

# The Pig Genome Database (PiGenome): an integrated database for pig genome research

Dajeong Lim · Yong-Min Cho · Kyung-Tai Lee ·  
Yeonkyung Kang · Samsun Sung · Jungrye Nam ·  
Eung-Woo Park · Sung-Jong Oh · Seok-Ki Im · Hee-bal Kim

Received: 29 April 2008 / Accepted: 28 October 2008 / Published online: 10 December 2008  
© Springer Science+Business Media, LLC 2008

**Abstract** We established the Pig Genome Database (PiGenome) for pig genome research. The PiGenome integrates and analyzes all publicly available genome-wide data on pigs, including UniGenes, sequence tagged sites (STS) markers, quantitative trait loci (QTLs) data, and bacterial artificial chromosome (BAC) contigs. In addition, we produced 69,545 expressed sequence tags (ESTs) from the full-length enriched cDNA libraries of six tissues and 182 BAC contig sequences, which are also included in the database. QTLs, genetic markers, and BAC end-sequencing information were collected from public databases. The full-length enriched EST data were clustered and assembled into unique sequences, contigs, and singletons. The PiGenome provides functional annotation, identification of transcripts, mapping of coding sequences, and SNP information. It also provides an advanced search interface, a disease browser, alternative-splicing events, and a comparative gene map of the pig. A graphical map view and genome browser can map ESTs, contigs, BAC contigs (from the National Institute of Animal Science), Sino-Danish Pig Genome Project transcripts, and UniGene onto pig genome sequences which include our 182 BAC contigs and publically available BAC sequences of the Wellcome

Trust Sanger Institute. The PiGenome is accessible at <http://pigenome.nabc.go.kr/>.

## Introduction

The pig is an important model for evolutionary research (Fang et al. 2005) and is used as a biomedical model for studying reproduction, tissue degeneration/biological maintenance, stem cells, and immune responses (Rothschild 2003; Vodicka et al. 2005). The genome-wide analysis of the pig has grown rapidly in recent years and is expected to impact both human medicine and pork production. In the pig genomics study, full-length enriched cDNA is a valuable resource for functional genomics and for annotating functional genes on the pig genome (Stapleton et al. 2002). For example, candidate genes related to meat quality can be searched using gene expression profiling of EST sequences from a full-length enriched cDNA library of porcine backfat (Kim et al. 2006b).

Genome-wide linkage maps of the pig were developed by the European PiGMap initiative (Archibald et al. 1995) and USDA-MARC (Rohrer et al. 1996) and subsequently expanded to 2,700 markers (Womack 2005) for quantitative trait mapping. A genome database of the pig was constructed and has been included in the ArkDB of the Roslin Institute. The genome databases are a comprehensive public repository for genome mapping data from farmed and other animal species including cat, chicken, cow, deer, duck, horse, pig quail, salmon, sea bass, sheep, and turkey (Hu et al. 2001). The Pig EST Data Explorer (PEDE) database, in which data collections are derived from sequences assembled from porcine 5' ESTs from

---

The first three authors should be regarded as joint First Authors.

---

D. Lim · Y. Kang · S. Sung · J. Nam · H. Kim (✉)  
Laboratory of Bioinformatics and Population Genetics,  
Department of Agricultural Biotechnology, Seoul National  
University (SNU), Seoul 151-742, South Korea  
e-mail: hee-bal@snu.ac.kr

D. Lim · Y.-M. Cho · K.-T. Lee · E.-W. Park · S.-J. Oh ·  
S.-K. Im  
Division of Animal Genomics and Bioinformatics, National  
Institute of Animal Science, Rural Development Administration,  
Omokchun-dong 564, Kwonsun-gu, South Korea

oligo-capped full-length cDNA libraries, was established to provide gene annotation information of large-scale EST sequence data (Uenishi et al. 2004). This database is also valuable for tissue-specific gene expression profiling. Genome-wide QTL data from pigs has been integrated in the Pig QTL Database (PigQTLdb), which contains published pig QTL information on major genes and markers that are associated with economically important traits. It allows a user to search by either chromosome or keywords on the basis of their publication information (Hu et al. 2005).

