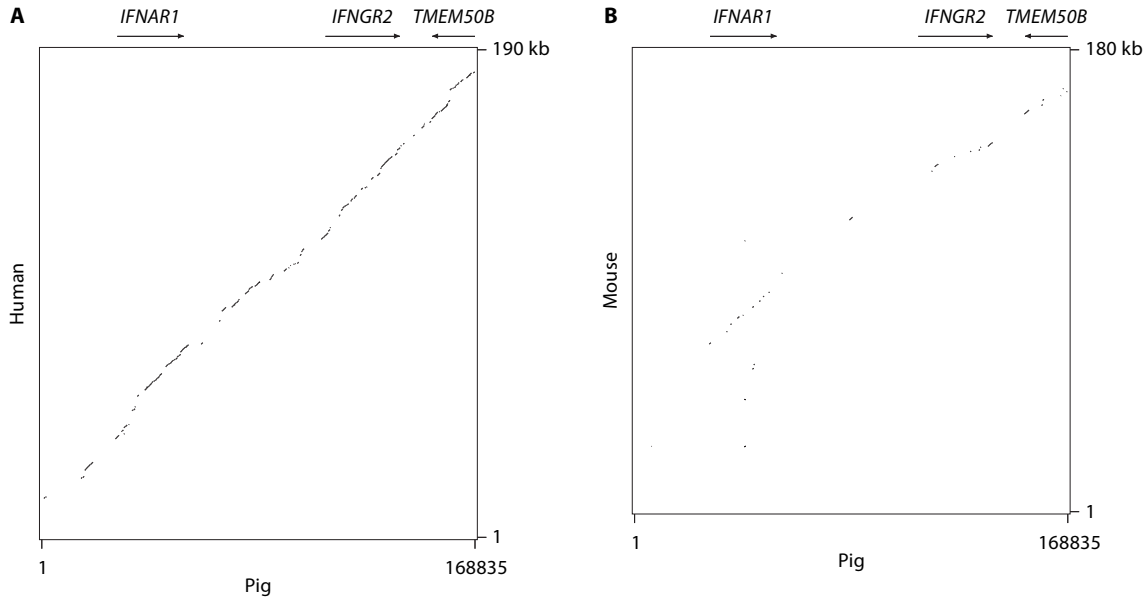

Fig. 1. Large scale sequence conservation between pig, human, and mouse. Dot plots were generated with the PiPMaker software (Schwartz et al., 2000). Repetitive sequences in the porcine sequence were masked. (**A**) Pig-human comparison. The analyzed porcine sequence was compared to the human 190-kb interval from 33.58 to 33.77 Mb on HSA21 (NCBI, build 35.1). The porcine sequence showed high overall sequence homology to about 167 kb of human genomic sequence. (**B**) Pig-mouse comparison. The analyzed porcine sequence was compared to the murine 180-kb interval from 91.47 to 91.65 Mb on MMU16 (NCBI, build 35.1). The pig-mouse overall homology was much lower than the pig-human homology. The covered interval in the murine genome was extrapolated at ~128 kb.

DNA sequencing

The insert of the BAC clone RP44-362L19 was sequenced as described (Leeb et al., 2004). Briefly, BAC DNA was prepared with the Qiagen Large Construct kit (Qiagen, Hilden, Germany). BAC DNA was sheared and a plasmid shotgun library with 2–5 kb insert size was prepared. Random shotgun clones were sequenced on a MegaBACE 1000 capillary sequencer (GE Healthcare, Freiburg, Germany) until 6× coverage of the insert was achieved. The sequences were assembled with Sequencher 4.5 (Genecodes, Ann Arbor, MI). Remaining gaps and single stranded regions were finished using a targeted primer walking strategy. The finishing reactions were performed either on Licor sequencers (Licor, Bad Homburg, Germany) or on an ABI 3730 sequencer (Applied Biosystems, Rotkreuz, Switzerland).

Assembly of cDNA sequences

Full-length cDNA sequences of the porcine *IFNAR1*, *IFNGR2*, and *TMEM50B* genes were assembled from public EST data. Putative cDNA sequences were put together from the determined genomic BAC sequence based on homology to the orthologous human genes. The putative porcine cDNA sequence was then used as query in a BLASTN search against the EST division of GenBank limited by the Entrez query 'Sus scrofa [ORGN]'. The BLASTN result was inspected visually, matching EST sequences were downloaded from GenBank and assembled using Sequencher 4.5. In the case of *IFNAR1*, a publicly available porcine cDNA sequence (accession AB116561) was also used for the assembly. The availability of porcine 5′-ESTs from full-length enriched cDNA libraries proved to be essential for this strategy (Uenishi et al., 2004).

