EVIDENCE FOR QUANTITATIVE TRAIT LOCI AFFECTING OVULATION RATE AND LITTER SIZE IN SWINE

by

Thomas A. Rathje

A DISSERTATION

Presented to the Faculty of

The Graduate College in the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Animal Science

Under the Supervision of Professor Rodger K. Johnson

Lincoln, Nebraska

November, 1995

DISSERTATION TITLE

Evidence for quantitative trait loci affecting ovulation rate and litter size in swine

BY

Thomas A. Rathje

SUPERVISORY COMMITTEE:

	APPROVED	DATE
Signature	James E. Kinder	11/15/95
Typed Name	James E. Kinder	
Signature	Gary Alan Rohrer	11/15/95
Typed Name	Gary Alan Rohrer	
Signature	Dale Van Vleck	11/15/95
Typed Name	Dale Van Vleck	
Signature	Rodger Johnson	11-15-95
Typed Name	Rodger Johnson	
Signature	John Osterman	11-15-95
Typed Name	John Osterman	
Signature	Merlyn K Nielsen	11-15-95
Typed Name	Merlyn K. Nielsen	

GRADUATE COLLEGE
UNIVERSITY OF NEBRASKA

EVIDENCE FOR QUANTITATIVE TRAIT LOCI AFFECTING OVULATION RATE AND LITTER SIZE IN SWINE

Thomas A. Rathje, Ph.D.

University of Nebraska, 1995

Advisor: Rodger K. Johnson

Improving reproduction in swine has been accomplished through selection for increased ovulation rate and litter size. Quantitative methodology relies upon estimating the cumulative effect of all loci in the genome. Identification of individual loci affecting reproduction will allow for increased accuracy of phenotypic selection. Fifty-eight microsatellite markers were scored in an F2 population of pigs developed at the University of Nebraska. The population was produced by crossing a line previously selected for ten generations using an index of ovulation rate and embryonic survival and a line selected at random. The lines were from a common base population and differed by 6.7 eggs ovulated and 3.1 fetuses at 50 d gestation following ten generations of index selection. Three replicates of F1 matings were made, but only the first replicate was used in this analysis (n=114 F2 females). Measurements were made for ovulation rate, number of fully-formed pigs and number of mummified pigs. Data were analyzed using the method developed by Haley et al. (1994) with a modification to include a random animal effect in the statistical model. Likelihood-ratio tests were performed to determine

significance of a quantitative trait locus (QTL) effect using the approximate method described by Haley et al. (1994) and by deviating the log-likelihood for the full model from the log-likelihood for the reduced model. A QTL for ovulation rate was found on chromosome eight (P<.001). Evidence was found that QTL affecting ovulation rate are on chromosomes 4 (P<.10), 13 and 15 (P<.05). These results were not significant for an experiment-wise threshold value of P<.001. Results for number of fully-formed pigs and number of mummified pigs were not consistent between the two methods of computing the likelihood-ratio making the interpretation of these results unclear. The addition of the second and third replicates will increase the F2 population size to approximately 400 females and should provide better evidence for presence of QTL affecting reproductive traits in this population.

Keywords: QTL, swine, ovulation rate, litter size

Acknowledgments

I would like to express my sincere gratitude to Dr. Rodger K. Johnson for his advice and guidance throughout my program. His influence upon my life goes far beyond my career, and I sincerely thank him for it. I would also like to acknowledge the other members of my committee: Dr. Merlyn Nielsen, Dr. Dale Van Vleck, Dr. Jim Kinder, Dr. Gary Rohrer and Dr. John Osterman. I especially want to thank Dr. Gary Rohrer who played an integral role in the success of my degree program. His patience and time spent training me in the skills and techniques required in the laboratory are greatly appreciated.

The crew at the swine research farm are greatly appreciated. If it was not for the efforts of Matt, Jeff, Darryll, Tom, Norm, Jane, Dan and Brian, this project could have never been completed. They were truly responsible for the day to day oversight of data collection and husbandry. I also want to thank the help of the laboratory technicians at the USDA Gene-Mapping laboratory. In addition, Dr. Deb Hammernik provided much instruction and insight during the early stages of this project.

