

발간등록번호

11-1543000-000166-01

한국재래돼지 종돈확립을 위한 기초축군 조성과 불량형질 향상을 위한 분자육종기법 개발

(Establishment of resource population for the breeding
stocks and development of molecular breeding technique for
improvement of low quality traits in Korean native pigs)

충남대학교

농림축산식품부

제 출 문

농림축산식품부 장관 귀하

이 보고서를 “한국재래돼지 종돈확립을 위한 기초축군 조성과 불량형질 향상을 위한 분자육종기법 개발” 과제의 보고서로 제출합니다.

2013 년 07 월 29 일

주관연구기관명 : 충남대학교

주관연구책임자 : 이 준 헌

세부연구책임자 : 이 준 헌

연 구 원 : 디아마하라니

연 구 원 : 정 우 영

연 구 원 : 서 동 원

연 구 원 : 진 실

연 구 원 : 정 은 지

연 구 원 : 유 채 경

연 구 원 : 이 재 봉

연 구 원 : 이 상 훈

협동연구기관명 : 환경대학교

협동연구책임자 : 이 득 환

연 구 원 : 윤 우 정

연 구 원 : 이 준 호

연 구 원 : 조 충 일

연 구 원 : 박 성 식

연 구 원 : 최 민 규

연 구 원 : 손 지 현

협동연구기관명 : 국립축산과학원

협동연구책임자 : 고 문 석

연 구 원 : 조 인 철

연 구 원 : 박 용 상

연 구 원 : 한 상 현

연 구 원 : 양 성 룡

연 구 원 : 배 종 하

연 구 원 : 장 권

연 구 원 : 김 호 균

연 구 원 : 고 기 석

연 구 원 : 김 희 수

연 구 원 : 김 민 우

연 구 원 : 김 판 규

연 구 원 : 강 용 준

연 구 원 : 강 시 내

연 구 원 : 김 상 금

요 약 문

I. 제 목

한국재래돼지 종돈확립을 위한 기초축군 조성과 불량형질 향상을 위한 분자유종기법 개발

II. 연구개발의 목적 및 필요성

1. 연구개발의 목표

본 연구의 최종 목표는 한국재래돼지 종돈의 계통확립 기반 조성 및 불량형질 개선을 위한 유전자기반 선발방법의 개발로 한국재래돼지 생산성 향상 및 보급용 종돈의 시범 생산이다.

2. 연구개발의 필요성

돼지고기는 대한민국 국민이 가장 즐겨하는 단백질 공급원이며 특히 제주도의 재래흑돼지는 유전자원으로서의 가치가 매우 높다. 그러나 제주재래흑돼지의 단점으로는 성장이 개량종에 비해 60% 수준이고 유두수는 랜드레이스에 비해 평균 1.8개가 작으며, 산자수는 평균 6-7두, 평균 육성률이 80-85%로 포육능력이 좋지 않다. 따라서 제주재래흑돼지의 불량형질을 개량하기 위하여 전통적인 통계육종방법에 분자유전학 기법을 적용한 제주재래돼지의 생산성 향상에 대한 연구가 필요한 실정이다.

III. 연구개발 내용 및 범위

1. 연구개발의 내용

- 한국재래돼지 종돈의 계통확립 기반 조성
- 한국재래돼지 종돈의 성장 및 포육능력 관련 표현형 분석 및 분자표지 마커의 대량 발굴 및 분석
- 시험 축군에 대한 60k chip 결과 분석을 통한 양적형질 유전자 연관 SNP 확보
- 확보된 성장 및 포육능력 QTL관련 SNP를 이용한 원인 SNP 확보 및 산업화 기반 조성

2. 연구개발의 범위

- 표현형 자료에 근거한 종돈의 선발 및 유지
- 성장 및 포육능력 관련 형질 분석 및 Illumina 60k chip을 이용한 SNP 마커 대량 분석
- 한국재래돼지의 맞춤형 모형개발 및 MBV(Molecular Breeding Value) 추정 알고리즘 개발
- 성장 및 포육능력 관련 형질 연관 마커 효과 추정
- 형질관련 SNP를 이용한 한국재래돼지 종돈의 선발 및 이를 이용한 산업화 기반 마련

IV. 연구개발결과

1. 한국재래돼지의 성장 및 포육능력관련 원인 유전자 변이 탐색

유두수 형질 (전체 유두수, 왼쪽 유두수, 오른쪽 유두수)과 Porcine SNP 60K beadchip의 데이터를 이용하여 GWAS 분석을 실시하였다. Genome-wide 임계수준을 bonferroni ($P = 1.61 \times 10^{-6}$) 방법을 이용해서 결정 했을 때, SSC7에서 유의성있는 38 SNP marker들을 발견하였다. 이중 strong positional candidate gene 5(HDDC3, LOC100156276, LOC100155863, ANPEP, SCAMP2)개를 선발하였다. 전체 유두에서는 SCAMP2 유전자의 g.25280 G>A (2.0×10^{-12}), 왼쪽 유두수에서는 HDDC3 유전자의 g.1319 G>A (2.3×10^{-7}), 오른쪽 유두수에서는 SCAMP2 유전자의 g.14198 G>A (4.7×10^{-12})에서 가장 유의성이 높았다.

또한 흉추수 형질 association 분석 결과 SSC7에서 유의성 있는 SNP marker들을 발견하여, ANPEP 유전자의 g.14837 C>T (4.37×10^{-12})의 유의성을 보였다. 유두수 형질과 흉추수(갈비대수) 형질과의 밀접한 관련이 있다고 생각되어 진다.

성장 관련 형질에 대한 association 분석을 140일령 체중과 도체중 형질에 대하여 분석하였다. SSC12번에서 각각 43 SNP marker와 12 SNP marker들을 발견하였다. 이 marker 중 이 두 형질에 대하여 overlap 되는 marker를 선정하여 2개의 유전자(LOC100621652, LOC100523510)를 후보 유전자로 선택하였다. LOC100621652 c.336A>G(I112M)에서 140일령 체중(6.48×10^{-9}), 도체중(4.04×10^{-8})의 유의성 있는 결과 값을 얻었다. 또한 LOC100523510 g.24 C>T에서 140일령 체중(2.89×10^{-9}), 도체중(4.08×10^{-8})의 유의성 있는 결과 값을 얻었다. 따라서, 본 연구의 랜드레이스와의 한국재래돼지 F2 집단을 이용하여 각 형질과 association을 수행한 결과 도체중 및 유두수 등 성장·번식 개량을 위한 선발 및 교배 실험을 통한 한국재래돼지의 개량에 유용하게 활용될 수 있을 것으로 사료되어 진다.

2. 한국재래돼지의 성장 및 포육능력관련 마커를 이용한 유전능력 평가 및 선발체계 확립

제주난지시험장 내 시험축으로 순종돈 36두(재래돼지(K): 19두, 랜드레이스(L): 17두)이며 이들간 교배를 통해 F1세대 91두(KL: 43두, LK: 48두)를 생산하였고, F1간 교배를 통해 F2세대 (KLKL: 31두, KLLK: 585두, LKKL: 489두)를 생성하였으며 이들 축군에 대해 번식자료, 성장자료 및 도축자료를 수집하여 연구에 이용하였다. 통계분석모형 설정을 위한 환경요인 분석 결과 유두수 형질의 경우 모산차와 검정차수가 유의하게 영향을 미치는 것으로 나타났으며($p < 0.05$), 성장형질인 체중에서 시기별 차이를 보였으나 전반적으로 교배품종, 성별, 모산차, 검정차수의 요인이 영향을 미치는 것으로 나타났고($p < 0.05$), 도체형질과 관련된 형질의 경우 검정차수와 도축시 일령의 효과가 다른 환경요인에 비해 보다 유의하게 영향을 미치는 것으로 분석결과 나타났다. 상기의 총 1,232두의 검정두수에 대하여 유전체 전장의 단일염기다형(SNP)을 분석하였다.

SNP을 통한 유전체 정보는 Illumina Porcine 60k DNA chip을 이용하여 수집하였다. 수집된 SNP자료에 대하여 이상치에 대한 품질평가(QC)를 실시한 결과, 18개 염색체 내 52,574개의 SNP마커가 존재하였고, 이 중 10%이상 결측된 표지인자가 1,830개, 시험축군에서 다형성이 없는 표지인자가 6,349로 나타나 이들을 제거한 후 44,395개의 유전정보를

이용하였으며 이들 SNP마커간 평균 맵 거리는 47.8kb로 나타났다.

집단의 유전적 다양성에 대한 지표인 이형접합형 및 다형성정보량에 대해 염색체별 평균을 계산해 본 결과, 각각 0.38 및 0.37로 나타났으며, 근친도(F-statistics)는 -0.03로 동형접합체에 대한 기대치와 관측치가 유사한 것으로 나타났다.

또한 모든 SNP마커간 조합을 통하여 연관불평형(LD)를 계산한 결과, 0.1 Mb이하인 것이 95,189개 조합이 존재하였으며, 이들의 r^2 값은 0.318였고, 0.1-0.2 Mb 거리에 있는 SNP 마커 조합이 98,025개로 이들의 r^2 는 0.259로 상당히 높은 연관 불평형의 값을 보였다.

모든 SNP마커간 조합으로 만들어진 r^2 값을 이용하여 유효집단의 크기를 추정된 결과 현재 유효집단의 크기(1 - 5세대)가 19두로 추정되어 초기 순종돈 집단의 적었던 현재 상황과 유사한 결과를 보였다.

MBV추정의 효율성을 검증하기 위한 일환으로 개체간 유전적 유사도를 혈연정보 및 유전체 정보를 이용하여 추정된 유전체 육종가 추정치(Genomic estimated breeding values: GEBV)를 얻기 위한 방법으로 GBLUP방법을 이용하여 비교 분석하였다. 또한 조밀한 SNP마커를 이용하여 SNP와 QTL간 연관불평형의 크기로 설명되는 효과를 통해 개체의 유전능력을 예측하는 Bayesian접근방법을 이용하여 통계 프로그래밍을 실시하였다.

유전체 분석 프로그램의 검증을 위해 QMSim 프로그램(Sargolzaei and Schenkel, 2009)을 이용하여 모의시험 자료를 생성한 후 개발된 프로그램을 활용한 MEBV 및 GEBV를 추정하였다. 본 연구의 시험축군과 유사한 형태로 가상의 유전체 정보, 혈통정보, 표현형 정보 생성하였고 이를 정보를 활용하여 개발된 프로그램을 통해 추정하였다.

분석 결과, 참조집단에서 BayesB가 0.735의 정확도로 GBLUP(0.706), BayesA(0.56), BLUP(0.56)에 비해 높게 추정되었으며, 참조집단도 마찬가지로 BayesB(0.493), GBLUP(0.471), BayesA(0.387), BLUP(0.114) 순서로 정확도가 높게 나타났다. 이들 연구 결과는 타 연구와 비슷하게 유전체정보를 통해 추정하였을 때, 그렇지 않는 BLUP방법에 비해 높은 결과를 얻었으며, 특히 이들의 정확도 편차는 참조집단 축군에서 더 큰 것으로 나타났다.

Bayesian접근방법을 통하여 각 형질별 또는 성장단계별 효과를 추정하였으며, 각 형질들의 특성(양적형질, 질적형질)에 따라 다양한 분포의 특성을 보이는 것으로 나타났다. 또한 개별 SNP마커 추정 없이 직접 개체의 유전능력을 평가하는 GBLUP과 기존의 혈통정보를 통해 개체의 유전능력을 추정하는 BLUP간 MBV정확도를 비교 분석한 결과, 작게는 GBLUP이 BLUP에 비해 0.03 높게 나타났으며 많게는 0.19의 정확도 향상효과가 발생하는 것으로 나타났다(성장형질). 유두수에서도 이는 비슷하게 나타났으며 GBLUP을 통한 MBV추정시 0.116 - 0.118의 정확도 향상효과가 있는 것으로 분석되었다. 유전체자료를 이용한 선발 육종체계는 순종의 종축집단(F0)과 실용축집단(F2)으로 나누어 활용할 수 있다. 우선 순종의 경우 추정된 마커 또는 유전적 혈연적 유사도(G)를 이용하여 종축이 되고자 하는 후보돈에 대한 선발이 가능할 것이며 이는 종축으로써 후대에 상당한 가치를 창출할 것이다. 또한 실용축집단에서 MBV를 이용한 개체선발은 검정이 완료되기 전에 개체를 선발함으로써 사료비 및 기타 생산비에 대한 절감효과가 기대된다고 할 수 있다.

3. 한국재래돼지의 성장 및 포유능력을 위한 축군조성 및 산업화 기반 조성

돼지의 척추수에 있어서 흉추와 요추 수는 고정이지 않는 형질로 품종간 또는 개체

간의 차이가 심한 것으로 알려져 있다. 본 연구는 제주재래돼지와 랜드레이스 두 품종 간 교배집단 F2세대 자손에서 경추 수, 흉추 수, 요추 수, 척추 수 등 체형관련 형질과 도체 형질의 상관관계를 살펴보기 위해 수행되었다. F2세대 자손에서 도축 직후 조사한 결과 경추 수의 변이는 나타나지 않았으나, 흉추와 요추 수 및 척추 수는 상이한 양상을 보였다. 도체형질의 성적과 요추 수만이 세 축군들 사이에서 통계적인 유의성을 나타내지 않았고($p>0.05$), 도체중, 육색, 근내지방도, 등심단면적, 등지방두께, 도체장, 흉추 수, 척추 수 등은 통계적인 유의성을 나타내었다($p<0.05$). 제주재래흑돼지는 흉추 수 14-16 개, 요추 수 5-6 개, 척추 수 27-29 개의 범위, Landrace는 흉추 수 15-16 개, 요추 수 5-7 개, 전체 척추 수 28-29 개, F2교배집단은 흉추 수 14-17, 요추 수 5-7 개, 전체 척추 수 27-30 개로 확인되었다. F2교배집단에서 흉추 수와 척추 수의 증가는 도체중, 도체장, 등지방두께의 평균이 유의적으로 다른 양상을 보인다($p<0.05$). 또한 척추 수의 증가는 요추 수의 증가보다 흉추 수의 증가에 더 많은 영향을 받는 것으로 나타났다. 흉추 수 및 척추 수의 증가가 등지방두께를 점점 두껍게 만들기는 하지만, 도체중과 도체장 등 경제성을 결정하는 핵심적인 도체형질을 향상시킨다는 점은 추후 제주재래돼지나 랜드레이스 품종을 활용한 양돈산업에서 체형개선을 통한 생산성 증가 특히 한국인이 선호하는 삼겹살량 증가로 농가 소득 향상에 좋은 전략이 될 것으로 판단된다.

V. 연구성과 및 성과활용 계획

- 형질관련 마커의 적용을 통한 한국재래돼지 종돈의 선발 및 유지
- 한국재래돼지 및 합성돈 사육농가에 종돈 보급 및 생산체계 기술이전 실시
- 한국재래돼지에서 유전체 육종가를 활용한 교체축 선발 및 도태계획 수립
- 개발 기술 및 연구결과의 국내외 전문 학술지 발표로써 과학대중화 유도
- 한국재래돼지의 국가적 유전자원 활용성 증대 및 국제적인 인지도 향상
- 국제수준의 분석 기술 및 유전자 표지인자에 대한 기술 선점

SUMMARY

(영문 요약문)

I. Title

Establishment of resource population for the breeding stocks and development of molecular breeding technique for improvement of low quality traits in Korean native pigs

II. Objective, Necessity of Research and Development

1. Objective

The final objective of this research is to establishment of the breeding stocks and development of the molecular breeding techniques for the improvement of low quality traits in Korean native pigs. Ultimately, the final goal is to improvement of the productivity and pilot production of breeding stocks using Korean native pigs.

2. Necessity

Pork is the most popular protein source in Korea and the Korean native pigs in Jeju island have very high reputation as the animal genetic resources. However, there are a number of limitation in this breed, which are the low growth rate (about 60% of the commercial breed) and small number of teats (1.8 less teat than Landrace), small litter size (average litter size of 6-7 pigs), less rate of raising (80-85%). Therefore, to improve the low quality traits in Korean native pigs, it is necessary to develop the method using both molecular and conventional breeding techniques in order to improve productivity in Korean native pigs.

III. Contents and Range of Research and Development

1. Contents

- The research for the breeding stock line establishment using Korean native pigs
- Phenotype analysis for growth, mothering ability in Korean native pigs and massive development and analysis of molecular markers
- Identifying linked SNPs for the quantitative traits using 60K chip analysis
- Identification of the causative SNPs for the QTL in growth and mothering ability and establishment of industrialization using these markers

2. Ranges

- Selection and maintenance of breeding stocks based on the phenotypic data

- Trait analysis for growth, mothering ability and massive development of SNP markers using Illumina 60K chip
- Customized model development for the Korean native pigs and Estimation algorithm development for MBV (Molecular Breeding Value)
- Estimation of molecular marker effects for the growth and mothering ability-related traits
- Selection of breeding stocks using trait-related SNPs and establishment of industrialization using these markers

IV. Results

1. Identification of the causative mutations for the growth and mothering ability in Korean native pigs

A. Identification of the causative mutations for teat numbers in Korean native pigs: Genome-wide threshold was determined using the Bonferroni method ($P = 1.61 \times 10^{-6}$), thirty-eight significant SNP markers were discovered in SSC7 in the three teat number traits (total teat number, left teat number, right teat number). Among these, five strong positional candidate genes (HDDC3, LOC100156276, LOC100155863, ANPEP, SCAMP2) were selected in these significant markers. The significant results obtained were g.25280 G>A (2.0×10^{-12}) SNP in SCAMP2 gene for total teat number, the g.1319 G>A (2.3×10^{-7}) SNP in HDDC3 for left teat number, and the g.14198 G>A (4.7×10^{-12}) SNP in SCAMP2 for right teat number. These results can be effectively used for the selection of animals with desirable teat numbers.

B. Identification of the causative mutations for number of vertebrae in Korean native pigs: The significant SNP markers were discovered in SSC7 in thoracic trait. The obtained significant SNP was g.14837 C>T (4.37×10^{-12}) in ANPEP gene. The results also indicated that thoracic (rib number) number may affecting teat number traits each other.

C. Identification of the causative mutations for body weight in Korean native pigs: The 140 day weight (d140_wt) and carcass weight (car_wt) growth-related trait association analysis were analyzed. Forty-three (d140_wt) and twelve (car_wt) significant SNP markers were discovered in SSC12. Two selected markers were from candidate genes (LOC100621652, LOC100523510) and these two genes are significant in two traits. The significant results obtained were c.336A>G (I112M) (6.48×10^{-9} , 4.04×10^{-8}) SNP in LOC100621652, and significant results obtained were g.24 C>T (2.89×10^{-9} , 4.08×10^{-8}) SNP in LOC100523510 for d140_wt, car_wt.

Therefore, this study investigated the GWAS for carcass weight and teat number traits in F2 population using the crosses between Landrace and KNP. The SNP

markers can greatly applied for improving carcass weights and teat numbers in Korean native pigs with further verifications.

2. Evaluation of genetic ability using genetic markers for the traits in growth and mothering ability and establishment of selection systems using Korean native pigs

A. Genetic characteristics of Purebred population using Genome-wide SNP information: For this study, Genome-wide SNP information was gotten by using Illumina Porcine 60k DNA chip from 1,232 animals (19 Korean Native pigs (K), 17 Landrace purebreds (L), 91 crossbred of their purebreds, and 1105 F2 crossbreds). After getting rid of outliers using by quality control of SNP, 52,574 SNP markers across 18 chromosomal region except sex chromosome, were identified. Among these markers, the number of markers with over 10% missing genotypes were 1,830 and the number of monomorphic markers were 6,349. After removing these marker, 44,395 markers were retained and used for association study and/or estimating genomic breeding values. The average distance between flanking markers was estimated of 47.8 kb. The estimates of heterozygosity and polymorphic information criteria with respect of indication on genetic diversity were average of 0.38 and 0.37 across of chromosomes, respectively. The observed homozygosity by way of inbreeding coefficients via F-statistics was little different with the expected homozygosity with -0.03 . From results of linkage disequilibrium between SNP markers, 95,189 marker combinations were less than 0.1Mb and average of r^2 on these combinations were estimated of 0.318. The number of effective population using LD was estimated of 19.

B. Algorithms and computing program for estimating marker breeding values: Marker breeding values and genomic breeding values were programmed based on genetic similarity between individuals using by markers and pedigree informations. Using linkage disequilibrium between markers and quantitative trait loci, several Bayesian approach via Bayes A, B, C, and C+ were developed and programmed for estimation of breeding values of animals.

C. Simulation for finding bug of program codes and accuracy of estimated marker breeding values: For testing our computer program, we estimated marker breeding values by our program using simulated data which were generated by QMSim (Sargolzaei and Schenkel, 2009) computer program and compared our estimates with parameters. From our results on reference population, the accuracies of breeding values were highest on BayesB with 0.735. Subsequently, same estimates on GBLUP, BayesA and BLUP was 0.706, 0.56, and 0.56, respectively.

D. Estimate marker effects and genomic breeding values: Marker effects on growth traits were estimated using Bayesian approach. From the results, distribution properties for genetic effects were shown of diversity according to the considered traits. According to comparing genomic estimated breeding values (GEBV) which were estimated based on pedigree and genomic information under linear regression function with maximum likelihood approach with GEBV with Bayesian approaches via estimating marker effects, The accuracy of GEBV for growth traits were high of 0.03~0.19 than those of estimated breeding values using only pedigree information. From these results, sows and boars can be selected based on GEBV rather than traditional estimated breeding values with higher accuracy of 0.12.

3. Establishment of Korean native pig population for the improvement of growth/mothering ability and establishment of industrial application

The number of thoracic and lumbar vertebrae has been known as un-fixed traits among mammals. This study was focused on the relationship between the numerical variations of the cervical (CER), thoracic (THO), lumbar (LUM), and total vertebrae (TNV) and the carcass traits in the Jeju Native Black pig (JBP), Landrace and their intercrossed F2 population. There were no numerical variations in CER. On the other hand, THO, LUM, and TNV were varied in all the three populations. Traits investigated within the three populations, only the means \pm SE of LUM had not showed the statistical significance ($p>0.05$), but the others including carcass weights (CW), meat color (MC), marbling score (MS), backfat thickness (BFT), carcass length (CLE), THO, and TNV showed statistical significance ($p<0.05$), respectively. JBP had 14-16 THO, 5-6 LUM, and 27-29 TNV, Landrace had 15-16 THO, 5-7 LUM, and 28-29 TNV, and F2 population had 14-17 THO, 5-7 LUM, and 27-30 TNV. In the F2 population, the increased numbers of THO and TNV indicated the significant increase of the levels of CW, CLE, BFT ($p<0.05$). Especially, the increase of TNV was caused to the increase of THO rather than that of LUM. Even the animals increased the THO/TNV had thicker level of backfat, they had longer CLE and heavier CW. Both traits of these are more economically critical and important to determine the productivity than backfat level. These results suggested that it may be one of great strategies for improving the productivity throughout genetic selection of the numerical increase of vertebrae, especially in Landrace, JBP and their related populations.

V. Research Outcome and Utilization Plan

- Maintenance and selection of breeding stocks for the Korean native pigs using trait-related marker application
- Supply the breeding stocks to the farms and technology transfer the production system
- Selection of replace animals and establishment of culling strategy for the Korean native

pigs

- Publish the research results to scientific journal at national and international levels
- Improve the utilization of Korean native pigs as the genetic resources and enhancement of the international reputation
- The prior occupation of international level of DNA marker analysis techniques

CONTENTS

(영 문 목 차)

Chapter 1 Summary of Research Project

Clause 1 Necessity of Research

Clause 2 Objective, Contents, and Ranges of Research

Chapter 2 Outlook of Technology Development

Clause 1 Domestic Status of Technology Development

1. Livestock Improvement using Genomic Information

2. Domestic Application of Genomic Selection Technique

Clause 2 International Status of Technology Development

1. Livestock Improvement using Genomic Information

2. International Application of Genomic Selection Technique

Chapter 3 Contents and Results of Research

Clause 1 Identification of the causative mutations for the growth and mothering ability in Korean native pigs

1. Techniques for the SNP chip analysis

2. Criteria for the genome wide association analysis using Plink and QxPak program

3. Comparison of the results from Plink and QxPak programs

4. Selection of significant SNPs and candidate genes using GRAMMAR

5. Association analysis between candidate genes and traits

6. Reestablishment of the database and selection of candidate genes after the publication of porcine genome sequencing

Clause 2 Evaluation of genetic ability using genetic markers for the traits in growth and mothering ability and establishment of selection systems using Korean native pigs

1. Development of statistical analysis model for breeding stocks of the Korean native pigs

2. Characterization of genetic structures in the population using genomic information

3. Development of MBV estimation algorithm and electronic program development

4. Estimation of genetic variation for the economic traits in breeding stocks using MBV program
5. Verification of MBV estimation program and accuracy measurement using genomic simulation study
6. Estimation of SNP marker effects and individual genetic ability

Clause 3 Establishment of Korean native pig population for the improvement of growth/mothering ability and establishment of industrial application

1. The characteristics of Korean native pigs and improvement of low quality traits
2. The phenotype measurement for teat numbers and its inheritance
3. The weekly body weights of the Korean native pigs
4. The teat number-related QTL region and candidate genes in Korean native pigs
5. The relationships between number of vertebrae and carcass length/carcass weight in Korean native pigs
6. The genotype effects of the candidate genes for the number of vertebrae QTL and population development using the genotypes

Chapter 4 Level of Achievement and Contribution to Related Area

Chapter 5 Research Outcomes and Plan to Utilization

Chapter 6 International Science and Technology Information during Research Progress

Chapter 7 Research Facilities and Equipments

Chapter 8 References

목 차

제 1 장 연구개발과제의 개요

- 제 1 절 연구의 필요성 및 목표
- 제 2 절 연구 목적, 내용 및 범위

제 2 장 국내외 기술개발 현황

- 제 1 절 국내 기술개발 현황
 - 1. 유전체 정보를 이용한 가축 개량기술
 - 2. 유전체 선발기술의 국내 적용현황
- 제 2 절 해외 기술개발 현황
 - 1. 유전체 정보를 이용한 가축 개량기술
 - 2. 유전체 선발기술의 해외 적용현황

제 3 장 연구개발수행 내용 및 결과

- 제 1 절 한국재래돼지의 성장 및 포육능력관련 원인유전자 변이 탐색
 - 1. SNP chip 분석 방법
 - 2. Plink와 QxPak을 이용한 Genome wide association 분석 기준
 - 3. Plink와 QxPak Program을 이용한 분석 결과의 비교
 - 4. GRAMMAR를 이용한 유효 SNP 및 후보 유전자의 선발
 - 5. 후보 유전자와 형질간의 연관성 분석
 - 6. 돼지 genome sequencing 발표 후 database의 수정 및 후보유전자 재탐색
- 제 2 절 한국재래돼지의 성장 및 포육능력관련 마커를 이용한 유전능력 평가 및 선발체계 확립
 - 1. 국산 재래돼지 종축의 맞춤형 통계분석모형 개발
 - 2. 유전체 정보를 활용한 집단유전학 특성 구명
 - 3. MBV추정 알고리즘 개발 및 전산프로그램 개발
 - 4. MBV 프로그램을 이용하여 돼지 종축의 각 경제형질에 대한 표지인자의 유전변이 추정
 - 5. 유전체 모의실험을 통한 MBV 추정 프로그램 검증 및 정확도 측정
 - 6. SNP Marker별 효과추정 및 개체의 유전능력추정
- 제 3 절 한국재래돼지의 성장 및 포육능력을 위한 축군조성 및 산업화 기반 조성
 - 1. 한국재래돼지의 특성 및 불량형질의 개선
 - 2. 한국재래돼지의 유두수에 대한 표현형과 유전양상
 - 3. 한국재래돼지의 주령별 체중변화

4. 한국재래돼지의 유두수 관련 QTL 영역 및 후보 유전자 탐색
5. 한국재래돼지의 갈비뼈 수와 도체장, 도체중의 상관관계
6. 한국재래돼지의 갈비뼈 수 관련 QTL 영역 내 후보유전자형의 효과와 이를 이용한 축군 조성

제 4 장 목표달성도 및 관련분야에의 기여도

제 5 장 연구개발 성과 및 성과활용 계획

제 6 장 연구개발과정에서 수집한 해외과학기술정보

제 7 장 연구시설·장비 현황

제 8 장 참고문헌

제 1장 연구개발과제의 개요

제 1절 연구개발의 필요성 및 목표

1. 연구개발의 필요성

가. 경제적 측면

- (1) 돼지고기는 대한민국 국민이 가장 즐겨하는 단백질 공급원(2008년 1인당 돼지고기 소비량 19.1 kg, 2009 농림수산물식품 주요통계)으로 전체의 육류 소비중 53%를 차지하고 매우 경제적 부가가치가 높은 산업으로 농가 소득 향상에 높은 기여를 하고 있음
- (2) 한·EU FTA 및 수입 축산물과의 경쟁을 위한 국내 양돈 산업의 차별화가 절실히 필요함
- (3) 돼지고기의 고급화 일환으로 육질이 우수한 흑돼지를 이용한 고급화 전략이 제주를 중심으로 일어나고 있으며, 흑돼지 비육농가가 점점 증가하고 있음
 - (가) 제주 흑돼지 사육농가 (111호) : '08년도 대비 '09년도 53% 증가
 - (나) 제주 흑돼지 사육두수 (61,047두) : '08년도 대비 '09년도 98.9% 증가 (전체 흑돼지 점유율: 12%)
- (4) 제주도에서만 생산되는 흑돈 돈육은 6,000여톤 생산되는 것으로 추정되며, 금전적으로 환산할 경우 년 280억 규모의 시장이 형성되고 있음.
- (5) 제주재래흑돼지의 장점은 먼저 고기내 조지방 함량이 평균 7% 이상으로 수입종에 비해 2-3배 높고, 육색중 적색도를 나타내는 CIE a(적색도)가 평균 13.3으로 쇠고기 수준의 육색을 나타냄.
- (6) 제주재래흑돼지의 단점으로는 성장이 개량종에 비해 60% 수준이고, 특히 비육후기에 체성장보다는 등지방이 지나치게 두꺼워지는 경향이 있고, 특히 일부개체의 경우 고기와 고기사이의 근간지방이 많이 축적되어 가식부위의 고기량이 적어 규격돈 생산이 힘들어 상품성이 떨어지고 있음. 또한 현재의 등급판정기준에서 육량지수는 도체중과 등지방을 고려한 등급기준에서 과대한 등지방 두께로 인하여 대부분 D등급을 받고 있는 실정임
 - (가) 제주재래흑돼지 20주령 평균체중 : 60 kg (개량종 대비 60% 성장률)
 - (나) 도축시 평균 등지방두께 : 50 mm, - 육량지수 도체등급 : C, D등급
- (7) 또한 제주재래흑돼지의 유두수는 랜드레이스에 비해 평균 1.8개가 작으며 제주재래흑돼지의 산자수는 평균 6-7두, 평균 육성률이 80-85%로 포육능력이 좋지 않음
- (8) 따라서, 제주지역 대부분 흑돼지 사육농가의 경우 제주재래흑돼지는 경제성이 전혀 없기 때문에 순종생산보다는 도입종인 버크셔종, 또는 버크셔종과의 교잡종을 이용 흑돼지를 생산하고 있고, 도축시 모색이 흑색으로 보이는 개체를 “흑돼지”로 판정하고 있음
- (9) 따라서 한국재래돼지 품종 보호 및 우량형질 고정을 위해 한국재래돼지 종돈 개량이 절실히 필요한 실정으로, 성장부분은 갈비수 증가에 따른 한국인이 선호하는 삼겹살과 갈비 고기량의 증가를 예측할 수 있고, 포육능력과 산자수는 유두수 증가로 단점을 보완할 수 있는 바, 전통적인 통계육종 보다는 분자유전학 기법을 이용하여 단기간에 형질고정을 한다면 농가소득 증대에는 기여할 수 있음

나. 산업적 측면

- (1) 한국재래돼지는 랜드레이스, 요크셔, 듀록 등 외래품종에 비해 지방질이 단단해 육질이 우수하나 체구, 성장률, 산자수 및 포유능력이 떨어져 종돈으로 활용도가 낮음
- (2) 2008년도 종돈 수입두수는 1,262두로 약 32억 6천만원의 외화가 소비되었음(2009 농림수산식품 주요통계)
- (3) 제주재래흑돼지는 제주도 축산진흥원에서 1985년경에 농가에서 외모심사를 통해 재래돼지에 가장 가까운 개체를 수집하여 25년간 inbred line을 유지한 축군이며, 이 축군을 국립축산과학원 난지축산시험장에서 구입하여 사육하고 있고, 최근 혈통과 외모심사를 통해 품종등록을 한 바 있음.
- (4) 현재 시판되는 한국재래돼지는 버크셔와의 교잡종으로 “흑돼지”와 혼돈되어 시장의 혼란을 가져오고 있음
- (5) 따라서 한국재래돼지의 산업적 바탕을 마련하고 사육농가를 보호하기 위하여 개량(형질고정)은 매우 절실함
- (6) 국내 돼지의 개량을 위한 분자유종 마커 개발은 아직도 미미하며 돼지 농가에 도움을 줄 실용화 연구는 전무한 실정임
- (7) 농림수산식품부에서 기획하고 있는 “종자산업” 계획안 중 한국 고유의 유전자원을 이용하여 사업화를 할 수 있는 유일한 돼지 품종인 한국재래돼지의 산업화 기반조성이 절실한 실정임
- (8) 한국재래돼지를 이용한 경제형질관련 마커의 개발은 농촌진흥청 바이오그린21사업 등 이미 수행된 적이 있으나 NCBI에 한국재래돼지(Korean native pig)를 key word로 검색한 결과 유효 논문은 많지 않음. 따라서 본 과제를 통하여 한국재래돼지를 이용한 학문적 성과 또한 높을 것으로 예상됨

다. 학술적 측면

- (1) 미국을 포함한 축산 선진국의 경우 돼지의 유전체 연구에 막대한 예산을 투입하여 그 결과로 논문 및 산업재산권을 확보하고 있음. 특히 유전체 연구를 바탕으로 확보된 경제형질관련 마커들은 현재 PIC(Pig Improvement Company)를 포함한 육종회사에서 종돈 선발을 위한 마커로 이용되고 있는 실정임
- (2) 따라서 국제적으로 돼지의 경제형질관련 마커의 선발 및 산업화 경쟁은 앞으로 더 치열할 것으로 예상됨
- (3) 국내 돼지의 개량을 위한 분자유종 마커 개발은 아직도 미미하며 돼지 농가에 도움을 줄 실용화 연구는 전무한 실정임
- (4) 한국재래돼지를 이용한 경제형질관련 마커의 개발은 농촌진흥청 바이오그린21사업 등 이미 수행된 적이 있으나 NCBI에 한국재래돼지(Korean native pig)를 key word로 검색한 결과 유효 논문은 본 연구팀이 소속된 축산과학원 난지축산시험장 팀이 유일함. 따라서 본 과제를 통하여 한국재래돼지를 이용한 학문적 성과 또한 높을 것으로 예상됨

라. 기술적 측면

- (1) UN산하 FAO의 주도로 domestic animal diversity information system(<http://dad.fao.org/>)이 개발되었으며 재래가축에 대한 자료 확보 및 이를 이용한

연구가 전세계적으로 확대되고 있음.

- (2) 특히 최근 유전자 전쟁이라고 부를 정도로 세계 각국은 유전자 확보에 많은 관심을 가지고 있음. 한국의 농림수산물부에서도 “종자산업”을 차세대 cash cow의 모델로 설정하고 연구, 인력양성 등 다양한 지원을 기획하고 있음
- (3) 2002년 human genome project의 완성으로 주요 가축의 유전체 연구가 빠르게 진행되고 있어 현재 소와 닭은 genome sequencing이 완성되었으며 돼지의 genome sequencing은 진행중에 있으면 2010년 완성을 목표로 마지막 마무리 작업을 하고 있음. 따라서 한국이 본 연구를 통해 한국재래돼지의 genomic selection을 이용한 육종 계획을 수립한다면 이 분야의 국제적 주도권을 장악할 수 있으며 한국재래돼지의 종축을 수출할 수 있는 바탕이 마련될 것으로 확신함
- (4) 돼지의 genome sequencing이 완성될 경우 돼지의 유전자 분석 및 경제형질관련 유전체 연구 및 이를 이용한 산업화는 매우 빠르게 발전할 것으로 예측됨
- (5) 본 연구에서 성장형질의 향상을 위해 최선호 부위인 삼겹살, 갈비살의 생산량을 결정하는 주요 경제형질인 척추수 변이에 대한 분자 표지인자 개발 및 이를 이용한 마커도움선발이 필요함
- (6) 한국재래돼지의 모계 형질이 포육능력 향상을 위한 유두수의 증가가 필요함
- (7) 돼지의 산자수 경우 유전력이 낮아(0.2 미만) 표현형 데이터만을 이용하여 검정자돈을 선발할 경우 정확도가 매우 낮음(약 10% 미만)
- (8) 돼지의 산자수를 마커도움선발을 이용할 경우 60%이상의 높은 정확도로 개량의 효율을 극대화할 수 있음
- (9) 또한 각 형질에 대한 SNP 마커 효과 추정 이후 표현형 데이터 없이 자돈에 대한 조기선발 가능 및 세대간격 단축으로 인한 검정 비용의 절감효과가 기대됨
- (10) 따라서 본 과제를 통하여 개발된 분자유종기술을 이용하여 한국재래돼지의 중돈 선발 및 개량에 이용은 한국의 돼지 유전 육종분야에 매우 큰 공헌을 할 것을 기대됨

제 2절 연구목적, 내용 및 범위

1. 연구개발의 목적

- 가. 한국재래돼지 종돈의 계통확립 기반 조성
- 나. 한국재래돼지 종돈의 불량형질(성장 및 포육능력) 개선 및 우량형질(육질) 보존을 위한 유전자 기반 선발 방법 개발
- 다. DNA 표지인자 기반 선발을 통한 한국재래돼지 생산성 극대화 및 보급용 종돈 시범 생산

2. 연구개발의 내용 및 범위

- 가. 한국재래돼지 기초축군 선발 및 교배
- 나. 한국재래돼지를 이용한 성장 및 포육능력 관련 형질 분석
- 다. 한국재래돼지 F2 상호교배 시험축군(약 600두)에 대한 SNP chip분석으로 형질관련 마커의 탐색
- 라. 형질관련 원인 유전자의 원인 변이(QTN) 규명 및 이를 이용한 표지인자 개발
- 마. 형질관련 마커의 유전자형에 따른 교배를 통해 성장 및 포육능력이 향상된 재래종돈 조성
- 바. 한국재래돼지 맞춤형 모형개발 및 MBV추정 알고리즘 개발과 이를 통한 컴퓨터 프로그램 개발
- 사. 발굴된 SNP마커를 이용한 잡종강세 효과 추정, 교배조합별 결합능력 추정 및 마커 효과 추정
- 아. 시험축군에서 추정된 SNP 마커별 효과를 바탕으로 한국재래돼지의 유전평가 및 선발
- 자. 경제형질관련 SNP를 이용한 한국재래돼지 종돈의 선발, 시범 보급 및 이를 이용한 산업화 방안 제시

제 2장 국내외 기술개발 현황

제 1절 국내 기술개발 현황

1. 유전체 정보를 이용한 가축 개량기술

가. 한우

현재 한우의 개량은 당대검정과 후대검정 두 단계로 나누어 선발을 실시하고 있다. 당대검정의 경우 자신의 체중 및 부모의 육종가 절반으로 계산된 근내지방도에 가중치를 부여하여 선발을 실시하고 있으며, 후대검정은 후대검정우당 약 10두 내외의 자손을 생성하고 이들을 도축시켜 자료를 얻음으로써 씨수소의 능력을 예측하여 선발한다.