Currently, several groups, including the Sanger Institute and the Sino-Danish Pig Genome Project, are sequencing the pig genome. However, the sequencing information is a small portion of the whole pig genome, thus analyses of the pig genome projects are dependent on comparative mapping against other mammals for which whole genomic information is available. Among these, the Pig Genomic Informatics System (PigGIS) allows biomedical researchers to locate pig genes using their human homologs and position single nucleotide polymorphisms (SNPs) with  $0.66 \times$  genomic reads and ESTs (Ruan et al. 2007). Complete genome sequencing data on the pig will be available in a few years. Although only limited genomic information is available on pigs, integration of all publicly available pig genomic data will accelerate the progress of pig genome research. Therefore, we collected pig data from public databases and generated an integrated map from genome informatics. The Pig Genome Database (PiGenome) displays mapped results and links them to other sources of mapping data. The database also provides transcript data on 69,545 porcine ESTs which we produced from full-length enriched cDNA libraries of six tissues. Users can also explore pig disease genes based on comparative mapping of homologous genes.

## Materials and methods

### Construction of EST data

The six tissues (abdominal fat, adipocyte, loin muscle, backfat, liver, and pituitary gland) used to construct the cDNA libraries were prepared from crossbred (Landrace  $\times$  Large White) pigs at 1, 7, 12, 18, and 24 weeks of age. A full-length enriched cDNA library was constructed using the oligocapped method (Suzuki et al. 1997). The cDNA inserts were sequenced once from the 5' end of clones using a BigDye<sup>®</sup> Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and a 3730 DNA analyzer (Applied Biosystems). The EST trace data were called bases using the Phred (Ewing and Green 1998)

and vector sequences were clipped by the cross-match program (Gordon et al. 1998). Vector-screened EST sequences were filtered for repetitive sequences and low-complexity regions using RepeatMasker (<http://www.repeatmasker.org/>). We excluded ESTs of less than 200 bp from our data set. The ESTs were clustered and assembled using the CAP3 program (Huang and Madan 1999). We retrieved 5,870 contigs and 7,669 singletons from 69,545 ESTs. We performed a homology search on 5,870 contigs using BLASTN (Altschul et al. 1997) against the UniGene Clusters of human (build No. 199), mouse (build No. 161), and pig (build No. 27). Similar to Uenishi et al. (2004), if a contig sequence contains a longer coding sequence compared with its homologous UniGene coding sequence with high similarity, it is considered to include a full-length coding sequence. Our ESTs have been submitted to NCBI dbEST and the identification numbers are FD587621-FD643263, FD698977-FD698990.

### BAC library screening and construction of BAC contigs

QTLs on pig chromosome 6q26–34 affect meat quality, including intramuscular fat content and backfat thickness (de Koning et al. 1999; Grindflek et al. 2001; Ovilo et al. 2000, 2002; Szyda et al. 2002). This region is syntenic with human chromosome 1 from 65 to 83 Mb and with human chromosome 18 from 0.61 to 40 Mb. We performed BAC physical mapping using the PCR-based four-dimensional BAC clone screening method with 19 microsatellite markers and BAC end-sequence (BES)-based screening markers. The BES-based screening markers were designed at intervals of 60 kb using the results of the BES corresponding to the region from the trace archive of Ensembl (<http://trace.ensembl.org>). BAC clones were screened from the Korean Native Pig BAC library (Jeon et al. 2003). The screened BAC clones were extracted without *Escherichia coli* genomic DNA contamination using the Large Construction Kit (Qiagen, Valencia, CA, USA). The extracted BAC DNAs were subjected to  $8 \times$  shotgun sequencing. Sequence data were assembled with Phred and Phrap (University of Washington, Seattle, WA, USA). The contigs of BAC clones were assigned to the human genome using GenomeVISTA (Couronne et al. 2003). A total of 182 BAC contigs were retrieved.