The efforts of my fellow graduate students were also greatly appreciated. Joe Cassady and Brad Freking provided many hours of labor and frank discussion to help with the completion of this project.

I want to thank my wife, Kristi, whose patience, support and understanding were always present and always appreciated throughout my

program. She has given up much to allow me to pursue this degree and I have failed to thank her enough for the role she has played in my success.

Last of all, none of this would have been possible without the guidance, wisdom and insight provided by our Creator, His son Jesus Christ and the gift of the Holy Spirit. We truly cannot understand this world through the scientific method without believing in His Word and incorporating His work in our lives and careers.

Dedication

This dissertation is dedicated to my wife, Kristi, and our children, Katelynn and Grant. Their patience, love and understanding of this endeavor can never be repaid. Thank you.

Table of Contents

Introduction

Improving reproductive efficiency in swine has been accomplished by selecting for improved breeding value for reproductive traits in the female. Several traits have been examined as selection criteria: ovulation rate, fetal survival, litter size and age at first estrus. Evidence from Leymaster and Johnson (1994) indicates that, for the selection objective of increased litter size, the selection criteria of ovulation rate and fetal survival at parturition (litter size), as the criteria on which selection pressure may produce the greatest response. Breeding value is commonly defined as a quantitative measure of the cumulative additive effects of all loci in the genome which influence a quantitative trait. In contrast, the development of applicable techniques in molecular biology will allow for the identification of individual genotypes with desirable effects on reproductive traits. Information on loci found to have effects on traits of economic importance can be included in a marker-assisted selection program to increase accuracy of identifying superior individuals and thus enhance response to selection for increased breeding value.

The use of molecular biology techniques to isolate individual loci that affect economically important traits in pigs has been successful (Haley et al., 1994; Rothschild et al., 1994). Rothschild et al. (1994) utilized crosses between Chinese Meishan and American Duroc, Landrace and Yorkshire. Haley et al.

(1994) utilized a population of pigs resulting from a cross between the European wild boar and Large White breeds. To date, no studies have utilized lines of pigs that have originated from the same base population, but with continuous phenotypic selection resulting in differences among lines in phenotypic measurements of economically important traits. The successful application of marker-assisted selection to reproductive traits requires identification of loci with large effects in outbreeding populations. The lack of such loci may limit the application of molecular biology to introgression of desirable alleles versus aiding in the accuracy of phenotypic selection. The objective of the study described in this dissertation was to identify loci that contribute to differences in ovulation rate and litter size between a line selected for ten generations on an index of ovulation rate and embryonic survival and a randomly-selected control line derived from the same base population.

Literature Review

A. Construction of the Experimental Design.

The initiation of a project to detect quantitative trait loci involves assessment of the availability of applicable molecular biology techniques, development of an appropriate mating design, selection of genetic markers, the minimization of costs to insure a high return on the investment in research and the overall size of experiment required to insure adequate power for detecting quantitative trait loci. The following subsections address these issues.

I. Application of Molecular Biology Techniques Useful for Swine Breeding.

The development of molecular biology techniques required for development of genetic markers has led to production of genetic linkage-maps for the pig. Several such maps have been developed (Archibald et al., 1995; Ellegren et al., 1994; and, Rohrer et al., 1994). Microsatellite markers, primarily of the $(CA:GT)_n$ type, provided an abundant class of highly polymorphic and codominant genetic markers useful for development of these linkage maps. Microsatellite markers used in the study described in this dissertation were primarily derived from the linkage map described by Rohrer et al. (MARC, 1994). The procedures used to develop the MARC map were outlined in detail in Rohrer et al. (1994) and are summarized here: Porcine DNA was digested with MboI restriction enzyme and fragments corresponding to 200- to 500- bp were cloned into M13mp18RF DNA and transfected into *E. Coli.*. The resultant

plaques were screened with $(GT)_{11}$ and $(CA)_{11}$ oligonucleotides. Positive phage were grown and single-stranded DNA was extracted and sequenced. Primers were designed for PCR-based genotyping. Genotyping of markers was done for 104 animals from a two generation pedigree. The population included two sires and eight dams and the resultant 94 progeny. Linkage computations were performed using CRIMAP 2.4 (Green et al., 1990) Markers were placed into linkage groups and ordered by maximizing the log-likelihood from multiple point linkage analyses.

