

Finding the Genes Expressed in Female Reproductive Tissues in Pigs

A.S. Leaflet R1950

Christopher K. Tuggle, Professor of Animal Science
for the
Midwest Consortium for Pig Reproduction Genomics

Summary and Implications

A molecular understanding of porcine reproduction is of biological interest and economic importance. Our Midwest Consortium has produced cDNA libraries containing the majority of genes expressed in major female reproductive tissues, and we have deposited into public databases 21,499 expressed sequence tag (EST) gene sequences from the 3' end of clones from these libraries. These sequences represent 10,574 different genes, based on sequence comparison among these data, and comparison to existing porcine genes indicate as many as 4,652 are novel. Computer analysis identified sequences that are expressed in specific pig tissues or organs, and confirmed the broad expression in pig for many genes ubiquitously expressed in human tissues. Furthermore, we have developed computer software to identify sequence similarity of these pig genes with their human counterparts, and to extract the mapping information of these human homologues from genome databases. We used this software to localize 61 genes on the porcine physical map of chromosomes 5, 10, and 14. Thus our sequence data is useful in accelerating mapping studies and will be useful in understanding pig reproductive biology.

Introduction

The reproductive process is central to pig production efficiency. Unfortunately there is loss of potential conceptuses during the first month of gestation (Perry, 1954). To increase the number and quality of offspring, we need to identify the genes that control ovulation rate, fertilization, conceptus quality, and the responsiveness of the dam to the conceptus.

A long-term goal of genetic mapping is the discovery of quantitative trait loci (QTL) for important physiological traits. Comparative data suggests that large regions are conserved between pig and human. Clearly the detailed mapping information available from the Human and Mouse Genome Projects will become extremely useful to find porcine QTL. Unfortunately, comparative maps are relatively limited in the pig. Mapping of individual genes suggests that gene order between species may be either conserved or quite divergent. A high resolution comparative map is therefore

a necessity, as human gene order is not always a good predictor of pig gene order.

The generation of Expressed Sequence Tags (ESTs), short DNA sequences from clones randomly picked from gene libraries constitutes an efficient strategy to identify and map genes. A deposit of 21,499 EST sequences has been made by the Midwest Consortium, which consists of scientists at Iowa State University (C.K. Tuggle, M.F. Rothschild), University of Iowa (M.B. Soares, T. Casavant), University of Missouri-Columbia (R. Prather), University of Nebraska-Lincoln (D. Pomp), and the National Center for Genomic Resources (W. Beavis). The data obtained by sequencing a large number of cDNAs derived from reproductive tissues were analyzed for expression patterns across tissues by using sequence frequency, and were used to identify and map ESTs with clear human matches to improve the comparative map between human and pig.

Materials and Methods

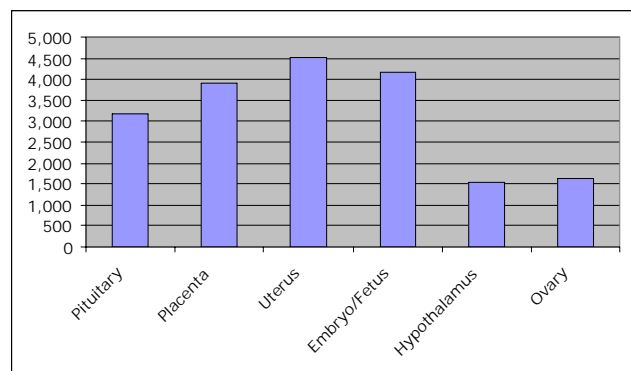
Porcine tissues used for this project included hypothalamus, ovary and anterior pituitary from gilts at day 0, 5 and 12 post-estrus, conceptuses from days 12 and 14 of pregnancy, and embryos at days 20 and 45 of pregnancy. RNA was isolated and used to produce cDNA libraries, which were sequenced at Iowa State University or University of Iowa. Sequence data was analyzed and submitted to Genbank. Gene sequences were analyzed to design specific primers for the physical mapping. Mapping was performed using the INRA-Toulouse DNA panel and database.

Results and Discussion

Library Production and Sequence Data Accumulation

Our Midwest Consortium has produced and analyzed cDNA libraries derived from porcine anterior pituitary, embryo, fetus, hypothalamus, ovary, and placenta collected at various stages of gestation or estrus. In Figure 1 we present the results of our sequencing of random

Figure 1. Total Pig Sequences by Tissue



clones from these libraries. In total, we have generated 21,499 sequences with sufficient quality score and length. All sequences have been submitted to the dbEST division of Genbank. Overall, the sequence quality and average length of submitted ESTs was very good. For further analysis, 2,287 sequences were removed from the dataset due to lower quality, and are not reflected in Figure 1.

Our goal was to generate at least 1,000 high quality sequences from each type of tissue. In Figure 1, we show this goal was met.

To determine the number of individual genes that these sequences represent, we have used clustering analysis to group (cluster) all ESTs with the same gene sequence. When the 19,218 sequences were compared, we identified 10,574 clusters. Of these 10,574 clusters, 7,286 (69%) represent single sequences and another 1,755 clusters have only 2 members. Thus, the sequencing data further indicate that we have highly complex libraries derived from a large set of expressed sequences in the pig.