### Gene Ontology (GO) annotation and SNP identification from transcripts

GO annotation and identification were performed with a sequence similarity search against the tentative consensus (TC) sequences of the SsGI release 12.0 (20 June 2006) using BLASTN. To categorize GO, we used the GO flat

files of the Gene Ontology Consortium (<http://www.geneontology.org>). The cutoff values for GO identification were 95% identity, 60% coverage, and an e-value of  $< 0.00001$ . To identify transcripts, we performed a BLASTX search against the nonredundant protein database of the National Center for Biotechnology Information (NCBI; <ftp://ftp.ncbi.nih.gov/blast/db/>; downloaded 4 February 2007). To determine the SNP positions on transcripts, we performed a BLASTN search against the NCBI SNP database ([ftp://ftp.ncbi.nih.gov/snp/organisms/pig\\_9823/ss\\_fasta/](ftp://ftp.ncbi.nih.gov/snp/organisms/pig_9823/ss_fasta/); cutoff value: identity  $> 90\%$ ).

#### Detecting the genomic locations of the pig data sets

The Sanger Institute's pig BES database (<ftp://ftp.sanger.ac.uk/pub/sequences/pig/>) and Ensembl map information ([http://pre.ensembl.org/Sus\\_scrofa/](http://pre.ensembl.org/Sus_scrofa/)) were downloaded. However, because the Pig Genome Project is not finished, we could not directly align our data set with the genome sequences; thus, we indirectly estimated the genomic regions of the data sets to generate a BAC-based physical map of the pig genome. Because end sequences of BACs provide highly specific markers for genome sequencing (Venter et al. 1998), we constructed the map via paired alignment of pig BES sequences with ESTs, contigs, UniGenes, and BAC contigs using BLAST with an e-value of  $1^{-e^{100}}$ .

We downloaded consensus and singleton sequences from 1,021,891 ESTs based on 97 non-normalized cDNA libraries (<http://pigest.ku.dk/download/index.html>) (Gorodkin et al. 2007) and aligned them against the genomic sequences. Only the highest-score alignment of each BLAST result was curated and stored in the database. The genomic regions of the data sets were estimated based on the chromosomal positions of BAC clones from the corresponding BESs using the BAC clone-BES pair database. We downloaded QTL data (e.g., chromosome, location, location span) from the PigQTL database. Marker locations were obtained from the Meat and Animal Research Center of the U.S. Department of Agriculture.

#### Identification of pig disease genes

We retrieved homogene data (<ftp://ftp.ncbi.nih.gov/pub/HomoloGene/build53/>) from NCBI and extracted the protein accession IDs from human and mouse data. To identify orthologs from the human, mouse, and pig homologues, we used the reciprocal best blast hits algorithm (Wall et al. 2003) to make comparisons between the human, mouse, and pig using BLASTX and TBLASTN (cutoff: e-value 0.00001). Candidate disease genes of humans and mice were obtained from the Mouse Genome Database (MGI; <http://www.informatics.jax.org/>).

