138–141 Nucleic Acids Research, 2003, Vol. 31, No. 1
DOI: 10.1093/nar/gkg061

The GENETPIG database: a tool for comparative mapping in pig (*Sus scrofa*)

Emmanuelle Karsenty[1,2], Emmanuel Barillot[1,2,3], Gwenola Tosser-Klopp[4], Yvette Lahbib-Mansais[4], Denis Milan[4], François Hatey[4], Susanna Cirera[5], Milena Sawera[5], Claus B. Jørgensen[5], Bhanu Chowdhary[5,6], Merete Fredholm[5], Klaus Wimmers[7], Siriluck Ponsuksili[7], Roberta Davoli[8], Luca Fontanesi[8], Silvia Braglia[8], Paolo Zambonelli[8], Daniele Bigi[8], Stefan Neuenschwander[9] and Joël Gellin[4,*]

[1]Infobiogen, 523 place des terrasses, F-91000 Evry, France, [2]Unité de Recherche Génomique-Info, INRA, 523 place des terrasses, F-91000 Evry, France, [3]Généthon, 1 bis rue de l'internationale, 91000 Evry, France, [4]Laboratoire de Génétique Cellulaire, INRA, 31326 Castanet-Tolosan, BP27, France, [5]The Royal Veterinary and Agricultural University, Department of Animal Science and Animal Health, Division of Animal Genetics, Buelowsvej 13, Copenhagen, Frederiksberg C, Denmark, [6]Department of Veterinary Anatomy and Public Health, Texas A&M University, College Station TX-77843, USA, [7]Institute of Animal Breeding Science, University of Bonn, Endenicher Allee 15, D-53115 Bonn, Germany, [8]Bologna University, DIPROVAL Sez. Allevamenti Zootecnici, Via F.lli Rosselli, 107, 42100 Coviolo-Reggio Emilia, Italy and [9]Institut für Nutztierwissenschaften, Tannenstrasse 1, ETH-Zentrum, CH-8092 Zürich, Switzerland

Received August 14, 2002; Revised and Accepted September 24, 2002

ABSTRACT

The GENETPIG database has been established for storing and disseminating the results of the European project: 'GENETPIG: identification of genes controlling economic traits in pig'. The partners of this project have mapped about 630 porcine and human ESTs onto the pig genome. The database collects the mapping results and links them to other sources of mapping data; this includes pig maps as well as available comparative mapping information. Functional annotation of the mapped ESTs is also given when a significant similarity to cognate genes was established. The database is accessible for consultation via the Internet at http://www.infobiogen.fr/services/Genetpig/.

INTRODUCTION

The GENETPIG database integrates the results of the European project 'GENETPIG' whose aim was to accelerate mapping and to identify GENes controlling Economic Traits in PIG. Deciphering the genetic mechanisms behind traits such as meat quality, litter size, or disease resistance is of high economic importance, but precise localization of these quantitative trait loci requires high-density genome maps. After localization, the logical approach is to use the high level of colinearity between mammalian genomes to propose positional candidate genes based on comparative mapping of known homologous genes.

The GENETPIG database collects the mapping results and links them to other sources of mapping data, such as pig maps, and comparative mapping results from the Mouse Genome Database (MGD) (1) and from human-pig bi-directional chromosome painting (2), or Zoo-FISH. Additionally, functional annotation of the mapped ESTs is provided when a significant similarity to cognate genes has been established. The GENETPIG database is, therefore, a novel and important resource for functional and comparative mapping of the pig genome. Database browsing is available at http://www.infobiogen.fr/services/Genetpig/.

GENETPIG DATABASE CONTENTS

GENETPIG database includes information on 747 ESTs, 632 of which have been mapped onto the pig genome (Table 1). Most of them (583) were mapped using a rodent porcine Somatic Cell Hybrid Panel (SCHP) (3,4) allowing rapid assignment at the cytogenetic level. A total of 372 ESTs were mapped onto the IMpRH radiation hybrid (RH) panel, resulting in a fine map that was integrated with the INRA-University of Minnesota Radiation Hybrid map (5–7; see also http://imprh.toulouse.inra.fr). Eleven ESTs were localized by Fluorescent In Situ Hybridization (FISH).

