Development of New Placental and Fetal Expressed Sequence Tags (EST) for Gene Discovery in Pig Reproduction

C.K. Tuggle[1], associate professor
S. Malchenko[2], research associate
R. Woods[3], research associate
K. Whitworth[3], research associate
J.A. Green[3], research assistant professor
R. Prather[3], professor
C. J. Fitzsimmons[1], research associate
C. A. Roberts[4], staff scientist
T. Casavant[4], professor
M. B. Soares[2], associate professor

[1] Department of Animal Science
Iowa State University

[2] Department of Pediatrics
University of Iowa

[3] Department of Animal Science
University of Missouri-Columbia

[4] Department of Electrical and
Computer Engineering
University of Iowa

ASL-R668

Summary and Implications

One major problem that has high economic impact on pig reproduction is the unexplained loss of potential porcine conceptuses during the first month of gestation. To better understand when and how these losses occur, it is imperative to investigate the underlying genetic regulatory mechanisms. We have recently initiated a large-scale cDNA sequencing project to provide molecular information regarding the genes expressed in female reproductive tissues. cDNA libraries are planned for ovary, hypothalamus, pituitary, placenta, uterus, and several stages of embryonic development. Sequence information will also be highly useful in developing sequence-tagged sites for physical mapping and developing comparative links between the human, mouse, and pig genome maps. We have previously reported the creation of two cDNA libraries, porcine fetal (day 20), and conceptus (day 17). Sequencing of these libraries produced 220 Expressed Sequence Tags (ESTs), with 180 sequences analyzed by clustering algorithms, and 139 clusters identified within these sequences. We now report the creation of two more libraries from porcine fetal (day 45) and placental tissues. The day 45 fetal library has 971,150 independent clones (average insert: 1.4 kb), whereas the placental library has 1,320,000 independent clones. Initial sequencing of the fetal library has produced 119 ESTs (81 clusters), whereas we have obtained 1411 ESTs (1056 clusters) from the placental library. After clustering all sequences thus far obtained, we have identified 1,233 unique clusters. Sequences obtained in this project will be deposited into Genbank dbEST, and all comparative homology information will be summarized on a public Website.

Introduction

Advances in gene mapping and genomics of farm animals have been considerable over the past five years. First-generation linkage and physical maps have been developed with worldwide efforts by dozens of scientists. These maps have been instrumental in identifying specific chromosomal regions and candidate genes associated with traits of economic importance. However, current pig maps are neither of high enough resolution nor informative enough at the comparative level for practical breed improvement or for studying biological mechanisms underlying economically important traits. An improved understanding of porcine reproductive biology is of crucial economic importance. Yet reproductive processes are poorly characterized at the molecular level, and reproduction is very difficult to genetically improve by classical breeding methods.

One basic need in molecular reproduction studies is identification and sequence analysis thousands of genes associated with physiology of reproduction. To develop this data, cDNA libraries need to be created and characterized (Figure 1). The clones in those libraries can then be sequenced to provide the pig-specific genetic information to efficiently improve the pig comparative map (1) and to provide reagents to measure gene expression in reproductive tissues.

Materials and Methods

Poly-A+ mRNA was isolated from porcine tissue, converted to cDNA fragments and cloned into pT7T3-*Pac* as described (2) at the University of Missouri-Columbia. Plasmids from random colonies were isolated by using alkaline lysis or the Washington University microwave procedure, and the 3'-untranslated region of each clone was sequenced at the ISU DNA Sequencing and Synthesis Facility or at the University of Iowa. Trace files were then processed in the UI Sequencing Pipeline (http://ratEST.uiowa.edu/).

Results and Discussion

We have developed four cDNA libraries, three representing embryo/fetal genes expressed from three different stages of embryogenesis and one from term

placental tissues (Table 1). We have generated a total of 339 sequences from the three fetal/embryo libraries and 1,411 sequences from the placenta library. By using the BLAST algorithmn (3) as well as computer programs developed at the University of Iowa, we have identified 1,233 clusters (different gene sequences) from a total of 1,683 sequences from placenta, fetal/embryo library sequence analyses (Table 2). The clustering analysis shows that the majority of the clones sequenced thus far from these libraries are derived from different genes, as their sequences are different from each other (Cluster size of 1). Thus, the placental library is sufficiently complex for normalization and data mining. The embryo libraries and the fetal libraries need additional sequencing to establish their suitability for long-term data mining.