#### Identification of tissue-specific alternative-splicing events

Our algorithm identifies novel alternative-splicing events based on sequence similarity searches between each EST data set and known gene set. A detailed description of the algorithm can be seen elsewhere (Kim et al. 2006a). Briefly, only the best hit result of a BLAST (Altschul et al. 1997) sequence similarity search against the known UniGene data set was retained. If the best alignment is a collinear form without deletion or insertion, it is regarded as an existing alternative-splicing event. On the other hand, if the best alignment contains an insertion or deletion of an EST query with more than two High-scoring Segment Pairs (HSPs), it is considered a new alternative-splicing form which has never been seen before, at least in the current pig gene data set. In the former case (case 1) event, an EST fragment is inserted in a sequence of current alternative-splicing isoform. For the latter case (case 2) event, an EST fragment is deleted from an existing isoform sequence that is already aligned to the UniGene data set. Thus, this event should be confirmed whether the case 2 fragment is a match with the sequence of the existing alternative-splicing isoform. The new alternative-splicing events were cross-checked using the SIM4 program (Florea et al. 1998). SIM4 uses a greedy algorithm to align two sequences and then makes the sequence alignment process reliable. In the case 1 event, case sequences were extracted from the UniGene sequence corresponding to the EST. In the case 2 event, case sequences were extracted from EST sequence corresponding to the UniGene sequence. Because the sequences of case events confirmed whether the case sequences are really novel in the existing UniGene data set, for sequences of tissue-specific alternative-splicing case events we revalidated the case sequences using the BLAST program. We accept only the case sequences that pass our filtering criteria: BLAST are 95% identity, e-value  $< 0.00001$ , length of HSP  $> 20$  bp considering length of the exon, and the SIM4 program.

#### Comparative gene mapping against the contigs of BAC clones in pigs

We constructed a comparative map between 182 porcine BAC contigs and other species, including human, mouse, cattle, and dog. The porcine BAC contigs were aligned against the BES data set of PigMAP using BLASTN, and the regional order (from lowest to highest genome position) was estimated based on the BLAST results. We also performed a BLASTX search against human, mouse, cattle, and dog protein sequences from NCBI and identified homologs. Our results increased the resolution of the current comparative gene maps between the pig and several other species.

### Database construction and implementation

The database and Web interface were developed using MySQL, PHP, HTML, and Javascript. The standard server requires an Apache 1.3.19, MySQL 4.0.21, and PHP 4.3.9 with GD library 2.0.33 and uses the FreeBSD 5.2 operating system. Data processing and analysis were written in Python. We will update the database quarterly and have modules for functional annotation using Python programs. After accumulating our new sequences data, we will indicate the last update date when the website is updated. A summary of the data processing pipeline is shown in Fig. 1. A more detailed outline of the data processing and data information can be viewed in data information menu of the database.

### Results

#### Contents of the PiGenome database

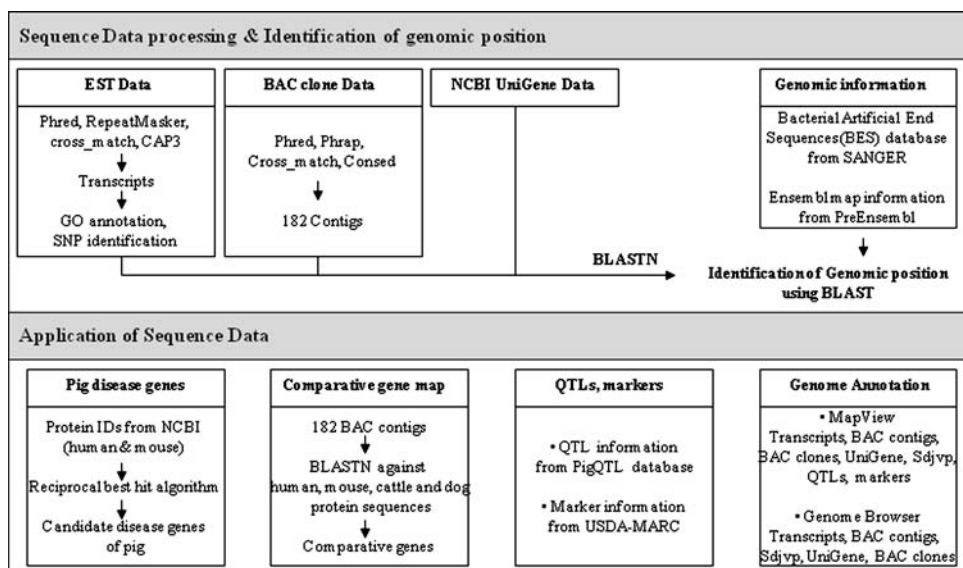
The PiGenome database is available at <http://piggenome.nabc.go.kr>. The PigGenome Web interface allows interactive use of the information related to the pig genome. The interface consists of five menus: Sequence Data, BLAST search, Search by, Application, and Genome Annotation. Users can obtain basic EST and transcript information from the Sequence Data menu. EST data can be searched by local ID, accession number, or description of homology against the nonredundant NCBI protein database based on the BLASTX or BLAST score and e-value. When a user clicks on a local ID, its details are reported on a new page, i.e., the BLAST results against the BAC End Sequence (BES) database of the Sanger Institute (<ftp://ftp.sanger.ac.uk/pub/sequences/pig/>), the source of