- (1) Lee (2012)는 한우 556두(후대검정우 495두, 후보씨수소 61두)에 대하여 Illumina Bovine 50K SNP chip을 이용하여 유전체 정보를 수집하였으며 도체형질에 대하여 BLUP과 GBLUP의 정확도를 비교하였으며 유전체 정보를 보유한 개체 집단에서 GBLUP이 BLUP에 비해 1.82 - 2.07%의 정확도 상승효과가 있었다고 보고하였으며, 유전체 정보가 없는 개체에 대해서도 형질별 1.41% - 2.29% 정확도 상승효과가 있었다고 보고함
- (2) Cho (2013)는 한우 1,008두에 대하여 Illumina Bovine 50K SNP chip (288두) 및 Bovine HD(777K) SNP chip(720두)을 이용하여 한우 후보검정우 및 후보씨수소에 대하여 유전체 정보를 수집하고 BLUP방법과 ssGBLUP방법간 유전체 정확도를 비교하였다. 그 결과, 자신의 표현형 기록이 없는 후보씨수소 집단의 경우 도체형질(등심단면적, 근내지방도, 등지방두께)에서 유전체 선발을 실시함으로써 약 1%의 정확도 증가를 보였으며, 도축되어 자신의 기록을 갖는 후보검정우에서는 BLUP과 ssGBLUP간 차이가 거의 없었으며, 아직 도축되지 않아 자신의 기록이 없는 후보검정우집단의 경우 유전체정보를 이용함으로써 약 4~6%의 정확도 상승효과가 있었다고 보고함
- (3) 타 연구과제에서 현재 한우에서 유전체 자료 수집, 분석, 검증하는 연구가 진행중에 있음

나. 돼지

Lee 등(2013)은 폐쇄 종돈집단의 랜드레이스 순종돈 448두에 대하여 Porcine SNP 60K chip을 이용하여 유전체 정보를 수집하여 G05, GMF, GOF, GOF* 및 GN의 5가지 방법을 이용하여 유전체 관계행렬을 구성하고 이를 이용하여 유전체 육종가를 추정하였다. GOF 방법에 의하여 계산된 혈연계수가 기존의 혈통정보를 이용한 혈연계수와 가장 작은 편차를 나타내고 평균소수대립유전자빈도를 이용하는 GMF 방법에서는 큰 차이가 나타나 대립유전자빈도 기준이 혈연계수의 평균이 동을 유발함을 확인하였으며, GOF*를 제외한 모든 방법에서 정규분포형태의 멘델리안샘플링이 나타나는 것을 확인하였다. 등지방두께 평균과 90 kg 도달일령에 대한 육종가 추정 모형을 설정하고 유전체 관계행렬을 이용하여 유전모수와 육종가를 추정한 결과 혈통정보를 이용한 육종가와 상관은 GOF 방법에서 가장 높게 나타났으며, 유전체 관계행렬의 척도(scale)에 베타함수를 이용한 GOF*의 경우 모든 형질에서 유전분산이 크게 추정되어 분모부분을 구성하는 척도는 유전모수 추정치 영향하는

것을 확인하였다. 동일한 표현형 정보량을 이용할 경우 유전체관계행렬을 이용한 육종가 추정의 정확도가 혈통정보를 이용한 육종가보다 높게 나타났으며, 90 kg 도달일령보다는 등지방두께 평균에서 그 차이가 더 크게 나타났다. 집단 내 누적 표현형자료가 부족한 경우, 외래 유전자원이 도입되어 집단 내 혈연관계가 부족할 경우 또는 멘델리안 분포가 전혀 고려되지 않는 어린 동복자손의 육종가를 예측해야 하는 경우에 유전체 정보를 활용하면 유전능력 평가의 정확성을 크게 향상시킬 수 있을 것으로 보고함

다. 젓소

- (1) 국내 젓소의 경우 2013년 2월 이전까지 유전체선발에 관한 연구가 전무한 실정이었다가 2013년 2월부터 젓소유전체선발을 위한 참조집단 축군조성을 위한 시료 확보 및 유전체자료 수집에 관한 연구가 수행중에 있으며 3년간 약 2,000여두의 참조축군을 조성할 예정에 있음

2. 유전체 선발기술의 국내 적용현황

국내 유전체 선발 기술은 참조집단 구성에 목적을 두고 진행중에 있으며 아직 산업에 적용된 사례는 없음

2절 해외 기술개발 현황

1. 유전체 정보를 이용한 가축 개량기술

가. 덴마크 듀록 : 덴마크의 듀록품종에서 일당증체량 및 사료요구율 형질에 대하여 기존의 혈통정보와 검정정보를 이용하여 추정된 BLUP방법과 유전체 정보를 갖는 개체에 대해서만 추정된 GBLUP, 그리고 모든 정보(유전체 정보 보유 및 미보유 개체 포함)를 하나의 모델에 적합하여 분석한 Single step GBLUP방법에 대한 정확도 및 회귀식 분석을 실시한 결과 유전체 정보를 이용하여 분석한 것이 혈통정보만 이용한 것에 비하여 정확도가 높다고 보고하였으며, Single step GBLUP을 이용하였을 때 유전체 정보를 갖지 않는 개체에 대해 BLUP추정방법에 비해 정확도를 더 향상시킬 수 있다고 보고함 (Christensen 등, 2012).

2. 유전체 선발기술의 해외 적용현황

가. 캐나다 돼지 개량 프로그램(Canadian Centre for Swine Improvement): 60K SNP Chip을 이용하여 유전체 정보 수집하여 검정데이터와 유전체정보를 모두 갖고 있는 개체의 집단(참조집단)에서 각 SNP의 효과를 추정하고 이를 검정정보없이 유전체정보만 갖는 개체의 집단(검정집단)에 적용하여 기존 BLUP방법과 유전체선발방법간 비교하였다. 그 결과, 등지방두께에서 유전체 정보 없이 육종가를 추정하였을 때 0.35의 정확도를 보였으며, 유전체 정보를 이용하였을 때 0.39의 정확도로 유전체 정보를 사용하였을 때 유전적 개량량이 11% 증가의 효과가 있다고 보고하였으며, 총산자수는 0.55와 0.66으로 유전적 개량량이 20%증가하였으며, 초음파 등심단면적의 경우 유전체 정보를 이용함으로써 67%의 유전적 개량량 향상효과가 있다고 보고함 (CCSI, 2013).

나. 국제적으로 유명한 종돈회사인 PIC는 후보돈에 대하여 300 - 400 개의 마커와 부모의 60K SNP Chip을 통해 Imputation하는 기술을 보유하고 있으며 이를 유전적 개량량의 가속화를 지속적으로 해오고 있다고 보고하였다. 총산자수에 대해 부모의 60K SNP chip정보만 가지고 후보돈에 대해 유전체 선발을 하였을 때 11%의 정확도가 향상된 반면 후보돈 자신의 60K SNP chip정보를 포함하여 분석하였을 때 68%의 증가효과를 보였다고 보고함(PIC, 2013)

다. 젓소의 대표품종인 홀스타인의 경우 전 세계에 걸쳐 사육되고 있기 때문에 국가 간 유전능력 평가를 참여국에 한하여 평가를 실시하고 있다. 2010년 이후로 이러한 젓소국제유전능력평가에 대하여 유전체 정보를 포함하여 분석하자는 분위기가 조성되었으며 유전체 선발에 대한 검증단계를 거쳐 현재 일부 참여국에 대해서 GMACE평가를 실시하고 있다. 이와 더불어서 CDN(Canadian Dairy Network)에서는 유전체정보를 활용하여 유전능력평가된 결과를 농가에 제공하고 있으며, 미국(USDA), 독일(VIT)등 국가에서 유전체 분석결과를 서비스하고 있음(그림 1, 그림 2)

TOP 100 HOLSTEIN YOUNG BULLS BY GENOMIC PARENT AVERAGE FOR LPI (GPA LPI) - APRIL 2013 GENETIC EVALUATIONS
100 MEILLEURS JEUNES TAUREAUX HOLSTEIN SELON LA MOYENNE DES PARENTS GÉNOMIQUE POUR L'IPV (IPV MPG) - ÉVALUATIONS GÉNÉTIQUES AVRIL 2013
(Based on young bulls at least 12 months old with semen available in Canada / Basée sur les jeunes taureaux âgés d'au moins 12 mois dont la semence est disponible au Canada)

RANK	GPA LPI	REGISTRATION	NAME	A.I. CONTROLLER	SEMEN CODE	AGE	MILK	FAT	PROT	%F	%P	SCS	CONF	MS	F&L	DS	RP	NAME of SIRE
RANG	IPV MPG	ENREGISTREMENT	NOM	CENTRE d'I.A.	CODE DE SEMENCE	ÂGE	LAIT	GRAS	PROT	%G	%P	CSS	CONF	SM	P&M	PL	Cr	NOM du PÈRE
1	3421	USA 71302825	MR LOOKOUT P ENFORCER-ET	200	0200H002870	18	1905	82	80	0.07	0.13	2.83	14	10	14	9	4	MARBRI FACEBOOK
2	3392	CAN 11397854	GENERATIONS LIQUID GOLD	2200	0250H001049	16	1546	102	82	0.43	0.28	2.87	14	13	9	10	0	MAPLE WOOD BOULDER
3	3382	CAN 9608054	BOLDI V S G ANTON	2200	0250H001073	12	1910	90	72	0.18	0.07	2.80	17	16	14	9	1	GENERATIONS EPIC
4	3349	USA 71411740	MR CHARTRAI ELOQUENT-ET	200	0200H002986	17	1740	106	86	0.41	0.27	2.79	12	10	14	7	3	GEN-BREQ LAVAMAN
4	3349	CAN 107848780	PARLE LOCARNO	200	0200H006516	15	1774	85	84	0.19	0.24	2.67	12	12	12	4	-2	DE-SU PIERRE 840-ET
6	3337	CAN 11420628	VELTHUS B AHEAD	200	0200H003006	12	1726	116	73	0.50	0.15	2.90	13	11	12	8	0	MISTY SPRINGS SUPERSONIC
7	3324	USA 69681349	SEAGULL-BAY SUPERSIRE-ET	7	0007H011351	27	2451	119	88	0.27	0.07	2.74	8	7	6	4	8	ROYLANE SOCRA ROBUST-ET
8	3312	USA 70940250	SEAGULL-BAY PLATINUM-ETS	7	0007H011618	21	2474	101	82	0.09	0.12	2.52	11	9	7	7	6	FLEVO GENETICS SNOWMAN
8	3312	USA 70940251	SEAGULL-BAY DIAMOND-ETS	7	0007H011617	21	2474	101	82	0.09	0.12	2.52	11	9	7	7	6	FLEVO GENETICS SNOWMAN
10	3306	USA 71302827	MR LOOKOUT P EMBARGO-ET	200	0200H002871	15	1648	96	66	0.33	0.11	2.85	15	11	12	12	4	MARBRI FACEBOOK
11	3291	CAN 11098958	GENERATIONS LEXOR	2200	0250H001010	33	1635	90	84	0.28	0.27	2.89	12	11	14	8	1	LONG-LANGS OMAN OMAN-ET
12	3278	USA 86960187	DE-SU JEROD 1223-ET	7	0007H011478	25	1969	97	68	0.21	0.03	2.64	13	12	9	4	6	DE-SU OBSERVER-ET
13	3272	USA 70801850	COYNE-FARMS JABRI-ET	200	0200H003877	15	1721	101	81	0.37	0.05	2.58	10	9	10	5	-2	MARBRI FACEBOOK
14	3265	CAN 107848788	ALPHAGEN ERSTEIN	200	0200H003008	12	1377	93	68	0.41	0.20	2.87	15	13	13	8	2	GENERATIONS EPIC
15	3260	USA 70885451	RMW ANCHOR-ET	200	0200H003883	18	1677	104	55	0.40	0.00	2.63	14	14	9	10	2	MARBRI FACEBOOK
16	3259	CAN 11347918	GENERATIONS BIG KAHUNA	2200	0250H001096	13	2167	80	78	0.01	0.64	2.82	14	13	11	8	2	GENERATIONS EPIC
17	3256	USA 70377846	TEEMAR OBSERVER ADAM-ET	7	0007H011504	28	2298	85	80	-0.01	0.05	2.64	13	13	9	2	5	DE-SU OBSERVER-ET
18	3255	USA 70371991	SULLY MCCORD 209-ET	7	0007H011837	19	1880	105	60	0.35	0.08	2.71	11	13	5	4	3	LADYS-MANOR RD GRAFEETH-ET
18	3252	CAN 11373111	STANTONS BOOKKEEPER	200	0200H003876	16	1802	68	76	0.01	0.16	2.79	14	12	11	9	6	DE-SU 521 BOOKEM-ET
20	3250	CAN 11420615	BOLDI ELECT	200	0200H003894	13	1376	91	66	0.37	0.19	2.66	13	10	14	9	3	VA-EARLY-DAWN SUDAN CR-ET

그림 1. 캐나다 홀스타인의 youngbull에 대한 유전체 유전평가결과

TOP 500 HOLSTEIN HEIFER CANDIDATES FOR GENOTYPING BY PARENT AVERAGE FOR LPI (P A LPI) - APRIL 2013 GENETIC EVALUATIONS
500 MEILLEURES GÉNÉSSES HOLSTEIN CANDIDATES POUR LE GÉNOTYPAGE SELON LA MOYENNE DES PARENTS L'IPV (MP IPV) - ÉVALUATION
(Based on heifers born in the past 30 months and have a Canadian owner / Basée sur les généssees nées au cours des 30 derniers mois et qui ont un propriétaire canadien)

HEIFER PARENT AVERAGES AND OTHER INFORMATION / MOYENNES DES PARENTS DES GÉNÉSSES ET D'AUTRES RENSEIGNEMENTS																		SIRE INFORMATION / RENSEIGNEMENTS			
RANK	PALPI	REGISTRATION	NAME	BIRTH DATE	AGE	MILK	FAT	PROTEIN	%FAT	%PROT	SCS	CONF	MS	F&L	DS	RP	REGISTRATION	NAME			
RANG	IPV MP	ENREGISTREMENT	NOM	NAISSANCE	ÂGE	LAIT	GRAS	PROTÈINE	%GRAS	%PROT	CSS	CONF	SM	P&M	PL	Cr	ENREGISTREMENT	NOM			
1	3225	CAN 101824930	VIEUXSAULE LEXOR CYN	09-Aug-12	8	1781	87	71	0.21	0.12	2.89	12	12	11	6	3	CAN 11098958	GENERATIONS LEXOR			
2	3212	CAN 11896410	STANTONS MCCUTCHEEN BLINK	26-Mar-13	0	2602	104	88	0.09	0.03	2.87	10	10	6	7	4	USA 69660138	DE-SU BKM MCCUTCHEEN 1174-ET			
3	3157	CAN 11218246	BRABANTDALE MGL SPLURGED	02-Dec-12	4	1799	95	63	0.27	0.04	2.81	12	12	10	6	4	840 3006972816	MOUNTFIELD SSI DCY MOGUL-ET			
4	3156	CAN 108032693	DESLACS SUPERSIRE MAYLA	15-Mar-13	1	2284	105	81	0.19	0.06	2.80	9	8	6	4	6	USA 69661349	SEAGULL-BAY SUPERSIRE-ET			
5	3152	CAN 11537535	STANTONS EPIC ESOR	04-Oct-12	6	1973	77	73	0.06	0.07	2.91	12	11	10	7	6	CAN 11104016	GENERATIONS EPIC			
6	3118	CAN 11230754	STANTONS BOOKEM ESTER	25-Jun-11	21	1694	80	76	0.16	0.18	2.85	10	10	8	4	4	USA 68638657	DE-SU 521 BOOKEM-ET			
7	3117	CAN 11896361	STANTONS HEADLINER HIGH	25-Feb-13	1	2077	100	80	0.24	0.11	2.79	9	9	6	4	2	USA 69661350	SEAGULL-BAY HEADLINER-ET			
8	3116	CAN 11591523	MAPLE WOOD SNOW KEPT SECRET	02-Mar-13	1	2307	83	85	-0.01	0.09	2.93	12	11	10	7	5	CAN 11294722	VELTHUS LET IT SNOW			
8	3116	CAN 11591524	OCONNORS SNOW KEEPSAKE	03-Mar-13	1	2307	83	85	-0.01	0.09	2.93	12	11	10	7	5	CAN 11294722	VELTHUS LET IT SNOW			
8	3116	CAN 11591525	MAPLE WOOD SNOW KEEP UP	05-Mar-13	1	2307	83	85	-0.01	0.09	2.93	12	11	10	7	5	CAN 11294722	VELTHUS LET IT SNOW			
11	3112	CAN 11538400	PENVIEW SCUBA	07-Jan-13	3	2265	72	74	-0.11	0.00	2.80	14	13	8	11	5	USA 69651907	SULLY HART MERIDIAN-ET			
12	3111	CAN 108176946	DELABERGE LEPA MASCALESE ELLA	24-Feb-13	1	2146	80	76	0.01	0.07	2.74	12	10	9	9	8	ITA 17960516801	ZANI BOLTON MASCALESE ET			
12	3111	CAN 108176948	DELABERGE LEPA MASCALESE ELLEN	24-Feb-13	1	2146	80	76	0.01	0.07	2.74	12	10	9	9	8	ITA 17960516801	ZANI BOLTON MASCALESE ET			
14	3109	CAN 11521703	EASTSIDE LEWISDALE PROVOCATEUR	18-Oct-12	6	1681	79	57	0.15	0.01	2.68	15	12	13	10	8	CAN 11104016	GENERATIONS EPIC			
15	3108	CAN 11538407	BENNER SILVA	04-Feb-13	2	2094	79	70	0.06	0.04	2.81	13	12	9	9	7	USA 69774730	MINNIGAN HILLS DAY-ET			
16	3106	CAN 11896362	STANTONS MCCUTCHEEN MAYBE	26-Feb-13	1	2131	91	76	0.12	0.07	2.85	11	11	7	8	3	USA 69660138	DE-SU BKM MCCUTCHEEN 1174-ET			
16	3106	CAN 11896374	STANTONS MCCUTCHEEN WENDEE	28-Feb-13	1	2131	91	76	0.12	0.07	2.85	11	11	7	8	3	USA 69660138	DE-SU BKM MCCUTCHEEN 1174-ET			
16	3106	CAN 11896401	STANTONS CLUTCH DACEY	17-Mar-13	1	2131	91	76	0.12	0.07	2.85	11	11	7	8	3	USA 69660138	DE-SU BKM MCCUTCHEEN 1174-ET			
19	3099	CAN 11850457	DONNANDALE MASCALESE LULLU	18-Feb-13	2	2281	78	82	-0.06	0.08	2.82	11	10	6	9	8	ITA 17960516801	ZANI BOLTON MASCALESE ET			
19	3099	CAN 11850467	DONNANDALE MASCALESE LULLU	19-Feb-13	1	2281	78	82	-0.06	0.08	2.82	11	10	6	9	8	ITA 17960516801	ZANI BOLTON MASCALESE ET			

그림 2. 캐나다 홀스타인의 heifer에 대한 유전체 유전평가결과

라. 미국의 앵거스협회는 암소에 대하여 유전체정보를 분석해주는 서비스를 유료로 진행하고 있으며, 유전체 정보를 통해 유전체유전능력분석뿐만 아니라 친자감정 및 일부 질병저항형질에 대한 결과를 농가에 제공하고 있음(Angus, 2013) (그림 3, 4)



그림 3. 농가에서 자가 DNA 샘플을 채취하는 방법에 대한 매뉴얼(앵거스)

AGI DNA Tests	Price	Genomic Tests	Price
Parentage	\$18.00	Zoetis® HD50K (includes Parentage)	\$75.00
Coat Color Test	\$18.00	Add Coat Color	\$18.00
Arthrogyposis Multiplex (AM) Test	\$23.00	Add AM Test	\$23.00
Neuropathic Hydrocephalus (NH) Test	\$23.00	Add NH Test	\$23.00
Contractural Arachnodactyly (CA) Test	\$23.00	Add CA Test	\$23.00
Dwarfism (D2) Test	\$23.00	GeneSeek® GGP-HD (includes Parentage)	\$75.00
Myostatin (M1) Test	\$23.00	Add Coat Color	\$5.00
		Add AM Test	\$8.00
		Add NH Test	\$8.00
		Add CA Test	\$8.00
		Add D2 Test	\$20.00
		Add M1 Test	\$18.00
Additional Sample handling fee of 2.00 will be charged per sample for samples submitted in blood tubes or hair not on proper hair cards.			

그림 4. DNA 분석 항목별 서비스 비용(앵거스)

제 3장 연구개발수행 내용 및 결과

제 1 절 한국재래돼지의 성장 및 포육능력관련 원인유전자 변이 탐색

1. SNP chip 분석 방법

가. 유의적인 QTL 영역 내 정밀지도 작성을 위해 PorcineSNP 60 BeadChip을 이용하여 참조 축군의 유전자형 분석을 실시하였으며 Sentrix® BeadChip Array 12-sample HD DNA Analysis BC PorcineSNP 60 (Illumina, USA) 사용하여 SNP chip 분석을 실시하였다(그림 5).

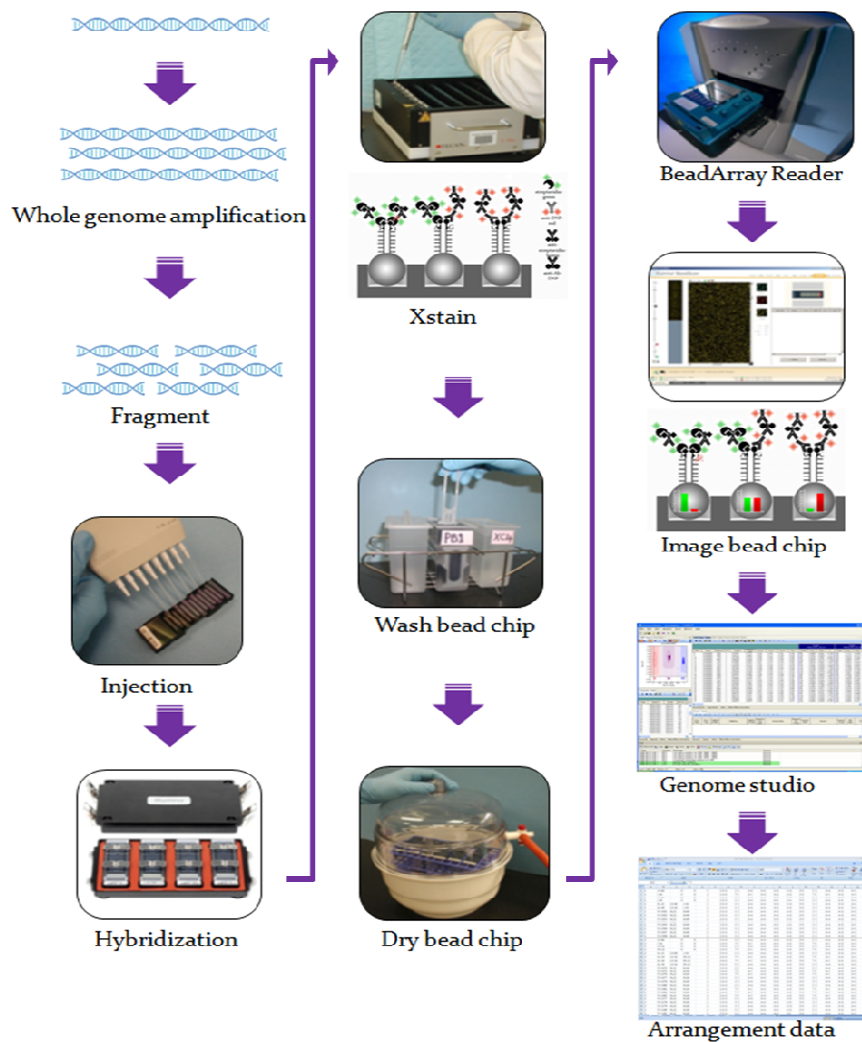


그림 5. PorcineSNP 60 Bead chip을 이용한 SNP분석을 위한 실험과정 모식도

나. Plink를 이용한 SNP chip 분석 방법

(1) Scan 하여 얻은 chip image data를 genome studio를 이용하여 final report를 생성하였으며 Final report의 62,163개의 SNP정보를 10,000개씩 편집하여 Plink program의 input data 형태로 편집하였다.

(2) 아래의 기준에 해당하는 SNP 마커를 제외하고 기초자료를 생성하였다.

(가) mendel error

(나) missing genotype 비율 : 10% 이상

(다) minor allele frequency : 5% 이하

(라) Hardy-Weinberg equilibrium error : P-value 0.001 이하

(마) Linkage disequilibrium (r2) : P-value 0.8 이상

(3) 그 결과 총 42,779개의 SNP 마커들이 기초데이터 생성에 이용되었다.

다. 전장 유전체 분석

(1) Plink를 이용한 각각의 성장 및 포육능력 형질에 대한 DNA chip에서 비롯된 각각의 SNP marker의 효과를 추정하기 위해서 다음의 다중 회귀 모형을 이용하였다.

$$y = b_0 + b_1X_{sex} + b_2X_{snp_a} + b_3X_{snp_d} + b_4X_{full-sib} + e$$

- ▶ y : 각각의 성장 형질
- ▶ b0 : 다중회귀식의 절편
- ▶ b1 : sex 효과에 대한 회귀계수
- ▶ X_{sex} : sex 효과에 대한 회귀변수
- ▶ b2, b3 : snp 표지의 상가적 (additive) 및 우성 (dominance) 효과들에 대한 회귀계수
- ▶ X_{snp_a}, X_{snp_d} : 상가적 및 우성효과들에 대한 회귀변수
- ▶ b4 : 전형매 (full-sib) 효과에 대한 회귀계수
- ▶ X_{full-sib} : 전형매 효과에 대한 회귀변수
- ▶ e : 임의 잔차 효과

(2) QxPak을 이용하여 각각의 성장 및 포육능력 형질에 대한 DNA chip에서 비롯된 각각의 SNP marker의 효과를 추정하기 위해서 다음의 혼합 모형을 이용하였다.

$$y = mean + sex + snp_a + snp_d + a_0 + e$$

- ▶ y : 각각의 성장 형질
- ▶ mean : 집단의 평균
- ▶ sex : 성별에 대한 고정효과
- ▶ snp_a, snp_d : snp 표지의 상가적 및 우성 효과들로서 모두 고정효과
- ▶ a0 : polygenic효과로서 개체에 대한 임의효과
- ▶ e : 임의 잔차 효과

2. Plink (②,④)와 QxPak (①,③)을 이용한 Genome wide association 분석 기준

가. A (additive), A+D (additive+Dominance) 모델을 사용하여 전장 유전체 분석을 실시하였으며 5% Genome wide significant threshold는 Boferoni correction에 의해서 결정 (red line) 하였다. 여기에 사용된 Bonferroni correction 값은 $0.05 / 42,779 = 1.2 \times 10^{-6}$ 이었다.

3. Plink와 QxPak Program을 이용한 분석 결과의 비교

Plink와 QxPak의 Genome wide 수준의 성장 및 포육능력 형질을 분석하여 유의적인 SNP 를 선발하였으나(그림 6 - 그림 16) 두 프로그램의 결과 비교에서 데이터들의 결과값이 상이하며, Q-Q plot의 결과 또한 정규분포선에서 많이 벗어나, sex와 임의 효과인 polygenic

effect가 보정된 residual을 구한 뒤 Plink 프로그램을 이용한 GWAS 즉, GRAMMAR(Genome-wide Rapid Association using Mixed Model And Regression)를 수행하여 성장 및 포육 능력 관련 후보유전자를 재 선별하였다.