Of the current EST dataset, 60% of the sequences are novel relative to the pig EST database; and 36% are novel relative to the human EST database (Table 3). Thus, the sequences that match human genes can be used to develop primers to map new genes onto the comparative map between human and pig. Such sequenced genes will be selected using available comparative information such as porcine:human chromosome painting data (4). Further, the novel gene sequences may be interesting to investigate to understand their role in pig biology, as no information will be available from human studies.

Acknowledgments

This work was supported by a PROMAG award from Iowa State University, Office of the Vice-Provost; USDA-NRI grant 95-37205-2189, and by the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 3596.

References

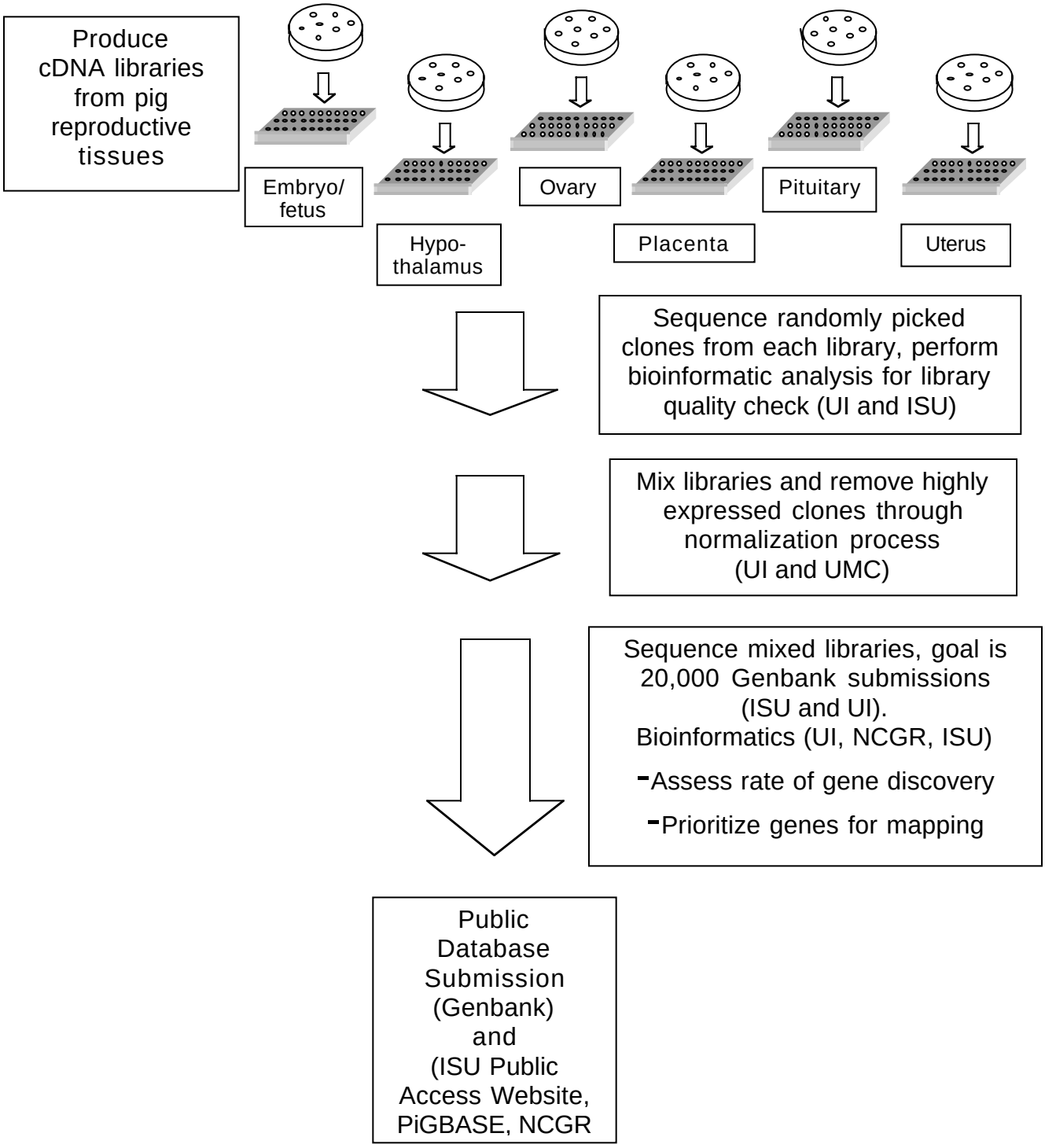
1. Tuggle, C. K. and C. B. Schmitz. (1994) Cloning and characterization of pig muscle cDNAs by an expressed sequence tag approach. Anim. Biotech. 5: 1–13.