In the GENETPIG database, 38 ESTs come from human specific ESTs (8) and 58 from Traced Orthologous Amplified

*To whom correspondence should be addressed. Tel: +33 5 61 28 54 30; Fax: +33 5 61 28 53 08; Email: gellin@toulouse.inra.fr

Table 1. Number of porcine ESTs produced and mapped by each laboratory. Some ESTs have been mapped by several methods

Laboratory	Number of ESTs	Tissue/origin	Mapped ESTs in pig			
			SCHP	RH	ISH	Total
Bonn	120	Liver	119	0	0	119
		Undetermined	1	0	0	1
Copenhagen	223	Small intestine/Jejunum	181	207	3	213
		Undetermined	10	9	0	10
Reggio Emilia (Bologna University)	235	Muscle	122	5	3	124
		Undetermined	1	0	0	1
Toulouse	166	Ovary	64	65	3	65
		Human/TOAST	83	86	1	96
Zürich	3	Fat	2	0	1	3
TOTAL	747		582	372	11	632

SCHP, somatic cell hybrid panel; RH, radiation hybrids; ISH, in situ hybridization.

Sequence Tags (TOASTs) (9,10). The porcine PCR products of these ESTs were sequenced and homology to the human sequence of origin was systematically checked before validation.

The 651 porcine ESTs were mainly, but not exclusively, selected for their preferential or exclusive expression in liver, skeletal muscle, ovarian follicular cells, small intestine and fat tissues. These tissues are of primary importance for livestock farming. For example, pig mortality is high in the perinatal period and problems of digestion and nutrition are often the causative factors. Therefore, elucidating genes that are of primary importance to the normal development of the gastro-intestinal tract would be of direct interest in QTL studies and selection. Also skeletal muscle is a target tissue for the identification of candidate genes for meat production and quality traits, while ovary plays a key role in litter size.

These porcine ESTs were selected from the following sources:

- A granulosa cell cDNA library (INRA Toulouse), and differential display experiments (11,12).
- A small intestine cDNA library and an assortment of ESTs for which a putative gene name could be assigned based on *in silico* analysis and the homologous gene in human has been mapped (Royal Veterinary and Agricultural University, Copenhagen) (13).
- By differential display PCR, used to isolate liver specific cDNA. Liver metabolism enzymes not mapped up to now were also selected from the literature and the public databases (Bonn University) (14).
- A subtracted porcine skeletal muscle cDNA library (Bologna University) (15,16).

Zoo-PCR (8) was carried out on a total of 451 ESTs, to assess primer amplification on different genomes: human, hamster, mouse, chicken, bovine, sheep, rabbit, goat and horse. Single Strand Conformation Polymorphism (SSCP) analysis of PCR products from different porcine breeds was also performed on some ESTs either for mapping experiments (discrimination of the rodent and the pig PCR products) or to assess polymorphism.

All the results from sequencing, mapping (FISH, SCHP, RH), Zoo-PCR tests, polymorphism analysis, bibliographical references and laboratory protocols (for sequencing, Zoo-PCR, localization) are stored in the GENETPIG database. Locus symbols for individual ESTs were allocated by the partners who mapped it, when the BLAST result gave enough evidence (264 ESTs). Moreover, all results about functional annotation and comparative mapping are reported in the database. Pig cytogenetic (17) and IMpRH (6,7) maps were also integrated into the GENETPIG database.

QUERYING THE GENETPIG DATABASE

Several levels of queries are offered in the GENETPIG database interfaces. These include:

- Statistics on EST localizations.
- Summaries of mapping data, locus symbol and comparative mapping, per EST or per chromosome. Alignment of individual locus with MGD human-mouse comparative mapping results . . .
- Queries on the different EST features: localization, Zoo-PCR, comparative mapping, human or computer annotations. Computer annotation was done with BLAST on a set of non-redundant nucleic and proteic databases at Infobiogen including GenBank/EMBL without HTG and GSS divisions, repbase, Pfam, SWISS-PROT, TREMBL, PIR, Genpept, NRL3D and PDB. Queries can also be made on the description of the EST hits.
- Visualization and comparison of the cytogenetic, RH, SCHP and FISH maps. This interface (Fig. 1) is based on the MappetShow viewer (18). It is noteworthy that SCHP mapping does not discriminate all chromosome segments drawn on the figure: if more than one region is proposed, mapped EST names will appear in red. Synteny checking has been implemented: if an EST has a porcine locus symbol, a localization onto the pig genome and a mapped human homolog in MGD (homolog meaning in this case with identical locus symbol), the software checks whether the human and porcine localizations are in correspondence according to the comparative Zoo-FISH map (2). Agreement is marked as '=' besides the EST name on the figure, and disagreement as '#'.
- Visualization of the cytogenetic, SCHP and FISH maps, with the comments of the partners who mapped the EST on the right of the EST name. The comment generally validates the human homolog.