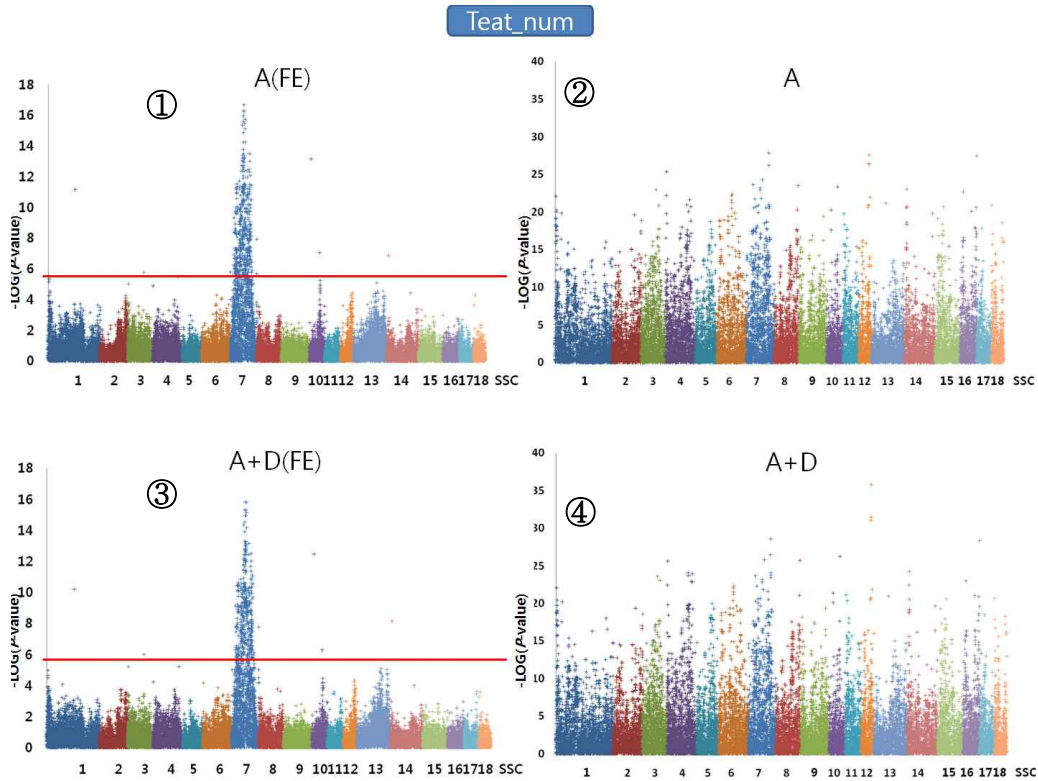


그림 6. Plink (②,④)와 QxPak (①,③)를 이용한 총 유두수에 대한 Genome wide association 분석 결과

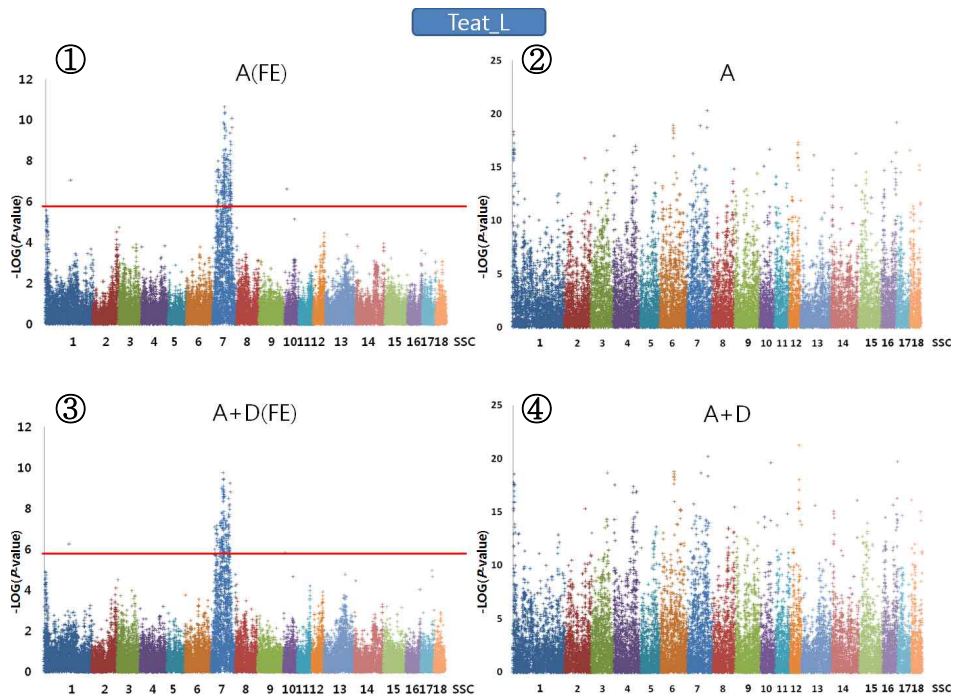


그림 7. Plink (②,④)와 QxPak (①,③)를 이용한 좌측 유두수에 대한 Genome wide association 분석 결과

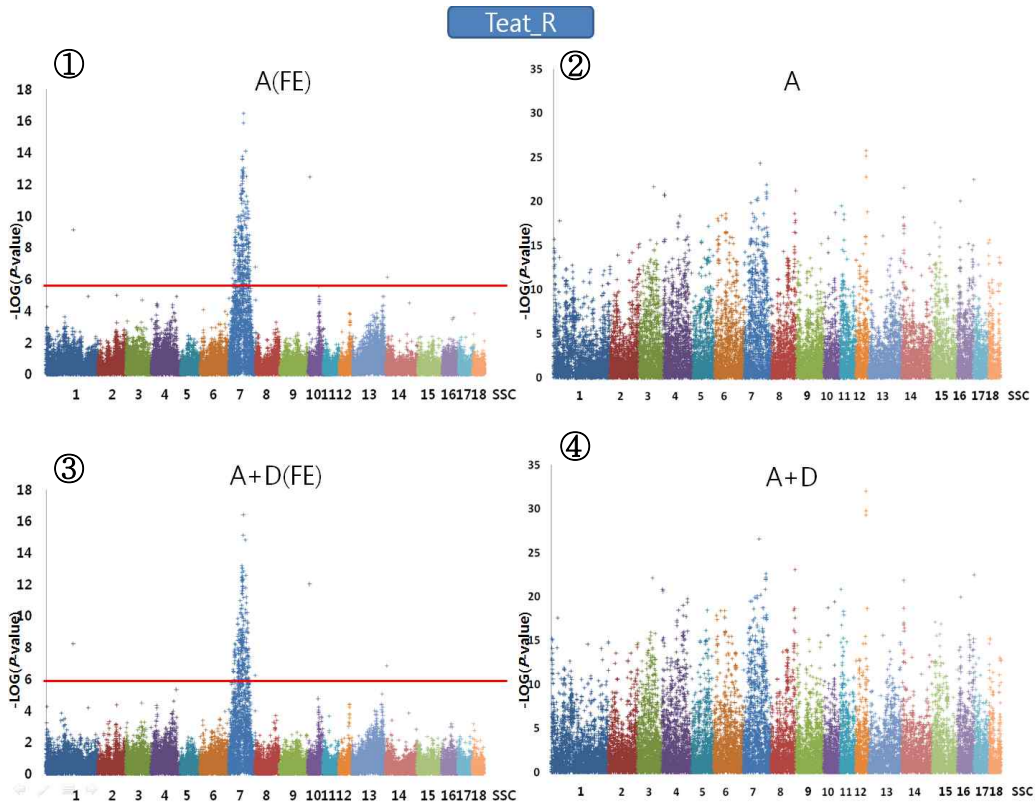


그림 8. Plink (②,④)와 QxPak (①,③)를 이용한 우측 유두수에 대한 Genome wide association 분석 결과

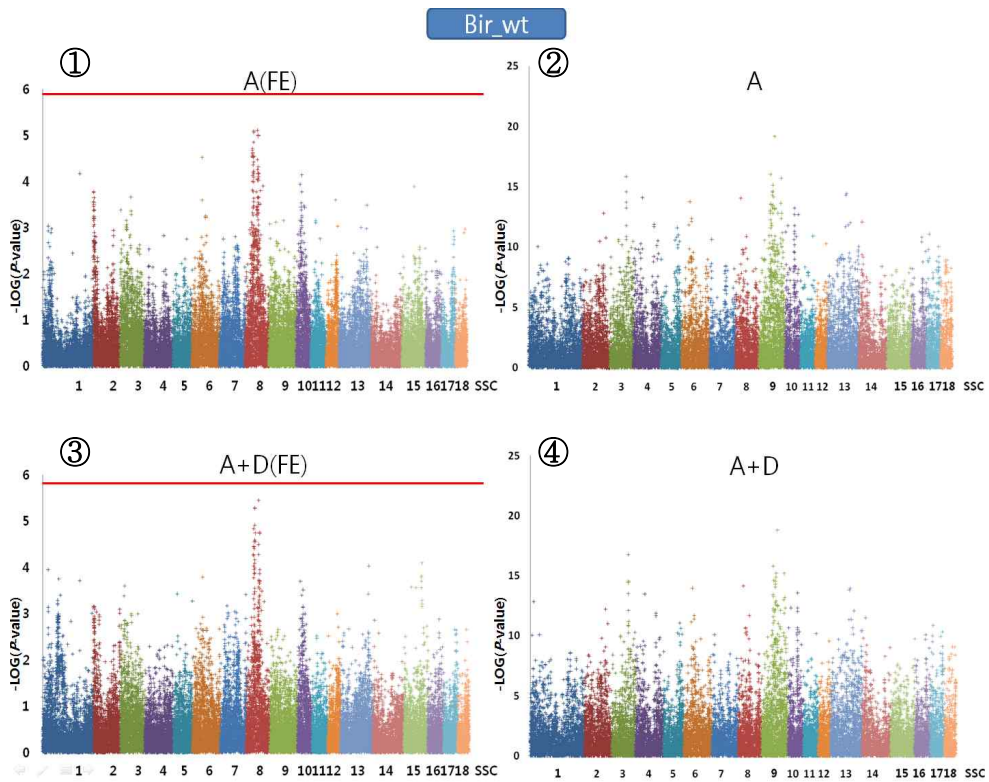


그림 9. Plink (②,④)와 QxPak (①,③)를 이용한 생시체중에 대한 Genome wide association 분석 결과

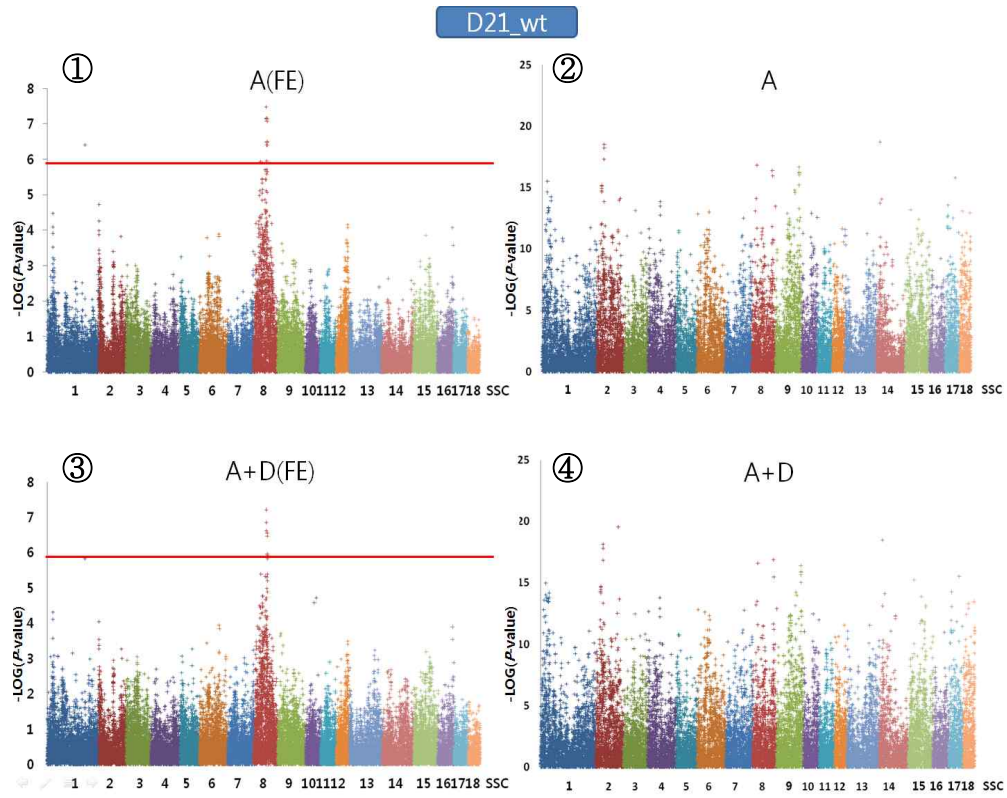


그림 10. Plink (②,④)와 QxPak (①,③)를 이용한 21일령 체중에 대한 Genome wide association 분석 결과

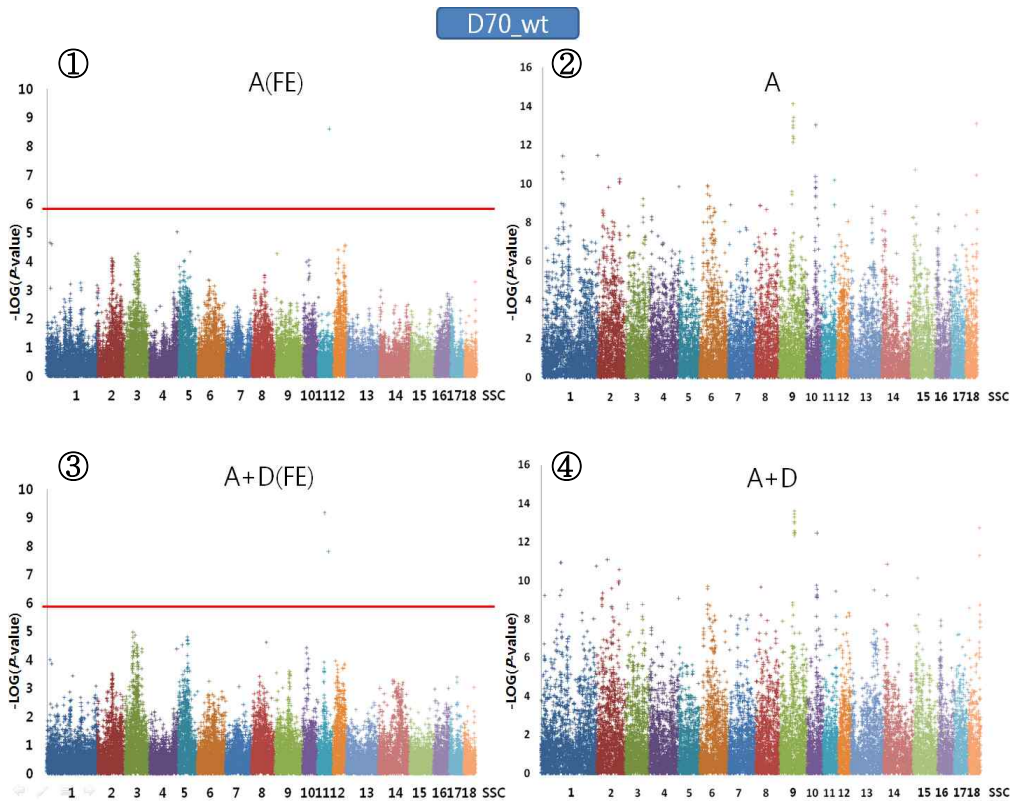


그림 11. Plink (②,④)와 QxPak (①,③)를 이용한 70일령 체중에 대한 Genome wide association 분석 결과

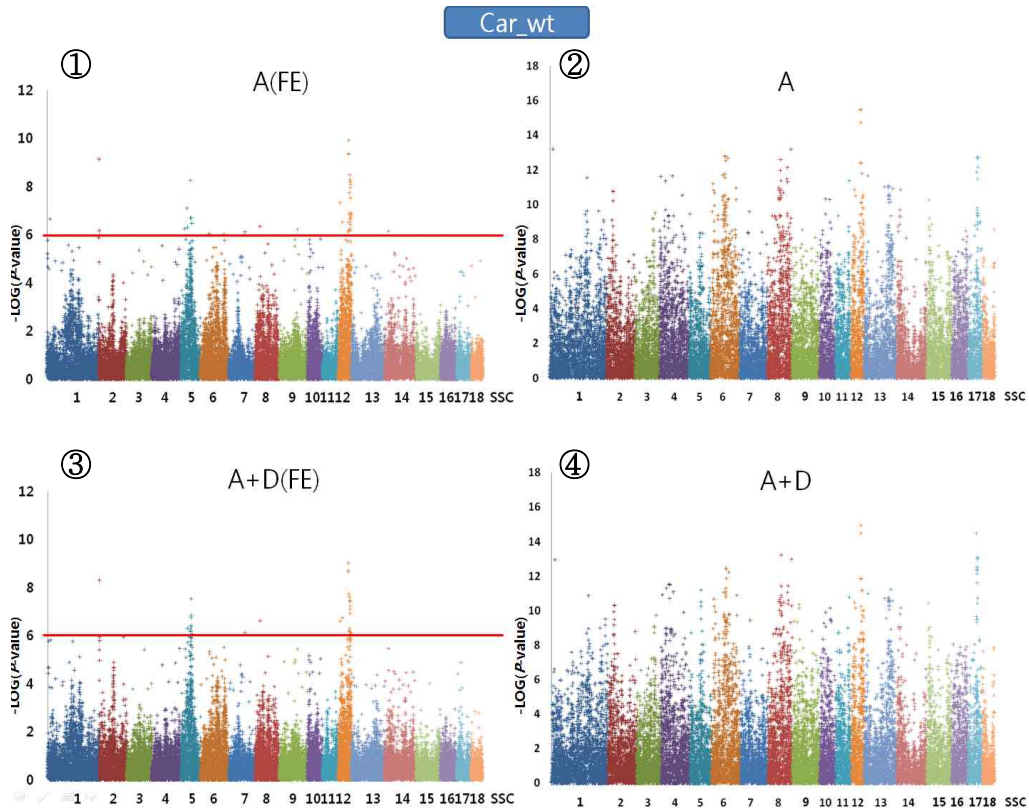


그림 12. Plink (②,④)와 QxPak (①,③)를 이용한 도체중에 대한 Genome wide association 분석 결과

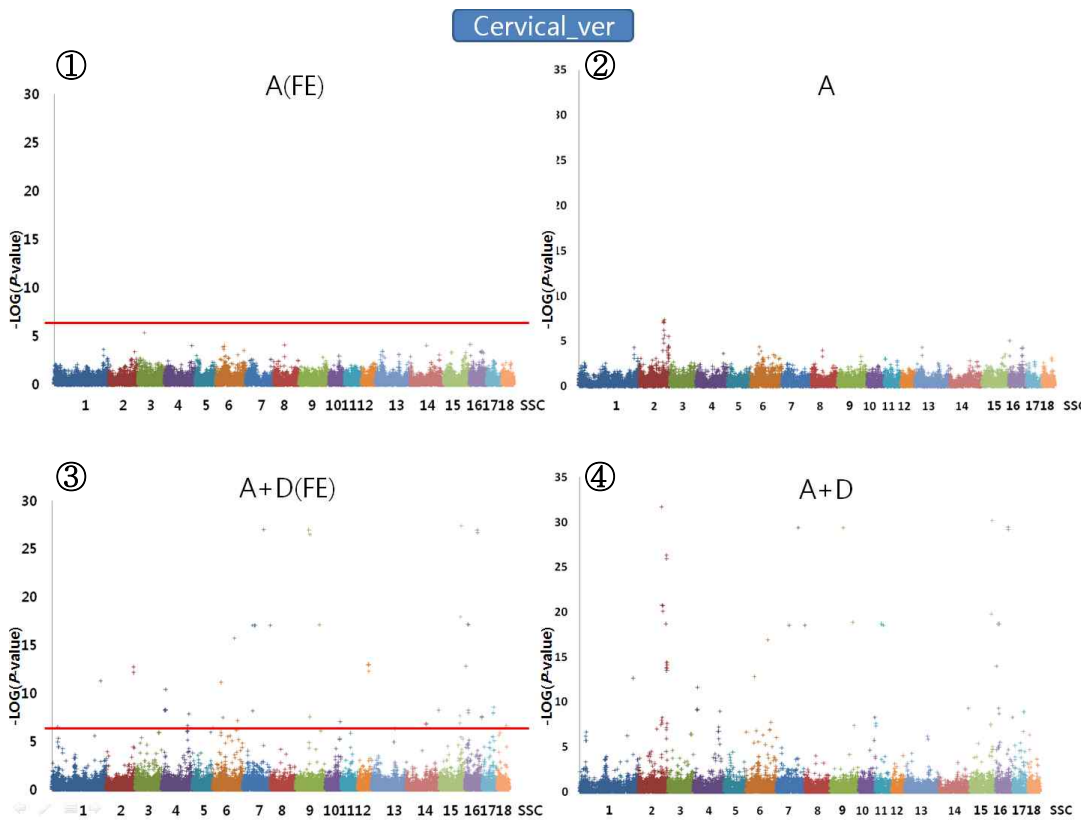


그림 13. Plink (②,④)와 QxPak (①,③)를 이용한 목 경추수에 대한 Genome wide association 분석 결과

thoracic_ver

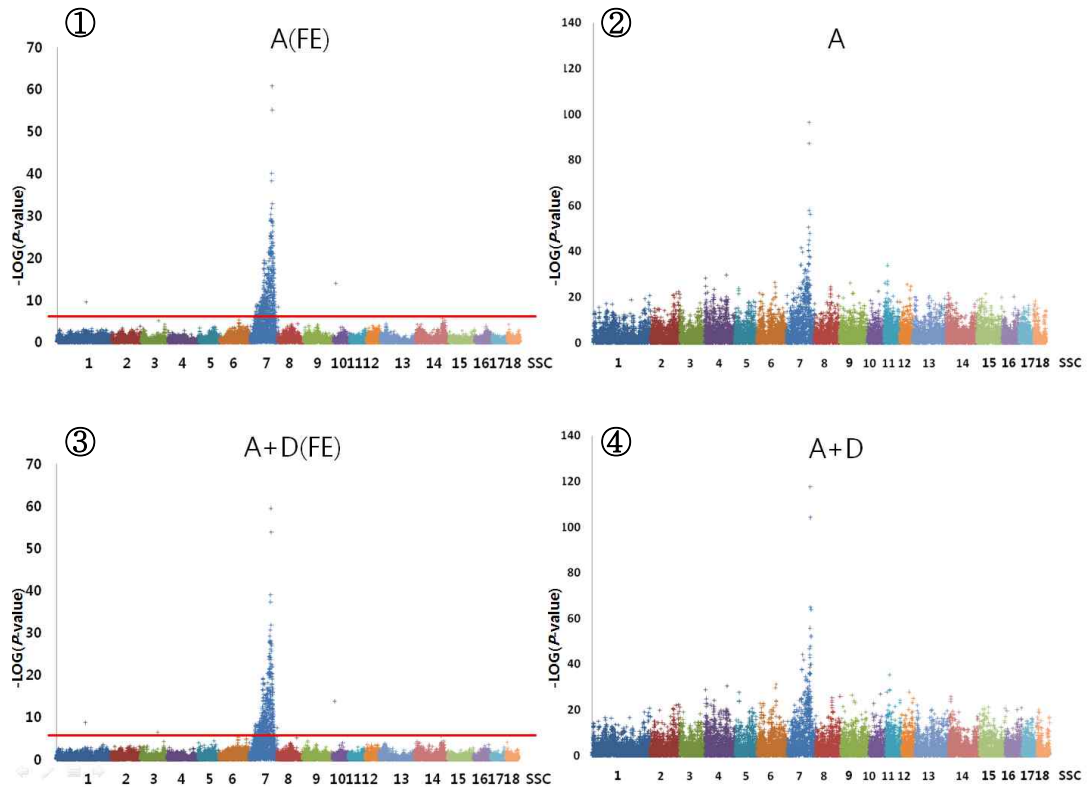


그림 14. Plink (②,④)와 QxPak (①,③)를 이용한 갈비뼈수에 대한 Genome wide association 분석 결과

Lumba_ver

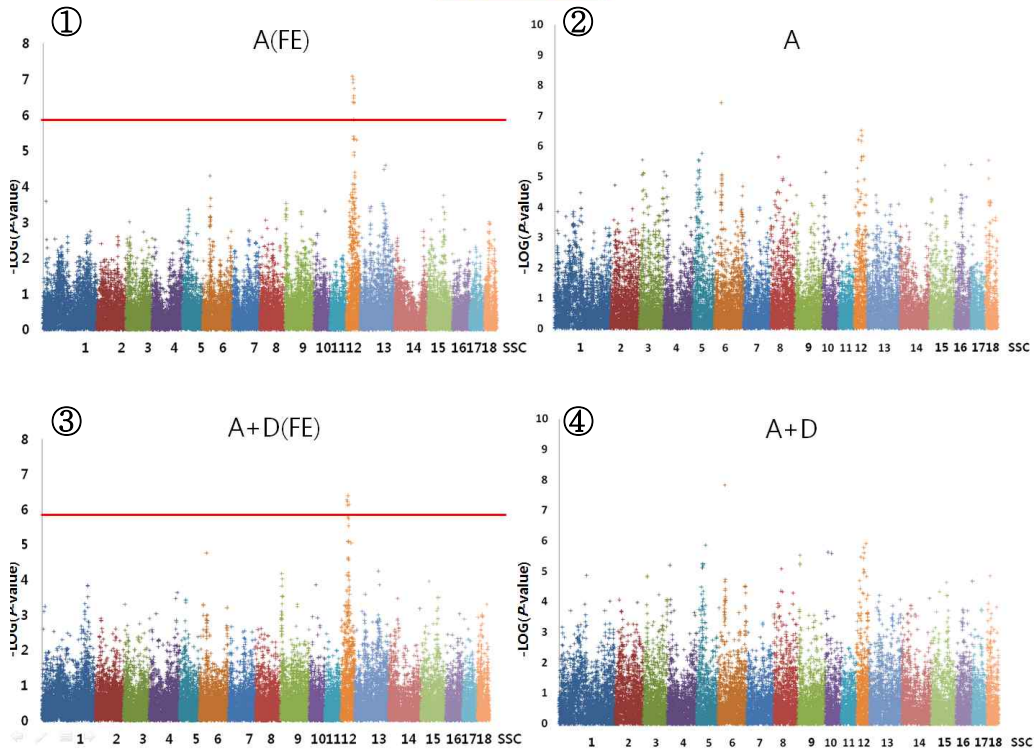


그림 15. Plink (②,④)와 QxPak (①,③)를 이용한 허리뼈수에 대한 Genome wide association 분석 결과

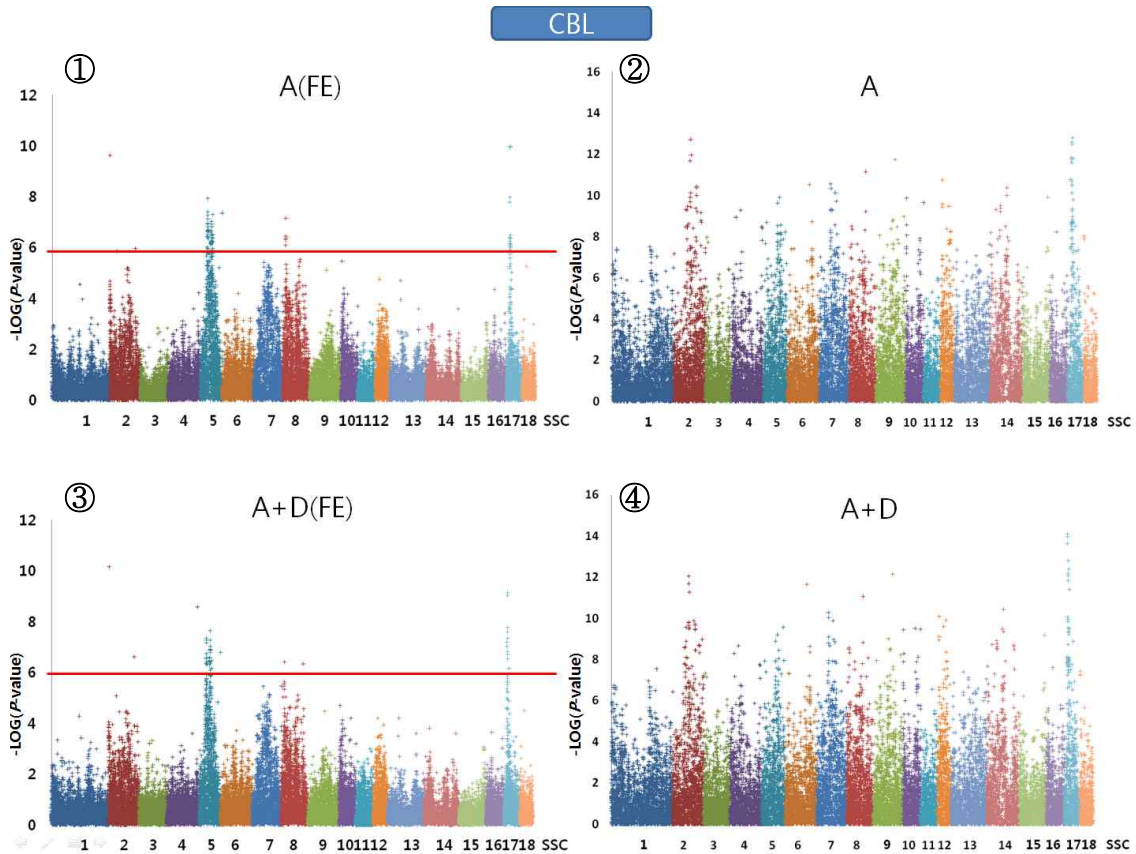


그림 16. Plink (②,④)와 QxPak (①,③)를 이용한 도체장에 대한 Genome wide association 분석 결과

3. GRAMMAR를 이용한 유효 SNP 및 후보 유전자의 선발

가. GRAMMAR를 이용한 유효 SNP 선발 방법

- (1) Linear regression analysis 옵션을 사용하여 P -value를 구하였으며 $-\log(P\text{-value})$ 를 취하여 manhattan plot을 완성후 Bonferroni correction($0.05/30,917=1.61 \times 10^{-6}$)을 이용하여 유효 SNP만을 도출하였다.
- (2) Bonferroni correction을 이용하여 유효 SNP가 도출된 형질들은 Additive model과 Additive + Dominance model 중 유효 SNP가 존재하는 model을 선택하였다.
- (3) 유효 SNP가 존재하는 형질의 경우 해당 염색체와 유효 SNP 주위에 관련 유전자가 있다고 추정하고 95% 신뢰구간에 존재하는 위치의 후보 유전자를 선발하였다.

나. 기 조성된 참조축군에 대한 성장 및 포육능력 관련 QTL 분석 결과와 GWAS 결과와의 비교를 통한 형질관련 QTL 위치 비교

- (1) 본 연구에 사용된 참조축군을 대상으로 선행연구를 통해 구축된 돼지 염색체별로 선발된 165개의 informative Microsatellite (MS) 마커를 이용하여 유전자형 분석을 실시하여 CRI-MAP version 2.4 software를 이용하여 연관 분석을 실시하였으며, 전체 길이가 2244.7 cM인 연관지도를 얻은 결과를 이용해 PorcineSNP 60 BeadChip 분석을 통해 획득한 유전자형 자료와 연관하여 QTL 분석에 활용하였다.
- (2) 유두수 관련하여 SSC7에서 유의적인 QTL을 확인(95% 신뢰구간, *S0102-SW147*)하였으

며 Basic Local Alignment Search Tool(BLAST) 분석을 통하여, 연관 지도상 마커 (*SO102-SW147*)의 물리적 위치(bp)를 확인하였다 (http://asia.ensembl.org/Sus_scrofa/Info/Index).

- (3) QTL(linkage analysis)과 GWAS(association analysis) 결과의 비교를 통하여 유두수 QTL 위치 확인 및 유의적인 마커를 선별하였다(그림 317).

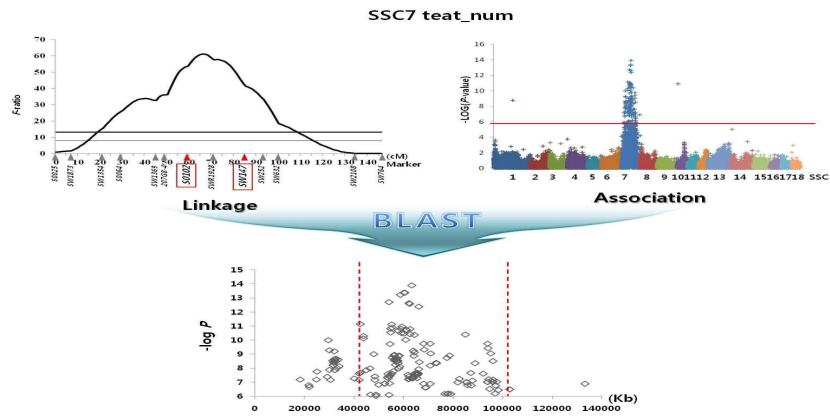


그림 17. QTL 분석 결과와 GWAS 결과를 비교한 유두수 QTL 위치 비교

다. 각 형질의 GWAS 분석 결과 및 후보유전자의 선정

- (1) 유두수 형질들을 GWAS 분석한 결과, 18개의 상염색체 중 7번 염색체(SSC7)에서 유의적인 결과를 얻을 수 있었으며(그림 18 - 그림 20), Ensembl version 10의 데이터를 이용하여 신뢰구간내에 존재하는 후보유전자를 선정하였다(표 1).

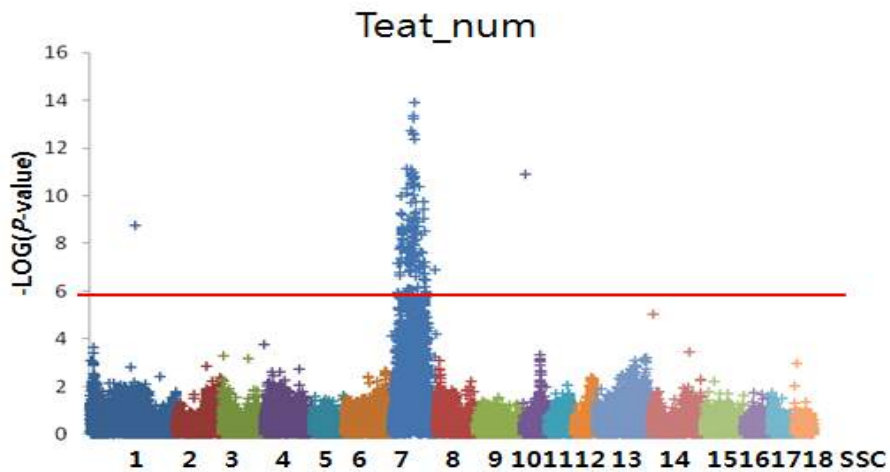


그림 18. Plink program의 GRAMMAR를 이용한 총 유두수 GWAS 결과.

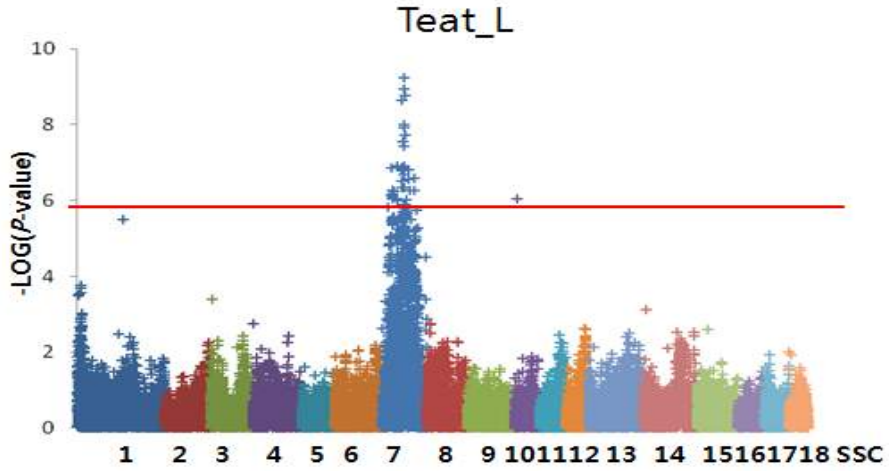


그림 19. Plink program의 GRAMMAR를 이용한 좌측 유두수 GWAS 결과.

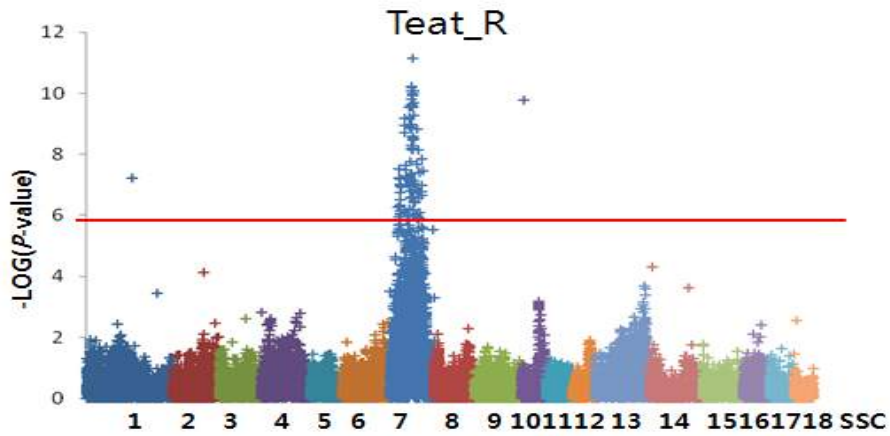


그림 20. Plink program의 GRAMMAR를 이용한 우측 유두수 GWAS 결과.

표 1. 7번 염색체에 위치한 유두수 관련 후보 유전자

SSC	GENE name	Locus Description	Position		Size (bp)
			Start	End	
7	CCND3	cyclin D3	42718333	42722900	4567
7	KIAA1199	KIAA1199	55181168	55364257	183089
7	ENSSSCG0000001830		59508735	59509625	890
7	C15orf26	Uncharacterized protein C15orf26	55501786	55514472	12686
7	CCDC37	coiled-coil domain containing 37	59105399	59135074	29675
7	ARNT2	aryl-hydrocarbon receptor nuclear translocator 2	54775587	54917283	141696
7	SNX33	sorting nexin 33	62889629	62898948	9319
7	NRG4	neuregulin 4	61026551	61147155	120604
7	ETFA	electron-transfer-flavoprotein, alpha polypeptide	61210791	61300284	89493
7	SCARNA15	Small Cajal body specific RNA 15	57934486	57934612	126

7	SCARNA15	Small Cajal body specific RNA 15	57934486	57934612	126
7	IGF2BP1	insulin-like growth factor 2 mRNA binding protein 1	59223322	59225004	1682
7	WDR93	WD repeat domain 93	60408127	60450712	42585
7	HMG20A	high-mobility group 20A	62615866	62642147	26281
7	ENSSSCG0 000002235		84982964	84983216	252
7	TBC1D21	TBC1 domain family, member 21	64412757	64426041	13284
7	PPP2R5D	protein phosphatase 2, regulatory subunit B', delta	43938486	43961327	22841
7	GNMT	glycine N-methyltransferase	43921225	43924263	3038
7	FBXO22	FBXO22 opposite strand	60994277	60998310	4033
7	BMP5	bone morphogenetic protein 5	29785282	29908279	122997
7	PLEKHG3	pleckstrin homology domain containing, family G (with RhoGef domain) member 3	94087305	94133693	46388
7	SLC25A21	solute carrier family 25 (mitochondrial oxodicarboxylate carrier), member 21	67971262	68290929	319667
7	EGLN3	egl nine homolog 3 (C. elegans)	70837098	70862950	25852
7	SPTB	spectrin, beta, erythrocytic	94138307	94215667	77360
7	NKX2-8	NK2 homeobox 8	68389739	68391134	1395
7	HMGCLL1	3-hydroxymethyl-3-methylglutaryl-CoA lyase-like 1	30098402	30118333	19931
7	CCDC33	coiled-coil domain containing 33	64130542	64161473	30931
7	KHDRBS2	KH domain containing, RNA binding, signal transduction associated 2	32025961	32335083	309122
7	ENSSSCG0 0000018472		95325356	95325482	126
7	NPAS3	neuronal PAS domain protein 3	70940224	71092954	152730
7	MEP1A	mepripin A, alpha (PABA peptide hydrolase)	48064931	48101774	36843
7	ENSSSCG0 0000001791		56464588	56485797	21209
7	STXBP6	syntaxin binding protein 6 (amisyn)	78620287	78972830	352543
7	ENSSSCG0 0000001810		58147647	58325321	177674
7	PML	promyelocytic leukemia	64086247	64088426	2179
7	CPEB1	cytoplasmic polyadenylation element binding protein 1	58067202	58119739	52537
7	5S_rRNA	5S ribosomal RNA	77626812	77626937	125
7	TM6SF1	transmembrane 6 superfamily member 1	57467901	57584680	116779
7	PRIM2	primase, DNA, polypeptide 2 (58kDa)	32783074	33088972	305898
7	ENSSSCG0 0000001793		56585189	56585292	103
7	ENSSSCG0 0000001792		56485881	56534482	48601
7	SH3GL3	SH3-domain GRB2-like 3	57004218	57137357	133139

(2) 21일령 체중 형질을 GWAS 분석한 결과, 8번 염색체(SSC8)에서 유의적인 결과를 얻을 수 있었으며(그림 21), 신뢰구간내에 존재하는 두 개의 후보유전자를 선정하였으나 이 두 유전자 모두 현재까지 기능이 알려져 있지 않은 유전자임이 확인되었다(표 2).

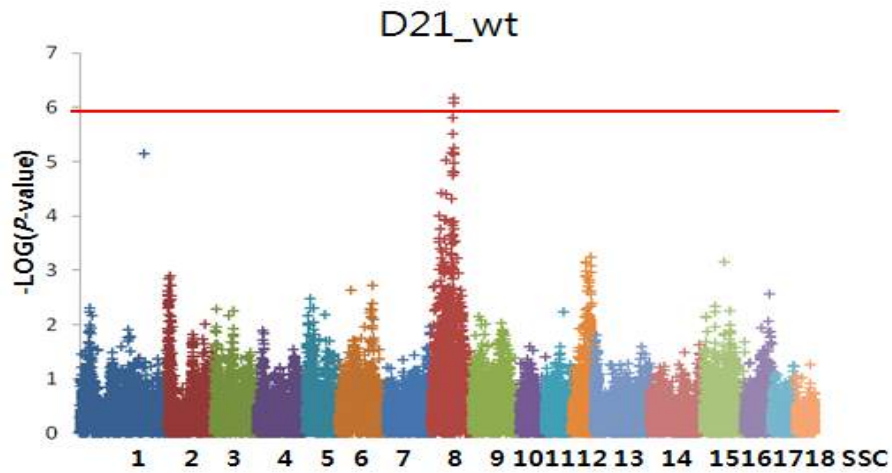


그림 21. Plink program의 GRAMMAR를 이용한 21일령 체중의 GWAS 결과.

표 2. 21일령 체중 관련 후보 유전자

SSC	GENE name	Locus Description	Position		Size (bp)
			Start	End	
8	ENSSSCG0000009027	-	84705334	84740892	35558
8	ENSSSCG0000009028	-	84744077	84775299	31222

(3) 140일령 체중 형질을 GWAS 분석한 결과, 18개 상염색체중 4개의 염색체(SSC1, SSC5, SSC8, SSC12)에서 유의적인 결과를 얻을 수 있었으며(그림 22), 신뢰구간내에 존재하는 후보유전자를 선정하였다(표 3).

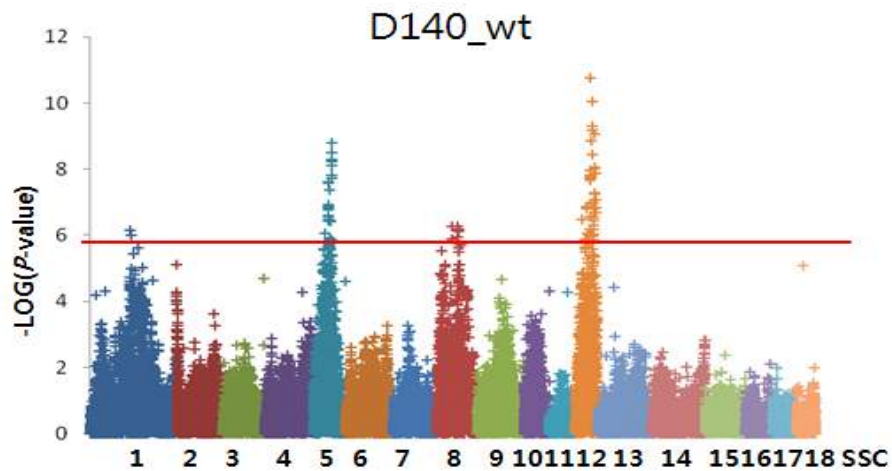


그림 22. Plink program의 GRAMMAR를 이용한 140일령 체중의 GWAS 결과.

표 3. 140일령 체중 관련 후보 유전자

SSC	GENE name	Locus Description	Position		Size (bp)
			Start	End	

1	ENSSSCG0 0000000596	-	109656164	109656469	305
1	SNX1	sorting nexin 1	111891339	111912700	21361
5	ANO2	anoctamin 2	65606148	65792035	185887
5	CCND2	cyclin D2	66783428	66800689	17261
5	RERGL	RERG/RAS-like	56684048	56691935	7887
5	PIK3C2G	phosphoinositide-3-kinase, class 2, gamma polypeptide	56036540	56437196	400656
5	GRIN2B	glutamate receptor, ionotropic, N-methyl D-aspartate 2B	60319131	60377552	58421
5	DERA	deoxyribose-phosphate aldolase (putative)	58186765	58236711	49946
5	ENSSSCG0 0000000596	-	57494638	57494992	354
5	EPS8	epidermal growth factor receptor pathway substrate 8	58475496	58544560	69064
5	CAPZA3	capping protein (actin filament) muscle Z-line, alpha 3	55968272	55969168	896
5	PLCZ1	phospholipase C, zeta 1	55969467	56013691	44224
5	ATF7IP	activating transcription factor 7 interacting protein	59598393	59692824	94431
8	ARFIP1	ADP-ribosylation factor interacting protein 1	79998755	80025983	27228
8	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	44149286	44277631	128345
8	ENSSSCG0 0000009027	-	84705334	84740892	35558
8	ENSSSCG0 0000009028	-	84744077	84775299	31222
8	TRIM2	tripartite motif-containing 2	79412505	79657444	244939
12	ABR	active BCR-related gene	49405379	49452237	46858
12	NXN	nucleoredoxin	49656874	49661419	4545
12	GOSR1	golgi SNAP receptor complex member 1	48899153	48951254	52101
12	CPD	carboxypeptidase D	46190757	46191961	1204
12	TUSC5	tumor suppressor candidate 5	49054052	49167629	113577
12	DNAH2	dynein, axonemal, heavy chain 2	55412156	55580859	168703
12	ACCN1	amiloride-sensitive cation channel 1, neuronal	43990589	44180280	189691
12	ENSSSCG0 0000018012	-	59113281	59119520	6239
12	ENSSSCG0 0000017990	-	56857023	56918670	61647
12	ARHGAP44	Rho GTPase activating protein 44	60299474	60663769	364295
12	ENSSSCG0 0000018641	-	54707338	54707406	68
12	SCPEP1	serine carboxypeptidase 1	34292966	34321478	28512
12	ZZEF1	zinc finger, ZZ-type with EF-hand domain 1	52220579	52258457	37878
12	TMEM100	transmembrane protein 100	32917755	32918159	404
12	CCL8	chemokine (C-C motif) ligand 8	42913034	42915054	2020
12	RABEP1	rabaptin, RAB GTPase binding effector protein 1	53759621	53837247	77626
12	NSF	N-ethylmaleimide-sensitive factor	18329585	18472697	143112
12	SNORA11	Small nucleolar RNA SNORA11	26141808	26141937	129

12	ATP2A3	ATPase, Ca ⁺⁺ transporting, ubiquitous	52291776	52301270	9494
12	SPNS3	spinster homolog 3 (Drosophila)	52577890	52661287	83397
12	C17orf102	Uncharacterized protein C17orf102	42546872	42627378	80506
12	ENSSSCG0 0000017624	-	35271294	35271681	387
12	ANKFY1	ankyrin repeat and FYVE domain containing 1	52366443	52419982	53539
12	DPH1	DPH1 homolog (S. cerevisiae)	50418000	50429494	11494
12	ENSSSCG0 0000017834	-	50967346	50978247	10901

(4) 도체중 형질을 GWAS 분석한 결과, 12번 염색체(SSC12)에서 유의적인 결과를 얻을 수 있었으며(그림 23), 신뢰구간내에 존재하는 7개의 후보유전자를 선정하였다(표 4).

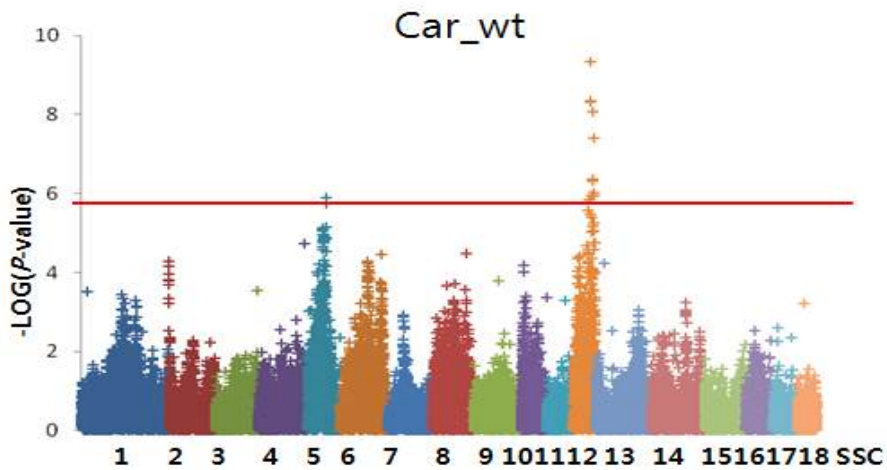


그림 23. Plink program의 GRAMMAR를 이용한 도체중의 GWAS 결과.