The map developed by Rohrer et al. (1994) included 376 microsatellite loci and seven Restriction Fragment Length Polymorphism (RFLP) loci. These 383 markers are in 24 linkage groups and covered approximately 1997 cM of the porcine genome. This map provided the basis for the genetic analysis of the inheritance of quantitative traits described in this study.

II. Structure of Populations Used for Mapping Quantitative Trait Loci.

Backcross and F2 populations are often produced for mapping quantitative trait loci. Both mating systems produce second generation progeny segregating for quantitative trait loci and marker alleles. Each progeny type is developed by crossing two genetically distinct populations which differ substantially in phenotype for a character under study. The F1 progeny of the cross of the two populations are intermated to produce the F2 (F1 x F1) or Backcross (F1 x Parent) generation. Second-generation progeny which contain segregating alleles are scored for a variety of genetic markers. A major

advantage of the F2 population is that approximately half the number of progeny are required to achieve the same power for detecting the effects of QTL as compared with a backcross design (Lander and Botstein, 1989). This is because twice the number of meioses are produced during the matings to create the F2 population as compared to backcross (due to the use of a recurrent parent). For fixed financial and physical resources, greater power can be achieved using the F2 population than the backcross. An F2 mating design also allows estimation of dominance effects for QTL (Haley et al., 1994). The quality of the F2 design can also be evaluated by considering family structure.

Elsen et al. (1994) examined the family structure of populations and its effect on minimizing the variance of recombination rate among markers. The criterion of minimizing variance can be utilized when setting up line crosses and ultimately, an F2 population. A population was defined as containing F sire families of D dams per sire and P progeny per dam. Therefore, the size of the population was M = F + FD + FDP. Estimations of recombination rate and associated precision is dictated by N = FDP, the total number of progeny. The results of their study showed that markers with a greater number of alleles provide more precise estimates of recombination rate than markers with only two alleles. Also, a mixture of full- and half-sib families was superior to strictly half-sib family designs by a factor of approximately two (variance of recombination rate was halved).

Elsen et al. (1994) suggest optimum designs for linkage analysis consist of a small number of reasonably large full-sib families, due partly to the use of the dam's genotype in estimating recombination rate. Also, the precision of the estimate of recombination rate is not much improved by full-sib families larger

than five progeny per dam. The authors suggested the use of the variance estimate of recombination rate as the equivalence criterion when applying their methodology to designing a population to be used for quantitative dissection of a trait versus linkage analysis. However, the approximation of this variance depends upon the number and frequency of marker alleles, the recombination rate between markers, and knowledge of sire and dam phases as determined by grandparent information. Each of these factors varies between marker intervals. Therefore, derivation of an optimum design with regard to family structure is difficult, if not impossible, when many intervals are being tested throughout the genome. Nevertheless, the general conclusions of these authors are that a population consisting of a mixture of half- and full-sib families with full-sib family size greater than five will provide the most precise estimates of recombination rate among linked markers.

van der Beek and van Arendonk (1993) examined the effect of individual family-types on the accuracy of linkage estimates. These authors examined the types of full-sib (or half-sib) families produced based upon the gametes originating from the parents. Because a set of families (i.e., a population) is used regardless of the suitability of those families for a specific set of markers, the overall quality of a design is affected by the frequency of the individual family-types.

Initially, van der Beek and van Arendonk (1993) determined seven classes of progeny: 1) two gametes of unknown type, 2) one gamete of unknown type, one non-recombinant, 3) one gamete of unknown type, one recombinant, 4) two non-recombinant gametes; 5) one non-recombinant and one recombinant gamete; 6) two recombinant gametes; 7) two gametes non-

recombinant or two gametes recombinant. Using these seven classes and parental phase, seven types of families are possible (Table 1).