**Fig. 2** Simple query and output format of the PiGenome. Users can search for individual genes and sequence information using search options. **a** Disease Browser enables a user to obtain putative disease transcripts of the pig using the OMIM accession ID or a disease term. The comparative gene map can know the evolutionary relationship between mammalian species in the pig genome (composed of 182 BAC contigs) using an orthologous relationship. **b** The Genome Annotation page is divided into two parts, map view (c) and genome browser (d). Users can see the output by map view according to data types: EST, contigs, markers, QTL, BAC clone, UniGene, National Institute of Animal Science (NIAS) BAC contigs from the NIAS, consensus sequences, and singletons of Sino-Danish Pig Genome Project. The genome browser provides genomic alignment of all data types within a specific genomic region

the library, and the sequence view. Transcript information can be searched in a similar way but it displays Gene Ontology (GO) annotation, SNP information (dbSNP accession number and alleles), and assembled ESTs lists. If a SNP exists in a transcript, its position is marked in red.

The BLAST search page is useful for searching sequence similarity against the pig database for our porcine abdominal fat, adipocyte, loin muscle, backfat, liver, and pituitary gland. BLASTN, TBLASTN, and TBLASTX are available which allow external users to compare their own sequences against our database with a user options of e-value and filtering. The input sequence must be in FASTA format or sequence-only format to allow searches of nucleotide or amino acid sequences. BLAST results are shown on a new page in an output format similar to that of the NCBI site, and they provide gene information on the basis of matched sequences. The PiGenome also contains simple query interfaces for EST, transcript, markers, QTL, and gene. A user can easily mine the information from each database. The PiGenome includes an advanced search interface, a disease browser, and a pig QTL comparative

**Fig. 1** Data processing pipeline of the PiGenome



**(A)**

**Search the vocabulary**  
 - Enter any text string or OMIM accession ID

**Terms indexed by beginning character**  
 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z 0-9

---

**Human Disease and Mouse & PIG Model Detail**

Human Disease OMIM ID: 608747  
 Human Disease Term: Insulin-Like Growth Factor I Deficiency

| Mouse Gene | Human Gene | Human-Pig Ortholog | Mouse-Pig Ortholog |
|------------|------------|--------------------|--------------------|
| Igf1       | IGF1       | BX672216           | CN157588           |

**(B)**

1. Select Chromosome or BAC contig name  
 Chromosome:  chr6  BAC contig name:

2. Select species  
 ALL  Human  Mouse  Cattle  Dog

3. Display the result by option

| FIG's Chr_6 | BAC contig name | Human              | Mouse                 | Dog               | Cattle                     |
|-------------|-----------------|--------------------|-----------------------|-------------------|----------------------------|
| 10M         | 1044G1_seq      | chr10: NP_071351.1 | chr10: NP_573500.2    | chr7: XP_537333.2 | scaffold1039: XP_001254966 |
|             | 992H3_seq       | chr10: Q9787       | chr10: XP_921509.2    | chr7: XP_209767.1 | chr8: NP_776393.1          |
|             | 244B5_seq       | chr10: XP_950805.1 | chr4: NP_035790.1     | chr7: XP_547674.2 | chr8: NP_776393.1          |
| 6M          | 217F2_seq       | chr10: XP_950805.1 | chr4: NP_035790.1     | chr7: XP_547674.2 | chr8: NP_776393.1          |
|             | 1091E06_seq     | chr10: EX01641.1   |                       |                   |                            |
| 11M         | 1061B4_seq      | chr19: AF086231.1  | chr2: BNE22391.1      |                   |                            |
|             | 921E08_seq      | chr9: EW07668.1    | chr12: XP_001003086.1 |                   |                            |
|             | 1129G6_seq      | chr19: AF027666.1  |                       |                   | chr16: XP_869427.2         |