표 4. 도체중 후보 유전자

SSC	GENE name	Locus Description	Position		Size (bp)
			Start	End	
12	ABR	active BCR-related gene	49405379	49452237	46858
12	NXN	nucleoredoxin	49656874	49661419	4545
12	GOSR1	golgi SNAP receptor complex member 1	48899153	48951254	52101
12	ENSSSCG0 0000018012	-	59113281	59119520	6239
12	DNAH2	dynein, axonemal, heavy chain 2	55412156	55580859	168703
12	ARHGAP44	Rho GTPase activating protein 44	60637019	60663769	26750
12	SNORA11	Small nucleolar RNA SNORA11	54685114	54685243	129

(5) 흉추 형질을 GWAS 분석한 결과, 7번 염색체(SSC7)에서 유의적인 결과를 얻을 수 있었으며(그림 24), 신뢰구간내에 존재하는 후보유전자를 선정하였다(표 5).

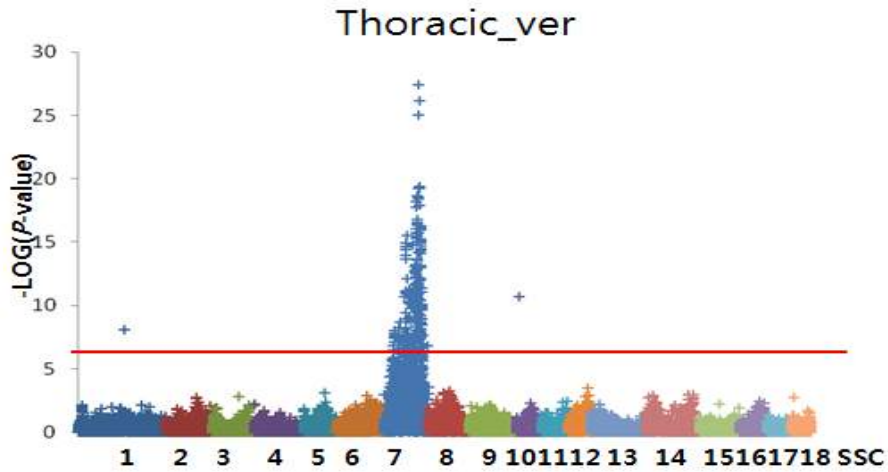


그림 24. Plink program의 GRAMMAR를 이용한 흉추의 GWAS 결과.

표 5. 흉추의 후보 유전자

SS C	GENE name	Locus Description	Position		Size(bp)
			Start	End	
7	NPC2	Niemann-Pick disease, type C2	103015142	103024543	9401
7	LTBP2	latent transforming growth factor beta binding protein 2	103032016	103049725	17709
7	FOS	FBJ murine osteosarcoma viral oncogene homolog	103806218	103809651	3433
7	ENSSSCG0 0000002331		100438672	100438834	162
7	VASH1	vasohibin 1	105338094	105351561	13467
7	ANGEL1	angel homolog 1 (Drosophila)	105368305	105375052	6747
7	C14orf181	Putative uncharacterized protein C14orf181	97910581	97910889	308
7	MED6	mediator complex subunit 6	100121356	100138686	17330
7	TTC9	tetratricopeptide repeat domain 9	100049645	100076943	27298
7	ENSSSCG0 0000018472		95325356	95325482	126
7	ESRRB	estrogen-related receptor beta	105064660	105121325	56665
7	SPTB	spectrin, beta, erythrocytic	94138307	94215667	77360
7	RAD51L1	RAD51-like 1 (S. cerevisiae)	97078068	97121755	43687
7	ARG2	arginase, type II	96828246	96871260	43014
7	VTI1B	vesicle transport through interaction with t-SNAREs homolog 1B (yeast)	96871224	96899333	28109
7	RDH11	retinol dehydrogenase 11 (all-trans/9-cis/11-cis)	96901995	96938051	36056
7	ZFYVE26	zinc finger, FYVE domain containing 26	97006134	97076681	70547
7	ADCK1	aarF domain containing kinase 1	106390882	106458616	67734
7	C14orf145	Uncharacterized protein C14orf145	108986597	109160948	174351

7	PLEKHG3	pleckstrin homology domain containing, family G (with RhoGef domain) member 3	94087305	94133693	46388
7	KIAA0317	KIAA0317	103168023	103175480	7457
7	ENSSSCG0000002284		95640239	95777459	137220
7	GPHN	gephyrin	96005710	96320864	315154
7	SCAMP2	secretory carrier membrane protein 2	63480022	63497161	17139
7	FNTB	farnesyltransferase, CAAX box, beta	94464431	94506219	41788
7	ENSSSCG0000002304		98450768	98451142	374
7	HDDC3	HD domain containing 3	58870630	58872078	1448
7	ENSSSCG0000002253		87201446	87201541	95
7	MESP1	mesoderm posterior 1 homolog (mouse)	60466115	60467168	1053
7	ANPEP	alanyl (membrane) aminopeptidase	60494911	60509828	14917
7	WDR93	WD repeat domain 93	60408127	60450712	42585
7	ENSSSCG0000018850		100747051	100747132	81
7	ENSSSCG0000001830		59508735	59509625	890
7	LINGO1	leucine rich repeat and Ig domain containing 1	62596677	62598521	1844
7	ENSSSCG0000001869		62220409	62276748	56339
7	MTHFD1	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase	93844974	93907324	62350
7	ENSSSCG0000002260		93093837	93107380	13543
7	MPP5	membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5)	96363565	96528831	165266
7	7SK	7SK RNA	97978781	97979137	356
7	ST8SIA2	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 2	91363093	91395293	32200
7	SIPAIL1	signal-induced proliferation-associated 1 like 1	103362224	100923898	144482
7	SEC23A	Sec23 homolog A (S. cerevisiae)	65942951	66011356	68405
7	PCNX	pecanex homolog (Drosophila)	100253441	100366787	113346
7	FUT8	fucosyltransferase 8 (alpha (1,6) fucosyltransferase)	94817723	95104273	286550
7	MCTP2	multiple C2 domains, transmembrane 2	89386290	89388660	2370

(6) 체장 형질을 GWAS 분석한 결과, 5번 염색체(SSC5)와 17번 염색체(SSC17)에서 유의적인 결과를 얻을 수 있었으며(그림 25), 신뇌구간내에 존재하는 후보유전자를 선정하였다(표 6).

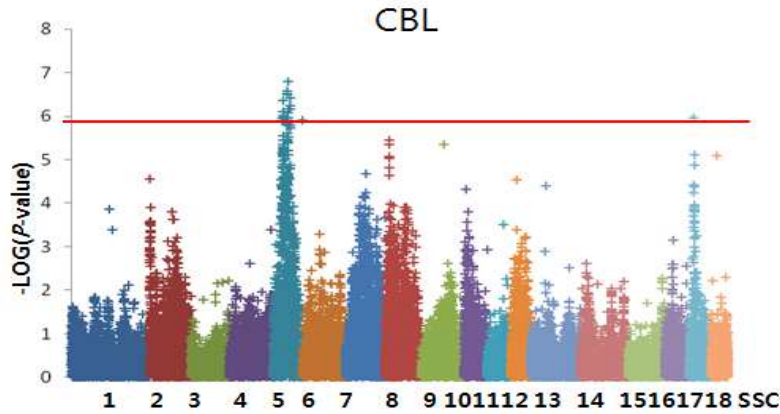


그림 25. Plink program의 GRAMMAR를 이용한 체장의 GWAS 결과.

표 6. 체장의 후보 유전자

SSC	GENE name	Locus Description	Position		Size(bp)
			Start	End	
5	GRIN2B	glutamate receptor, ionotropic, N-methyl D-aspartate 2B	60319131	60377552	58421
5	PIK3C2G	phosphoinositide-3-kinase, class 2, gamma polypeptide	56036540	56399781	363241
5	RERGL	RERG/RAS-like	56684048	56691935	7887
5	ANO2	anoctamin 2	65606148	65711636	105488
5	ENSSSCG0000018696	-	32518436	32518555	119
5	ENSSSCG000000957	-	34006825	34006892	67
5	GRIP1	glutamate receptor interacting protein 1	33045508	33368692	323184
5	RAB3IP	RAB3A interacting protein (rabin3)	35736559	35737983	1424
5	LEMD3	LEM domain containing 3	31845689	31889119	43430
17	PSD3	pleckstrin and Sec7 domain containing 3	14310395	14678894	368499

4. 후보 유전자와 형질간의 연관성 분석

가. 가. 유두수 관련 후보 유전자인 TBC1D21과 형질간의 연관성 분석

(1) TBC1D21 유전자는 TBC1 domain family중의 일원으로 Rab family protein (a member of the Ras superfamily of monomeric G protein) 의 GTPase의 활성을 증가시키는 역할을 하며, 외배엽 유래 간엽성 세포는 폐지의 유두 생성과정에서 관여를 하는 것으로 알려져 있으며, 유선이 형성되는 동안 상피조직의 간엽세포 상호작용으로 신호를 조절하는 것으로 알려져 있다. 따라서 이 유전자는 유두의 생성과 밀접한 관계가 있는 것으로 이미 mouse의 연구에 의해서 확인되었으며 유두수 관련 가능성이 가장 높은 QTL 영역내 후보 유전자로 판단되었다.

(2) TBC1D21 유전자의 exon 1에서 확인된 A>G mutation을 증폭하기 위한 primer 정보는 다음과 같다.

(가) TBC1D21-mini : CAGGATCAGGTCCAATGGGG

(나) TBC1D21-R : TATTCCTCGTTCCCTGCATC

(다) TBC1D21-F : [5BIOTIN]TCTAGCCCTGGCATTCTCTC

(3) 선정된 TBC1D21 유전자 내 SNP와 유두수 형질과의 연관성 분석에 앞서 유전자형의 분포 조사를 실시하였으며 한국재래돼지 19두와 랜드레이스 17두의 유전자형 선행 조사 결과는 표 7과 그림 26에 나타난 바와 같다. 즉, 한국재래돼지와 랜드레이스 교잡축군의 F0 개체들의 분포를 살펴본 결과 한국재래돼지의 TBC1D21 유전자는 A/A, A/G, G/G 유전자형이 각 5, 1, 10두의 비율로 나타났으며, 랜드레이스는 각 13, 3, 0두의 비율로 나타나 교배양식을 확인했을 때, F1과 F2에서 충분한 정보력이 있는 것으로 판단되었다. 또한 TBC1D21 유전자의 exon 1 SNP와 유두수 형질간의 연관성을 분석한 결과 유의적인 결과를 보여 이 SNP는 유두수의 개량에 이용될 수 있는 마커로서의 가능성이 있다고 판단되었다(표 8).

표 7. 한국재래돼지 × 랜드레이스 교잡 축군 F0 개체들의 TBC1D21 유전자형 분석 결과

Gene	Genotype	Korean native pig	Landrace
<i>TBC1D21</i> (Exon 1)	A/A	5	13
	A/G	1	3
	G/G	10	0

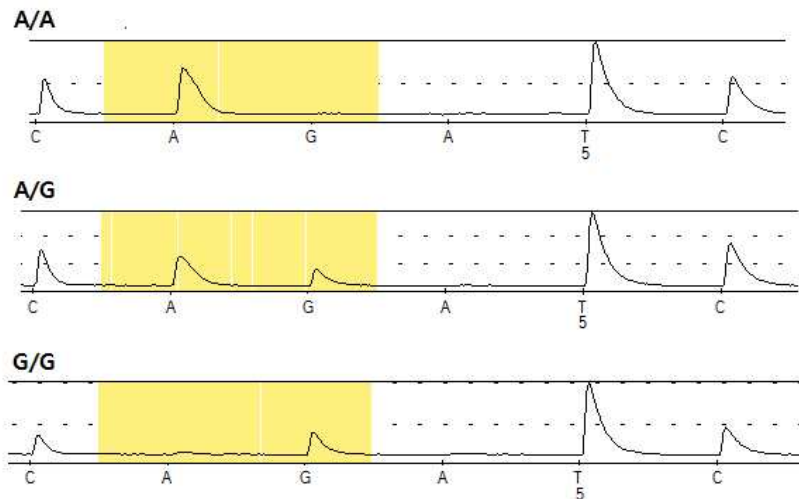


그림 26. TBC1D21의 exon 1 SNP 유전자형을 확인하기 위한 pyrosequencing 결과.

표 8. TBC1D21 유전자의 association 분석 결과

Trait	SSC	Location	Allele ¹	MAF ²	Model ³	P-value
Teat_num	7	64412799	G/A	0.3173	Add	2.69×10 ⁻⁴
Teat_R	7	64412799	G/A	0.3173	Add	2.59×10 ⁻⁴
Teat_L	7	64412799	G/A	0.3173	Add	1.62×10 ⁻²

¹major/minor allele; ²minor allele frequency; ³additive model.

나. 140일령 체중 관련 후보 유전자인 ANO2와 형질간의 연관성 분석

- (1) ANO2는 calcium-activated chloride channels (CaCCs) family중 하나로 알려져 있으며, 다양한 조직에서 발현됨이 확인되었다.
- (2) ANO2 유전자의 exon 1에서 확인된 A>G mutation을 증폭하기 위한 primer 정보는 다음과 같다.
 - (가) ANO2-mini : CCCAGGCTACTTGATGAACT
 - (나) ANO2-R : [5'BIOTIN]CCCACAGCGTAACAGACTCA
 - (다) ANO2-F : TGGCCTCTGTTCTCCTGTTT
- (3) 선정된 ANO2 유전자 내 SNP와 성장 형질과의 연관성 분석에 앞서 유전자형의 분포 조사를 실시하였으며 한국재래돼지 19두와 랜드레이스 17두의 유전자형 선형 조사 결과는 표 9과 그림 27에 나타난 바와 같다. 즉, 한국재래돼지의 ANO2 유전자는 C/C, C/G, G/G 유전자형이 각 7, 7, 5두의 비율로 나타났으며, 랜드레이스는 각 17, 0, 0두의 비율로 나타나 교배양식을 확인했을 때, F1과 F2에서 충분한 정보력이 있을 것으로 판단되었다. 또한 ANO2 유전자의 exon 1 SNP와 성장 형질간의 연관성을 분석한 결과 유의적인 결과를 보여 이 SNP는 성장형질 개량에 이용될 수 있는 마커로서의 가능성이 있다고 판단되었다 (표 10).

표 9. 한국재래돼지 × 랜드레이스 교잡 축군의 F0 개체들의 ANO2 유전자형 분석 결과

Gene	Genotype	Korean native pig	Landrace
ANO2(Exon 1)	C/C	7	17
	C/G	7	0
	G/G	5	0

표 10. ANO2 유전자의 association 분석 결과

Trait	SSC	Gene	Location	Allele ¹	MAF ²	Model ³	P-value
D140_wt	5	ANO3	65606210	G/C	0.2845	Add	1.19×10 ⁻³

¹major/minor allele; ²minor allele frequency; ³additive model.

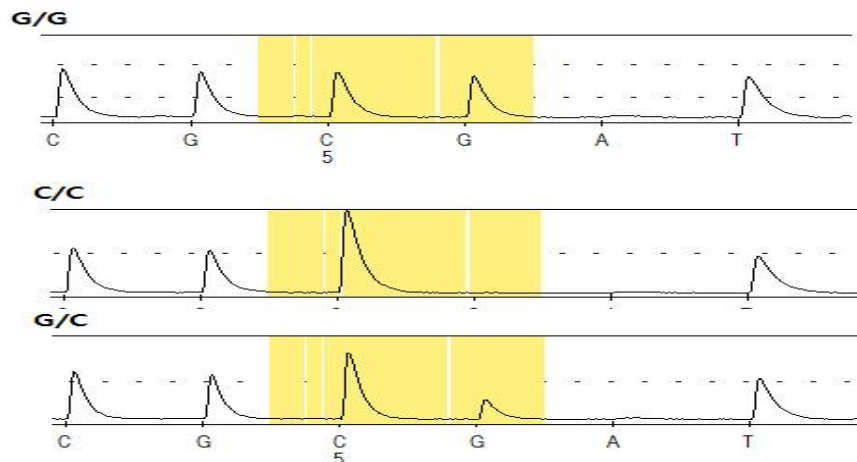


그림 27. ANO2의 exon 1 SNP 유전자형을 확인하기 위한 pyrosequencing 결과.

나. 갈비대수 관련 후보 유전자의 확보와 SNP 발굴

(1) 제주재래흑돼지와 랜드레이스 기반 참조축군을 활용하여 갈비대수와 관련된 QTL이 돼지 7번 염색체상에서 확인되었고, 그 영역에는 VRTN, LTBP2 유전자가 후보 유전자로 추정되었다. 기초축군에 대한 유전자형 분석결과에서 기존에 후보유전자로 보고된 VRTN 보다는 LTBP2 유전자가 갈비대수에 더 영향을 주는 것으로 확인되어 LTBP2 유전자에 염기서열 변이 발굴과 발굴된 SNP를 대상으로 형질간의 연관성을 확인하였다.

(2) LTBP2 유전자의 염기서열상의 돌연변이를 검출하기 위하여 제주재래흑돼지 5두와 랜드레이스 5두의 DNA를 pooling 후 각각 돼지 whole genome re-sequencing을 reference sequence와 비교하여 그 유전자 영역에서 SNP를 검출하였고, SNP에 대한 검정단계로서 PCR후 DNA sequencing에 의해 재차 돌연변이를 검출하였다(그림 28, 그림 29). LTBP2 유전자의 exon 13 - intron 13영역과 exon 32영역을 분석하기 위하여 프라이머 정보는 다음과 같다.

(가) LTBP2 exon 13에서 확인된 G>A mutation을 증폭하기 위한 primer 정보는 다음과 같음

- ① pLTe13F : 5'-CACGGCTACACCTACTCAAGC-3'
- ② B-pLTe13R : 5'-Biotin-GCTGAGTTGCTCCAATCTCC-3' ggagattggagcaactcagc
- ③ s-pLT13 : 5'-GGCACCTGGTTGGAGGCGG-3'
- ④ Program 입력 : [G/A]GACCAT

(나) LTBP2 exon 32에서 확인된 A>C mutation을 증폭하기 위한 primer 정보는 다음과 같음

- ① pLTe32F : 5'-GGAAACTGAGGCAAGGAGAA-3'
- ② B-pLTe32R : 5'-Biotin-TGTCCTCACCTCACACTTG-3'
- ③ s-pLTe32 : 5'-CGAGTGTGTGATGTTTCGGGC-3'
- ④ Program 입력 : [A/C]TGCTCT

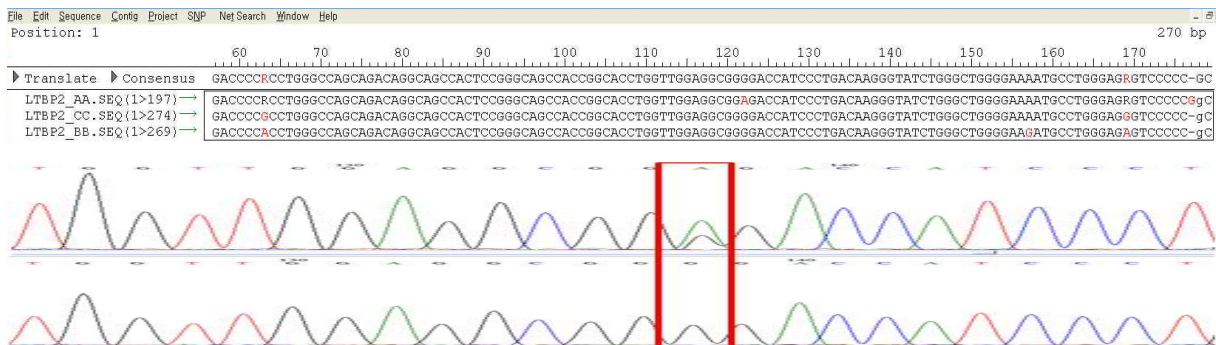


그림 28. LTBP2 exon 13에서 확인된 SNP.

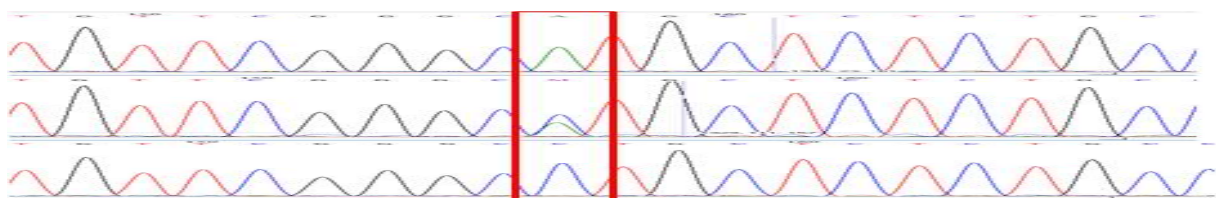


그림 29. LTBP2 exon 32에서 확인된 SNP.

5. 돼지 genome sequencing 발표후 database의 수정 및 후보유전자 재탐색

가. 돼지 genome sequencing 결과에 근거한 database의 수정

- (1) Porcine 60K chip내 SNP map position의 수정 보완하였다.
- (2) 기존 Ensembl(www.ensembl.org/index.html) ver.10에서 NCBI(www.ncbi.nih.gov) ver.10.2로 Porcine 60K chip map position을 변경하였다.

나. 후보유전자 재탐색 실시

- (1) 수정된 map에 따라 각 형질의 후보유전자 재탐색을 실시하였으며 대상 형질은 유두수 (좌측, 우측 및 전체), 체중 (생시, 21일령, 70일령, 140일령 및 도체중), 도체장 및 흉추수 등 10개 형질의 후보유전자 탐색을 실시하였다.
- (2) 그 결과 SSC5(140일령 *SW1482-SW963*, 도체장 *SW963-SW967*), SSC7(유두수 전체 *S0102-SW147*, 왼쪽·오른쪽 *207G8-4-SW147*, 흉추수 *SWR1928-SW632*), SSC8(140일령·도체중 *SW1345-S0086*), SSC12(140일령 *SW168-SE259162*, 도체중 *SW1962-SWR1021*)의 QTL region에서 GWAS 분석 결과 유의성이 높은 SNP 마커를 확인하였으며 확인된 신뢰구간 내의 후보 유전자는 표 11 및 표 12에 나타난 바와 같다.

표 11. 유두수(Teat_L, Teat_R, Teat_num) 및 흉추수(Thoracic_ver) 관련 후보 유전자

Traits*	SSC	GENE name	Locus Description	Position		Size (bp)
				Start	End	
		<i>SCAMP2</i>	Secretory Carrier Membrane Protein 2	63361065	63387733	26669
		<i>LOC100156276</i>	WD repeat-containing protein 93-like	60153369	60205210	51842
		<i>LOC100155863</i>	mesoderm posterior protein 1-like	60211072	60212402	1331
		<i>ANPEP</i>	Alanyl (Membrane) Aminopeptidase	60240145	60255061	14917
		<i>HDDC3</i>	HD Domain Containing 3	58557178	58559084	1907
		<i>BCL2A1</i>	BCL2-Related Protein A1	54395400	54405839	10440
		<i>PEAK1</i>	Pseudopodium-Enriched Atypical Kinase 1	61961070	62180499	219430
		<i>HMG20A</i>	High Mobility Group 20A	62233324	62261716	28393
		<i>LOC100738147</i>	uncharacterized LOC100738147	65750368	65868070	117703
		<i>CCND3</i>	Cyclin D3	42489107	42493235	4128
Teat_num (전체, 좌측, 우측)	7	<i>LOC100738461</i>	protein KIAA1199 homolog	55313175	55375488	62313
		<i>LOC100152264</i>	myeloid-associated differentiation marker-like	59241812	59261546	19735
thoracic_ver		<i>C7H15orf26</i>	Chromosome 15 Open Reading Frame 26	55527119	55540522	13404
		<i>CCDC37</i>	Coiled-Coil Domain Containing 37	58791150	58820990	29841
		<i>LOC100512114</i>	40S ribosomal protein S24-like	54994461	55049618	55158
		<i>LOC100155878</i>	outer dense fiber protein 3-like protein 1-like	62627387	62630699	3313
		<i>LOC100516621</i>	pro-neuregulin-4, membrane-bound isoform-like	60799892	60869908	70017
		<i>ETFA</i>	Electron-Transfer-Flavoprotein, Alpha Polypeptide	60946251	61036017	89767
		<i>LOC100156419</i>	homer 2	57489012	57584450	95439
		<i>ARNT2</i>	Aryl-Hydrocarbon Receptor Nuclear Translocator 2	54768574	54955626	187053

*Teat_num : 유두수, Thoracic_ver : 흉추수

표 12. 체중(D140_wt, CBL, Car_wt) 관련 후보 유전자

Traits*	SSC	GENE name	Locus Description	Position		Size (bp)
				Start	End	
D140_wt CBL	5	<i>ANO2</i>	Anoctamin 2	67262865	67280819	17955
		<i>CCND2</i>	Cyclin D2	68314784	68331944	17161
		<i>NTF3</i>	Neurotrophin 3	67204186	67204959	773
		<i>LOC100153016</i>	ras-related and estrogen-regulated growth inhibitor-like protein-like	58237478	58245717	8239
		<i>RECQL</i>	RecQ Protein-Like (DNA Helicase Q1-Like)	58163512	58164743	1232
		<i>GRIN2B</i>	Glutamate Receptor, Ionotropic, N-Methyl D-Aspartate 2B	61956145	62272470	316326
		<i>PLEKHA5</i>	Pleckstrin Homology Domain Containing, Family A Member 5	57056542	57298868	242327
		<i>PIK3C2G</i>	Phosphatidylinositol-4-Phosphate 3-Kinase, Catalytic Subunit Type 2 Gamma	57747752	58107859	360108
		<i>LOC100525576</i>	60S ribosomal protein L18-like	59826275	59836755	10481
		<i>LOC100154654</i>	centrin-2-like	59187848	59189626	1779
		<i>EPS8</i>	Epidermal Growth Factor Receptor Pathway Substrate 8	60303025	60391705	88681
		<i>LOC100525748</i>	epidermal growth factor receptor kinase substrate 8-like	60212994	60223404	10411
		<i>LOC100738345</i>	uncharacterized LOC100738345	33473741	33539724	65984
		<i>LEMD3</i>	LEM Domain Containing 3	32770431	32896643	126213
		<i>WIF1</i>	WNT Inhibitory Factor 1	32656315	32717690	61376
		<i>LOC100514460</i>	cullin-associated NEDD8-dissociated protein 1-like	34956247	34995104	38858
		<i>LOC100514286</i>	glutamate receptor interacting protein 1	33971320	34295811	324492
<i>LOC100156985</i>	uncharacterized LOC100156985	31262271	31417436	155166		
<i>TMEM5</i>	Transmembrane Protein 5	31433527	31462384	28858		
D140_wt car_wt	8	<i>LOC100518024</i>	ADP-ribosylation factor interacting protein 1	80737492	80791567	54076
		<i>KIT</i>	V-Kit Hardy-Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog	43550231	43601377	51147
		<i>NR3C2</i>	Nuclear Receptor Subfamily 3, Group C, Member 2	85761679	85798666	36988
		<i>LOC100521464</i>	rho GTPase-activating protein 10-like	85901094	86044524	143431
		<i>TRIM2</i>	Tripartite Motif Containing 2	80265928	80393652	127725

		<i>LOC100049650</i>	myosin	43931996	44246596	314601
		<i>LOC100523510</i>	nucleoredoxin-like	49189585	49239689	50105
		<i>LOC100621652</i>	uncharacterized LOC100621652	49052343	49053360	1018
		<i>LOC100521749</i>	uncharacterized LOC100521749	48801703	48804484	2782
		<i>LOC100736982</i>	myosin-13-like	57808734	57933831	125098
		<i>WSB1</i>	WD Repeat And SOCS Box Containing 1	45667728	45686850	19123
		<i>LOC100517676</i>	Golgi SNAP receptor complex member 1-like	48631022	48683129	52108
		<i>DNAH2</i>	Dynein, Axonemal, Heavy Chain 2	55123644	55190268	66625
		<i>LOC100621981</i>	myosin-10-like	56343507	56433106	89600
		<i>LOC100627824</i>	amiloride-sensitive cation channel 1, neuronal-like	42601940	42602829	890
D140_wt car_wt	12	<i>MYHC</i>	Myosin, Heavy Chain 8, Skeletal Muscle, Perinatal	58110959	58133067	22109
		<i>PIK3R6</i>	Phosphoinositide-3-Kinase, Regulatory Subunit 6	56662053	56723008	60956
		<i>MYOCD</i>	Myocardin	59867285	60048521	181237
		<i>BCL6B</i>	B-Cell CLL/Lymphoma 6, Member B	54708169	54711360	3192
		<i>LOC100628008</i>	amiloride-sensitive cation channel 1, neuronal-like	43444507	43633460	188954
		<i>SCPEP1</i>	Serine Carboxypeptidase 1	33993049	34027377	34329
		<i>LOC100621291</i>	zinc finger ZZ-type and EF-hand domain-containing protein 1-like	51941150	51997480	56331
		<i>LOC100736688</i>	uncharacterized LOC100736688	32244531	32268860	24330
		<i>LOC100038010</i>	chemokine (C-C motif) ligand 11	42424290	42427337	3048
		<i>NSF</i>	N-Ethylmaleimide-Sensitive Factor	17490814	17655336	164523

*D140_wt : 140일령 체중, CBL : 도체장, car_wt : 도체중

다. 신뢰 구간 내의 형질 관련 후보 유전자, SNP의 선발 및 형질과의 연관성 분석

- (1) 각 형질별 신뢰 구간 내의 유전자를 대상 형질과 연결하여 최종 13개의 후보 유전자를 선발하였다. 이들 후보 유전자의 정보는 표 13에 나타난 바와 같으며 각 후보유전자의 exon 영역 내 변이 탐색을 위하여 direct sequencing 실시하였으며 그 결과는 표 14에 나타난 바와 같다. 본 연구에서는 우선적으로 아미노산의 변이를 가져오는 missense mutation을 확인하였는데 총 7개의 유전자에서 11개의 missense mutation을 확인할 수 있었다. 이 missense mutation을 포함하여 총 49개의 변이가 veracode SNP genotyping chip을 이용한 유전자형 분석 이용되었다.
- (2) 후보 유전자내 SNP와 형질과의 연관성을 분석하기 위하여 Bonferroni method에 따라 5% genome-wide threshold를 $P\text{-value} = 1.6 \times 10^{-5}$ 로 설정하였으며 GRIN2B_SNP07, ANO2_SNP06, LOC100621652_SNP05 및 LOC100523510_SNP08은 형질과 관련성이 유의적이지 않았다(표 15).
- (3) 유두수 관련 형질(Teat_num, Teat_L, Teat_R)은 SSC7에서 11개의 마커(HDDC3_SNP02, LOC100156276_SNP04, LOC100156276_SNP05, LOC100156276_SNP09, LOC100156276_SNP15, LOC100155863_SNP01, LOC100155863_SNP02, ANPEP_SNP01,

ANPEP_SNP02, SCAMP2_SNP01, SCAMP2_SNP02)가 유의적으로 확인되었으며 흉추 (Thoracic) 또한 SSC7에서 11개의 동일한 마커가 유의적으로 확인되었다(표 16).

- (4) 140일령 체중(D140_wt) 및 도체중(Car_wt)과 관련하여 SSC8에 존재하는 SNP는 연관이 없는 것으로 확인되었으며 SSC12에서 LOC100621652_SNP01와 LOC100523510_SNP01가 유의적인 연관성이 있는 것으로 확인되었다(표 17, 표 18).

표 13. SNP genotyping chip 제작에 선정된 후보유전자와 대상 형질

SSC	GENE name	Position		Target trait
		Start	End	
5	<i>GRIN2B</i>	61956145	62272470	D140_wt, CBL
	<i>NTF3</i>	67204186	67204959	
	<i>ANO2</i>	67262865	67280819	
7	<i>HDD3</i>	58557178	58559084	Teat_num, Teat_L, Teat_R, Thoracic_ver
	<i>LOC100156276</i>	60153369	60205210	
	<i>LOC100155863</i>	60211072	60212402	
	<i>ANPEP</i>	60240145	60255061	
8	<i>SCAMP2</i>	63361065	63387733	D140_wt, Car_wt
	<i>NR3C2</i>	85761679	85798666	
12	<i>LOC100521464</i>	85901094	86044524	D140_wt, Car_wt
	<i>LOC100049650</i>	43931996	44246596	
	<i>LOC100621652</i>	49052343	49053360	
	<i>LOC100523510</i>	49189585	49239689	

표 14. 선발된 후보 유전자 내 단백질 변이 SNP 정보

SSC	Gene	Description	Missense mutation (ea)
5	<i>ANO2</i>	Anoctamin 2	2
	<i>GRIN2B</i>	Glutamate Receptor, Ionotropic, N-Methyl D-Aspartate 2B	2
7	<i>ANPEP</i>	Alanyl (Membrane) Aminopeptidase	1
	<i>HDDC3</i>	HD Domain Containing 3	1
	<i>LOC100156276</i>	WD repeat-containing protein 93-like	1
12	<i>LOC100049650</i>	myosin	3
	<i>LOC100621652</i>	uncharacterized LOC100621652	1

표 15. SSC5에서 선발된 후보 유전자와 각 형질과의 연관성 분석 결과

Locus Name	Position	Allele ¹	MAF ²	P-value	
				D140_wt	CBL
GRIN2B_SNP01	61956250	A/C	0.4405	5.9×10 ⁻⁴	8.4×10 ⁻⁵
GRIN2B_SNP02	61956444	G/A	0.3421	0.029	0.002
GRIN2B_SNP07	62261653	A/0	0	NA	NA
GRIN2B_SNP10	62270641	A/G	0.4459	0.002	2.9×10 ⁻⁴
GRIN2B_SNP11	62271274	G/A	0.0092	0.851	0.178
GRIN2B_SNP13	62272223	G/A	0.3421	0.080	0.003
NTF3_SNP01	67204243	A/G	0	0.033	3.4×10 ⁻⁴
NTF3_SNP02	67204866	G/A	0.0476	0.183	0.089
ANO2_SNP06	67263202	A/0	0	NA	NA
ANO2_SNP09	67277712	G/A	0.3222	2.8×10 ⁻⁴	6.4×10 ⁻⁴
ANO2_SNP10	67279725	G/C	0.3222	3.5×10 ⁻⁴	4.9×10 ⁻⁴

¹major/minor allele; ²minor allele frequency

표 16. SSC7에서 선발된 후보 유전자와 각 형질과의 연관성 분석 결과

Locus Name	Position	Allele ¹	MAF ²	P-value			
				Teat_num	Teat_L	Teat_R	Thoracic
HDDC3_SNP01	58557751	G/A	0.2614	3.4×10 ⁻⁵	6.8×10 ⁻⁴	4.1×10 ⁻⁴	0.023
HDDC3_SNP02	58558358	A/G	0.4583	1.1×10 ⁻¹⁰	2.1×10 ⁻⁶	1.6×10 ⁻⁹	2.0×10 ⁻⁹
LOC100156276_SNP04	60158278	A/C	0.45	1.4×10 ⁻¹¹	4.1×10 ⁻⁷	4.1×10 ⁻¹⁰	1.6×10 ⁻⁹
LOC100156276_SNP05	60159359	G/A	0.4271	5.0×10 ⁻¹⁰	4.3×10 ⁻⁶	5.7×10 ⁻⁹	3.3×10 ⁻⁹
LOC100156276_SNP09	60190906	G/A	0.4583	6.1×10 ⁻¹¹	1.4×10 ⁻⁶	8.5×10 ⁻¹⁰	1.3×10 ⁻⁹
LOC100156276_SNP15	60195826	G/A	0.45	7.5×10 ⁻¹²	3.6×10 ⁻⁷	1.9×10 ⁻¹⁰	7.9×10 ⁻¹⁰
LOC100155863_SNP01	60211081	A/C	0.45	1.6×10 ⁻¹¹	2.9×10 ⁻⁷	8.1×10 ⁻¹⁰	3.1×10 ⁻⁹
LOC100155863_SNP02	60211298	G/A	0.45	4.9×10 ⁻¹¹	1.3×10 ⁻⁶	6.7×10 ⁻¹⁰	4.2×10 ⁻⁹
ANPEP_SNP01	60240180	G/A	0.2788	2.7×10 ⁻¹⁰	4.2×10 ⁻⁶	2.5×10 ⁻⁹	4.4×10 ⁻¹²
ANPEP_SNP02	60240311	A/G	0.45	4.4×10 ⁻¹²	2.2×10 ⁻⁷	1.8×10 ⁻¹⁰	3.2×10 ⁻⁹
ANPEP_SNP03	60240423	C/G	0.09677	2.3×10 ⁻⁴	0.009	3.4×10 ⁻⁴	5.9×10 ⁻⁵
ANPEP_SNP04	60249084	G/A	0.1058	0.396	0.455	0.502	0.626
ANPEP_SNP05	60254918	G/A	0.01818	0.460	0.836	0.136	0.914
SCAMP2_SNP01	63375389	G/A	0.4894	1.5×10 ⁻¹¹	3.2×10 ⁻⁶	2.6×10 ⁻¹¹	5.0×10 ⁻¹⁰
SCAMP2_SNP02	63386471	G/A	0.4894	1.2×10 ⁻¹¹	9.6×10 ⁻⁷	1.0×10 ⁻¹⁰	3.7×10 ⁻¹⁰
SCAMP2_SNP03	63386732	A/G	0	0.008	0.080	0.007	0.011
SCAMP2_SNP04	63386883	C/G	0.04464	0.188	0.166	0.424	0.004
SCAMP2_SNP05	63387533	A/G	0.2234	0.010	0.149	0.004	1.9×10 ⁻⁴

¹major/minor allele; ²minor allele frequency

표 17. SSC8에서 선발된 후보 유전자와 각 형질과의 연관성 분석 결과

Locus Name	Position	Allele ¹	MAF ²	P-value	
				D140_wt	Car_wt
NR3C2_SNP01	85761775	G/A	0.3571	0.056	0.116
LOC100521464_SNP01	86038187	A/G	0.4444	0.004	0.392
LOC100521464_SNP02	86044262	C/G	0.5	0.008	0.554
LOC100521464_SNP03	86044390	G/A	0.4255	0.001	0.136

¹major/minor allele; ²minor allele frequency

표 18. SSC12에서 선발된 후보 유전자와 각 형질과의 연관성 분석 결과

Locus Name	Position	Allele ¹	MAF ²	<i>P</i> -value	
				D140_wt	Car_wt
LOC100049650_SNP01	43932005	G/A	0.4898	0.348	0.698
LOC100049650_SNP10	43986161	A/G	0.0273	0.703	0.895
LOC100049650_SNP11	43992701	A/G	0.4674	0.257	0.758
LOC100049650_SNP16	44075868	G/A	0.1818	0.002	0.110
LOC100049650_SNP17	44083467	C/A	0.4574	0.244	0.742
LOC100049650_SNP18	44084471	G/A	0.2442	0.026	0.039
LOC100049650_SNP19	44131400	A/G	0.3448	0.020	0.003
LOC100049650_SNP20	44246569	G/A	0.2949	0.005	0.005
LOC100621652_SNP01	49052355	A/G	0.3111	6.5×10^{-9}	4.0×10^{-8}
LOC100621652_SNP05	49053130	G/G	0	NA	NA
LOC100523510_SNP01	49189608	G/A	0.3111	2.9×10^{-9}	4.1×10^{-8}
LOC100523510_SNP07	49233897	A/G	0.1489	5.0×10^{-5}	4.2×10^{-5}
LOC100523510_SNP08	49238356	A/C	NA	1	1
LOC100523510_SNP09	49238480	G/C	0.1321	1.6×10^{-5}	4.9×10^{-5}
LOC100523510_SNP11	49238850	A/C	0.1321	1.1×10^{-5}	9.4×10^{-6}
LOC100523510_SNP12	49239102	A/C	NA	0.030	0.518

¹major/minor allele; ²minor allele frequency

제 2 절 한국재래돼지의 성장 및 포육능력관련 마커를 이용한 유전능력 평가 및 선발체계 확립

1. 국산 재래돼지 종축의 맞춤형 통계분석모형 개발

가. 공시재료

본 연구에 사용된 공시축은 제주난지시험장 내 시험축으로 순종돈 36두(재래돼지(K) : 19두, 랜드레이스(L) : 17두)이며 이들간 교배를 통해 F1세대 91두 (KL : 43두, LK : 48두)를 생산하였고, F1간 교배를 통해 F2세대 (KLKL : 31두, KLLK : 585두, LKKL, 489두)를 생성하였고 이들 축군에 대해 번식자료, 성장자료 및 도축자료를 수집하여 연구에 이용하였다. 여기서 F1집단에서 KL은 모계가 재래돼지(K), 부계가 랜드레이스(L)에서 태어난 자손 축군을 지칭한다.