To determine how many of these 10,574 clusters were novel relative to existing sequence entries, we used each cluster to search the public sequence databases. These comparisons are provided at: (<http://pigest.genome.iastate.edu/pub/MammGenome.html>). We found a significant percentage of our clusters have a match to entries in the full database (47%) or to the EST database (71%). Thus the vast majority of our clusters likely represent bona fide porcine genes based on a high frequency of matches to expressed sequences in other organisms. With respect to novelty compared to existing pig ESTs, 44% of our clusters did not have a match to existing porcine ESTs, indicating we have identified as many as 4,652 new pig genes

“Virtual” expression analysis

Our library production approach tagged each sequence when the library is made. This allows us to recognize differential tissue expression patterns “virtually”, i.e., through calculating frequency of sequences found for a specific gene cluster per tissue. As all sequences were obtained through random selection of clones, the frequency of sequences obtained for a particular gene is proportional to the level of expression of that gene. Thus we have analyzed the frequency of sequences in these libraries to identify highly expressed genes that are either broadly expressed in many tissues or specific to a tissue. We identified clusters with at least 15 members originating from a specific tissue source, or those found in multiple tissues and many libraries. At the above website, we list examples of widely expressed ESTs. Overall, 37 clusters of size 9-130 were represented in at least 9 different libraries. Of these 37 clusters, 26 match known “house-keeping” genes due to their broad expression in humans.

While it is more difficult to definitively identify tissue-specific expression patterns, in our sequence data it

was possible to clearly identify genes with high levels of expression within a specific tissue. As expected, we find that the genes for preprolactin and growth hormone are expressed exclusively in pituitary libraries (Table 1).

We also were interested to find genes expressed at the crucial stage of development for embryos that is at implantation. As we had libraries produced from mRNA collected from just before and just after implantation, we wanted to identify genes with differential expression between these two stages. We compared the frequencies of clusters in the E7 (early embryo) library to frequencies seen in the E4 (late embryo) library. We identified four genes with patterns indicating differential expression at implantation (Table 1). Interferon- γ (IFNG) and interleukin 1-beta are known to be expressed in the peri-implantation porcine embryo. We also identified two novel genes that are induced in expression at implantation (Table 1). We have confirmed the high level of placental expression for one of these, MI-P-E4-aii-c-03, through additional analysis (S. Zhao and C. Tuggle, unpublished results), and it would be of interest to characterize this gene and the other unknown placentally-expressed genes for their roles in reproduction. More generally, these analyses indicate our data is useful in understanding the genes expressed at critical times during pig reproduction, and the sequences obtained here will be an important resource for all reproduction biological research.

Development of EST selection software and use in comparative mapping

An important use of these sequences is in efficient comparative map development. Mapping requires the design of PCR primers from pig ESTs that have been selected to most efficiently improve the comparative map. To select such ESTs, we developed a set of computer scripts that automatically perform sequence match analysis to human sequences and collect information on the human matches obtained. This information is then used to develop lists of porcine ESTs that have a strong match to a human locus for which there is unambiguous, consistent mapping information. Using this software, we have compiled a list of over 1,600 ESTs with significant similarity to human genes with consistent mapping localizations. These data are available (http://pigest.genome.iastate.edu/cgiperl/map_table4-1.pl).

To demonstrate the utility of these programs and of the available EST sequences for comparative mapping, we selected ESTs for mapping with matches to human loci on human chromosomes 10, 12, and 22. These three human chromosomes have regions which are conserved with porcine chromosomes 5, 10 and 14. We identified 133 porcine ESTs with matches to loci mapping to these human chromosomes, and determined that mapping reagents for 64 of these would be useful for mapping. We mapped 61 of the 64 genes to specific locations on the pig

genome map (see above website for complete list of genes and mapping results). For those 61 loci, 56 (92%) were mapped to a location predicted by comparison to known human gene locations (see <http://www.toulouse.inra.fr/lgc/pig/cyto/cyto.htm>).

These data indicate both the value of the existing comparative map information to make accurate prediction of porcine gene position, as well as the need for further

mapping to define the exceptions to the predicted conserved regions.

Acknowledgements

This work was supported by USDA-NRICGP 99-35205-8370 and by the office of the Vice-Provost, Iowa State University.

Table 1. "Virtual" Expression Patterns: Clusters found exclusively in one tissue/organ

<u>Cluster Sequence Code</u>	<u>Human match gene name</u>	<u>Specific Tissue</u>	<u>Total Cluster size</u>
MI-P-A2-aey-e-11-1-UM	preprolactin	A.Pit	94
MI-P-E4-aik-e-11-1-UM*	interferon gamma	Conceptus	92
MI-P-E7-agm-h-09-1-UM*	prointerleukin-1 beta	Fetus	85
MI-P-E4-aia-c-03-1-UM*	no hits to human ESTs	EE Mem#	55
MI-P-A2-afh-g-07-1-UM	growth hormone mRNA	A.Pit	34
MI-P-E3-aal-e-09-1-UM	fetuin	Fetus	32
MI-P-E3-aam-b-03-1-UM	alpha-fetoprotein	Fetus	26
MI-P-E3-aal-a-08-1-UM	alpha-1-antitrypsin	Fetus	26
MI-P-E4-ahd-h-11-1-UM	pregn-assoc. glyco.prot 2/4/6	EE Mem	23
MI-P-HO-afv-d-09-1-UM	plp gene	hypothalamus	22
MI-P-A1-aao-g-12-1-UM	Follicle stim hormone (FSHB)	A. Pit	20
MI-P-AY1-nqb-b-03-1-UM	no hits to human ESTs	Placenta	16
MI-P-E7-aia-f-09-1-UM*	sim to CAA12352 membrane gp36	Conceptus	17
MI-P-AY1-nqb-b-03-0-UI	no hits to human ESTs	Placenta	16
MI-P-E3-agp-e-01-1-UM	no hits to human ESTs	Fetus	15

EE Mem = Placenta and day 14 embryo which includes developing extra-embryonic membranes (pre-pituitary tissue). A. Pit = anterior pituitary.

* ESTs marked with asterisk are significantly ($P < .001$) differentially expressed in comparison between pre-implantation (E4) and postimplantation (E7) libraries.