**(C)**

1. Select Chromosome and karyotype band range:  
 Chromosome Karyotype band:  chr1  p1.1  User-defined Range from  to

2. Select data type:  
 EST  Contig  Marker  QTL  BAC clone  UniGene  NLR1\_BAC\_clone  Sd\_jvp

Genetic Map (chr1) Physical Map Karyotype band : q2.2  
 QTL : 0(0)/29  
 Marker : 1/106  
 EST : 44  
 Contig : 7  
 UniGene : 51  
 BAC clone : 651  
 NLR1 BAC clone : 8  
 Sd\_jvp : 22

**(D)**

Click on a chromosome for a closer view:

Jump to region 6 : 83635504 to 84635503

chromosome band:

FFC contig

contig: EST more than 1 only 1, unigene more than 1 only 1

Sd\_jvp map

NLR1 BAC map  
 BAC map

Legend: Free (blue), Selected (red), Committed (green), Accessed (yellow), Finished (grey)

map. A user can select a disease name, OMIM accession ID, or type in specific text, and the orthologs of the human, mouse, and pig are summarized for the disease term. Output format of the database is shown in Fig. 2.

### Tissue-specific alternative splicing

The results of tissue-specific alternative-splicing (AS) events are available at the Alternative-splicing page. We found 717 sequences of insertion in a specific tissue (case 1) and 255 sequences of deletion in a specific tissue (case 2). We also detected 16 common cases that were shown in two tissues, seven case 1 types, and nine case 2 types. Most of the alternative-splicing events (65%) were detected in translated regions. We have also observed frequent alternative splicing in untranslated regions (UTR), 8% of 5' UTR and 3% of 3' UTR in total case events. When a user selects tissues and case types, the result will be displayed in a table.

### Comparative gene map and genome annotation

A comparative gene map of the pig genome (composed of 182 BAC contigs) is useful for investigating the evolutionary relationships among mammalian species. To conduct such an analysis, a user selects the chromosome or BAC contigs and then chooses the species. Comparative candidate genes can be displayed in two ways, map view and table view, along with their IDs, description, and chromosome name. The Genome Annotation page consists of two sections: map view and genome browser. On the Map View page, a user selects a chromosome and a karyotype band (or types in a specific genomic position). Then the user chooses from the data types EST, contigs, markers, QTL, BAC clone, UniGene, BAC contigs from the National Institute of Animal Science (NIAS), consensus sequences, and singletons of the Sino-Danish Pig Genome Project (Gorodkin et al. 2007). Only the markers and QTL defined by the genetic position and the physical map are displayed. The amount of data for a specific genomic region is shown on the right of the output page, and a summary of the results is given by data type. Clicking on any genomic location opens a Genome Browser page that provides genomic alignment data corresponding to the genomic region. We developed our own genome browser, which is similar to PreEnsembl. Users can click a chromosome and then specify a genomic position. By default, genomic alignment shows a region of 1 Mb. The genome browser also features zoom (in/out) buttons for 0.2, 0.5, 1, 2, 5, 10, and 50 Mb. Users can adjust the range by clicking on the chromosome or specifying a start and end position. Yellow bars represent contigs placed on the fingerprint contig (FPC) map. The EST sequences are colored according to alignment number. Pink bars indicate that only one EST is

aligned to a genomic region. Red bars represent several ESTs on the chromosome. Clicking on any bar opens a pop-up window that provides more detailed information on the data. BAC clones are displayed in different colors according to their status.