나. 연구방법

(1) 공시축의 표현형 데이터에 대한 기초통계량 계산

기초통계량은 빈도, 산술평균, 표준편차, 최소값, 최대값을 이용하여 계산하였으며 보편적으로 사용하는 방법이기 때문에 수식에 대한 정보는 별도로 표기하지 않았다.

(2) 표현형 형질별 환경요인 분석

집단 개념의 개체별 유전능력을 평가하기 위하여 각 표현형 형질별 영향하는 환경요인에 대하여 알아보고 유전능력평가 모형을 설정하기 위하여 분석을 실시하였다.

환경요인분석을 위하여 SAS V. 9.2의 일반선형모형(proc glm)을 사용하였으며 분석을 위한 통계모형은 형질에 따라 아래와 두 가지형태로 구분하여 설정하였다.

(가) 체중 및 유두수 형질

$$\text{분석모형} : y = \mu + \text{sex} + \text{PAbreed} + p + t + e$$

여기서, y : 관측치 벡터(유두수형질, 체중형질)

μ : 전체평균

sex : 개체성별효과

PAbreed : 부모(F1)의 품종효과

p : 모 산차효과

t : 검정차수효과

e : 잔차효과

(나) 도체형질

$$\text{분석모형} : y = \mu + \text{sex} + \text{PAbreed} + p + t + \beta_1 \text{ca} + e$$

여기서, y : 관측치 벡터(도체형질)

μ : 전체평균

sex : 개체성별효과

PAbreed : 부모(F1)의 품종효과

p : 모 산차효과

t : 검정차수효과

β1 : 도축시 일령에 대한 공변량 효과

다. 연구결과

(1) 공시축의 표현형 데이터에 대한 기초통계량 계산

형질별 표현형 자료의 이해 및 이상치 자료 제거를 위해 기초통계량을 분석하였으며, 그 결과는 표 19와 같다

표 19 재래돼지의 성장형질 및 포육능력형질에 대한 기초통계량

Traits	N	MEAN	SD	MIN	MAX
유두수	1,196	13.67	1.30	10	18
좌측 유두수	1,196	6.81	0.74	5	10
우측 유두수	1,196	6.85	0.74	5	9
생시체중	1,196	1.51	0.29	0.6	2.6
21일령 체중	1,196	6.32	1.39	1.5	10.2
70일령 체중	1,193	25.46	5.44	8	41
140일령 체중	1,194	72.82	10.54	31	115
도체중	1,081	79.17	12.48	29	128
갈비뼈 수	1,069	15.25	0.64	14	17
허리뼈 수	1,065	6.03	0.39	5	7
도체장	1,079	102.64	5.79	82	123
도축연령	1,049	198.55	10.34	165	215

성장형질(체중) 및 유두수의 경우 F0, F1에서 측정이 가능한 반면 도체관련 형질(도체장, 허리뼈 수 등)은 도축되지 않은 F0, F1에서 관측되지 않았다. 유두수의 경우 최소 10개에서 최대 18개로 분포하였으며, 체중의 경우 외래돼지품종(랜드레이스, 요크셔 등)에 비해 낮게 나타났다.

(2) 표현형 형질별 환경요인 분석

가축의 경제적 형질(체중, 일당증체량, 등지방 두께, 산자수 등)은 크게 유전적인 요인과 환경적인 요인에 의하여 영향을 받는다. 환경적요인(사양관리, 계절효과 등)은 농장주, 기후 및 지리적 요건 등에 의하여 영향을 받을 수 있기 때문에 일반적으로 개체의 능력은 부모로부터 물려받는 유전부분의 수치를 육종가로 정의하여 개체를 평가한다. 따라서 표현형 능력에서 환경요인에 대한 요인을 제거하는 것은 상당히 중요하다. 개체 유전평가 시 환경요인에 대하여 보정이 이루어지지 않을 시 설명되지 않는 환경요인이 유전능력에 포함되어 추정된 육종가의 정확도가 떨어지게 될 것이다.

표 20 전체 유두수, 좌측 유두수 및 우측 유두수에 대한 분산분석표

Source	DF	Mean Square		
		전체유두수	좌측유두수	우측 유두수
교배품종	1	5.06	0.58	2.2 *
성별	1	4.05	2.14	0.3
모산차	2	13.82 **	2.96 **	4.02 **
검정차수	21	3.33 **	1.09 **	0.89 *

** : significant (P<0.01), * : significant (P<0.05)

유두수(좌, 우, 전체)의 경우 모산차 및 검정차수 효과에 영향을 받는 것으로 P<0.05로 유의하게 나타났으며, 우측 유두수의 경우 모산차, 검정차수뿐 아니라 교배품종효과에서도 P<0.05로 유의하게 나타났다. 교배품종이 KL×LK일 때, 우측유두수의 평균은 6.88, LK×KL일 때 6.81로 교배품종이 KL×LK일 때 우측유두수의 성적이 우수한 것으로 나타났으며 이는 좌측유두수 및 전체유두수에서도 마찬가지로 KL×LK 품종 LK×KL에 비해 상대적으로 성적이 우수한 것으로 나타났다(표 20)

표 21. 생시체중, 21일 체중, 70일 체중 및 140일 체중에 대한 분산분석표

Source	DF	Mean Square			
		생시체중	21일령 체중	70일령 체중	140일령 체중
교배품종	1	0.78 **	25.41 **	92.54 *	152.52
성별	1	0.74 **	16.86 **	3.43	968.07 **
모산차	2	0.74 **	7.42 **	55.68 *	74
검정차수	21	0.18 **	10.91 **	350.67 **	322.71 **

** : significant (P<0.01), * : significant (P<0.05)

생시체중, 21일체중의 경우 교배품종, 성별, 모산차 및 검정차수의 효과가 표현형에 영향을 받는 것으로 나타났으며(P<0.05), 70일령 체중의 경우 성별효과에서만 유의차가 나타나지 않았으며(P<0.05), 140일령 체중의 경우 성별 및 검정차수에서 유의적 차이를 보였다(P<0.05). 140일령체중을 제외하고 나머지 체중에서 교배품종이 유의적으로 영향을 받는 것으로 나타났으며, 유의차가 나타난 체중에 대해 Duncan 분석한 결과 유두수와 마찬가지로 KLLK교배품종이 LKKL교배품종에 비해 체중이 우수한 것으로 나타났다(표 21).

표 22. 도체중, 갈비뼈 수, 허리뼈 수 및 도체장에 대한 분산분석표

Source	DF	Mean Square			
		도체중	갈비뼈 수	허리뼈 수	도체장
교배품종	1	133.25	2.55 *	0.39	33.84
성별	1	236.62	0.39	0.54	128.18 *
모산차	2	10.97	3.06 **	1.39 **	25.27
검정차수	21	498.61 **	0.9 **	0.4 **	282.13 **
도축시 연령	1	4641.27 **	0.28	0.78 *	708.59 **

** : significant (P<0.01), * : significant (P<0.05)

도축형질은 도축시 연령에 대해 영향을 받기 때문에 유두수나 체중형질과 다르게 도축시 연령을 추가적으로 모델이 설정하여 분산분석을 실시하였다. 검정차수효과의 경우 모든 형질에서 유의차를 보였으며(P<0.01), 도축시 연령의 경우 도체중, 허리뼈 수 및 도체장에 영향을 미치는 것으로 나타났다(P<0.05)(표 22).

위에서 분석된 재래돼지의 성장 및 포육능력형질에 영향하는 환경요인들은 유전변이 및 개체별 육종가 추정모델로 사용가능하며, 또한 양적형질좌위 발굴을 위한 표현형과의 연관성 분석에서 환경요인 보정을 위한 모델로도 사용가능하다.

2. 유전체 정보를 활용한 집단의 유전학적 특성 구명

가. 공시재료

(1) 공시축

제3장 1절의 공시재료의 내용과 동일

(2) 유전체 정보

본 연구에서 개체의 유전체 정보를 얻고자 Illumina Porcine 60k DNA chip을 이용하여 개체로부터 얻어진 DNA를 통해 약 6만 2천여개의 SNP에 대한 정보를 수집하였다.

나. 연구방법

(1) 유전체 데이터 가공 및 기본정보 분석

(가) 유전체 데이터에 대한 사전가공은 아래와 같은 항목에 대하여 실시하였다.

- ① 다형성이 없는 SNP 마커 제거
- ② 유전자형의 결측률이 10%이상인 SNP 마커 제거
- ③ 남은 결측데이터에 대해 fastPHASE(Scheet 등, 2006)프로그램 이용하여 대치(Imputation)
- ④ 문자화 되어있는 염기정보를 분석에 이용 가능하도록 숫자화 함

(나) 각 염색체 내 SNP마커간 거리 및 마커수를 파악하고자 SNP마커의 물리적 거리를 이용하여 기초통계량 분석을 실시하였다.

(2) 유전체 정보활용 집단유전학적 특성 파악

공시축군에 대한 집단유전학적 특성을 알아보기로 아래와 같이 분석하였다.

(가) 기대이형접합형 : $\hat{H} = \frac{N}{N-1} \left(1 - \sum_{i=1}^l p_i^2 \right)$

\hat{H} : 기대이형접합량

N : 전체 자료수(개체수)

p_i : i번째 유전자빈도(본 연구에 사용되는 표지인자는 모두 biallelic 이므로 $i=1, 2$)

기대이형접합량은 1에서 유전자 빈도로 계산한 동형접합체 빈도를 뺀 값

(나) 하디-바인버그 평형 카이 자승값 : $\chi^2 = \sum \frac{(O - E)^2}{E}$

χ^2 : 하디-바인버그 평형 카이 자승값

O : 관측된 유전자형 빈도

E : 유전자빈도로 계산된 유전자형 빈도의 기댓값

(가) 다형성정보량 : $PIC = 1 - \sum_{i=1}^l p_i^2 - 2 \sum_{i=2}^l \sum_{j=1}^{i-1} (p_i^2 p_j^2)$

다형성정보량은 기대이형접합량에서 $2 \sum_{i=2}^l \sum_{j=1}^{i-1} (p_i^2 p_j^2)$ 만큼 뺀 값

(나) 근친도 : $F = 1 - \frac{O(f(Aa))}{E(f(Aa))}$

F : 근친도

O(f(Aa)) : 관측된 이형접합자 빈도

E(f(Aa)) : 유전자빈도로 계산한 이형접합자 빈도의 기댓값

(3) 연관불평형 및 유효집단크기 추정

(가) EM알고리즘을 이용한 유전자 일체배형의 빈도 추정

$$\Pr\left(\frac{A_1 B_1}{A_2 B_2} | A_1 A_2 B_1 B_2\right) = \frac{2f(A_1 B_1) \times f(A_2 B_2)}{2f(A_1 B_1) \times f(A_2 B_2) + 2f(A_1 B_2) \times f(A_2 B_1)}$$

$$\Pr\left(\frac{A_1 B_2}{A_2 B_1} | A_1 A_2 B_1 B_2\right) = \frac{2f(A_1 B_2) \times f(A_2 B_1)}{2f(A_1 B_1) \times f(A_2 B_2) + 2f(A_1 B_2) \times f(A_2 B_1)}$$

A와 B 두 개의 표지인자가 각각 1, 2의 대립유전자가 존재하고 두 표지인자가 모두 이형일 경우의 일체배형 빈도는 상인상태($\frac{A_1 B_1}{A_2 B_2}$), 상반상태($\frac{A_1 B_2}{A_2 B_1}$)의 조건부 확률을 계산하고 적용하여 변화량이 일정량 이하로 감소할 때 까지 반복 계산하여 추정하였다.

(나) 일체배형의 빈도를 이용한 연관불평형 계산

$$D = f(A_1 B_1) \times f(A_2 B_2) - f(A_1 B_2) \times f(A_2 B_1)$$

D : 두 표지인자가 모두 이형일 경우의 일체배형 빈도 불평형량

$$r^2 = \frac{D^2}{f(A_1) \times f(A_2) \times f(B_1) \times f(B_2)}$$

r^2 : 연관불평형량

(다) 유효집단크기 추정

$$E(r^2) = \frac{1}{(4N_t \times C) + 1}, \quad t = \frac{1}{2C}$$

$E(r^2)$: 연관불평형량의 기댓값

C : morgan 단위의 거리

(단, 본 연구에서 1 morgan = 100 Mbs는 동일하다고 가정하였다)

t : 세대수

N_t : t 세대 이전의 유효집단 크기

(라) 품종간 교배효과의 유전체 분석

유전자 빈도, 유전자형 빈도, Wright's F-statistic를 랜드레이스 순종(F0_L), 재래돼지 순종(F0_K), 1대 잡종(F1) 및 2대 잡종(F2)의 4개 그룹으로 분류하여 비교하는 연구를 수행하였다. 또한 재래돼지와 랜드레이스종 고유의 다형성을 탐색하기 위하여 재래돼지 종에서는 모두 동형이면서 랜드레이스에서만 다형성이 존재, 랜드레이스 종에서는 모두 동형이면서 재래돼지에서만 다형성이 존재, 재래돼지와 랜드레이스종 모두에서 다형성이 존재하지 않으나 서로 유전자형이 다른 표지인자들을 탐색하였다. 이러한 품종별 고유 다형성을 나타내는 표지인자들은 품종별 유전체 정보의 유전적 특성을 대표할 수 있으며, 품종 간 잡종 교배에 많은 영향을 미칠 수 있다. 또한 모든 표지인자에 대하여 재래돼지 품종과 랜드레이스 품종간의 유전자 빈도 차이를 비교하였으며, 이러한 유전자 빈도 차이가 잡종형성 시 구명된 유전체 정보에 어떠한 형태로 반영되는 지 확인하고자 유전자 빈도 차이와 잡종세대의 이형접합성을 비교분석하였다.

다. 3. 연구결과

(1) 유전체 데이터 가공 및 기본 정보 분석

Illumina Porcine 60k DNA chip은 다양한 돼지품종을 이용하여 개체 간 변이성을 나타내는 약 6만 2천여개의 SNP표지인자를 함축시켜 만들었다. 때문에 다른 축종에서 변이가 나타났으나, 재래돼지 집단에서는 그 변이가 나타나지 않을 수 있다. 그러므로, 개체간의 차이가 없는 SNP표지인자는 사전에 제거해야 하며, 실험과정 및 DNA chip에서 발생하는 결측데이터에 대해 보정을 하거나 이것이 전체의 10%이상 된다면 자료의 정확도 측면을 고려하여 사전에 제거를 실시해야 한다. 또한 Illumina Porcine 60k DNA chip에 대해 염색체별 표지인자수 및 표지인자간 밀도에 대한 기본정보가 필요하다. 왜냐하면, 염색체별 표지인자간 밀도는 QTL mapping, 유전체 선별 등의 분석 정확도에 영향을 미치기 때문에 본 유전분석에 앞서 필요한 사전작업이라고 볼 수 있다.

각 표지인자별 또는 개체별 결측률이 10%이상인 것에 대해 사전제거 후 fastPHASE(Scheet 등, 2006)을 이용하여 남아있는 표지인자 결측데이터에 대해 보정을 실시하였다. 이 후 A, T, G, C의 문자화된 코드를 분석에 용이하게 숫자로 변형 및 유전분

석에 용이한 형태로 가공하였다 (그림 30).

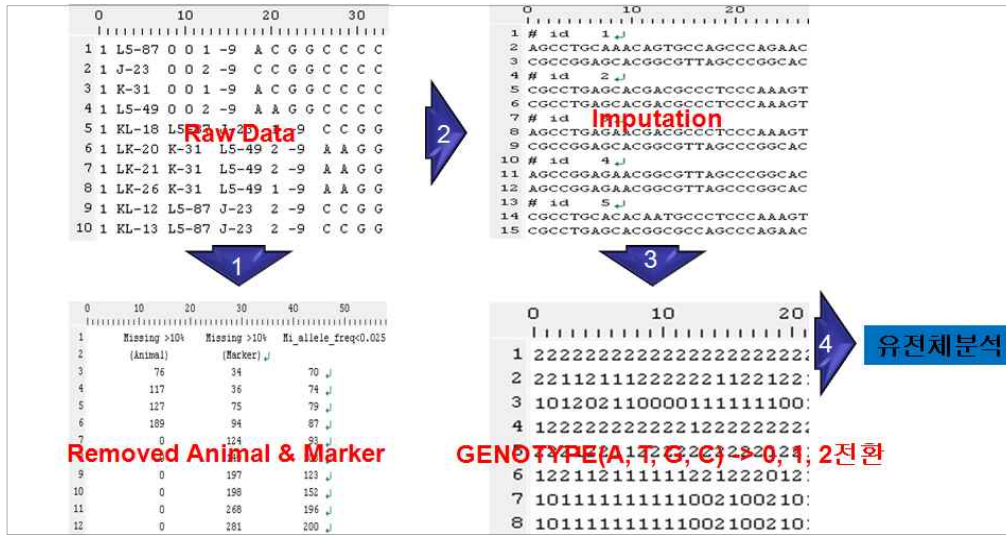


그림 30. 유전체 사전분석 순서

유전체 분석을 위한 사전작업결과 성염색체(X, Y)를 제외한 상염색체내(18쌍) 총 52,574개의 표지인자 중 10%이상 결측된 표지인자가 1,830개로 나타났으며, 재래돼지집단에서 다형성이 나타나지 않은 표지인자는 6,349개로 나타났다. 본 분석을 위해 이들에 대해 사전제거를 실시하였으며, 그 결과 총 44,395개의 표지인자에 대하여 유전체 분석을 실시하였다(표 23).

표 23. 유전체 분석을 위해 사전 제거된 표지인자 및 선별된 표지인자 수

SSC No.	SNP genotyped	Missing > 10%	Monomorphic	SNP selected for use in analysis	Available SNP ratio
1	6,947	173	1,019	5,755	0.83
2	3,349	161	435	2,753	0.82
3	2,820	96	291	2,433	0.86
4	3,762	111	515	3,136	0.83
5	2,452	96	315	2,041	0.83
6	2,970	103	246	2,621	0.88
7	3,581	139	499	2,943	0.82
8	2,674	102	269	2,303	0.86
9	3,198	111	338	2,749	0.86
10	1,665	71	184	1,410	0.85
11	1,972	83	222	1,667	0.85
12	1,548	62	146	1,340	0.87
13	3,668	115	441	3,112	0.85
14	4,125	143	528	3,454	0.84
15	2,861	85	387	2,389	0.84
16	1,888	60	166	1,662	0.88
17	1,757	70	183	1,504	0.86
18	1,337	49	165	1,123	0.84
Total	52,574	1,830	6,349	44,395	0.84

사전 데이터 가공작업을 통해 제거 되고 남은 표지인자 40,851에 대해 각 표지인자간 거리에 대해 평균 및 표준편차를 계산하여 표 24에 나타내었다.

표 24. 염색체별 선별된 표지인자간 맵 거리

SSC No.	Number of SNPs	Average Interval size(kb)	Standard deviation(kb)
1	5,755	51.3	57.1
2	2,753	50.8	76.4
3	2,433	50.7	69.6
4	3,136	43.5	50.8
5	2,041	49.3	58.2
6	2,621	46.9	56.3
7	2,943	46.3	59.7
8	2,303	51.8	63.1
9	2,749	48.2	57.2
10	1,410	46.6	55.1
11	1,667	47.9	48.9
12	1,340	42.8	49.9
13	3,112	46.7	44
14	3,454	43	42.7
15	2,389	56.2	70.7
16	1,662	46.6	50.4
17	1,504	42.6	40.1
18	1,123	48.3	50.8
전체	44,395		
평균		47.8	55.6

전장의 유전체 정보에 대한 표지인자의 수는 44,395로 나타났으며, 이들의 평균 맵 거리는 47.8(kb)로 나타났다. 염색체별 표지인자에 대한 정보를 보면 1번 염색체내 표지인자 수가 5,755개로 가장 많이 나타났으며, 14번 염색체에서 표지인자간 밀도가 가장 높은 것으로 나타났다.

(2) 유전체 정보활용 집단유전학적 특성 파악

집단 내 유전자원에 대한 변이는 기대이형접합형(H_{exp}) 또는 다형성정보량(PIC) 및 근친도로 나타낼 수 있다. 집단내 유전변이의 크기는 개량에 대한 가능성을 나타내며, 염색체별 유전변이의 크기를 추정함으로써 근친도 및 다형성을 고려한 선발계획이 가능할 것이다. 기대이형접합형 및 다형성정보량은 집단의 변이를 추정한 값이며, 근친도는 집단내 동형접합체에 대한 값으로 나타낸다.

표 25. 염색체별 기대이형접합형, 근친도 및 하디-바인베르크 평형 < 0.001 의 표지인자수

SSC No.	Hexp	PIC	Inbreeding	Number of Hardy-weinberg (χ^2)<0.001
1	0.37	0.36	-0.03	563
2	0.37	0.36	-0.03	272
3	0.39	0.38	-0.04	323
4	0.38	0.37	-0.03	329
5	0.38	0.37	-0.03	213
6	0.39	0.38	-0.03	213
7	0.38	0.37	-0.02	344
8	0.38	0.36	-0.05	293
9	0.38	0.38	-0.01	372
10	0.39	0.37	-0.03	148
11	0.37	0.36	-0.02	159
12	0.38	0.37	-0.02	153
13	0.38	0.37	-0.04	391
14	0.39	0.37	-0.04	535
15	0.37	0.36	-0.03	312
16	0.38	0.37	-0.04	172
17	0.37	0.36	-0.01	159
18	0.35	0.34	-0.04	112
Total				5,063

(H_{exp} =이형접합형, PIC=다형성정보량, Inbreeding=근친도, Hardy-weinberg (χ^2)<0.001 =하디-바인베르크 불평형 표지인자수)

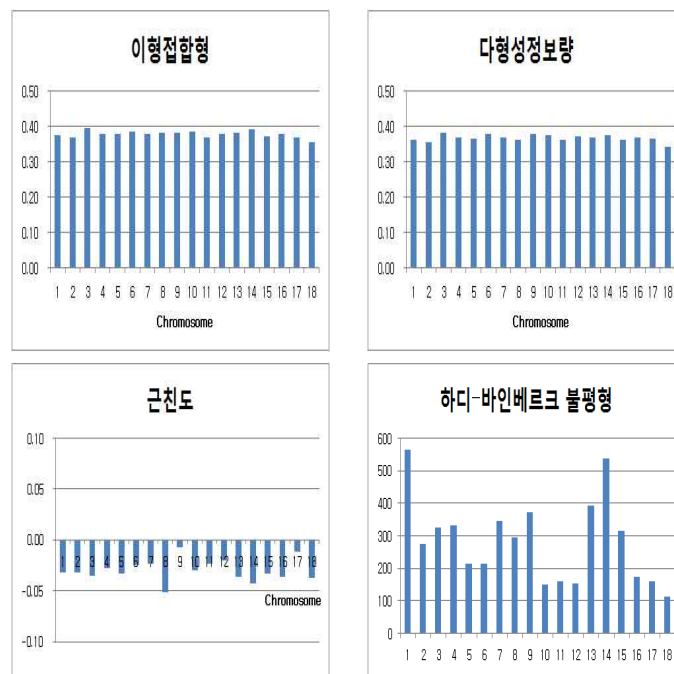


그림 31. 염색체별 기대이형접합형, 근친도 및 하디-바인베르크 평형 < 0.001 의 표지인자수

이형접합체 및 다형성 정보량은 집단의 변이폭을 나타내는 값으로 이형접합체가 높을수록(0.7이상) 유전적 변이가 크다고 할 수 있으며, 반대로 이형접합체가 낮으면 집단내 개체들이 근친되었다고 말할 수 있다(표 25). 재래돼지 대해 이형접합체 및 다형성정보량에 대해 염색체별 평균을 계산해 본 결과 각각 0.38, 0.37로 나타났다. 또한 유전학적 근친도는 -1~1사이의 범위로 1에 가까울수록 근친이 높다고 할 수 있는데 돼지에 대해 근친도를 계산해 본 결과 -0.03로 근친이 높게 되어있지 않음을 볼 수 있었다. 또한 하디-바인베르크 평형에 벗어나는(유의수준 0.001이하) 표지인자가 5,063개 나타났다(그림 31).

(3) 연관불평형 및 유효집단크기 추정

(가) 연관불평형 추정

가축에서 연관불평형(Linkage Disequilibrium)은 동일품종 또는 이품종 간 유전적 다형성의 정도를 나타내 줄 수 있으며, 이는 SNP(Single Nucleotide Polymorphisms) 표지인자 또는 MS(Microsatellite) 표지인자를 통해 다양하게 측정 가능하다. 또한 연관불평형 추정은 가축에서 경제형질에 영향하는 양적형질좌위(QTL)의 발굴 및 유전체 선발의 타당성을 간접적으로 증명해 주는 도구로 사용된다(Hayes 등, 2008; Su 등, 2010; VanRaden 등, 2009)

가축의 형질 발현에 영향하는 실제의 양적형질좌위는 그 수가 무수히 많고 큰 효과를 갖는 유전자는 적고 작은 효과를 갖는 유전자는 많은 감마분포 형태 (Hayes and Goddard, 2001)를 이루기 때문에 모든 유전자를 탐색하여 그 효과를 추정하는 것은 매우 어렵다. 따라서 표지인자를 통해 간접 예측하게 되는데, 이 때의 양적형질좌위와 표지인자간 연관불평형 크기는 표지인자가 양적형질좌위의 분산을 얼마나 설명할 수 있는지에 대한 정도로 나타내게 되기 때문에 (Hayes, 2008), 양적형질좌위 발굴에 앞서 표지인자간 연관불평형 크기의 추정은 상당히 중요하며, 꼭 필요한 사전작업이라고 할 수 있다.

재래돼지의 염색체 14번의 경우 연관불평형의 정도가 다른염색체에 비해 높은 것으로 나타났으며, 특히 표지인자간 거리가 0.1이하 일 때의 연관불평형 값이 0.383로 높게 나타났다. 반면 16번 염색체의 경우 0.1이하의 거리에서 연관불평형의 값이 0.255로 가장 낮게 추정되었다(그림 32, 표 26). 연관불평형의 값으로 나타나는 r^2 은 또한 양적형질좌위(QTL)발굴에 대한 검정력에 영향을 미치며, Luo (Luo, 1998)의 논문에 연산공식을 이용해 보면 표지인자와 양적형질좌위간의 r^2 이 0.3이고, 표현형 데이터를 갖는 개체수가 250두 일 때, 전체분산의 5%를 설명하는 QTL을 찾는 것에 대한 검정력이 약 35%였으며, 0.3의 r^2 에서 95%이상의 검정력을 갖기 위해서 1,000두 이상의 검정자료가 필요할 것으로 추정되었다(그림 33).

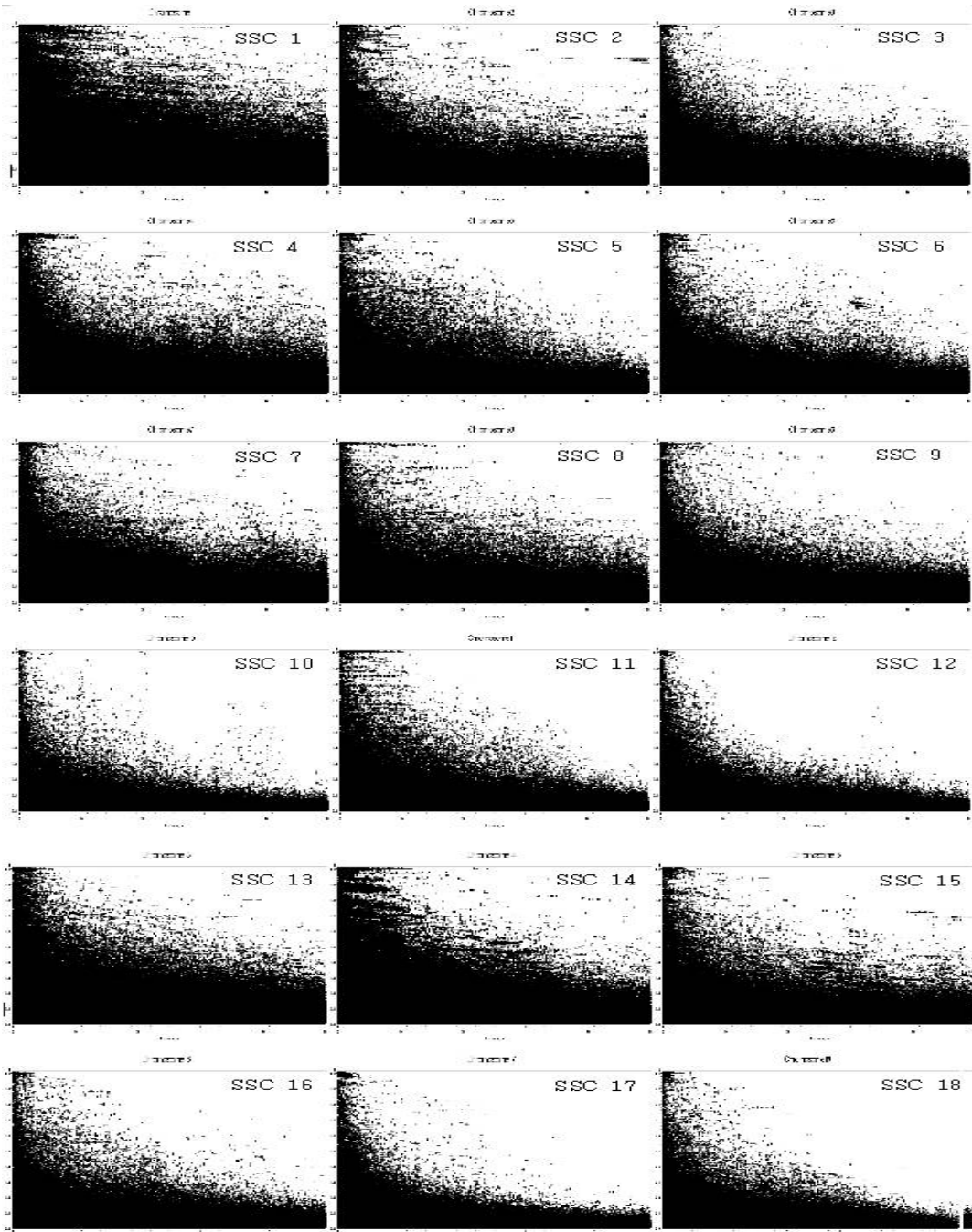


그림 32. 염색체별 0 Mb -50 Mb 내 SNP 표지인자쌍간 연관불평형(r^2)

표 26. SNP 거리별 표지인자쌍에 대한 r^2 의 평균

SSC No.	Distance(Mb)	No of SNP pairs	r^2	SSC No.	Distance(Mb)	No of SNP pairs	r^2
1	0.0-0.1	11,350	0.359	11	0.0-0.1	3,466	0.296
	0.1-0.2	11,834	0.291		0.1-0.2	3,646	0.243
	0.2-0.3	11,391	0.264		0.2-0.3	3,616	0.216
	0.3-0.4	11,278	0.243		0.3-0.4	3,546	0.202
	0.4-0.5	11,074	0.236		0.4-0.5	3,541	0.174
	>0.5	15,291,642	0.037		>0.5	1,314,713	0.046
2	0.0-0.1	6,157	0.334	12	0.0-0.1	3,351	0.303
	0.1-0.2	6,146	0.265		0.1-0.2	3,124	0.230
	0.2-0.3	6,092	0.240		0.2-0.3	3,067	0.203
	0.3-0.4	5,945	0.228		0.3-0.4	3,144	0.189
	0.4-0.5	5,855	0.210		0.4-0.5	3,137	0.181
	>0.5	3,565,026	0.037		>0.5	828,527	0.044
3	0.0-0.1	5,211	0.311	13	0.0-0.1	6,185	0.325
	0.1-0.2	5,337	0.241		0.1-0.2	6,593	0.267
	0.2-0.3	5,163	0.214		0.2-0.3	6,569	0.240
	0.3-0.4	4,982	0.197		0.3-0.4	6,409	0.223
	0.4-0.5	5,050	0.184		0.4-0.5	6,278	0.211
	>0.5	2,810,028	0.036		>0.5	4,397,740	0.045
4	0.0-0.1	7,184	0.316	14	0.0-0.1	7,666	0.383
	0.1-0.2	7,611	0.264		0.1-0.2	8,223	0.320
	0.2-0.3	7,362	0.240		0.2-0.3	7,908	0.295
	0.3-0.4	7,262	0.222		0.3-0.4	7,804	0.278
	0.4-0.5	7,268	0.210		0.4-0.5	7,740	0.262
	>0.5	4,597,803	0.039		>0.5	5,620,589	0.049
5	0.0-0.1	4,289	0.295	15	0.0-0.1	4,710	0.338
	0.1-0.2	4,396	0.236		0.1-0.2	4,832	0.289
	0.2-0.3	4,247	0.203		0.2-0.3	4,634	0.249
	0.3-0.4	4,111	0.195		0.3-0.4	4,587	0.225
	0.4-0.5	4,057	0.179		0.4-0.5	4,584	0.206
	>0.5	1,953,977	0.040		>0.5	2,722,649	0.038
6	0.0-0.1	5,926	0.316	16	0.0-0.1	3,507	0.255
	0.1-0.2	5,753	0.251		0.1-0.2	3,647	0.211
	0.2-0.3	5,551	0.230		0.2-0.3	3,567	0.186
	0.3-0.4	5,476	0.211		0.3-0.4	3,443	0.180
	0.4-0.5	5,458	0.197		0.4-0.5	3,572	0.169
	>0.5	3,209,076	0.035		>0.5	1,283,955	0.042
7	0.0-0.1	6,496	0.302	17	0.0-0.1	3,323	0.298
	0.1-0.2	6,803	0.251		0.1-0.2	3,500	0.249
	0.2-0.3	6,625	0.222		0.2-0.3	3,424	0.220
	0.3-0.4	6,508	0.222		0.3-0.4	3,353	0.214
	0.4-0.5	6,405	0.207		0.4-0.5	3,382	0.189
	>0.5	4,087,048	0.040		>0.5	1,053,934	0.041
8	0.0-0.1	4,797	0.292	18	0.0-0.1	2,335	0.315
	0.1-0.2	4,811	0.241		0.1-0.2	2,459	0.235
	0.2-0.3	4,778	0.221		0.2-0.3	2,370	0.214
	0.3-0.4	4,758	0.208		0.3-0.4	2,352	0.211
	0.4-0.5	4,631	0.191		0.4-0.5	2,321	0.189
	>0.5	2,492,871	0.040		>0.5	597,019	0.055
9	0.0-0.1	6,078	0.288	Total	0.0-0.1	95,189	0.318
	0.1-0.2	6,210	0.228		0.1-0.2	98,025	0.259
	0.2-0.3	6,088	0.204		0.2-0.3	95,476	0.232
	0.3-0.4	5,791	0.187		0.3-0.4	93,794	0.218
	0.4-0.5	5,745	0.174		0.4-0.5	93,154	0.203
	>0.5	3,608,341	0.034		>0.5	60,375,208	0.040
10	0.0-0.1	3,158	0.258				
	0.1-0.2	3,100	0.207				
	0.2-0.3	3,024	0.180				
	0.3-0.4	3,045	0.168				
	0.4-0.5	3,056	0.151				
	>0.5	940,270	0.035				

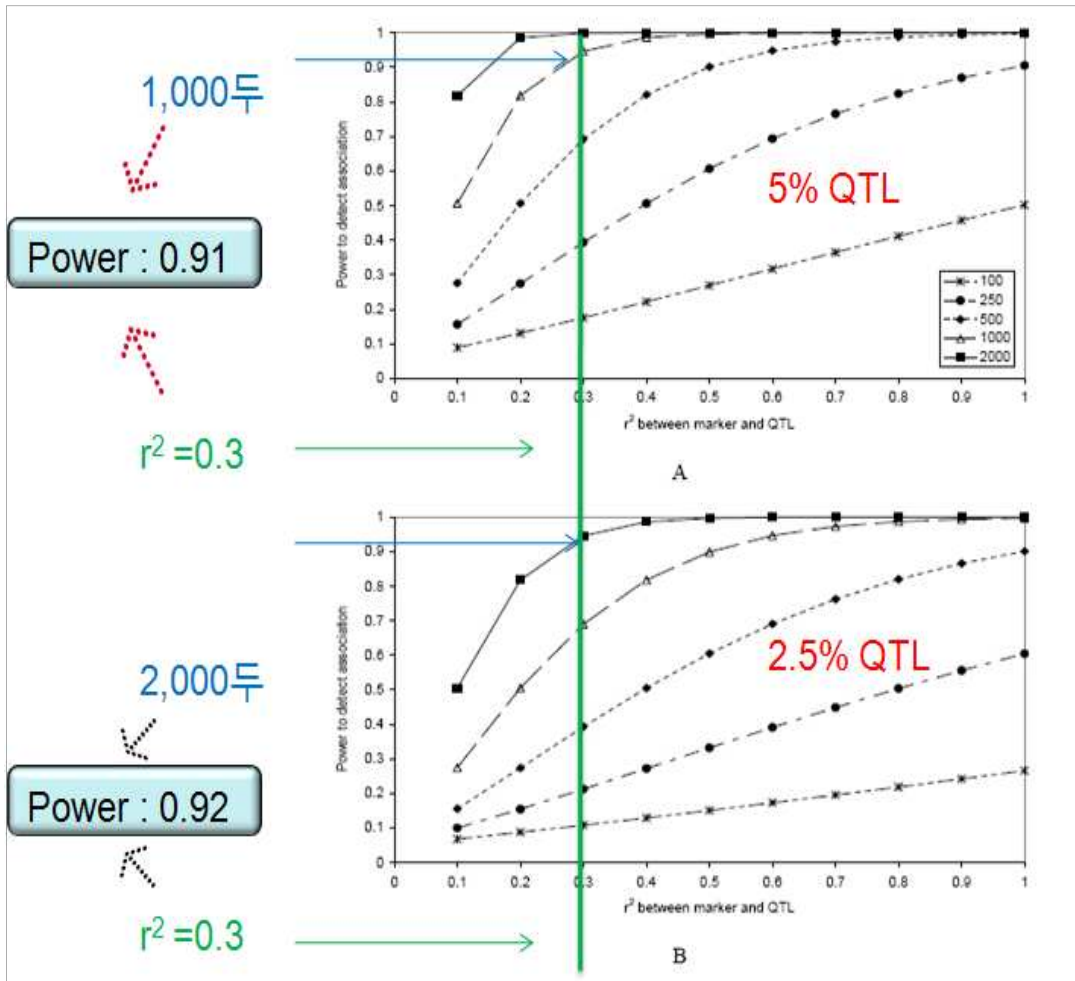


그림 33. 표현형 분산의 5%(위) 또는 2.5%(아래)를 설명하는 QTL을 발견하기 위한 검정력 (Luo, 1998)

(나) 유효집단크기 추정

연관불평형은 돌연변이 및 재조합(Recombination)에 의해 발생되며, 재조합의 경우 염기서열간의 거리가 먼 좌위쌍(locus pair)에서 발생할 확률이 높기 때문에, 거리와 연관 불평형의 값을 안다면 돌연변이가 없다는 가정하에 Sved가 1971에 제안한 아래 공식을 통해 세대에 따른 유효집단의 크기를 추정할 수 있다.

$$E(r_c^2) = \frac{1}{1 + 4N_{e,t}c}$$

여기서, t 는 이전세대(Generation ago), N_e 은 유효집단 크기, c 는Morgan단위에서 표지인자간의 물리적 거리(distance)이며, 세대별 N_e 추정은 $t = \frac{1}{2c}$ 에 의해 추정할 수 있다. 즉, 표지인자간 거리가 보다 가까운 위치에서 발생한 연관불평형 값은 보다 더 오래전 세대에 발생했음을 나타낸다고 할 수 있다(Hayes 등, 2003; Sargolzaei 등, 2008). 재래돼지의 세대별 유효집단의 크기는 세대의 간격을 설정하고 해당 세대의 r^2 값의 평균값으로 유효집단의 크기 추정하였다

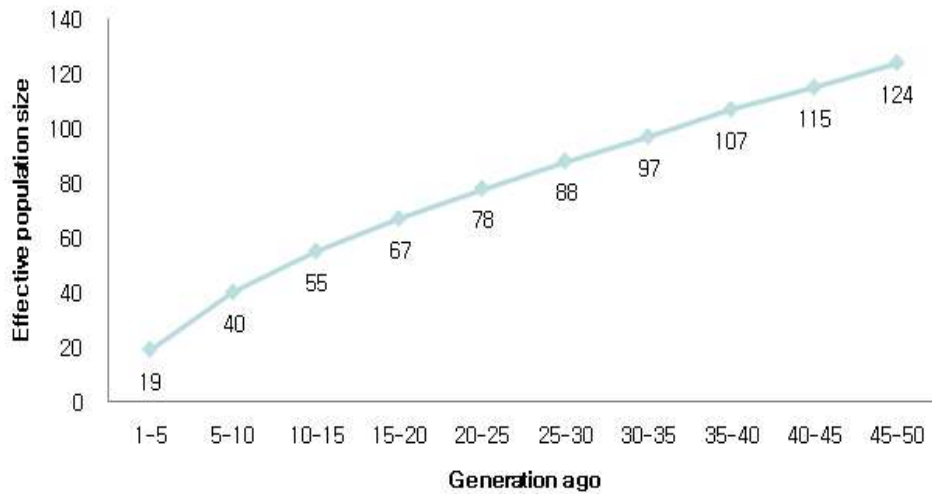


그림 34. 재래돼지 유효집단 크기 추정

그림 34에 재래돼지의 세대별 유효집단 크기에 대해 나타내었다. 현재 재래돼지의 유효집단의 크기는 19두로 나타났으며, 지금으로부터 약 50세대 이전의 유효집단 크기는 124두로 추정되어 현재 세대에 가까울수록 유효집단의 크기가 적게 나타났다. 유효집단 크기의 감소는 유전자원의 손실을 가져올 수 있기 때문에, 개체선발 및 교배시 유전자원의 다양성을 극대화 할 수 있는 방안을 모색해야 할 것으로 사료된다.