Table 1. Gamete Type Inherited by offspring for the seven possible family types.[A]

	Type of gametes[B]						
Family type	Un, un	Un, non	Un, rec	Non, non	Non, rec	Rec, rec	2 non, 2 rec
1	X						
2	X	X	X				
3		X	X				
4					X		X
5				X	X	X	X
6				X	X	X	X
7				X	X	X	

[A]Reproduced from Table 3, van der Beck and Van Arendonk (1993)
[B]Definitions: Un = unknown gamete type; Rec = recombinant gamete; Non = non-recombinant gamete

The family types are the result of the following matings:

1) None of the parents heterozygous for both loci
2) Single backcross (e.g., A_1B_1/A_2B_2 x A_1B_1/A_1B_2)
3) Double or single backcross (e.g., A_1B_1/A_2B_2 x A_1B_1/A_1B_1 or A_1B_1/A_2B_2 x A_1B_1/A_1B_3)
4) Intercross with unequal phase (e.g., A_1B_1/A_2B_2 x A_1B_2/A_2B_1)
5) Intercross with equal phase (e.g., A_1B_1/A_2B_2 x A_1B_1/A_2B_2)
6) Intercross with equal alleles for a single locus (e.g., A_1B_1/A_2B_2 x A_1B_1/A_2B_3)
7) Intercross with one equivalent allele for each locus (e.g. A_1B_1/A_2B_2 x A_1B_1/A_3B_3)

Family type one, the cross between homozygous parents with equivalent alleles, is the least favorable as no information on linkage is provided. Family type seven is the most favorable as all progeny contribute to estimation of recombination rate. Families of type seven increase in frequency as the number of alleles at a marker locus increases.

Ellegren et al. (1994) found 63% of their markers (both microsatellite and RFLP) to be heterozygous in F1 parents. Rohrer et al. (1994) found heterozygosity levels ranging from 54.4% to 83.9% for various crosses of American and Chinese breeds of pig. These levels of heterozygosity indicate most microsatellite markers fall in the range of two to four alleles (van der Beek and van Arendonk, 1993). Given that a random set of markers from the previously listed linkage maps were chosen, most families would be of type one, two or three. According to van der Beek and van Arendonk (1993) two equivalent alleles would result in 56% of families to be type one and uninformative for mapping. With three and four alleles, type one families make up 31 and 19% of the population, respectively. Type three families would be the dominate family comprising 41 and 46% of the population for markers with three and four alleles, respectively. Type seven families are nonexistent for two-allele systems and make up 9% and 22% of the population for three or four equiprobable alleles. Family type three has approximately half the number of informative gametes as family type seven.

Another conclusion from van der Beek and van Arendonk (1993) was that parental phase as determined by genotyping grandparents, is worthwhile for recombination distances greater than 0.2 between loci. This was shown by increased accuracy of the estimate of recombination rate (e.g., $r = E(r)$, where r is the estimated recombination rate between markers). Comparisons of full- and half-sib family structures showed increased accuracy for a mixture of full and half-sib families. Their results also showed maximum accuracy for designs that minimize the number of families. However, designs which minimize the number of families also maximize the probability of having no information (e.g., type one

families) and therefore a minimum number of families is required to insure P(Type one family) does not exceed some threshold. For example, with two equiprobable alleles, four families are required to insure that the probability of no information is 0.1 (n = number of families = $-1/\log_{10}[P(\text{type 1 family})] = 1/\log_{10}(.56) = 4$).

III. Optimum Spacing of Genetic Markers for Detection of QTL.

In addition to designing a mating structure to maximize the probability of detecting QTL, the selection and spacing of genetic markers can also improve the opportunity for detecting QTL.