Annotation of the genomic sequence

The determined genomic sequence was compared with the assembled porcine full-length cDNA sequences using the program spidey, which identifies exon-intron boundaries (http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/). The output of spidey was manually inspected for plausibility and used for the annotation of the EMBL entry AM229679.

Results and discussion

The ongoing porcine genome project provides excellent resources that facilitate the investigation of specific genes. To obtain a porcine BAC clone harboring the *IFNAR1* and *IFNGR2* genes we initially retrieved a 400-kb HSA21 sequence fragment from the public databases, which contained the human orthologous *IFNAR1* and *IFNGR2* genes. After masking repetitive sequences this sequence was used as query in a BLASTN search against the publicly available porcine BAC end sequences. The BLAST results indicated that the T7 end of porcine BAC clone RP44-362L19 had a unique and significant match to HSA21 at 33.76 Mb in reverse complementary orientation. Thus, it seemed likely that this clone contained the porcine orthologs of the human *IFNAR1* and *IFNGR2* genes, which are located approximately 140 and 60 kb proximal of the BLASTN match, respectively. We obtained the clone RP44-362L19 and determined the complete sequence of the 168,835 bp insert. The genomic sequence was deposited in the EMBL nucleotide database under accession AM229679.

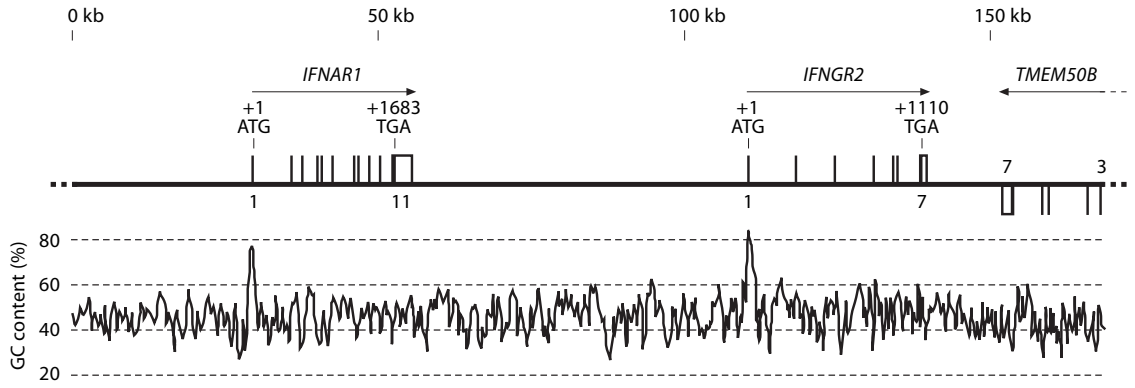

Fig. 2. Genomic organization of the porcine *IFNAR1*, *IFNGR2*, and *TMEM50B* genes. The sequenced BAC clone RP44-362L19 is represented by a horizontal line. Exons are represented by vertical lines above or below the horizontal line according to the transcription orientation. First and last exons are numbered. In the lower part of the figure the GC-content is illustrated. The GC content was calculated using CpG plot and a 300 bp window (http://www.ebi.ac.uk/emboss/cpgplot/). Note the striking correlation between elevated GC content and the promoters of the *IFNAR1* and *IFNGR2* genes.

The sequence had an overall GC content of 46.0% and contained 51.0% repetitive sequences. Comparative analysis with the finished human genome sequence revealed homologies to the *IFNAR1*, *IFNGR2*, and *TMEM50B* (previously *C21orf4*) genes. Interestingly, and consistent with the relatively high repeat content in the porcine sequence, the porcine sequence covered a slightly smaller interval in the human genome (~169 kb pig vs ~167 kb human, Fig. 1). In previous comparative investigations we had observed slightly smaller porcine genomic distances compared to the human genome (Leeb and Müller, 2004). The corresponding mouse genome interval is only ~128 kb or roughly 75% of the pig and human sequences.