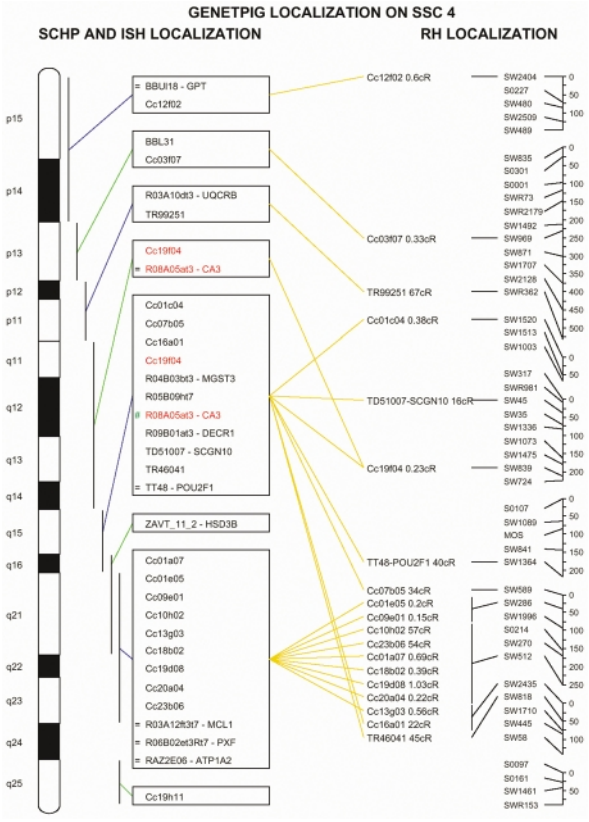

Figure 1. Comparing cytogenetic, SCHP and ISH, RH maps from the pig genome. The marker R08A03at3-CA3 appears in red because two localizations or more are proposed for this marker. On this chromosome two localizations are proposed: 4q11-q14 and 4q15-q16, because the SCHP mapping does not allow discrimination between these two segments. But this marker has a mapped human homolog, CA3 (carbonic anhydrase 3), that maps in 8q13-q22 (Unigene) and according to the human-porcine bi-directional painting, 8q13-q24 in human hybridizes 4pter-q14 but not 4q15-q16 in pig (4q15-q16 is hybridized by human chromosome 1). The marker 4q11-q14 occurrence is labelled a = symbol, and the 4q15-q16 occurrence a # symbol to express respectively the concordance and the discordance of the SCHP mapping with the human-porcine bi-directional painting.

Various web links to other databases such as the Human Gene Nomenclature Database (19), Pigbase (http://iowa.thearkdb.org), Pigmap (http://www.projects.roslin.ac.uk/pigmap/ecpigmap.html), MGD (1) (Mammalian Homology section), LocusLink (20), Unigene (20), Entrez (20) or EMBL (21) are integrated into the GENETPIG database.

IMPLEMENTATION

The GENETPIG environment was developed on a Sun/Solaris platform using the EyeDB object-oriented database management system (22) from Sysra (http://www.sysra.com). All data management software programs and most interfaces are written in C++ (forms) or Java (graphical interfaces) using the EyeDB Application Programming Interface, and some CGI are written in Perl. The database is accessible via the Internet at http://www.infobiogen.fr/services/Genetpig/.

ACKNOWLEDGEMENTS

We thank Eric Viara for his support in developing the GENETPIG environment, Frédéric Guyon for his help in adapting and using the graphical interface, and Christophe Cussat-Blanc for system management. The GENETPIG project was funded by the European Commission BIO4-CT98-0237 contract.