### Discussion

In this article we described PiGenome, an integrated genome database for determining gene annotation, orthologous disease genes, tissue-specific alternative splicing, comparative gene maps, and genomic alignment of pig data sets. PiGenome includes UniGenes, markers, QTLs, transcripts, BAC contigs, consensus sequences, and singletons of the Sino-Danish Pig Genome Project and is a useful tool for studying genomic and biological mechanisms of the pig. It has been suggested that the pig is the best model for human disease because it has similar body size and physiologic conditions. Although the pig is an important model for human disease, there are no well-established pig-related disease browsers or databases for disease. For example, OMIA was developed based on inherited disorders and other familial traits, collectively called “phenes.” It provides 214 pig genes with human homologs. We predicted 1,420 human-pig and 1,384 mouse-pig ortholog candidate disease genes using the human disease browser of MGI. A disease browser provides information for human or pig diseases. In addition, the database contains our 69,545 EST data which were annotated and integrated into Sequence Data, Disease Browser, and Alternative-splicing pages of the database. The pig has been studied as an important economic animal using fat content data. Our data in particular are related to fat contents. For example, about 80% of the ESTs were obtained from fat- or immune-related tissues. Most of BAC clones were obtained from SSC6q32–34 which contains the QTL region for meat quality, intracellular fat content, and backfat thickness.

We analyzed the comparative map for four mammalian species: human, mouse, dog, and cattle. We think that the pig QTL comparative map provides positional candidate genes associated with QTL on chromosome 6 that affect fat deposition in pigs. We also collected our sequencing data (5,870 contigs and 7,669 singletons) and Sino-Danish Pig Genome Project clusters (48,629 consensus sequences and 73,171 singletons). Utilizing the consensus sequences of this gene index in the genome draft could lead to many more exonic sequences, which would significantly facilitate current gene-oriented annotation efforts. Exonic sequences provide useful information for the discovery of additional relationships among genomic sequences, hypothetical genes, and conserved functional domains (Zhuo et al. 2001). BACs and ESTs and other public sequences

are mapped against the pig genome, which increases the coverage of the pig genome and makes a high-coverage clone map available. In the near future, the PiGenome will be updated quarterly with new pig genome data, including BAC sequences, fingerprint contig (FPC) map information, whole-genome shotgun (WGS) data, and more QTLs and markers. WGS data require substantially more computation time for assembly.

In the past decade, genome research on livestock has focused on map linking and QTL identification. Currently, it has been focused on genome-wide information on genes. PiGenome has biological significance for both agriculture and biomedicine. The database provides sequence mapping and QTL characterization, information that is useful for identifying genes associated with economically important traits and for studying the functional genomics of the pig and for researching genetic breeding.

**Acknowledgments** This work was supported by a grant from the National Institute of Animal Science (H. Kim) and from the Biogreen21 Project (20050301034467) of the Korean Rural Development Administration (S.-K. Im).