2. Bonaldo M.F., Lennon, G., and Soares, M.B. (1996) Normalization and Subtraction: Two approaches to facilitate gene discovery. Genome Research 6: 791-806.

3. Altschul S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.

4. Goureau A., Yerle M., Schmitz A, J. Riquet, D. Milan, P. Pinton, G. Frelat, and J. Gellin. (1996) Human and porcine correspondence of chromosome segments using bidirectional chromosome painting. Genomics 36: 252–262.

Figure 1. Schematic showing expressed sequence tag (EST) project* approach toward goals of sequencing 20,000 genes from the pig.



*Plan for larger EST project which was funded by the USDA, project 99- 99-35205-8370 , awarded to Tuggle, C.K. (Project Leader), Prather, R., Soares, M.B., Casavant, T., Pomp, D., Beavis,W., Green, J., Rothschild, M., Day, B., Lamberson, W., Lucy, M.

Table 1. Pig cDNA library information.

Tissue Source	Number of primary Clones	Average insert size (bp)	Gene sequences
Day17/20 embryo	N.A.	N.A.	220
Day 45 Fetus	970,000	1,400	119
Placenta (Term)	1,320,000	1, 300	1,411

N.A. not available.

Table 2. Cluster size distribution of ESTs In current dataset of 1,683 sequences.*

Cluster Size	Frequency (No. of ESTs)
1	1068
2	100
3	25
4	13
5	5
6	6
7	1
8	3
10	2
11	1
12	2
14	2
16	2
18	1
25	1
38	1

Cluster is defined as those sequences that have high sequence identity, thus likely represent the same gene.
(Entire Data Set; n=1233 clusters,
1683 total sequences)
* Totals from Table 1 do not sum to 1683 as a small number were eliminated due to redundancy upon further analysis.

Table 3. Pig EST novelty rate relative to public pig or human ESTs.

Comparison to Pig ESTs		Comparison to Human ESTs	
Score 0 - 50	737	Score 0 - 50	444
Score 50 - 100	28	Score 50 - 100	147
Score 100 - 150	31	Score 100 - 150	126
Score 150 - 200	36	Score 150 - 200	88
Score 200 - 250	37	Score 200 - 250	74
Score 250 - 300	29	Score 250 - 300	78
Score 300 - 350	36	Score 300 - 350	57
Score 350 - 400	39	Score 350 - 400	41
Score 400 - 450	42	Score 400 - 450	34
Score 450 - 500	33	Score 450 - 500	28
Score 500 - 550	37	Score 500 - 550	30
Score 550 - 600	32	Score 550 - 600	32
Score 600 - 650	23	Score 600 - 650	19
Score 650 - 700	29	Score 650 - 700	17
Score 700 - 750	21	Score 700 - 750	8
Score 750 - 800	15	Score 750 - 800	4
Score 800 - 850	7	Score 800 - 850	2
Score 850 - 900	5	Score 850 - 900	3
Score 900 - 950	9	Score 900 - 950	0
Score 1000 +	7	Score 1000	1

Pig-Pig Novelty rate: 737/1233= 60%
Pig-Human Novelty rate: 444/1233= 36%

Cluster representative sequence was used in sequence alignments (alignment score is given above). A score <50 indicates no homology to genes in the database; while a score > 50 indicates possible homology to known genes in the database exists. Left side: BLAST results (3) only for pig ESTs were enumerated (text search for “scrofa” used as criteria; dbEST data as of July 16, 2000 used).
Right side: BLAST results only for human ESTs were enumerated. BLAST analysis against human ESTs in dbEST (July 16, 2000 database).