REFERENCES

1. Blake,J.A., Richardson,J.E., Bult,C.J., Kadin,J.A., Eppig,J.T. and the Mouse Genome Database Group (2002) The Mouse Genome Database (MGD): The model organism database for the laboratory mouse. *Nucleic Acids Res.*, **30**, 113–115.
2. Goureau,A., Yerle,M., Schmitz,A., Riquet,J., Milan,D., Pinton,P., Frelat,G. and Gellin,J. (1996) Human and porcine correspondence of chromosome segments using bidirectional chromosome painting. *Genomics*, **36**, 252–262.
3. Robic,A., Riquet,J., Yerle,M., Milan,D., Lahbib-Mansais,Y., Dubut-Fontana,C. and Gellin,J. (1996) Porcine linkage and cytogenetic maps integrated by regional mapping of 100 microsatellites on somatic cell hybrid panel. *Mamm. Genome*, **7**, 438–445.
4. Yerle,M., Echard,G., Robic,A., Mairal,A., Dubut-Fontana,C., Riquet,J., Pinton,P., Milan,D., Lahbib-Mansais,Y. and Gellin,J. (1996) A somatic cell hybrid panel for pig regional gene mapping characterized by molecular cytogenetics. *Cytogenet. Cell Genet.*, **73**, 194–202.
5. Yerle,M., Pinton,P., Robic,A., Alfonso,A., Palvadeau,Y., Delcros,C., Hawken,R., Alexander,L., Beattie,L.B., Milan,D. and Gellin,J. (1998) Construction of a whole genome radiation hybrid panel for high-resolution gene mapping in pigs. *Cytogenet. Cell Genet.*, **82**, 182–188.
6. Hawken,R.J., Murtaugh,J., Flickinger,G.H., Yerle,M., Robic,A., Milan,D., Gellin,J., Beattie,C.W., Schook,L.B., and Alexander,L.J. (1999) A first generation porcine whole genome radiation hybrid map. *Mamm. Genome*, **10**, 824–830.
7. 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. and Gellin,J. (2000) IMpRH server: an RH mapping server available on the Web. *Bioinformatics*, **6**, 558–559.
8. Lahbib-Mansais,Y., Dalias,G., Milan,D., Yerle,M., Robic,A., Gyapay,G. and Gellin,J. (1999) A successful strategy for comparative mapping with human ESTs: 65 new regional assignments in the pig. *Mamm. Genome*, **10**, 145–153.
9. Jiang,Z., Priat,C. and Galibert,F. (1998) Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. *Mamm. Genome*, **9**, 577–587.
10. Lahbib-Mansais,Y., Leroux,S., Milan,D., Yerle,M., Robic,A., Jiang,Z., Andre,C. and Gellin,J. (2000) Comparative mapping between humans and pigs: localization of 58 anchorage markers (TOASTs) by use of porcine somatic cell and radiation hybrid panels. *Mamm. Genome*, **11**, 1098–1106.
11. Clouscard-Martinato,C., Mulsant,P., Robic,A., Bonnet,A., Gasser,F. and Hatey,F. (1998) Characterization of FSH-regulated genes isolated by mRNA differential display from pig ovarian granulosa cells. *Anim. Genet.*, **29**, 98–106.
12. Lahbib-Mansais,Y., Tosser-Klopp,G., Leroux,S., Cabau,C., Karsenty,E., Milan,D., Barillot,E., Yerle,M., Hatey,F. and Gellin,J. (2003) Contribution to High-resolution mapping in pigs with 101 Type I markers and Progress in comparative map between Humans and Pigs. *Mamm. Genome*.
13. Cirera,S., Jørgensen,C.B., Sawera,M., Raudsep,T., Chowdhary,B. and Fredholm,M. (in preparation). Comparative mapping in the pig: localization of 214 expressed sequence tagged sites.
14. Ponsuksili,S., Wimmers,K., Yerle,M. and Schellander,K. (2001) Mapping of 93 porcine ESTs preferentially expressed in liver. *Mamm. Genome*, **12**, 869–872.
15. Davoli,R., Bigi,D., Fontanesi,L., Zambonelli,P., Yerle,M., Zijlstra,C., Bosma,A.A., Robic,A. and Russo,V. (2000) Mapping of 14 expressed sequence tags (ESTs) from porcine skeletal muscle by somatic cell hybrid analysis. *Anim Genet.*, **31**, 400–403.

16. Davoli,R., Fontanesi,L., Zambonelli,P., Bigi,D., Gellin,J., Yerle,M., Milc,J., Braglia,S., Cenci,V., Cagnazzo,M. and Russo,V. (2002) Isolation of porcine expressed sequence tags for the construction of a first genomic transcript map of the skeletal muscle in pig. *Anim Genet.*, **33**, 3–18.
17. Yerle,M., Lahbib-Mansais,Y., Robic,A., Pinton,P., Goureau,A., Milan,D. and Gellin,J. (1997) The cytogenetic map of the Pig (*Sus scrofa* Domestica). *Mamm. Genome*, **8**, 592–607.
18. Guyon,F., Vaysseix,G. and Barillot,E. (2000) MappetShow: Non-Linear Visualization for genome data. *Pacific Symposium on Biocomputing*, Hawaii, pp. 215–226.
19. Wain,H.M., Lush,M., Ducluzeau,F. and Povey,S. (2002) Genew: The Human Gene Nomenclature Database. *Nucleic Acids Res.*, **30**, 169–171.
20. Wheeler,D.L., Church,D.M., Lash,A.E., Leipe,D.D., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Tatusova,T.A., Wagner,L. and Rapp,B.A. (2002) Database resources of the National Center for Biotechnology Information: 2002 update. *Nucleic Acids Res.*, **30**, 13–16.
21. Stoesser,G., Baker,W., van den Broek,A., Camon,E., Garcia-Pastor,M., Kanz,C., Kulikova,T., Leinonen,R., Lin,Q., Lombard,V. Lopez,R., Redaschi,N., Stoehr,P., Tuli,M.A., Tzouvara,K. and Vaughan,R. (2002) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **30**, 21–26.
22. Viara,E., Barillot,E. and Vaysseix,G. (1999) The EyeDB OODBMS. *International Database Engineering and Applications Symposium*, Montreal. IEEE Publications, pp. 390–402.