(4) 품종간 교배효과의 유전체 분석

각 그룹별 소수유전자 빈도를 그림35에 나타내었으며, 품종 특이적 변이를 나타내는 표지인자를 표 27에 나타내었다.

표 27. 품종 특이적 다형성을 보이는 SNP표지인자의 염색체별 빈도

SSC	Unique in Korean native	Totally opposite	Unique in Landrace
1	768	2	742
2	292	4	380
3	212	2	268
4	355	0	470
5	277	0	280
6	217	0	198
7	391	2	410
8	290	1	236
9	214	0	341
10	164	0	172
11	280	0	181
12	81	1	116
13	357	0	446
14	377	0	571
15	320	0	274

16	123	0	181
17	153	0	181
18	105	1	197
Total	4,976	13	5,644

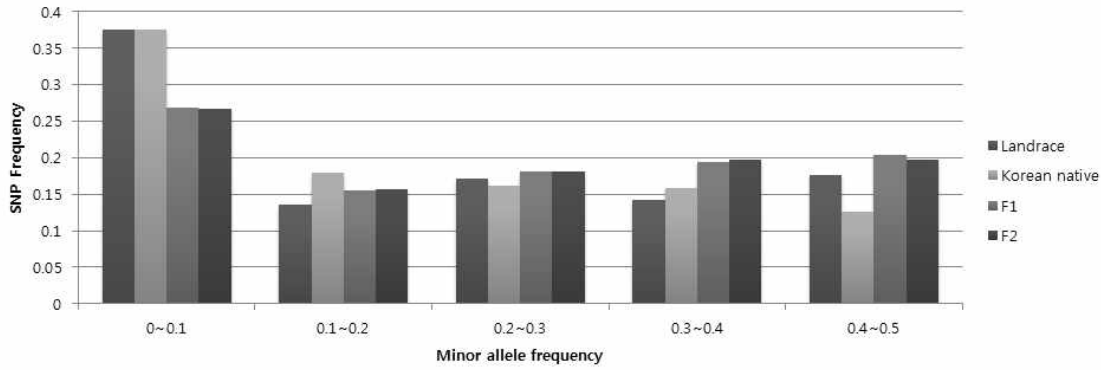


그림 35. 각 그룹에 따른 소수유전자 빈도 분포

재래돼지에서는 모두 동형이면서 랜드레이스에서만 다형성이 존재하는 표지인자는 4,976 개, 랜드레이스에서는 모두 동형이면서 재래돼지에서만 다형성이 존재하는 표지인자는 5,644개가 발견되었다. 재래돼지와 랜드레이스 모두에서 다형성이 존재하지 않으나 서로 유전자형이 다른 표지인자들은 13개로 1, 2, 3, 7, 8, 12 및 18번 염색체에서 발견되었으며, 이러한 상반된 유전자형을 나타내는 표지인자들은 잡종세대에서의 유전적 특이성을 나타내는 것으로 나타났다. 특이적 표지인자의 예로 1번 염색체에 존재하는 한 표지인자는 재래돼지에서는 모든 개체에서 유전자형이 AA로 나타나지만 랜드레이스의 모든 개체들은 GG형으로 나타났으며, 이러한 유전자형은 1대 잡종과 2대잡종에서 이형접합성과 근친도(Wright's F-statistic)를 크게 변화시키게 된다(그림 36).

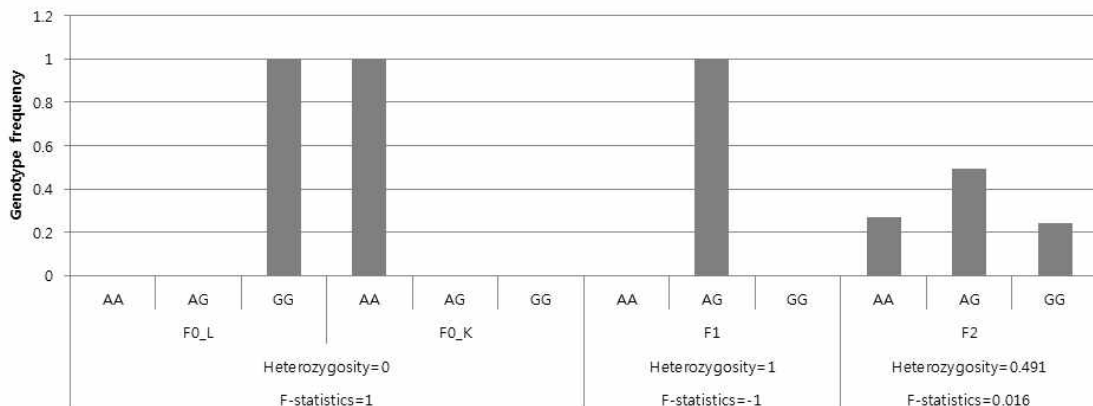


그림 36. 품종 특이적 표지인자의 유전자 빈도, 이형접합성, 근친도의 세대별 변화

이러한 특성은 소수유전자 빈도에서도 나타나는데 빈도 0-0.1의 표지인자들은 순종돈들에서 빈도가 높게 나타나지만 잡종세대에서는 낮게 나타난다. 이는 한 품종에서 유전자 빈도가 낮았으나 다른 품종에서 유전자 빈도가 높은 경우가 교배되었을 때 나타날 수 있는 현상으로 품종 상호간의 유전자 빈도를 상쇄시키는 원인으로 사료된다.

품종간 교배는 잡종강세를 목적으로 하며, 이러한 잡종강세는 형질에 관여하는 유전자가 이형일 때 발생하는 우성효과에 가장 크게 영향을 받는다. 따라서 품종간 교배에 따라 이형접합성이 얼마나 증가할 수 있는지 확인하는 것이 매우 중요하다. 이론적으로 잡종교배시 이형접합성이 증가함을 예상할 수 있으나 실제적인 분석이 어려워 실행되지 않았다. 그러나 단기간에 대용량의 유전체 정보를 형성할 수 있는 DNA chip의 개발로 이러한 이형접합성의 실제적 증가를 가늠할 수 있게 되었다. 이형접합성의 증가량에 가장 크게 영향을 미치는 요인은 각 품종의 유전자 빈도 차이이다. 재래돼지와 랜드레이스종의 평균 유전자 빈도 차이는 0.283으로 나타났으며, 표지인자별로 다양하게 나타났다(그림 37).

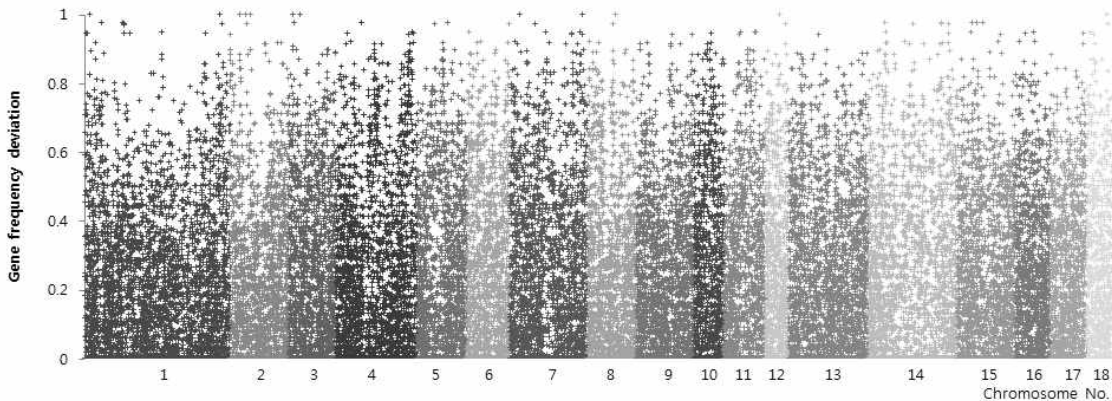


그림 37. 재래돼지와 랜드레이스종의 유전자 빈도 차이의 표지인자별 도화 결과

표 28. 각 그룹 조합별 유전자 빈도 차이 및 증가된 이형접합성

between group	Gene frequency deviation	Increased heterozygosity
F0_L & F0_K	0.283	
F0_L & F1	0.142	0.081
F0_K & F1	0.145	0.108
F0_L & F2	0.144	0.019
F0_K & F2	0.145	0.046

1대 잡종과 재래돼지순종, 1대 잡종과 랜드레이스 순종의 증가된 이형접합성은 각각 0.081, 0.108로 1대 잡종과 재래돼지순종에서 다소 높게 나타났다. 순종대비 증가된 이형접합성은 2대 잡종에서 다소 감소하는 것으로 나타나 하디-바인버그 평형을 이루면서 잡종강세의 효과가 감소하는 것으로 사료된다. 이러한 현상은 그림 38와 같이 근친도에

서도 2대 잡종에서 하디-바인버그 평형을 이루는 것이 나타나며 이형접합성의 극대화는 1대 잡종에서 나타나는 것을 확인할 수 있다.

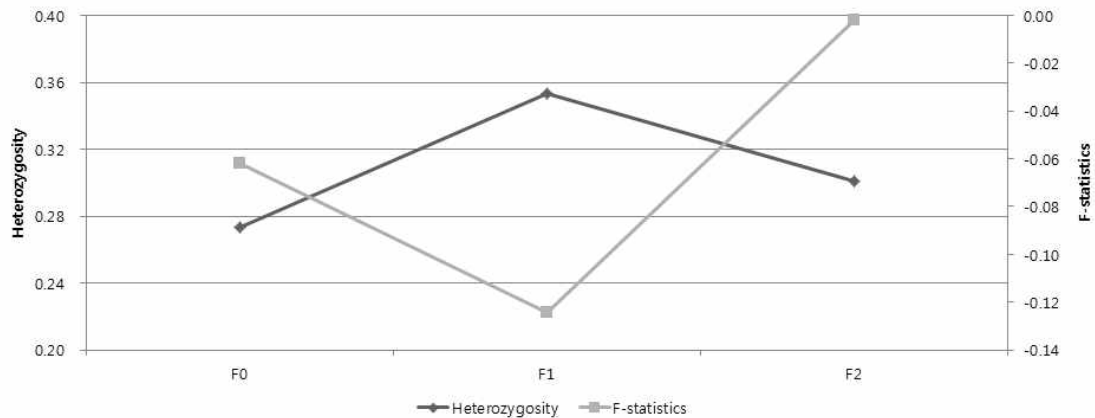


그림 38. 세대별 이형접합성 및 Wright's F-statistic 변화 추세

재래돼지와 랜드레이스종의 1대 교잡종의 표지인자별 이형접합성 증가량은 그림 13와 같이 두 품종사이의 유전자 빈도 차이와 함수관계를 나타내고 있으며, 품종간 유전자 빈도 차이가 낮을수록 변이가 크게 나타나는 추세를 나타내고 있다. 이러한 함수관계는 이론적으로 다음과 같이 유도될 수 있다(Falconer and Mackay, 1996). 랜드레이스 품종에서 한 표지인자의 유전자 빈도를 각각 p와 q로 하고 재래돼지의 경우를 p'과 q'이라 하면, 유전자 빈도 차이 d는 아래와 같다.

$$d = p - p' = q' - q.$$

랜드레이스의 이형접합성은 $2pq$ 이고 재래돼지는 $2p'q'$ 이며 이를 $2pq + 2pd - 2qd - 2d^2$ 로 나타낼 수 있다. 1대 잡종의 이형접합성은 $(p+q) \times [(p-d) + (q-d)]$ 으로 $2pq + pd - qd$ 이다. 재래돼지로부터 1대잡종까지의 증가된 이형접합성은 아래와 같다.

$$(2pq + pd - qd) - (2pq + 2pd - 2qd - 2d^2) = -pd + qd + 2d^2$$

랜드레이스부터 1대잡종까지의 증가된 이형접합성은 아래와 같다.

$$(2pq + pd - qd) - (2pq) = pd - qd$$

따라서 두 품종간 교배에 의하여 증가된 이형접합성 평균치는 d^2 가 되며 각 품종대비 증가량의 합은 그림 39와 같이 $2d^2$ 가 된다.

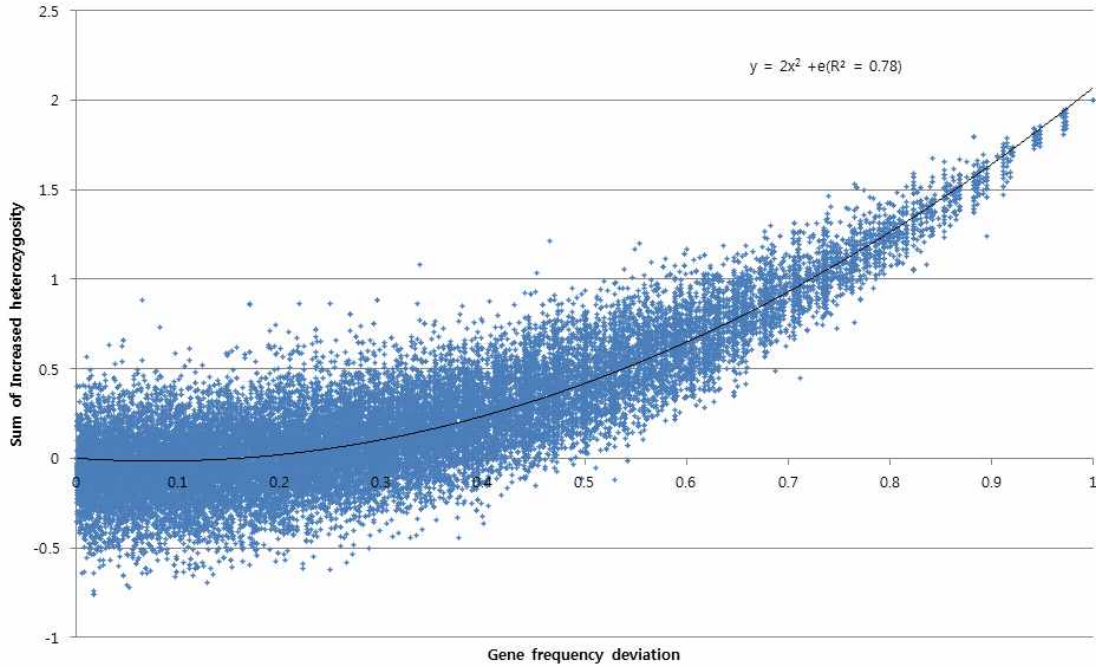


그림 39. Plot of increased heterozygosity by gene frequency deviation

이러한 관계식을 통하여 이형접합성을 극대화하기 위해서는 각 품종의 유전자 빈도 차이를 최대화 하여야 하며, 고밀도 유전체 정보의 활용으로 실제적 증가량을 예측할 수 있을 것으로 사료된다.

3. MBV추정 알고리즘 개발 및 전산프로그램 개발

가. 유전체 육종가 추정프로그램 개발

(1) BayesA

(가) 알고리즘

BayesA의 기본 알고리즘 및 절차는 다음과 같다.

$$y_i = \mu + \sum_j X_{ij} \alpha_j + e_i$$

여기서 y_i 는 I번째 개체의 보정된 형질별 관측치, μ 는 표지인자 효과의 절편값, X_{ij} 는 i번째 개체의 j번째 allele coding(이형접합자는 1, 서로 다른 동형접합자는 각각 0과 2), α_j 는 j번째 표지인자의 효과이며, e_i 는 i번째 관측치의 잔차이다.

베이지 추론을 위한 사전분포 가정은 다음과 같다 (Gianola 등, 2009; Meuwissen 등, 2001).

$$(e_i | \sigma_e^2) \sim (\text{iid})N(0, \sigma_e^2); \sigma_e^2 \sim \nu_e S_e^2 \chi_{\nu_e}^{-2}$$

분산이 σ_e^2 인 조건에서의 각 관측치별 잔차 e_i 는 평균이 0이고 분산이 σ_e^2 인 독립이며 동일한(Independent and identically distributed : iid) 정규분포를 따른다. 또한 σ_e^2 는 degree of freedom이 ν_e 이고 scale parameter가 S_e^2 인 scaled inverse chi-square분포를 따른다.

$$(\alpha_j | \sigma_j^2) \sim (\text{iid})N(0, \sigma_j^2); \sigma_j^2 \sim \nu_\alpha S_\alpha^2 \chi_{\nu_\alpha}^{-2}$$

j 번째 표지인자의 분산 σ_j^2 의 조건에서 j 번째 표지인자의 효과 추정치 α_j 는 평균이 0이고 분산이 σ_j^2 인 정규분포를 따르며, σ_j^2 는 degree of freedom이 ν_α 이고 scale parameter가 S_α^2 인 scaled inverse chi-square분포를 따른다.

비조건분포(unconditional distribution) α_j 는 평균이 0, scale parameter가 S_α^2 , degree of freedom이 ν_α 인 독립이며 동일한 univariate-t 분포를 따른다.

$$\alpha_j \sim (\text{iid})t(0, S_\alpha^2, \nu_\alpha)$$

표지인자 효과 분산 σ_a^2 과 상가적 유전분산(육종가 분산) σ_a^2 의 관계는 다음과 같다.

$$\sigma_a^2 = \frac{\sigma_a^2}{\sum_j 2p_j q_j} \text{ 즉, 표지인자 효과 분산은 상가적 유전분산을 이항분포형태로 나타나는 유전자빈도 분산의 합으로 나눈 것이다. 어떠한 분산 } \sigma^2 \text{이 scaled, inverse}$$

chi-square 분포를 따를 경우 분포함수의 평균은 다음과 같으며, 이를 변형하여 S^2 를 계산할 수 있다.

$$E(\sigma^2) = \frac{\nu S^2}{\nu - 2} \text{ 따라서 } S^2 = \frac{(\nu - 2)E(\sigma^2)}{\nu} \text{ 이다.}$$

이를 σ_a^2 에 적용할 경우 scale parameter S_α^2 는 다음과 같다.

$$S_\alpha^2 = \frac{(\nu_\alpha - 2)\sigma_a^2}{\nu_\alpha (\sum_j 2p_j q_j)}$$

표지인자 효과의 degree of freedom ν_α 는 4를 적용하였으며, S_α^2 를 계산하기 위하여 REML방법을 이용하여 추정된 형질별 유전모수를 위의 공식에 적용하여 prior로 설정하였다.

위와 같은 사전분포 가정을 통한 표지인자 효과 추정방법을 BayesA라 하며 사후분포 추정을 위한 Gibbs sampling 절차는 다음과 같다.

→ σ_e^2 sampling

σ_e^2 의 full conditional 확률밀도함수는 다음과 같이 scaled inverse chi-square 분포를 따른다

$$f(\sigma_e^2 | y, \mu, \alpha) \sim \tilde{\nu}_e \tilde{S}_e^2 \chi_{\tilde{\nu}_e}^{-2}$$

여기서 degree of freedom $\tilde{\nu}_e = n + \nu_e$ 이고 (n=표현형 수)

scale parameter $\tilde{S}_e^2 = \frac{SSE + \nu_e S_e^2}{\tilde{\nu}_e}$ 이다.

(위에서 SSE는 $e'e$, $e = \mathbf{y} - \mathbf{1}\mu - \Sigma_j \mathbf{x}_j' \alpha_j$)

$$X \sim \chi^2(\nu) \text{ 이면 } \frac{1}{X} \sim \chi^{-2}(\nu), \quad \frac{\nu S^2}{X} \sim \nu S^2 \chi^{-2}(\nu, S^2)$$

따라서 σ_e^2 sampling은 $\frac{SSE}{\text{random-}\chi^2(\tilde{\nu}_e)}$ 방식으로 진행

($\text{random-}\chi^2(\tilde{\nu}_e)$ 는 카이분포함수에서 생성하는 랜덤넘버)

→ μ sampling

$$f(\mu | \mathbf{y}, \alpha, \sigma_e^2) \sim N\left(\frac{\mathbf{1}'(\mathbf{y} - \mathbf{X}\alpha)}{n}, \frac{\sigma_e^2}{n}\right)$$

→ σ_j^2 sampling

$\xi = [\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2]$, ξ_{j-} 를 j번째 표지인자 분산을 제외한 나머지 분산이라 할 경우 표지인자 분산의 full conditional 확률밀도함수는 다음과 같다.

$$f(\sigma_j^2 | y, \mu, \alpha, \xi_{j-}, \sigma_e^2) \sim \tilde{\nu}_\alpha \tilde{S}_\alpha^2 \chi_{\tilde{\nu}_\alpha}^{-2}$$

여기서 degree of freedom $\tilde{\nu}_\alpha = 1 + \nu_\alpha$ 이고

scale parameter $\tilde{S}_\alpha^2 = \frac{\alpha_j^2 + \nu_\alpha S_\alpha^2}{\tilde{\nu}_\alpha}$ 이다.

→ α_j sampling

표지인자 효과의 full conditional 확률밀도함수는 다음과 같으며 정규분포를 따른다.

$$f(\alpha_j | \mathbf{y}, \mu, \alpha, \sigma_e^2) \sim N\left(\hat{\alpha}_j, \frac{\sigma_e^2}{c_j}\right)$$

여기서 $\hat{\alpha}_j = \frac{\mathbf{x}_j' \mathbf{w}}{c_j}$, $\mathbf{w} = \mathbf{y} - \mathbf{1}\mu - \sum_{j' \neq j} \mathbf{x}_{j'} \alpha_{j'}$, $c_j = (\mathbf{x}_j' \mathbf{x}_j + \frac{\sigma_e^2}{\sigma_j^2})$

위의 ① ~ ④ 까지의 난수 생성을 20,000번 반복하고 초기 5,000회의 자료를 제외한 나머지 sampling 결과의 평균을 이용하여 모수를 추정하였으며 FORTRAN을 이용하여 프로그램을 개발하고 이를 이용하여 분석하였다.

추정된 표지인자 효과 추정치를 이용하여 개체별 육종가를 다음과 같이 추정하였다.

$$\mathbf{g}_i = \mu + \mathbf{x}_i' \alpha$$

여기서 \mathbf{g}_i 는 i 번째 개체의 유전체 육종가, μ 는 절편, \mathbf{x}_i' 는 i 번째 개체의 allele coding 벡터, α 는 표지인자 효과 추정치 벡터이다.

(나) 프로그램 코드

Fortran 으로 프로그래밍한 Bayes A의 소스는 아래와 같다.

```
! Name : BayesA
! Programed by : Joonho Lee - PhD working in Hankyong National Univ.(Korea) zoonolee@gmail.com,
+82-10-3408-2895
! Version : Linux 6.0 (2012-10-11)
! Contents :          1. Estimation of marker effect and breeding value using BayesA
program BayesA
use ranlib
!variables
integer::sol_no=0,LNO
integer::NOM,MNO,NOFE,FENO,NORE,RENO,NOP,PNO,NOS,SNO !number of markers, fixed effect,
phenotypes and samplings(NOM,NOP,NOS) and marker, fixed effect, phenotype and sampling
NO.(MNO,RNO,SNO) for do_loop
integer::NO0,NO1,NO2,BIP,FPOSI,RPOSI,MPOSI !counting 0, 1, 2 in the allelic coding data,
burn-in period
integer,allocatable::EQ(:,:),LOFE(:),FE(:),LORE(:),RE(:) !EQUATION,Level of fixed effect, fixed
effect position in equation
real(8)::AGV,EAGV,EAGV2,EAGV3,EAGV4,EAGV5,EV,RV,meanRV,SOA,SOE ! additive genetic variance,
estimated additive genetic variance, error variance, residual variance, mean of residual
variance, scale parameter of alpha(marker effect) and e(residual)
real(4)::NU_A,Nu_E !degree of freedom for alpha(marker effects) and e(residual)
real(4)::SOEH=0,RHS=0,LHS=0,invLHS=0,MU=0 ! sum of expected heterozygosity, right-hand-side,
left-hand-side, inverse left-hand-side, MU:intercept
real(8)::GF1=0,GF2=0 !gene frequencies
real(8),allocatable::SOL(:),PHE(:),VBM(:),ADJ_y(:) !solution vector(intercept(1) and marker
effect(rest)), phenotype vector, variances by markers, adjusted y
real(8),allocatable::EH(:),OH(:),OV(:) ! expected and observed heterozygosity by markers,
observed genotype frequency variation
real(8),allocatable::meanSOL(:),meanVBM(:),GEBV(:)
real(8),allocatable::SRV(:),meanSRV(:)
character(len=2)::HE !choice of gene frequency variation for scale parameter for alpha(EH, OH,
OV)
character(len=3),allocatable::NAFE(:),NARE(:),FETYPE(:),RETYPE(:) !name and type of fixed and
random effect(cro:cross, cov:covariable)
character(len=12)::OFT !output file tag
character(len=220)::GDD,PDD,ODD,FDD,RDD !genomic, phenotype and output file directory
character(len=30)::GDFN,PDFN,FDN,RDFN !genomic data file name and phenotype data file name
character(len=250)::GDFNIP,PDFNIP,ODFNIP,FDNIP,RDFNIP,PF !genomic, phenotype and output data
file name including path
!reading parameters
print
print
"*****"
*****"
print *, "FBA_Lv6 : FORTRAN(version) Bayes A Linux version 6.0"
print *, "Developed by Joonho Lee : zoonolee@gmail.com, +82-10-3408-2895"
print *, "Genomic data should be allelic coding(genotype -> 0, 1, 2) & no missing data allowed"
```

```

print *, 'Please enter the name of parameter file including path(ex) "./parGRDH"'
print                                                                                                     *,
"*****"
*****"
read *, PF
PF=adjustl(PF)
open (99, file=PF)
read (99,*) OFT
print *, "file tag ", OFT
read (99,*) NOP
print *, "number of phenotypes", NOP
read (99,*) NOM
print *, "number of markers", NOM
!reading prarmeter for fixed effect
read (99,*) NOFE
print *, "number of fixed effects", NOFE
print *, "effects & leves"
allocate(NAFE(NOFE),LOFE(NOFE),FETYPE(NOFE))
do FENO=1, NOFE
  read (99,*) NAFE(FENO), LOFE(FENO),FETYPE(FENO)
  print *, NAFE(FENO), LOFE(FENO)
enddo
!reading parameter for random effect
read (99,*) NORE
print *, "number of random effects", NORE
print *, "effects & leves"
allocate(NARE(NORE),LORE(NORE),RETYPE(NORE))
do RENO=1, NORE
  read (99,*) NARE(RENO), LORE(RENO),RETYPE(RENO)
  print *, NARE(RENO), LORE(RENO)
enddo
MPOSI=sum(LOFE)+sum(LORE)
RPOSI=sum(LOFE)
read (99,*) AGV
print *, "additive genetic variance", AGV
read (99,*) NU_A
print *, "degree of freedom for marker effect variance", NU_A
read (99,*) EV
print *, "residual variance", EV
read (99,*) NU_E
print *, "degree of freedom for residual variance", NU_E
read (99,*) NOS
print *, "number of samplings", NOS
read (99,*) BIP
print *, "burn-in period", BIP
read (99,*) GDD
print *, "directory of genomic data file ", GDD
read (99,*) GDFN
print *, "name of genomic data file(0, 1, 2 converted, example : Genotype012_OUT_TAG.out) ",
GDFN
read (99,*) FDD
print *, "directory of fixed effect data file ", FDD
read (99,*) FDFN
print *, "name of fixed effect data file ", FDFN
read (99,*) RDD
print *, "directory of simple random effect data file ", RDD
read (99,*) RDFN
print *, "name of simple random effect data file ", RDFN
read (99,*) PDD

```



```

print *, "directory of phenotype data file ", PDD
read (99,*) PDFN
print *, "name of phenotype data file ", PDFN
read (99,*) ODD
print *, "directory of output file ", ODD
close(99)

allocate(EQ(NOP,NOM+MPOS1),SOL(NOM+MPOS1),meanSOL(NOM+MPOS1),PHE(NOP),ADJ_y(NOP),FE(NOFE),RE(NORE),SRV(NORE)&
&
,meanSRV(NORE),VBM(NOM+MPOS1),meanVBM(NOM+MPOS1),GEBV(NOP),EH(NOM),OH(NOM))
ODD=adjustr(ODD)
GDD=adjustr(GDD)
PDD=adjustr(PDD)
FDD=adjustr(FDD)
RDD=adjustr(RDD)
GDFNIP=GDD//GDFN
GDFNIP=adjustl(GDFNIP)
open(98,file=GDFNIP,status='old')
PDFNIP=PDD//PDFN
PDFNIP=adjustl(PDFNIP)
open(97,file=PDFNIP,status='old')
FDFNIP=FDD//FDFN
FDFNIP=adjustl(FDFNIP)
open(91,file=FDFNIP,status='old')
RDFNIP=RDD//RDFN
RDFNIP=adjustl(RDFNIP)
open(89,file=RDFNIP,status='old')
ODFNIP=ODD//"JBAYES__A_MEFF"//OFT//".out"
ODFNIP=adjustl(ODFNIP)
open(96,file=ODFNIP,status='unknown')
ODFNIP=ODD//"JBAYES__A_MVAR"//OFT//".out"
ODFNIP=adjustl(ODFNIP)
open(95,file=ODFNIP,status='unknown')
ODFNIP=ODD//"JBAYES__A_GEBV"//OFT//".out"
ODFNIP=adjustl(ODFNIP)
open(94,file=ODFNIP,status='unknown')
ODFNIP=ODD//"JBAYES__A_STAT"//OFT//".out"
ODFNIP=adjustl(ODFNIP)
open(93,file=ODFNIP,status='unknown')
ODFNIP=ODD//"JBAYES__A_VOGF"//OFT//".out"
ODFNIP=adjustl(ODFNIP)
open(92,file=ODFNIP,status='unknown')
ODFNIP=ODD//"JBAYES__A_CONV"//OFT//".out"
ODFNIP=adjustl(ODFNIP)
open(90,file=ODFNIP,status='unknown')
write(90,*) "Convergence"
write(90,*) "Sampling", "Var_R", "SOL_F", "SOL_M"
call initseed()
EQ=0
!reading genomic data
do PNO=1,NOP
  read(98,*) EQ(PNO,(MPOS1+1):)
  read(91,*) FE(:)
  do FENO=1, NOFE
    if(FENO.eq.1) then
      if(FETYPE(FENO).eq.'cro') then
        EQ(PNO,FE(FENO))=1
      else if (FETYPE(FENO).eq.'cov') then

```

```

EQ(PNO,1)=FE(FENO)
else
  print *, "unknown effect type! neither cro nor cov"
endif
else
  if(FETYPE(FENO).eq.'cro') then
    EQ(PNO,FE(FENO)+sum(LOFE(: (FENO-1))))=1
  else if (FETYPE(FENO).eq.'cov') then
    EQ(PNO,sum(LOFE(: (FENO-1)))+1)=FE(FENO)
  else
    print *, "unknown effect type! neither cro nor cov"
  endif
endif
enddo
read(89,*) RE(:)
do RENO=1, NORE
  if(RENO.eq.1) then
    if(RETYPE(RENO).eq.'cro') then
      EQ(PNO,RE(RENO)+RPOSI)=1
    else if (RETYPE(RENO).eq.'cov') then
      EQ(PNO,RPOSI+1)=RE(RENO)
    else
      print *, "unknown effect type! neither cro nor cov"
    endif
  else
    if(RETYPE(RENO).eq.'cro') then
      EQ(PNO,RE(RENO)+RPOSI+sum(LORE(: (RENO-1))))=1
    else if (RETYPE(RENO).eq.'cov') then
      EQ(PNO,RPOSI+sum(LORE(: (RENO-1)))+1)=RE(RENO)
    else
      print *, "unknown effect type! neither cro nor cov"
    endif
  endif
enddo
enddo
EQ(:, (MPOSI+1):)=EQ(:, (MPOSI+1):)-1
do MNO=1, NOM+MPOSI
  NO0=0; NO1=0; NO2=0
  do PNO=1, NOP
    if(MNO.gt.MPOSI.and.EQ(PNO,MNO).eq.(-1)) NO0=NO0+1
    if(MNO.gt.MPOSI.and.EQ(PNO,MNO).eq.0) NO1=NO1+1
    if(MNO.gt.MPOSI.and.EQ(PNO,MNO).eq.1) NO2=NO2+1
  enddo
  if(MNO.gt.MPOSI) then
    EH(MNO-MPOSI)=real((NO0*2+NO1)*(NO2*2+NO1),8)/real(NOP**2*2,8)
    SOEH=SOEH+EH(MNO-MPOSI)
    write(92,*) EH(MNO-MPOSI)
  endif
! OH(MNO)=real(NO1,8)/real(NOP**2*2,8)
! print *, SOEH,NO0,NO1,NO2
enddo
SQL=0
print *, SOEH
SOA=((NU_A-2)/NU_A)*AGV/SOEH ! scale parameter of alpha(marker
effect)=(v-2)*AGV/v*NOM*mean2pq(or v*SOEH)
SOE=((NU_E-2)/NU_E)*EV ! scale parameter of alpha(marker effect)=(v-2)*EV
! print *, SOA
!reading phenotype data
do PNO=1,NOP

```

```

    read(97,*) PHE(PNO)
  enddo
! adjust y for the first time with mu
! ADJ_y=PHE-sum(PHE)/real(NOP)
  ADJ_y=PHE
! mcmc sampling
do SNO=1,NOS
call set_seed(SNO)
! sample residual variance
  RV=(dot_product(ADJ_y,ADJ_y)+SOE*NU_E)/genchi(real(NOP,4)+NU_E)
  if(SNO.gt.BIP) meanRV = meanRV + RV
! sample fixed effects
  FPOSI=0
  do FENO=1, NOFE
    do LNO=1, LOFE(FENO)
      FPOSI=FPOSI+1
      ADJ_y=ADJ_y + EQ(:,FPOSI)*SOL(FPOSI)
      RHS=dot_product(EQ(:,FPOSI),ADJ_y)/RV
      LHS=dot_product(EQ(:,FPOSI),EQ(:,FPOSI))/RV
      invLHS=1.0/LHS
      MU=invLHS*RHS
      SOL(FPOSI)=gennor(MU,sqrt(invLHS))
      ADJ_y=ADJ_y - EQ(:,FPOSI)*SOL(FPOSI) ! adjust y for the new value of fixed effect
      if(SNO.gt.BIP) meanSOL(FPOSI) = meanSOL(FPOSI) + SOL(FPOSI)
    enddo
  enddo
! sample random variance
  RPOSI=FPOSI
  do RENO=1, NORE
    SRV(RENO)=(dot_product(ADJ_y((RPOSI+1):(RPOSI+LORE(RENO))),
      ADJ_y((RPOSI+1):(RPOSI+LORE(RENO)))))/genchi(real(LORE(RENO),4)-2)
    RPOSI=RPOSI+LORE(RENO)
    if(SNO.gt.BIP) meanSRV(RENO) = meanSRV(RENO) + SRV(RENO)
  enddo
! sample simple random effects
  RPOSI=FPOSI
  do RENO=1, NORE
    do LNO=1, LORE(RENO)
      RPOSI=RPOSI+1
      ADJ_y=ADJ_y + EQ(:,RPOSI)*SOL(RPOSI)
      RHS=dot_product(EQ(:,RPOSI),ADJ_y)/RV
      LHS=dot_product(EQ(:,RPOSI),EQ(:,RPOSI))/RV + 1.0/SRV(RENO)
      invLHS=1.0/LHS
      MU=invLHS*RHS
      SOL(RPOSI)=gennor(MU,sqrt(invLHS))
      ADJ_y=ADJ_y - EQ(:,RPOSI)*SOL(RPOSI) ! adjust y for the new value of fixed effect
      if(SNO.gt.BIP) meanSOL(RPOSI) = meanSOL(RPOSI) + SOL(RPOSI)
    enddo
  enddo
! sample variance for each locus
  do MNO=MPOSI+1,NOM+MPOSI
    VBM(MNO)=(SOA*NU_A+SOL(MNO)*SOL(MNO))/genchi(NU_A+1.0)
    if(SNO.gt.BIP) meanVBM(MNO) = meanVBM(MNO) + VBM(MNO)
  enddo
! sample effect for each locus
  do MNO=MPOSI+1,NOM+MPOSI
    ADJ_y=ADJ_y + EQ(:,MNO)*SOL(MNO) ! unadjust y for this locus
    RHS=dot_product(EQ(:,MNO),ADJ_y)/RV
    LHS=dot_product(EQ(:,MNO),EQ(:,MNO))/RV + 1.0/VBM(MNO)
  enddo