The detailed annotation of the porcine genomic sequence revealed a conserved genomic organization with respect to the human genome sequence (Fig. 2). The porcine *IFNAR1* gene contained 11 exons spanning 26 kb of genomic sequence. All exon-intron boundaries conformed to the GT-AG rule. The porcine *IFNAR1* cDNA contained a rather large 3′-UTR with 2.7 kb. However, the porcine *IFNAR1* 3′-UTR is still smaller than the corresponding human 3′-UTR at 4.3 kb. The porcine *IFNAR1* gene encoded a precursor polypeptide of 560 amino acids and a calculated weight of 63,292 Da, which showed 66.8% identity to the corresponding 557 amino acid human IFNAR1 protein and 44.8% identity to the murine Ifnar1 protein. The promoter of the porcine *IFNAR1* gene was GC-rich and did not contain a TATA-box motif.

The porcine *IFNGR2* gene contained seven exons spanning 29 kb of genomic DNA. Intron 4 of this gene belonged to the rare class of GC-AG introns (Burset et al., 2001), the other introns were of the common GT-AG type. The porcine *IFNGR2* gene contained an open reading frame for 369 amino acids encoding a precursor polypeptide of 41,392 Da with 60.0% identity to the human 337 amino acid IFNGR2 and 48.8% identity to the murine Ifngr2 protein. The promoter of the porcine *IFNGR2* gene was extremely GC-rich (80% GC over 700 bp) and did not contain a TATA-box motif. The promoters of the *IFNAR1* and *IFNGR2* genes were associated with extended (~1,200 bp) CpG islands. However, 34 additional 200–700-bp segments that also fulfilled the sequence composition criteria of CpG islands were present on the sequenced BAC clone.

Finally, the putative exons 3–7 of the *TMEM50B* gene were contained in the analyzed sequence. The entire porcine *TMEM50B* ORF sequence could be assembled from the publicly available ESTs. This gene encoded a putative 158 amino acid transmembrane protein that showed 100% identity to the orthologous human TMEM50B protein and 98.1% identity to the murine Tmem50b protein.

In conclusion, our analysis provides the basis for further functional studies of the porcine *IFNAR1* and *IFNGR2* genes. Typical for immune defense-related genes, these genes showed relatively weak sequence conservation across mammalian species compared to most other protein encoding genes. The employed method of identifying a gene-specific BAC by using an 'in silico library screen' represented an efficient approach compared to a conventional experimental library screen using either PCR screening of hierarchical DNA-pools or radioactive hybridization of high-density colony filters.

Acknowledgements

The technical assistance of Tina Ueffing and Heike Klippert-Hasberg is gratefully acknowledged.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 215:403–410 (1990).

Briscoe J, Kohlhuber F, Müller M: JAKs and STATs branch out. Trends Cell Biol 6:336–340 (1996).

Burset M, Seledtsov IA, Solovyev VV: SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res 29:255–259 (2001).

Kuiper H, Martins-Wess F, Lassnig C, Distl O, Müller M, Leeb T: Assignment of the porcine janus kinase 1 gene *(JAK1)* to chromosome 6q34→q35 by fluorescence in situ hybridization and radiation hybrid mapping. Cytogenet Genome Res 101:92D (2003).

Leeb T, Müller M: Comparative human-mouse-rat sequence analysis of the *ICAM* gene cluster on HSA19p13.2 and a 185-kb porcine region from SSC 2q. Gene 343:239–244 (2004).

Leeb T, Martins-Wess F, Kuiper H, Lassnig C, Distl O, Müller M: Molecular characterization of the porcine *TYK2* gene on SSC2q1.3→q2.1. Cytogenet Genome Res 107:103–107 (2004).

Ottenhoff TH, Verreck FA, Lichtenauer-Kaligis EG, Hoeve MA, Sanal O, van Dissel JT: Genetics, cytokines and human infectious disease: lessons from weakly pathogenic mycobacteria and salmonellae. Nat Genet 32:97–105 (2002).

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, et al: PipMaker – a web server for aligning two genomic DNA sequences. Genome Res 10:577–586 (2000).

Uenishi H, Eguchi T, Suzuki K, Sawazaki T, Toki D, et al: PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. Nucleic Acids Res 32:D484–488 (2004).

Announcement

International System for Human Cytogenetic Nomenclature (ISCN)

The newly elected committee for the International System for Human Cytogenetic Nomenclature (ISCN) has been formed. The committee consists of eleven members from six geographical locations. The following individuals will serve on the ISCN committee for 2007–2011: Myriam Chaabouni (Africa/Middle East); Yoshimitsu Fukushima and Prochi Madon (Asia); Lynda Campbell (Australia), Christine Harrison, Nils Mandahl, and Albert Schinzel (Europe); Kathleen Rao, Marilyn Slovak, and Lisa Shaffer (North America); and Carla Rosenberg (Latin America).