Darvasi and Soller (1994) examined the optimum spacing of genetic markers from the standpoint of minimizing total costs for a desired level of power (i.e., Type II error rate). The number of animals in an experiment can be considered to be a function of two components. The first h(r) is a function of recombination rate and equal to $1/(1-2r)^2$ for an F2 design. The second, K, is a function of QTL effect, error variance, and Type I and Type II error rates. The equation for K includes all other parameters determining experimental size except recombination rate. The parameter K is equal to: $8\sigma^2(Z_{\alpha/2} + Z_{\beta})^2/d^2$ where σ^2 is the population variance, Z is the ordinate of the normal curve for $\alpha/2$ and β, the Type I and II error rates, and d is the allele substitution effect in units of standard deviation. The above expressions were derived from the following result from Soller et al. (1976) for an F2 design:

$$N = 8\sigma^2 \frac{(Z_{\alpha/2} + Z_{\beta})^2}{d^2(1-2r)^2}$$

where N is the number of animals in the F2 generation of the mating design.

Darvasi and Soller (1994) then derived a cost function for the experiment based upon the above parameters: $f(M) = K(e^{2M}-1)(GC + M)/2M^2$ where M is the marker spacing in morgans, C is the cost of collecting phenotypic data relative to the cost of genotyping individuals and G is total genome size. This equation was optimized for a given level of power (i.e., K) for the experiment. They showed that K varies little. As M decreases, the Type I error rate must be decreased (lower α) which would necessitate a corresponding increase in Type II error rate (higher β) to account for the increased number of markers. Type I error decreases because more markers in the vicinity of a QTL increase the likelihood of a significant result. Thus, f(M) can be calculated independent of K.

The results of Darvasi and Soller (1994) indicate that for a cost ratio, C, of 0.005 the optimum marker spacing is approximately 30 cM. These results correspond closely to the cost ratio for the study described in this dissertation (e.g. $1.50 cost per genotype versus $300 data collection cost per animal). The results obtained by Darvasi and Soller (1994) were for α=.05 and β=0.8.

IV. Experimental Size and Detection of QTL Effects.

One final aspect of the design process is to determine the size of experiment required to ensure an adequate probability of detecting QTL. The size of experiments is often limited by financial and physical resources. The ideal experimental size is often unattainable. For instance, the results of Soller

ideal experimental size is often unattainable. For instance, the results of Soller et al. (1976) indicate F2 populations approaching 1000 individuals are required to detect QTL of very small effect (e.g. 0.25 standard deviation effects). The population used for the experiment described in this dissertation will approach 425 F2 individuals following the completion of three replicates of F1 matings. As only the first replicate is used for this study, one can calculate the lowest QTL effect detected for the 130 animals used in the pilot study described in this dissertation from the formula of Soller et al. (1976), a 30 cM linkage map, α equal to .05 and β equal .8 so that:

$$d = \frac{\sqrt{8\sigma^2 (Z_{\alpha/2} + Z_{\beta})^2}}{N(1-2r)^2}$$

$$r = \frac{1 - e^{-2M}}{2}$$

$$= \frac{1 - e^{-2(.3)}}{2}$$

$$= -.23$$

σ^2 = 7.17 for litter size (total number born) and 19.71 for ovulation rate;

$Z_{\alpha/2}$ = the ordinate at $\alpha/2$ = .0584;

Z_{β} = the ordinate at β = .278

Resulting values for various levels of power and populations of 130 and 425 animals are presented in Table 2. The population size of 425 corresponds to the approximate size of the total F2 population when all three replicates are included.

Table 2. Results for the size of effect, in units of standard deviation, detectable at various levels of power and population size using phenotypic variance, with α=.05.

	Population Size			
	130		425	
Power	Ovulation Rate	Litter Size	Ovulation Rate	Litter Size
.8	5.73	3.46	3.17	1.91
.5	4.00	2.41	2.21	1.33
.3	2.92	1.76	1.61	0.97

Because the analysis of data for the experiment described herein will be done using an animal model, the phenotypic variance will be partitioned into additive genetic variance and error variance. The most likely QTL location and estimate of effect is determined by minimizing the residual error. Therefore, it is more appropriate to calculate the power of detection for a QTL effect based upon residual rather than phenotypic variance. Table 3 contains results using error variance. The error variance results from using a heritability for ovulation rate and total number born of 0.3 and 0.1, respectively. The resulting values for error variance are 13.8 for ovulation rate and 6.45 for total born.