```

```

    invLHS=1.0/LHS
    MU=invLHS*RHS
    SOL(MNO)=gennor(MU,sqrt(invLHS))
    ADJ_y=ADJ_y - EQ(:,MNO)*SOL(MNO) ! adjust y for the new value of this locus
    if(SNO.gt.BIP) meanSOL(MNO) = meanSOL(MNO) + SOL(MNO)
  enddo
  if(SNO.le.BIP) write(90,*) SNO, RV, SOL(1), SOL(MPOSI+1)
  if(mod(SNO,100).eq.0) print *, SNO,"th sampling : residual variance is ",RV
enddo
meanSOL=meanSOL/real(NOS-BIP)
meanVBM=meanVBM/real(NOS-BIP)
meanRV=meanRV/real(NOS-BIP)
do MNO=1,NOM+MPOSI
  write(96,*) meanSOL(MNO)
  write(95,*) meanVBM(MNO)
enddo
GEBV=matmul(EQ(:,MPOSI:),meanSOL(MPOSI:))
do PNO=1,NOP
  write(94,*) GEBV(PNO)
enddo
EAGV=dot_product(meanVBM(2:), EH)
EAGV2=dot_product(meanSOL(2:)**2, EH)
EAGV3=dot_product(meanVBM(2:), (EH-1.0)*(-1.0))
EAGV4=dot_product(meanVBM(2:), (EH-1.0)*(-1.0))+EAGV2
EAGV5=dot_product(meanVBM(2:), EH)+dot_product(meanSOL(2:)**2, EH)
write(93,*) "*** basic values"
write(93,*) "output file tag "
write(93,*) OFT
write(93,*) "number of equations"
write(93,*) NOP
write(93,*) "number of markers"
write(93,*) NOM
write(93,*) "number of samplings"
write(93,*) NOS
write(93,*) "burn-in period"
write(93,*) BIP
write(93,*) "*** input prior"
write(93,*) "additive genetic variance"
write(93,*) AGV
write(93,*) "sum of gene frequency variance"
write(93,*) SOEH
write(93,*) "scale-parameter of marker effects variance as prior"
write(93,*) SOA
write(93,*) "degree of freedom for variance of marker effect as prior"
write(93,*) NU_A
write(93,*) "residual variance"
write(93,*) EV
write(93,*) "scale-parameter of residual variance as prior"
write(93,*) SOE
write(93,*) "degree of freedom for variance of residual as prior "
write(93,*) NU_E
write(93,*) "*** estimates"
write(93,*) "estimated additive genetic variance1(dot_product(var(j),2pq(j)))"
write(93,*) EAGV
write(93,*) "estimated additive genetic
variance2(dot_product(var(j),2pq(j))+dot_product(alpha(j)^2,2pq(j)))"
write(93,*) EAGV5
write(93,*) "estimated additive genetic variance3(dot_product(alpha(j)^2,2pq(j)))"
write(93,*) EAGV2

```

```

write(93,*) "estimated additive genetic variance4(dot_product(var(j),1-2pq(j)))"
write(93,*) EAGV3
write(93,*) "estimated additive genetic
variance5(dot_product(var(j),1-2pq(j))+dot_product(alpha(j)^2,2pq(j)))"
write(93,*) EAGV4
write(93,*) "estimated residual variance"
write(93,*) "heritability1(using EAGV1)"
write(93,*) EAGV/(EAGV+meanRV)
write(93,*) "heritability2(using EAGV2)"
write(93,*) EAGV5/(EAGV2+meanRV)
write(93,*) "heritability3(using EAGV3)"
write(93,*) EAGV2/(EAGV2+meanRV)
write(93,*) "heritability4(using EAGV4)"
write(93,*) EAGV3/(EAGV3+meanRV)
write(93,*) "heritability5(using EAGV5)"
write(93,*) EAGV4/(EAGV4+meanRV)
write(93,*) meanRV
do RENO=1, NORE
  write(93,*) "estimated variance of simple random effect", NARE(RENO)
  write(93,*) meanSRV(RENO)
enddo
write(93,*) "estimated fixed effect"
do FENO=1,NOFE
  do LNO=1, LOFE(FENO)
    sol_no=sol_no+1
    write(93,*) NAFE(FENO),LNO, meanSOL(sol_no)
  enddo
enddo
write(93,*) "estimated simple random effect"
do RENO=1,NORE
  do LNO=1, LORE(RENO)
    sol_no=sol_no+1
    write(93,*) NARE(RENO),LNO, meanSOL(sol_no)
  enddo
enddo
end program BayesA
!-----
! Initialize Random number generate by seed
!-----

subroutine initseed()
use ranlib
character*8 phrase
integer:: iseed1,iseed2,a1, r1, r2
integer:: now(3)
integer :: iz,i1,i2
call itime(now)
  iz=ichar('0')
  i1=floor(now(3)/10.)+iz
  i2=mod(now(3),10)+iz
  r1=int(rand()*9.9)+iz
  r2=int(rand()*9.9)+iz
  phrase(2:2)=achar(i1)
  phrase(3:3)=achar(i2)
  phrase(1:1)=achar(r1)
  i1=floor(now(2)/10.)+iz
  i2=mod(now(2),10)+iz
  phrase(4:4)=achar(i1)
  phrase(5:5)=achar(i2)
  phrase(8:8)=achar(r2)

```

```

        i1=floor(now(1)/10.)+iz
        i2=mod(now(1),10)+iz
        phrase(6:6)=achar(i1)
        phrase(7:7)=achar(i2)
!   phrase = clock(): by JEP
call phrtsd(phrase,iseed1,iseed2)
a1=ignuin(1,2)
if(a1 == 1) then
    call set_seed(iseed1)
else
    call set_seed(iseed2)
endif
end subroutine
!-----
! Seeting seed
!-----
subroutine set_seed(n)
use ranlib
! sets seed for all random number generators in this module

integer::n
real::x
call inrgcm()
x=ranf() !need to generate at least one sample before setting the seed
call setsd(n,n)
end subroutine

```

(2) GBLUP

(가) 알고리즘

개체의 유전체 유전능력을 평가하기 위하여 GBLUP방법을 이용하였으며 혼합모형방정식은 다음과 같다 (VanRaden, 2008).

$$\begin{bmatrix} 1'1 & 1'Z \\ Z'1 & Z'Z + G^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} 1'y \\ Z'y \end{bmatrix}$$

여기서, y는 관측치에 대한 벡터이며, μ 는 전체평균, Z는 계수행렬, \hat{u} 는 임의효과로써 개체의 상가적 유전체효과에 대한 벡터, e는 임의잔차효과이다. σ_a^2, σ_e^2 는 각각 개체의 상가적 유전분산과 잔차분산, G는 유전체 정보를 이용한 개체간 혈연계수 행렬이며 다음과 같이 계산된다.

개체의 수를 n, 좌위(locus)의 수를 m이라고 하면, 행의 크기는 n이고 열의 크기는 m인 M의 행렬이 된다. M행렬의 각 원소는 소수 대립유전자(Miner allele)의 동형접합체에서 -1, 다수 대립유전자(Major allele)의 동형접합체의 경우 1 그리고 이형접합체는 0로 구성된 행렬이다. MM' 행렬을 구성하면 대각선의 값은 각 개체에 대한 동형접합체의 수이며, 비대각선의 값은 두 개체간 공동으로 소유하는 대립유전자의 수이다. 이에 반해서 M'M의 대각선은 각 좌위에 대한 동형접합체의 수, 비대각선은 서로 다른 좌위에서 공동으로 소유하는 대립유전자의 수이다. i번째 좌위에 두 번째 대립유전자의 빈도를 p_i 라고 하고, P행렬의 i번째 열을 $2(p_i-0.5)$ 로 정의하면 P는 0을 평

균으로 갖는 기대유전자형에 대한 빈도값이 된다. 관측된 M의 대립유전자 행렬에서 기대대립 유전자형에 대한 값으로 나타나는 P에 대한 행렬의 값을 빼 주게 되면 대립 유전자 효과의 평균이 0이 되는(Centered value) Z행렬을 구성할 수 있다. 위에서 계산된 Z행렬을 이용하여 아래와 같은 수식에 의해 G행렬을 구성할 수 있다(VanRaden, 2007; VanRaden, 2008).

$$P_{ij} = 2(p_i - 0.5)$$

$$Z = M - P$$

$$G = \frac{ZZ'}{2\sum p_i(1-p_i)}$$

(나) 프로그램 코드 (R)

```

data=read.table("./geno.dat")
ped=read.table("./pedigree.dat")
test_pop_size = 200
startanimal = 1
data[((nrow(data)-test_pop_size+1):nrow(data)),2] = 0
X = desgn(rep(1,nrow(data)))
sigma_a = 0.3
sigma_e = 0.7
alpha = sigma_e / sigma_a
Z = desgn(data[,1])
y = data[,2]
y=as.matrix(y)
ai = ainvs(ped)
LHS = rbind(cbind(t(X) %*% X, t(X) %*% Z),cbind(t(Z) %*% X, t(Z) %*% Z + ai * alpha))
RHS = rbind(t(X) %*% y, t(Z) %*% y)
gi_LHS = ginv(LHS)
sol = gi_LHS %*% RHS
gped=data[,c(-1,-2,-3,-4)]
Z = diag(1,nrow(data))
gmatrix = g_matrix(gped)
gi = ginv(gmatrix)
LHS = rbind(cbind(t(X) %*% X, t(X) %*% Z),cbind(t(Z) %*% X, t(Z) %*% Z + gi * alpha))
RHS = rbind(t(X) %*% y, t(Z) %*% y)
gi_LHS = ginv(LHS)
sol1 = gi_LHS %*% RHS
tbv=data[,3]
ebv1 = sol1[((startanimal+1):nrow(sol1)),]
ebv2 = sol1[-c(1),]
ref_cor_ebv = cor(ebv1,tbv)
ref_cor_gebv = cor(ebv2,tbv)
test_ebv1 = ebv1[((length(ebv1)-test_pop_size+1):length(ebv1))]
test_ebv2 = ebv2[((length(ebv2)-test_pop_size+1):length(ebv2))]
test_tbv = tbv[((length(tbv)-test_pop_size+1):length(tbv))]
test_cor_ebv = cor(test_ebv1,test_tbv)
test_cor_gebv = cor(test_ebv2,test_tbv)
library("MASS")
desgn <- function(v) {
  if(is.numeric(v)){
    va=v
  }
}

```

```

    mrow = length(va)
    mcol = max(va)
  }
  if(is.character(v)){
    vf = factor(v)
    va = as.numeric(vf)
    mrow = length(va)
    mcol = length(levels(vf))
  }

  X = matrix(data=c(0), nrow=mrow, ncol=mcol)

  for(i in 1:mrow){
    ic = va[i]
    X[i,ic] = 1
  }
  return(X)
}

g_matrix=function(gped){
  library(MASS)
  M1=gped
  M= M1[,2:ncol(M1)]-1
  p1=round((apply(M,2,sum)+nrow(M))/(nrow(M)*2),3)
  p=2*(p1-.5)
  P = matrix(p,byrow=T,nrow=nrow(M),ncol=ncol(M))
  Z = as.matrix(M-P)
  b=1-p1
  c=p1*b
  d=2*(sum(c))
  ZZt = Z %*% t(Z)
  G = (ZZt/d)
  return(G)
}

ainv = function(pedi){
  n = nrow(pedi)
  Ainv = matrix(c(0), nrow = n, ncol = n)
  for ( i in 1:n) {
    animal = pedi[i,1]
    sire = pedi[i,2]
    dam = pedi[i,3]
    if (sire == 0 & dam == 0) {
      alpha = 1
      Ainv[animal,animal] = alpha + Ainv[animal,animal]
    }
    else if (sire != 0 & dam == 0) {
      alpha = 4/3
      Ainv[animal,animal] = alpha + Ainv[animal,animal]
      Ainv[animal,sire] = -alpha/2 + Ainv[animal,sire]
      Ainv[sire,animal] = -alpha/2 + Ainv[sire,animal]
      Ainv[sire,sire] = alpha/4 + Ainv[sire,sire]
    }
    else if (sire == 0 & dam != 0) {
      alpha = 4/3
      Ainv[animal,animal] = alpha + Ainv[animal,animal]
      Ainv[animal,dam] = -alpha/2 + Ainv[animal,dam]
      Ainv[dam,animal] = -alpha/2 + Ainv[dam,animal]
      Ainv[dam,dam] = alpha/4 + Ainv[dam,dam]
    }
  }
  else {

```



```

alpha = 2
Ainv[animal,animal] = alpha + Ainv[animal,animal]
Ainv[animal,sire] = -alpha/2 + Ainv[animal,sire]
Ainv[sire,animal] = -alpha/2 + Ainv[sire,animal]
Ainv[animal,dam] = -alpha/2 + Ainv[animal,dam]
Ainv[dam,animal] = -alpha/2 + Ainv[dam,animal]
Ainv[sire,sire] = alpha/4 + Ainv[sire,sire]
Ainv[sire,dam] = alpha/4 + Ainv[sire,dam]
Ainv[dam,sire] = alpha/4 + Ainv[dam,sire]
Ainv[dam,dam] = alpha/4 + Ainv[dam,dam]
}
}
return(Ainv)
}
a_matrix=function(ped){
NRM=matrix(0,nrow(ped),nrow(ped))
for(i in 1:nrow(ped)){
a=ped[i,1]
s=ped[i,2]
d=ped[i,3]
if(s > 0 && d > 0){
for(j in 1:(a-1)){
NRM[j,a] = 0.5*(NRM[j,s]+ NRM[j,d])
NRM[a,j] = NRM[j,a]
}
NRM[a,a] = 1 + 0.5 * NRM[s,d]
} else{
if (s+d > 0 && s*d == 0){
for(j in 1:(a-1)){
NRM[j,a] = 0.5 * NRM[j,s+d]
NRM[a,j] = NRM[j,a]
}
}
NRM[a,a] = 1.0
}
}
return(NRM)
}

```

4. MBV 프로그램을 이용하여 돼지 종축의 각 경제형질에 대한 표지인자의 유전변이 추정

가. 공시재료

3장 1절의 공시재료의 내용과 동일

나. 연구방법

BayesA 방법을 이용하였으며, 모형식은 3절과 동일

다. 연구결과

BayesA 방법을 이용하여 대상 형질발현에 관여하는 모든 표지인자에 대한 효과를 추정하고 이를 이용하여 유전체 육종가를 추정하였다. 유두수에(좌측, 우측 및 전체) 대한 각 표지인자 효과 추정결과 및 상관분석 결과는 그림39 및 표 29와 같다.

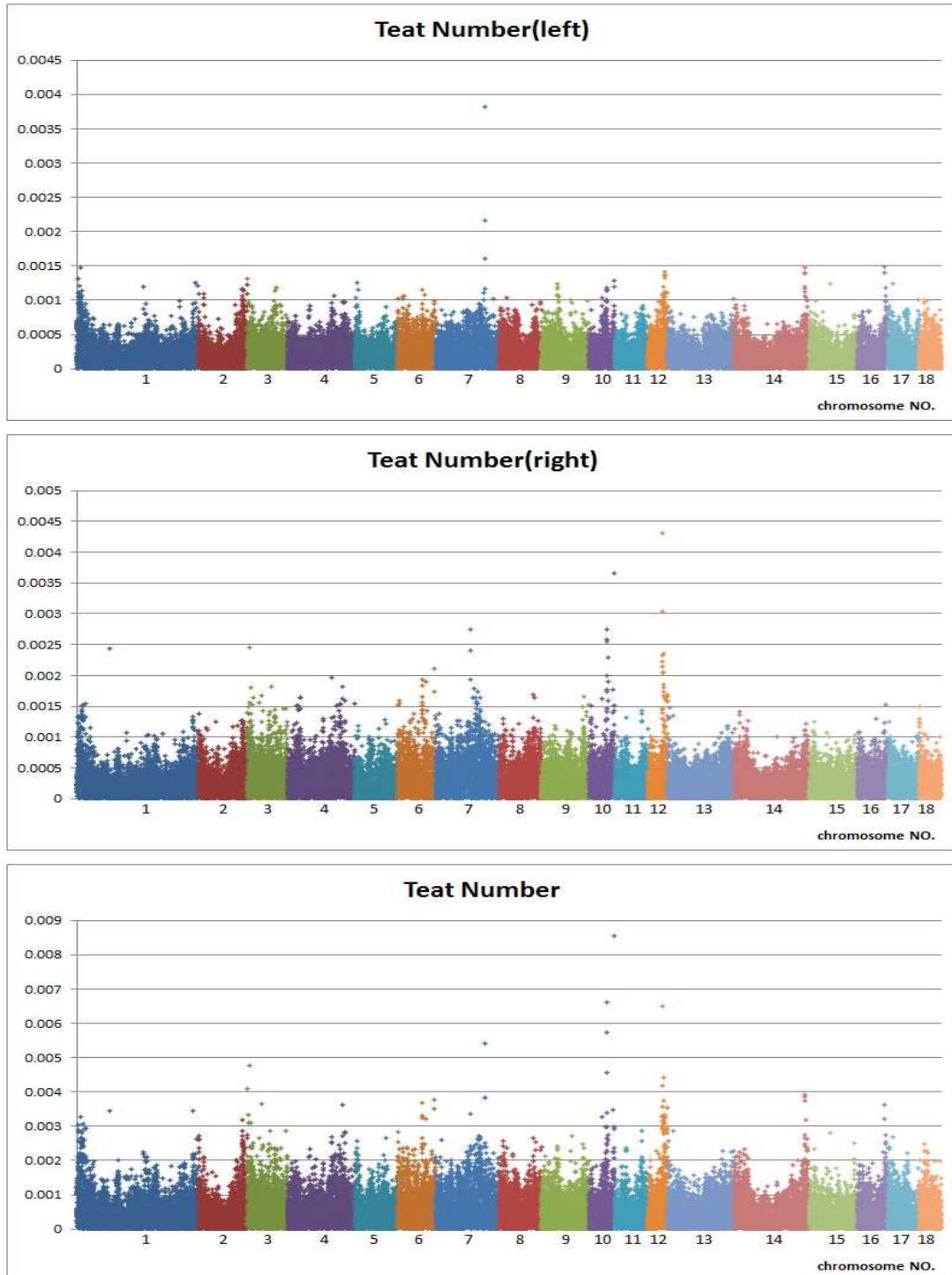
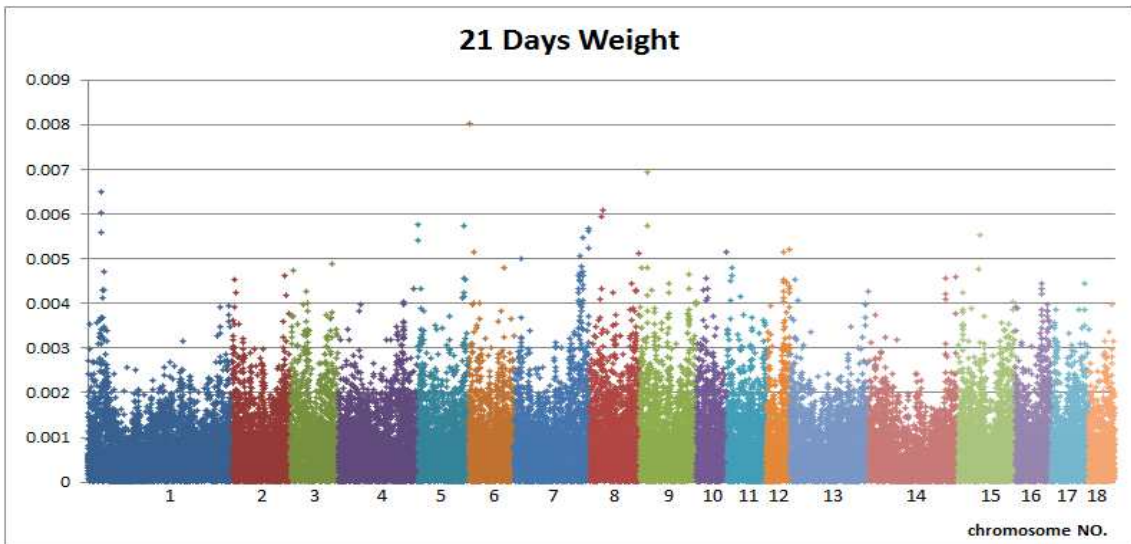
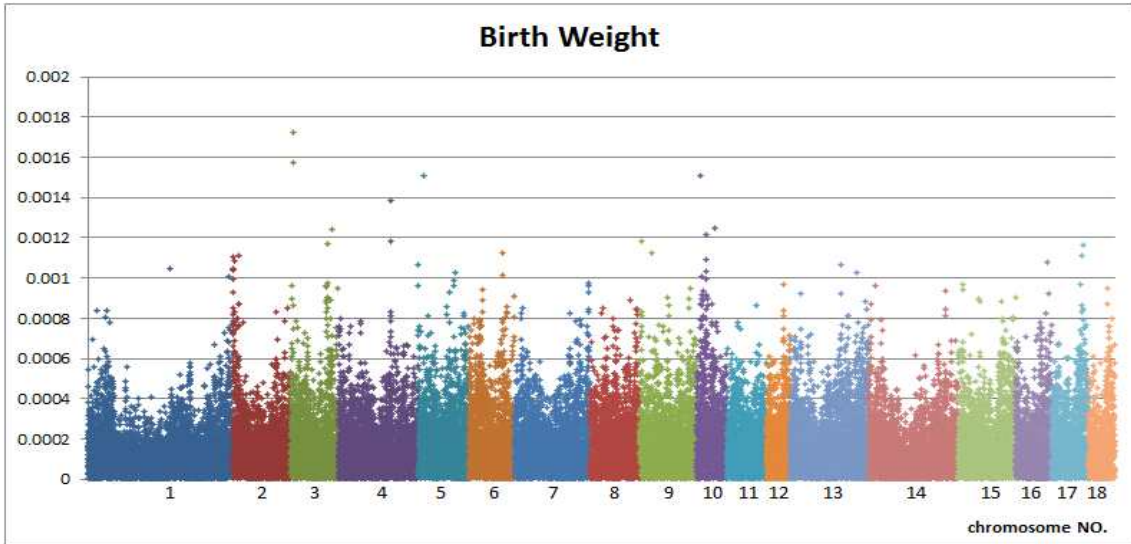


그림 39. 좌측, 우측 및 전체 유두수에 대한 표지인자 효과 추정 결과

유두수의 경우 일부의 표지인자에서 특이적으로 큰 효과를 나타내며, 총 유두수의 경우 좌우 유두수에서 효과가 추정된 표지인자들이 중복되어 나타나는 것을 확인할 수 있다. 이러한 패턴은 표지인자 효과의 상관분석 결과와 유사하게 나타난다.

표 29. 좌측, 우측 및 전체 유두수에 대한 표현형(상삼각 행렬) 및 표지인자 효과(하삼각 행렬)에 대한 상관분석 결과

	좌측유두수	우측유두수	총유두수
좌측유두수	1.000	0.525	0.872
우측유두수	0.510	1.000	0.875
총유두수	0.865	0.859	1.000



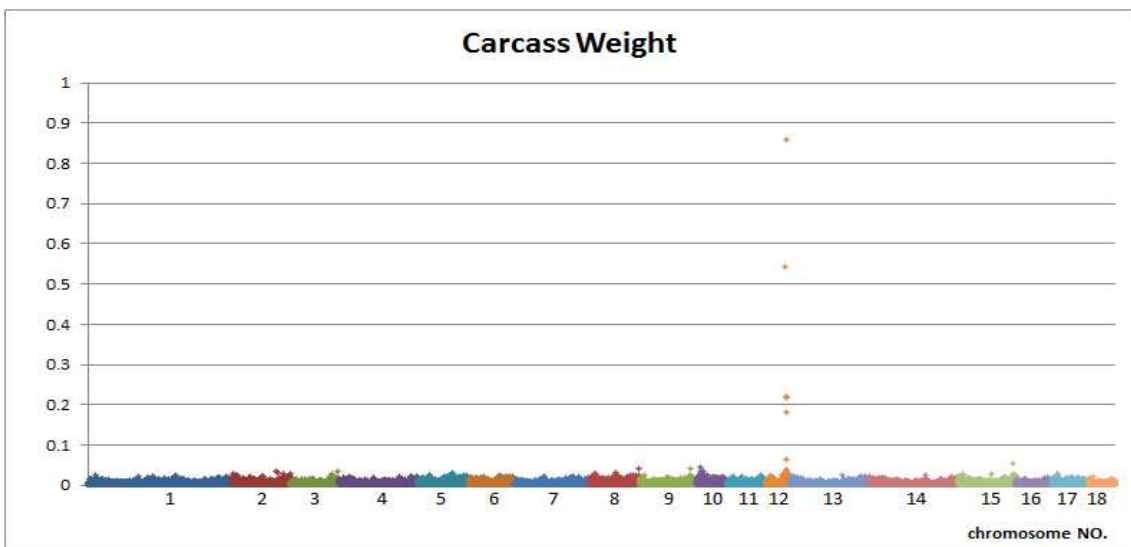
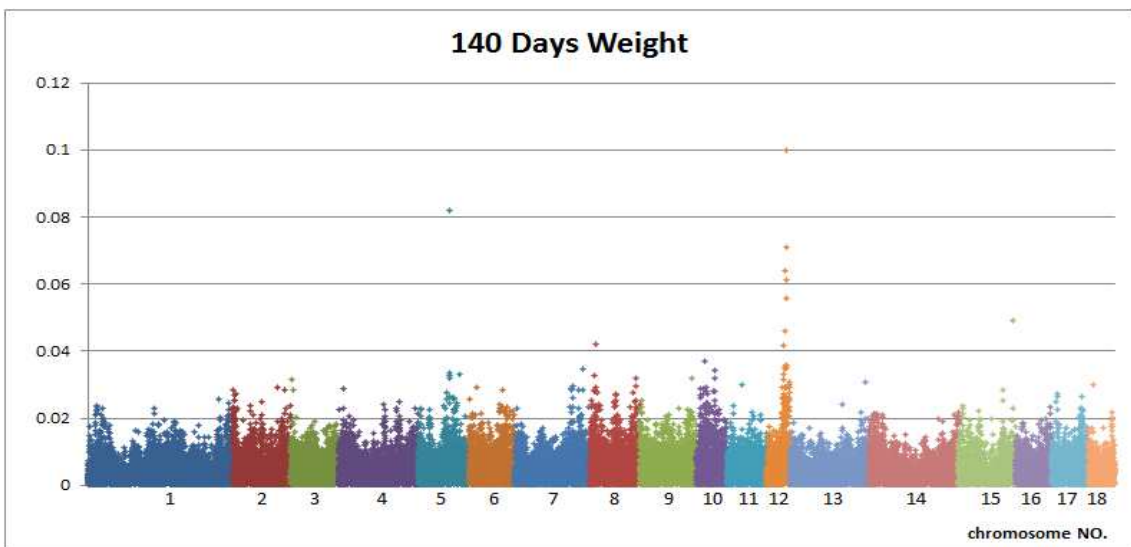
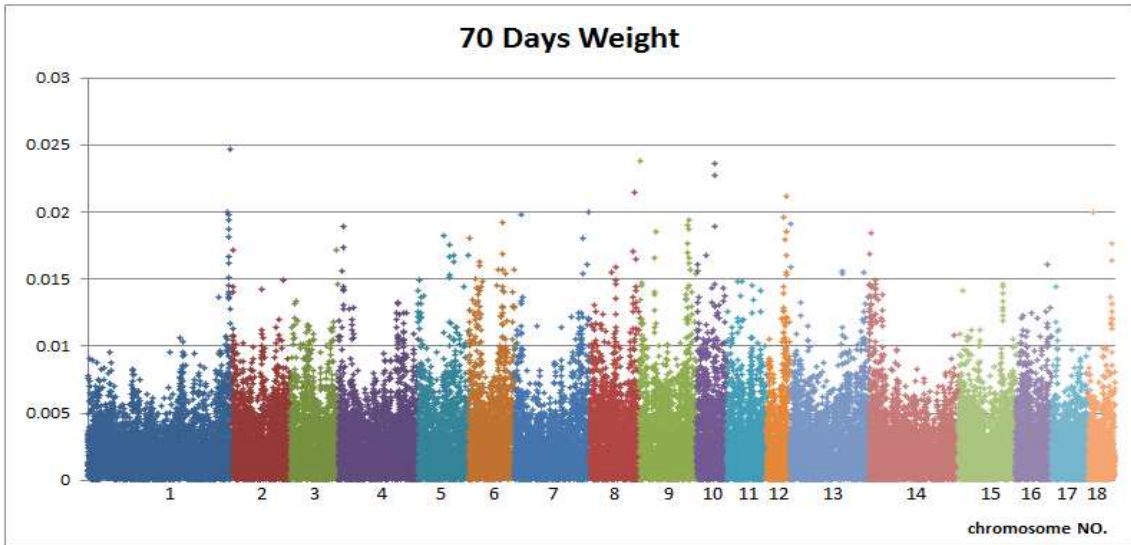


그림 40. 연령별 체중에 대한 표지인자 효과 추정 결과

그림 40는 연령별 체중에 관련된 표지인자 효과 추정자료이며, 어린 연령에서는 체중에 관여하는 유전자가 유전체 전장에 고르게 분포하다가 성장이 완료되어가면서 소수의 표

표지인자들이 큰 효과를 나타내는 패턴을 보이고 있다. 특히 12번 염색체에 존재하는 표지인자들에서 이러한 현상이 두드러지게 나타나고 있다.

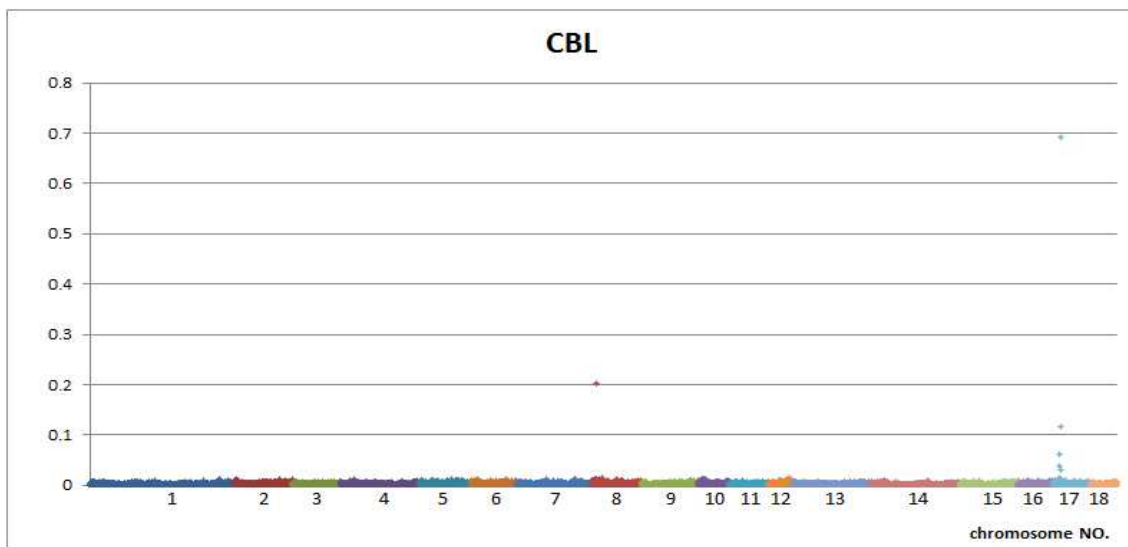


그림 41. 도체장에 대한 표지인자 효과 추정 결과

그림 41는 도체장에 관련된 표지인자 효과 분석자료이며, 소수의 표지인자들이 큰 효과를 나타내고 있다.

표 30. 혈통정보를 이용하여 추정된 육종가와 유전체 정보를 이용하여 추정된 육종가의 (순위) 상관 비교

	Person	Rank
총유두수	0.910	0.910
좌측유두수	0.890	0.894
우측유두수	0.902	0.896
생시체중	0.859	0.863
21일령 체중	0.833	0.822
70일령 체중	0.825	0.819
140일령 체중	0.868	0.859
도체중	0.714	0.738
도체장	0.807	0.801

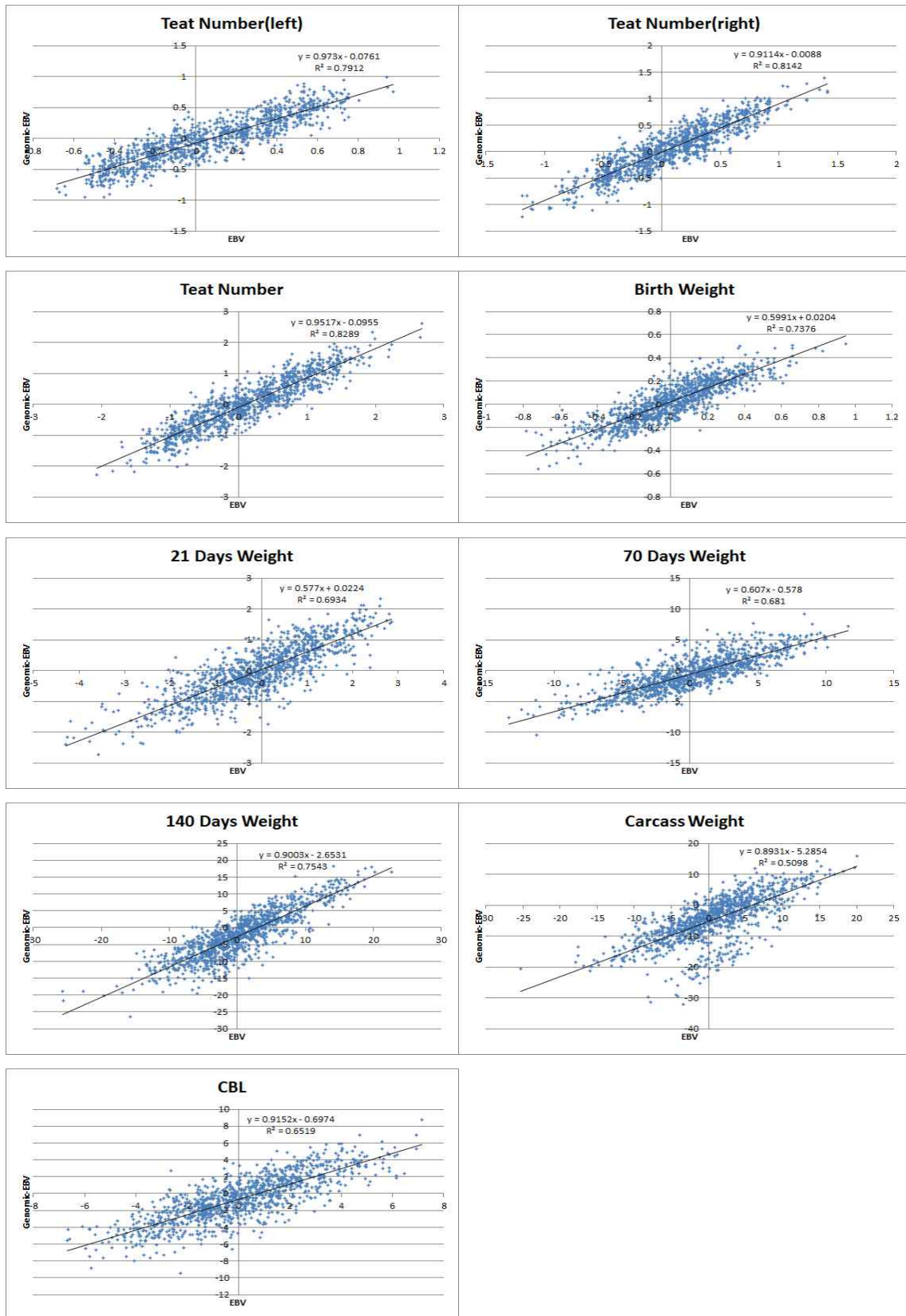


그림 42. 혈통정보를 이용하여 추정된 육종가와 유전체 정보를 이용하여 추정된 육종가의 도화 (plotting) 결과.

표 30과 그림 42와 같이 혈통정보를 이용하여 추정된 육종가와 유전체 정보를 이용하여

추정한 육종가의 상관은 형질별로 다르나 0.7-0.9정도로 나타나고 있어 서로 다른 정보를 이용하지만 유사한 육종가가 추정되는 것을 확인할 수 있다. 유전체 정보를 이용할 경우 혈연관계가 존재하지 않더라도 육종가를 추정하여 비교할 수 있으며, 추정된 표지 인자를 이용하여 표현형 자료가 결여된 개체를 조기에 선발할 경우 혈통정보를 이용할 때 보다 정확하게 육종가를 예측할 수 있다.

5. 유전체 모의실험을 통한 MBV 추정 프로그램 검증 및 정확도 측정

가. 모의실험

유전체 분석 프로그램의 검증을 위해 QMSim 프로그램을 이용하여 시뮬레이션하였다 (Sargolzaei and Schenkel, 2009). 본 연구의 시험축군과 유사한 형태로 가상의 유전체 정보, 혈통정보, 표현형정보 생성하였고 이를 정보를 활용하여 개발된 프로그램을 통해 추정하였다.