## References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Archibald AL, Haley CS, Brown JF, Couperwhite S, McQueen HA et al (1995) The PiGMap consortium linkage map of the pig (*Sus scrofa*). *Mamm Genome* 6:157–175
- Couronne O, Poliakov A, Bray N, Ishkhanov T, Ryaboy D et al (2003) Strategies and tools for whole-genome alignments. *Genome Res* 13:73–80
- de Koning DJ, Janss LL, Rattink AP, van Oers PA, de Vries BJ et al (1999) Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (*Sus scrofa*). *Genetics* 152:1679–1690
- Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186–194
- Fang M, Hu X, Jiang T, Braunschweig M, Hu L et al (2005) The phylogeny of Chinese indigenous pig breeds inferred from microsatellite markers. *Anim Genet* 36:7–13
- Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res* 8:967–974
- Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202
- Gorodkin J, Cirera S, Hedegaard J, Gilchrist MJ, Panitz F et al (2007) Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags. *Genome Biol* 8:R45
- Grindflek E, Szyda J, Liu Z, Lien S (2001) Detection of quantitative trait loci for meat quality in a commercial slaughter pig cross. *Mamm Genome* 12:299–304
- Hu J, Mungall C, Law A, Papworth R, Nelson JP et al (2001) The ARKdb: genome databases for farmed and other animals. *Nucleic Acids Res* 29:106–110
- Hu ZL, Dracheva S, Jang W, Maglott D, Bastiaansen J et al (2005) A QTL resource and comparison tool for pigs: PigQTLDB. *Mamm Genome* 16:792–800
- Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9:868–877
- Jeon JT, Park EW, Jeon HJ, Kim TH, Lee KT et al (2003) A large-insert porcine library with sevenfold genome coverage: a tool for positional cloning of candidate genes for major quantitative traits. *Mol Cells* 16:113–116
- Kim H, Lim D, Han BK, Sung S, Jeon M et al (2006a) ChickGCE: a novel germ cell EST database for studying the early developmental stage in chickens. *Genomics* 88:252–257
- Kim TH, Kim NS, Lim D, Lee KT, Oh JH et al (2006b) Generation and analysis of large-scale expressed sequence tags (ESTs) from a full-length enriched cDNA library of porcine backfat tissue. *BMC Genomics* 7:36
- Ovilo C, Perez-Enciso M, Barragan C, Clop A, Rodriguez C et al (2000) A QTL for intramuscular fat and backfat thickness is located on porcine chromosome 6. *Mamm Genome* 11:344–346
- Ovilo C, Clop A, Noguera JL, Oliver MA, Barragan C et al (2002) Quantitative trait locus mapping for meat quality traits in an Iberian × Landrace F2 pig population. *J Anim Sci* 80:2801–2808
- Rohrer GA, Alexander LJ, Hu Z, Smith TP, Keele JW et al (1996) A comprehensive map of the porcine genome. *Genome Res* 6:371–391
- Rothschild MF (2003) From a sow's ear to a silk purse: real progress in porcine genomics. *Cytogenet Genome Res* 102:95–99
- Ruan J, Guo Y, Li H, Hu Y, Song F et al (2007) PigGIS: Pig Genomic Informatics System. *Nucleic Acids Res* 35:D654–D657
- Stapleton M, Carlson J, Brokstein P, Yu C, Champe M et al (2002) A Drosophila full-length cDNA resource. *Genome Biol* 3:RESEARCH0080
- Suzuki Y, Yoshitomo-Nakagawa K, Maruyama K, Suyama A, Sugano S (1997) Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene* 200:149–156
- Szyda J, Liu Z, Grindflek E, Lien S (2002) Application of a mixed inheritance model to the detection of quantitative trait loci in swine. *J Appl Genet* 43:69–83
- Uenishi H, Eguchi T, Suzuki K, Sawazaki T, Toki D et al (2004) PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. *Nucleic Acids Res* 32:D484–D488
- Venter JC, Adams MD, Sutton GG, Kerlavage AR, Smith HO et al (1998) Shotgun sequencing of the human genome. *Science* 280:1540–1542
- Vodicka P, Smetana K Jr, Dvorankova B, Emerick T, Xu YZ et al (2005) The miniature pig as an animal model in biomedical research. *Ann NY Acad Sci* 1049:161–171
- Wall DP, Fraser HB, Hirsh AE (2003) Detecting putative orthologs. *Bioinformatics* 19:1710–1711
- Womack JE (2005) Advances in livestock genomics: opening the barn door. *Genome Res* 15:1699–1705
- Zhuo D, Zhao WD, Wright FA, Yang HY, Wang JP et al (2001) Assembly, annotation, and integration of UNIGENE clusters into the human genome draft. *Genome Res* 11:904–918