- (1) 시뮬레이션 모수 : 시뮬레이션은 단일형질로 가정하였으며 돼지의 체중형질로 가정하여 유전력을 0.3로 가정하였다.
- (2) 유전체 데이터 : Illumina Porcine 60K SNP Chip v2의 정보를 참고하여 상염색체(18쌍) 내 SNP마커의 물리적거리 정보를 이용하여 생성하였다(표 13)(Illumina, 2013).
- (3) 초기집단 : 유효집단의 크기는 100두로 가정하였다. 초기 집단은 유전적변이가 존재하지 않는 monomorphic marker를 갖는 세대로 가정하여 100세대를 진행하였으며, 세대를 진행하는 동안 돌연변이 및 교차를 발생시켜 다형성 있는 집단의 데이터를 생성하였다.
- (4) 세대성장 : 100세대의 유전정보를 이용하여 F0세대로써 제주재래돼지 19두(수 : 8두, 암 : 11두), 랜드레이스 17두(수 : 8두, 암 : 9두)를 생성하였으며, F0간 교배를 통하여 F1세대를 생성하였으며 (KL : 43두, LK : 48두), F1간 교배를 통하여 약 1,020여두의 F2집단의 데이터를 생성하였다(그림 43).

본 연구의 시뮬레이션 데이터 생성을 위하여 QMSim 프로그램(Sargolzaei and Schenkel, 2009)을 이용하였으며, 시뮬레이션 데이터생성은 다음과 같이 크게 3단계로 구분된다. 첫 번째는 초기 상태의 연관불평형(Linkage disequilibrium) 및 mutation-drift equilibrium을 구성하기 위해 선조집단(historical population)을 돌연변이와 이주 두 가지 조건을 통해 다형성을 만드는 단계로 임의의 개체간 선별을 통해 교배를 하였으며 임의 배우자(gamete)간 조합을 통해 자손의 유전정보를 얻었다. 또한 자손에 대한 성비는 동수비율로 고려하였고 초기 유효집단의 크기를 100두로 설정하여 100세대를 진행시키면서 2,000두로 집단을 성장시켰다.

두 번째는 현재집단(recent population)을 구성하는 단계로 선조집단의 마지막 세대로부터 제주재래돼지 19두(수: 8두, 암: 11두), 랜드레이스 17두(수: 8두, 암: 9두)의 F0세대를 생성하였으며, F0간 교배를 통하여 F1세대를 생성하였으며 (KL: 43두, LK: 48두), F1간 교배를 통하여 약 1,020여두의 F2집단의 데이터를 생성하였다(그림 43).

마지막 단계로 시뮬레이션에서 개체의 유전체정보는 젓소의 특성을 고려하여 상염색체를 제외한 나머지 18개 상염색체에 대해 Illumina Porcine 60K SNP Chip V2의 염색체 길이 및 염색체 내 마커의 수를 참고하여 생성하였으며 이에 대한 정보는 표 31에 나타냈다 (Illumina, 2013).

표 31. 시뮬레이션에 사용된 유전체 정보

SSC	Length of chromosome	Number of SNP markers
1	296	6,583
2	162	2,775
3	145	2,278
4	144	3,627
5	110	2,258
6	171	2,195
7	135	3,491
8	149	2,310
9	154	2,709
10	79	1,424
11	84	1,865
12	68	1,165
13	217	3,691
14	153	4,029
15	169	2,709
16	87	1,607
17	68	1,681
18	59	1,176

Porcine 60k SNP panel													
Progeny	Sire	Dam	G	Sex	NMPrg	NFPrg	F	Homo	Phen	Res	Polygene	QTL	Final_EBV
1	0	0	0	M	11	9	0.000000	0.636979	-0.384037	-0.740269	-0.000000	+0.356233	+0.026078
2	0	0	0	M	7	3	0.000000	0.635234	+2.213482	+1.977973	+0.000000	+0.235509	+0.317519
3	0	0	0	M	7	3	0.000000	0.631955	-0.704926	-1.166653	+0.000000	+0.461727	-0.191385
4	0	0	0	M	10	10	0.000000	0.636411	-0.258813	+0.413956	-0.000000	-0.672769	+0.331721
5	0	0	0	M	3	7	0.000000	0.638345	-0.538922	-0.696793	+0.000000	+0.157871	-0.140319
6	0	0	0	M	6	4	0.000000	0.635697	-1.467436	-0.898114	+0.000000	-0.569322	-0.565506
7	0	0	0	M	9	11	0.000000	0.625691	-0.511137	+0.083030	-0.000000	-0.594167	-0.744081
8	0	0	0	M	6	4	0.000000	0.632964	-0.434250	-0.598779	+0.000000	+0.164529	+0.169873
9	0	0	0	F	4	6	0.000000	0.633469	-0.045095	+0.451226	-0.000000	-0.496321	-0.128392
10	0	0	0	F	6	4	0.000000	0.640931	+0.491800	+0.367758	-0.000000	+0.124042	+0.005298
11	0	0	0	F	5	5	0.000000	0.635381	+0.298079	+0.437697	-0.000000	-0.139618	+0.031447
12	0	0	0	F	3	7	0.000000	0.633742	+1.353915	+2.299211	+0.000000	-0.945296	+0.373023
13	0	0	0	F	5	5	0.000000	0.632439	-0.137426	-0.194330	+0.000000	+0.056904	+0.221040
14	0	0	0	F	7	3	0.000000	0.632481	-0.392246	+0.819703	-0.000000	-1.211949	-0.389161
15	0	0	0	F	5	5	0.000000	0.637483	+0.268292	-0.564058	+0.000000	+0.832350	+0.524066
16	0	0	0	F	5	5	0.000000	0.630526	+1.273254	+1.862132	-0.000000	-0.588878	-0.143987
17	0	0	0	F	6	4	0.000000	0.633027	-0.908909	-0.111709	+0.000000	-0.797200	-0.414032
18	0	0	0	F	6	4	0.000000	0.635634	-0.840990	-0.587386	-0.000000	-0.253603	+0.059565
19	0	0	0	F	7	3	0.000000	0.640027	-1.080728	-0.586998	+0.000000	-0.493729	-0.293303
20	0	0	0	M	3	7	0.000000	0.631346	-0.754236	-0.773598	-0.000000	+0.019363	-0.180345
21	0	0	0	M	6	4	0.000000	0.639018	+0.849809	+0.803465	+0.000000	+0.046344	+0.527917
22	0	0	0	M	6	4	0.000000	0.634036	+2.314608	+1.053196	+0.000000	+1.261412	+0.723458
23	0	0	0	M	6	4	0.000000	0.629685	+1.345246	+1.358787	-0.000000	-0.013541	+0.017015
24	0	0	0	M	3	7	0.000000	0.631451	-0.298765	-1.042184	-0.000000	+0.743419	-0.162593
25	0	0	0	M	7	3	0.000000	0.634267	+0.253923	-0.436411	+0.000000	+0.690334	+0.643615
26	0	0	0	M	12	8	0.000000	0.632439	-0.191048	+0.021459	-0.000000	-0.212507	-0.009095
27	0	0	0	F	6	4	0.000000	0.637147	-0.681709	-0.887733	+0.000000	+0.206024	+0.057706
28	0	0	0	F	3	7	0.000000	0.627730	-0.707151	-0.975286	+0.000000	+0.268135	-0.165889
29	0	0	0	F	3	7	0.000000	0.629138	-0.967425	-0.189650	+0.000000	-0.777775	-0.367887
30	0	0	0	F	5	5	0.000000	0.632712	+1.561226	+1.379758	+0.000000	+0.181467	+0.215136
31	0	0	0	F	6	4	0.000000	0.637357	-1.241204	-0.580174	+0.000000	-0.661030	-0.368257
32	0	0	0	F	7	3	0.000000	0.629012	-0.181479	-0.567462	+0.000000	+0.385983	+0.258039
33	0	0	0	F	6	4	0.000000	0.633616	+1.112615	+1.013572	+0.000000	+0.099043	-0.054408
34	0	0	0	F	7	3	0.000000	0.636033	+0.094909	-0.266531	-0.000000	+0.361440	+0.594794
35	7	16	1	M	0	0	0.000000	0.600894	-1.800078	-0.839647	+0.000000	-0.960431	-0.656762
36	7	16	1	M	0	0	0.000000	0.612476	-0.356274	+0.474468	+0.000000	-0.830741	-0.430267

그림 43. 시뮬레이션을 통해 생성된 혈통, 표현형 및 육종가 정보

나. 유전체 분석프로그램 정확도 검증

유전체 분석프로그램을 검증하기 위하여 참조집단과 검정집단으로 구분하여 비교 분석하였다. 전체 1256두 중 1056두를 참조집단으로 표현형과 유전체정보를 모두 갖는 집단으로 구성하였으며, F2세대에서 200두를 임의선별하여 표현형정보가 없고 오직 유전체정보만 갖는 집단으로 설정하였다.

(1) 모형

(가) BLUP

유전체 정보가 없이 표현형 기록 및 혈통정보를 통한 개체유전능력평가는 헨더슨의 혼합모형방정식을 이용한 BLUP(Best linear unbiased prediction)방법을 적용하여 추정하였으며 통계분석 모형은 다음과 같다 (Henderson, 1975).

$$y = 1'\mu + Za + e$$

여기서, y는 관측치에 대한 벡터이며, μ 는 전체평균, Z는 계수행렬, a는 임의효과로써 개체의 상가적 효과에 대한 벡터, e는 임의잔차효과이다. y에 대한 분산은 $ZAZ'\sigma_a^2 + I\sigma_e^2$ 이며, 여기서 A는 혈통정보를 이용한 개체간 혈연계수 행렬이며, σ_a^2, σ_e^2 는 각각 개체의 상가적 유전분산과 잔차분산이다. 이를 개체의 상가적 유전효과에 대한 식으로 전환하면 다음과 같다.

$$\left[Z'Z + A^{-1} \frac{\sigma_e^2}{\sigma_a^2} \right] [\hat{a}] = [Z'y]$$

위의 혼합모형방정식은 conjugate gradient method을 통해 추정하였다.

(나) GBLUP

개체의 유전체 유전능력을 평가하기 위하여 GBLUP방법을 이용하였으며 혼합모형방정식은 다음과 같다 (VanRaden, 2008).

$$\begin{bmatrix} 1'1 & 1'Z \\ Z'1 & Z'Z + G^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} 1'y \\ Z'y \end{bmatrix}$$

여기서, y 는 관측치에 대한 벡터이며, μ 는 전체평균, Z 는 계수행렬, \hat{u} 는 임의효과로써 개체의 상가적 유전체효과에 대한 벡터, e 는 임의잔차효과이다. σ_a^2, σ_e^2 는 각각 개체의 상가적 유전분산과 잔차분산, G 는 유전체 정보를 이용한 개체간 혈연계수 행렬이며 다음과 같이 계산된다.

개체의 수를 n , 좌위(locus)의 수를 m 이라고 하면, 행의 크기는 n 이고 열의 크기는 m 인 M 의 행렬이 된다. M 행렬의 각 원소는 소수 대립유전자(Miner allele)의 동형접합체에서 -1 , 다수 대립유전자(Major allele)의 동형접합체의 경우 1 그리고 이형접합체는 0 로 구성된 행렬이다. MM' 행렬을 구성하면 대각선의 값은 각 개체에 대한 동형접합체의 수이며, 비대각선의 값은 두 개체간 공동으로 소유하는 대립유전자의 수이다. 이에 반해서 $M'M$ 의 대각선은 각 좌위에 대한 동형접합체의 수, 비대각선은 서로 다른 좌위에서 공동으로 소유하는 대립유전자의 수이다. i 번째 좌위에 두 번째 대립유전자의 빈도를 p_i 라고 하고, P 행렬의 i 번째 열을 $2(p_i - 0.5)$ 로 정의하면 P 는 0 을 평균으로 갖는 기대유전자형에 대한 빈도값이 된다. 관측된 M 의 대립유전자 행렬에서 기대대립 유전자형에 대한 값으로 나타나는 P 에 대한 행렬의 값을 빼 주게 되면 대립유전자 효과의 평균이 0 이 되는(Centered value) Z 행렬을 구성할 수 있다. 위에서 계산된 Z 행렬을 이용하여 아래와 같은 수식에 의해 G 행렬을 구성할 수 있다(VanRaden, 2007; VanRaden, 2008).

$$P_{ij} = 2(p_i - 0.5)$$

$$Z = M - P$$

$$G = \frac{ZZ'}{2 \sum p_i(1 - p_i)}$$

(다) Bayesian

Bayesian은 BayesA와 BayesB 두 가지 방법을 통하여 추정하였다. Bayesian추정은 SNP마커추정과 육종가추정 두 단계로 구분하여 평가하였으며 SNP마커효과 추정을 위한 모형식은 아래와 같다.

$$y_i = 1_n' \mu + \sum X_{ij} g_j + e_i$$

여기서, y_i 는 i 번째 개체의 표현형 값, 1_n 는 원소가 1인 n 개의 행벡터, μ 는 전체평균, X_{ij} 는 i 번째 개체의 j 번째 좌위에서 Major allele의 수(0,1,2), g_j 는 j 번째 표지인자효과, e_i 는 i 번째 개체의 오차이다.

참조집단의 정보를 상기의 모형식에 적합하여 각 SNP마커의 효과를 추정하였으며, 여기서 추정된 SNP마커효과는 아래와 같은 수식에 의해 검정집단의 유전체 육종가를 추정하였다.

$$G\widehat{EBV}_i = X_{ij} \hat{g}_j$$

여기서, $G\widehat{EBV}_i$ 는 i 번째 개체의 추정육종가, X_{ij} 는 i 번째 개체의 j 번째 좌위에 SNP유전자형 표지인자효과(0,1,2)의 계수행렬, \hat{g}_j 는 j 번째 좌위에서 추정된 표지인자효과이다.

(라) 유전능력정확도

유전능력 추정방법별(BLUP, GBLUP, Bayesian) 개체의 육종가를 추정하였으며 (Estimated breeding value; EBV) 이를 참육종가(True breeding value; TBV)와 상관관계를 통하여 정확도 비교를 실시하였다.

$$accuracy_{(TBV,EBV)} = \frac{Cov(TBV,EBV)}{\sqrt{Var(TBV)Var(EBV)}}$$

(2) 결과

표 32에는 유전평가 방법별(BLUP, GBLUP, BayesA, BayesB) 유전평가의 정확도 나타났다. 표현형과 혈통기반의 BLUP은 유전체정보와 표현형정보를 모두 갖는 개체의 집단인 참조집단에서 0.56의 정확도로 추정되었으며, 유전체정보를 이용한 GBLUP, BayesA, BayesB의 정확도는 각각 0.71, 0.56, 0.74로 보여 BayesB방법으로 개체의 육종가를 추정하였을 때 가장 높은 추정의 정확도를 얻을 수 있었다. 자신의 표현형정보가 없고 유전체정보만 갖는 개체의 집단인 검정집단의 경우 BayesB가 0.49로 다양한 분석방법 중 육종가추정의 정확도가 가장 높은 것으로 나타났으며, GBLUP 0.47, BayesA 0.39, BLUP 0.11순으로 추정되었다.

그림 44에는 Bayesian방법을 이용하여 추정한 육종가와 참육종가간 상관에 대해 그래프로 나타냈다. 그림 44을 보면 BayesB로 개체의 육종가를 추정하였을 때 참육종가와 스케일 크기 및 산포도(퍼짐정도)가 선형적으로 잘 나타난 반면 BayesA의 경우 추정육종가의 스케일이 참육종가 크기와 다소 차이를 나타냈으며 산포도 역시 BayesB에 비하여 퍼짐의 정도가 더 큰 것으로 나타났다. 또한 Bayesian방법을 이용하여 추정된 SNP마커효과에 대한 분포를 그림 45에 나타냈다. BayesA로 마커효과를 추정하였을 때 대부분의 SNP마커들이 -0.1 에서 0.1의 값을 갖는 것으로 추정된 반면 BayesB는 대부분의 마커효과가 0에 가까운 것으로 추정되었다. 이는 BayesB로 마커효과를 추정할 때 $1-\pi$ 의 확률로 SNP마커를 선별하여 효과를 추정하기 때문에 나타난 결과로 사료된다.

본 시뮬레이션의 목적은 프로그램검증이였다. 시뮬레이션을 통하여 유전체 데이터, 표현

형 데이터, 혈통데이터 및 참육종가 정보를 생성하였고 이들 정보를 이용하여 BLUP, GBLUP, BayesA, BayesB방법으로 추정하였다. 그 결과 유전체 육종가의 정확도가 BLUP의 육종가 정확도에 비해 높게 나타나 Meuwissen 등(2001)과 유사한 결과를 얻었고 기타 다른 연구와 비슷한 연구결과를 나타내 유전체 유전능력평가 프로그램이 이론에 맞게 잘 구현되었다고 사료되며 시뮬레이션 결과 본 시험축군에서 유전체선발의 효과를 기대할 수 있을 것으로 사료되었다.

표 32. 유전평가 방법별, 집단형태별 육종가 정확도

Method	Reference Population	Test Population
	(N=1056)	(N=200)
BLUP	0.560	0.114
GBLUP	0.706	0.471
BayesA	0.560	0.387
BayesB	0.735	0.493

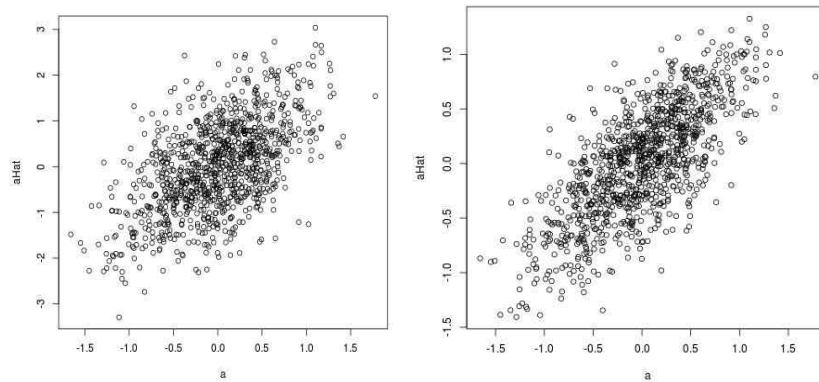


그림 44. 참조집단에서 참육종가와(x축) 추정육종가간(y축) 상관그래프(왼쪽 그래프 : BayesA, 오른쪽 그래프 : BayesB)

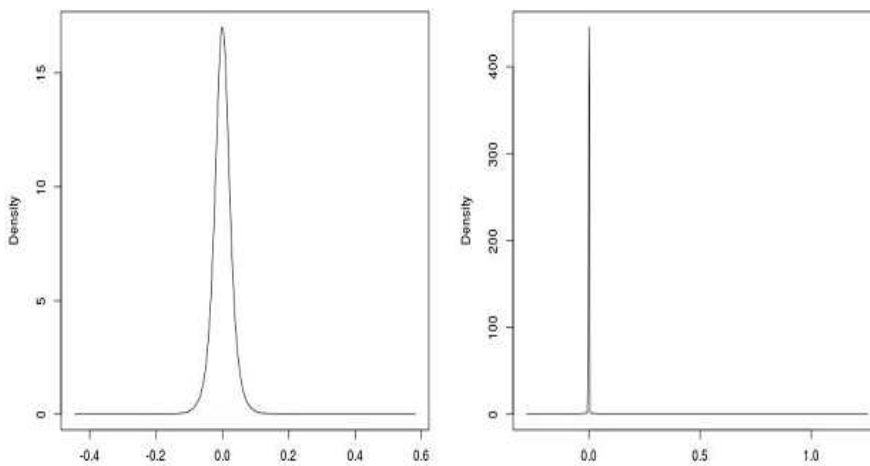


그림 45. Bayesian을 통해 추정된 마커효과들의 분포(왼쪽 그래프 : BayesA, 오른쪽 그래프 : BayesB)

다. 3. Cross Validation 정확도 평가 방법을 이용하여 유전체 육종가 정확도 검증

(1) Cross Validation(교차검증)

k-fold cross validation은 수집된 샘플들의 검증을 위한 통계적인 분석 방법이다. 전체 집합을 k개로 나눈 뒤 하나를 다른 것들과 비교하여 전체적으로 특이한 집합이 없는지 확인하는 방식이다. 현재 인공지능의 데이터 마이닝 연구에서 주로 사용되며, k-fold cross validation의 장점은 충분한 양의 데이터가 주어지지 않은 상황에서 주어진 데이터만을 가지고 검증할 수 있는 분석방법이며, 모든 observation이 training과 validation에 모두 사용된다는 것이다.

정확도 검증의 수행방법은 그림 46과 같다.

수집된 최초의 original sample을 k개의 subset sample로 나눈다. 이 가운데 k-1개의 subset을 사용해 training 한 후 나머지 1개의 subset을 validation 한다. subset sample들이 validation 데이터로 각각 한번씩 사용될 때까지 k개의 subset sample들은 cross-validation 프로세스 동안 k번 반복된다. 프로세스의 각 스텝마다 각 부분으로부터 나온 k개의 결과는 하나의 평가 지표로 만들기 위해 평균을 구하며, 이를 이용해 검증을 수행할 수 있다.

(2) 방법

재래돼지의 유전체 데이터를 10등분하였다. 그리고 그중에서 9그룹을 참조집단(Reference population) 학습에 사용하고, 나머지 1개 그룹은 검정집단(Validation population or Test population)으로 설정하여 MBV 추정의 정확도를 계산하였다.

참조집단으로부터 추정된 각각의 SNP 표지인자 효과를 검정집단의 개체들로부터 분석된 유전자형과 곱한 뒤 총합을 계산하여 육종가를 추정하고 개체들의 표현형 값을 예측한다. 예측된 표현형과 검정집단의 실제 표현형과의 correlation을 통하여 유전체 육종가의 정확도를 추정하였으며, 위의 과정을 60회 반복 실시하여 육종가의 정확도 평균을 계산하였다.

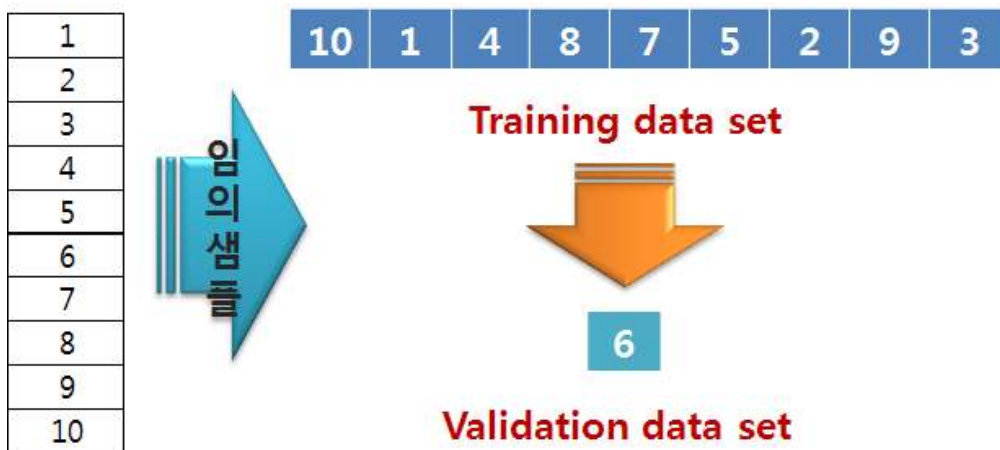


그림 46. 유전체 정보를 이용한 Cross validation 모식도

(3) 결과

표 33는 총유두수형질에 대하여 고정효과에 대하여 사전 보정한 값 및 Quality Control을 통해 가공된 SNP정보를 이용하여, 총 재래돼지를 임의로 10개의 sub set으로 구성하고, 9개 sub set으로부터 SNP의 효과를 추정하였으며, 여기서 추정된 SNP효과를 나머지 1개 sub set에 적용시켜 표현형과 추정값 사이의 상관을 나타냈으며, 위와 같은 과정을 60회 반복 실험하였다.

실험결과 추정된 MBV와 보정된 표현형 관측치와의 상관(정확도)이 0.592에서 0.752의 분포를 보였으며, 평균 0.592의 정확도를 나타내었다.

이는 표현형자료가 없는 개체에 대하여 BLUP 방법으로 추정된 육종가 정확도 평균보다 약 10%정도 다소 높게 추정된 결과이며, 전통적인 유전능력평가 방법보다 더욱 높은 정확도로 개체의 육종가를 추정 할 수 있을 것으로 사료된다.

또한 일정 수준 이상의 참조집단의 자료가 누적되었을 때 유전체 육종가와 혈통을 이용한 육종가의 상관비교, 교차 검증(Cross-Validation)등을 이용하여 유전체 육종가의 실제 적용에 대한 신뢰도를 추측할 수 있을 것으로 사료된다.

표 33. Cross validation을 이용하여 추정한 정확도 평균

N	Mean	SD	Min	Max
60	0.592	0.049	0.485	0.752

6. SNP Marker별 효과추정 및 개체의 유전능력추정

가. 표현형데이터와 유전체자료 활용 Marker효과 및 개체별 유전능력 추정

(1) 공시재료

(가) 공시축

제3장 1절의 공시재료의 내용과 동일

(나) 유전체 정보

본 연구에서 개체의 유전체 정보를 얻고자 Illumina Porcine 60k DNA chip을 이용하여 개체로부터 얻어진 DNA를 통해 약 6만 2천여개의 SNP에 대한 정보를 수집하였다.

(2) 연구방법

4절의 연구방법과 동일함 (BLUP, GBLUP)

(3) 연구결과

일반적으로 개체 육종가 정확도는 자신의 검정데이터 보유 여부, 혈연관계가 높은 개체의 검정데이터 보유 두수 등 다수의 요건이 작용하기 때문에 종축으로 사용되었으면서 자신의 기록을 갖지 않는 Sire(Sire without obs.)집단과 자신의 기록을 갖는 Progeny(Progeny with obs.)집단으로 구분하여 생시체중(BW), 3주령 체중(BW3), 10주령 체중(BW10), 20주령 체중(BW20)에 대한 육종가 및 육종가의 정확도를 표 16과 표 17에 각각 나타냈다. 표 34에서 품종별 육종가의 정확도를 보면 BLUP의 경우 BW와 BW3에서

제주재래돼지(KK)와 랜드레이스(LL)간 육종가의 큰 차이를 보이지 않다가 BW20에서 제주재래돼지가 랜드레이스에 비해 육종가가 높은 것으로 나타났다. 반면 GBLUP의 결과를 보면 생시체중부터 20주령 체중까지 랜드레이스의 육종가가 더 높은 것으로 나타났다. 이는 Han 등 (2007)의 연구결과와 유사한 결과로 랜드레이스 집단이 제주재래돼지 집단에 비해 높은 결과를 얻어 적은 자료 두수 및 세대간격이 짧은 축군에서 유전체 정보를 이용하여 개체의 유전능력을 추정된 GBLUP방법이 BLUP방법에 비해 보다 정확하게 추정되었음을 나타낸다고 사료된다. 집단 내 품종간 평균을 비교하여 보았을 때, 대부분의 성장형질에서 LL집단의 육종가가 가장 높게 나타났고, 이후 LK, F2, KL, KK 순으로 나타났다.

표 34. 4개의 체중형질에 대한 BLUP과 GBLUP방법간 육종가 비교

Type		BW		BW3		BW10		BW20	
		BLUP	GBLUP	BLUP	GBLUP	BLUP	GBLUP	BLUP	GBLUP
Sire without obs.	KK	-0.01	-0.21	0.03	-0.90	-0.33	-4.27	-0.21	-17.17
	LL	-0.02	0.13	-0.02	0.89	-0.54	3.92	-0.89	16.06
Progeny with obs.	KL	-0.02	-0.05	-0.13	-0.25	-1.79	-1.84	-2.02	-3.29
	LK	0.00	0.00	0.11	0.18	2.02	1.95	2.06	3.08
	F2	0.02	0.00	-0.01	0.00	0.04	-0.01	0.43	0.03
Total		0.02	0.00	-0.01	0.00	0.03	-0.02	0.35	-0.01

표 35는 성장형질별 집단구분별 분석방법별 육종가의 정확도를 비교하였다. 그 결과 혈통기반의 BLUP의 경우 Sire집단에서 성장형질별 0.57 - 0.59의 정확도를 보였으며, 유전체 정보를 포함하여 GBLUP으로 추정하였을 때 성장형질별 0.74 - 0.77로 GBLUP방법이 BLUP방법에 비해 0.16 - 0.19 육종가의 정확도가 높게 나타났다. 반면 개체가 표현형 정보를 갖는 Progeny세대에서는 BLUP이 0.76 - 0.85, GBLUP이 0.83 - 0.89로 유전체 정보를 사용함으로써 약 3 - 7%의 정확도 증가만을 보여 검증데이터가 없는 Sire집단에서 유전체 선발의 효율성이 증대하는 것으로 나타났다.

표 35. 4개의 체중형질에 대한 BLUP과 GBLUP방법간 육종가 정확도 비교

Type	BW			BW3			BW10			BW20			
	BLUP	GBLUP	Diff.	BLUP	GBLUP	Diff.	BLUP	GBLUP	Diff.	BLUP	GBLUP	Diff.	
Sire without obs.													
KK	0.58	0.77	0.19	0.58	0.77	0.19	0.58	0.76	0.18	0.57	0.75	0.18	
LL	0.59	0.77	0.17	0.59	0.77	0.17	0.59	0.75	0.17	0.58	0.74	0.16	
Progeny with obs.													
KL	0.85	0.89	0.04	0.85	0.89	0.04	0.84	0.88	0.04	0.83	0.87	0.04	
LK	0.85	0.88	0.03	0.85	0.88	0.03	0.84	0.87	0.04	0.83	0.86	0.04	
F2	0.79	0.85	0.06	0.79	0.85	0.06	0.77	0.84	0.07	0.76	0.83	0.07	
Total		0.78	0.85	0.07	0.78	0.85	0.07	0.77	0.84	0.07	0.76	0.83	0.07

표 36. 유두수형질에 대한 BLUP과 GBLUP방법간 육종가 정확도 비교

Traits	BLUP	GBLUP	Diff.
Teat_Left	0.733	0.851	0.118
Teat_Right	0.736	0.853	0.116
Teat_Total	0.735	0.852	0.117

나. 유전체자료 활용 순종 재래 종돈 선발 육종체계 제시

본 연구집단에서 표현형데이터 자료 및 유전체 정보를 이용하여 개별 SNP마커효과 추정 및 개체의 유전체 유전능력을 추정하였다. 여기서 추정된 SNP마커는 이 후 순종돈 선발을 위한 유전평가에 활용이 가능하며, 또한 추가된 순종돈 및 F1간 교배를 통해 생성된 F2에 대해 검정전에 유전체정보 획득을 통하여 개체의 유전능력을 추정함으로써 생산비 절감효과를 얻을 수 있을 것으로 기대된다 (그림 47). 이에 대한 효과는 앞서 분석한 전통 BLUP방법과 유전체 정보를 이용하여 추정한 GBLUP방법간 정확도의 차이 정도로 생각해 볼 수 있다. 본 연구의 최종목적은 유전체정보를 이용하여 개체의 유전능력을 추정하고 이를 통한 선발방법 제안까지 연구였으며 차후 연구에서 유전체 선발을 통하여 얻게 되는 추정에 정확도가 얼마만큼의 수익(지출 대비 수익)을 가져올 수 있는지에 대한 경제성에 대한 연구가 추가적으로 필요할 것으로 사료된다.

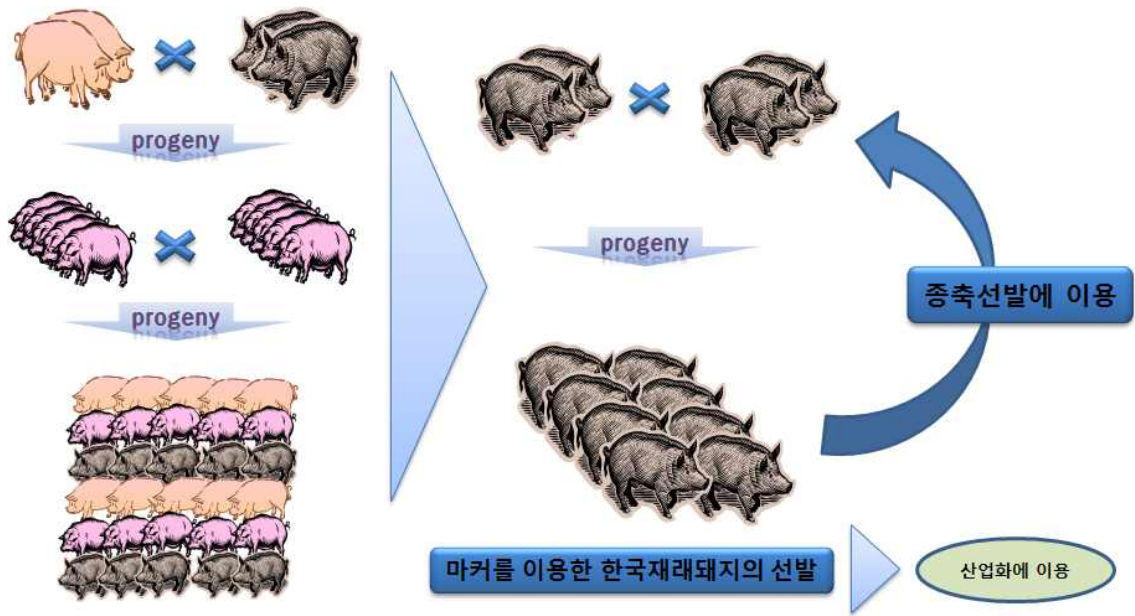


그림 47. SNP마커를 활용한 한국재래돼지 선발에 대한 선발체계

제 5 절 한국재래돼지의 성장 및 포유능력을 위한 축군조성 및 산업화 기반 조성

1. 한국재래돼지의 특성 및 불량형질의 개선

가. 한국재래돼지에 대한 육종이나 사양 시험에 대한 연구는 거의 없는 실정으로 단순히 국가 기관이나 지자체에서 품종 보존차원에서 보유하고 있는 실정이다. 재래돼지 사육두수가 적어 육종에 필요한 최소 두수 확보도 어려운 상황이라 할 수 있다. 따라서 재래돼지에 대한 전반적인 개량 및 불량형질을 개선하기 위해서는 전형적인 통계육종체계를 도입하기에는 여러 가지 제한이 따른다.

- (1) 한국재래돼지의 장점을 일반적으로 우리나라 기후에 적응성이 강해 강건성이 우수하며, 육질과 맛이 좋다.
- (2) 그러나 집단내에서 개체변이가 매우 크며 형질이 전혀 고정이 되어 있지 않다. 특히 도체형질중 등심단면을 절개해 보면 가식부위가 많은 개체와 가식부위가 거의 없어 거의 근간지방으로 가득 찬 개체들까지 다양하다. 산자수나 유두수, 체장 등에서도 육안으로 개체간 변이를 확인할 수 있다.

나. 본 연구과제를 통해서 한국재래돼지의 불량형질을 개선하기 위하여 전형적인 통계육종체계중 가계위주의 검정능력 자료를 참고로 하여 축군에 활용한 반면, 전반적인 재래돼지 육종 프로그램은 분자유전학적 기법을 도입하였다. 분자유전학적 기법 활용을 위하여 유용한 유전자 개발은 한국재래돼지와 랜드레이스 기반 참조축군에서 개발된 유전자 또는 DNA marker를 한국재래돼지 집단에 적용하여 비교적 짧은 기간내에 불량형질을 개선 및 산업화 축군을 조성하기 위하여 활용하고 있다.

- (1) 기존의 전통적인 통계육종 방법에 비해 형질연관 유전자가 동정될 경우 개체 단위에서 우수개체 선발이 가능하므로 단기간에 형질 고정이 가능한 장점이 있다.
- (2) 2차년도까지 한국재래돼지의 불량형질 개선을 위한 형질(번식형질과 성장형질) 원인 유전자 좌위가 확인되었지만 집단에 적용할 정도로 표현형을 설명해 줄 수 있는 DNA 마커의 정보력이 높지 않았으나 3차년도에는 구체적으로 표현형 연관 QTL영역내에서 대상 후보 유전자를 유전자 기능을 우선 선정하였으며 한국재래돼지 집단에 접목 가능한 마커 개발 위주로 연구과제가 추진이 되었다.
- (3) 한국재래돼지 집단에서 번식용으로 선발된 개체들 외에 비 선발개체들은 시험도축을 통하여 한국재래돼지의 불량형질 조사를 위하여 일반적인 도체조사와 함께 본 과제에서 조사 항목인 갈비뼈수 등이 조사가 되었다.
- (4) 한국재래돼지 순종 개체들을 대상으로 유전적 다양성을 유지하면서 번식용 우량 개체를 대량으로 생산하기 위하여 체형 및 능력검정과 더불어 DNA 마커 적용을 통하여 적절한 체격을 가진 장방향 개체를 산업화 기반을 위한 번식용 우량 재래돼지 집단으로 조성하고 있다.

가. 본 연구결과를 요약해 보면, 1차년도 1세대 자손에서 암컷은 전수 번식축으로 선정하여 교배가 진행되었고, 그중 수컷은 복당 1-2두만 번식축으로 활용되었다. 2차년도에는 장방향 체형의 암컷은 번식축으로 선발하였고 나머지 번식축에서 제외된 개체들은 도체 일반 분석 자료 습득을 위하여 시험도축되었으며 재래돼지의 우량형질과 열성형질을 발굴하는데 이

용되었으며, 육량조사는 대분할 방법을 통해 측정이 되었다. 3차년도에는 개발된 형질연관 유전자 또는 DNA 마커를 1세부과제와 협동하여 축군에 적용하여 유전자형과 표현형과의 효과를 검정하였다.



<방목중인 임신모돈>



<분만 자돈>



<자돈 육성>

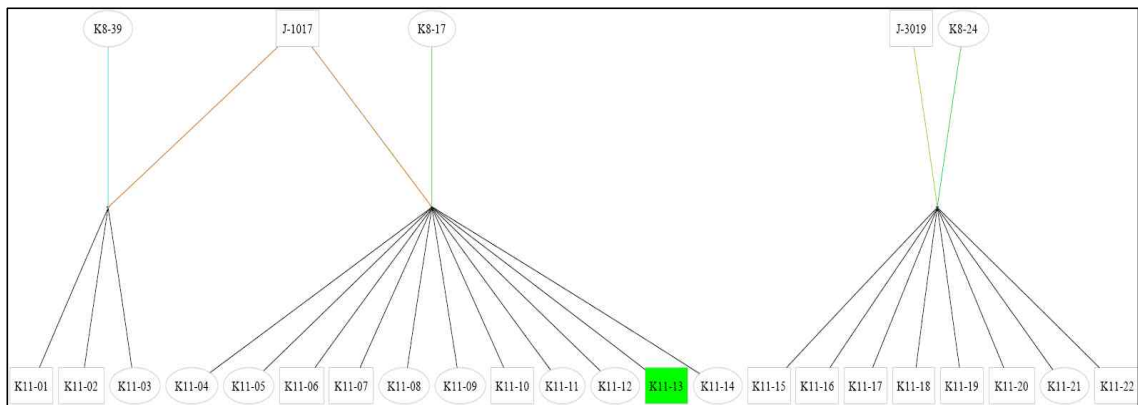


그림 48. 제주재래흑돼지 1세대 자손의 가계도

표 37. 한국재래돼지 유두수에 대한 표현형 조사

세대	성별	두수	유 두 수		
			좌	우	합계
기초세대	암	8	6.0±0.5	6.1±0.4	12.1±0.6
	수	12	6.1±0.5	6.2±0.4	12.3±0.8
	전체	20	6.1±0.5	6.2±0.4	12.2±0.7
1세대	암	10	6.7±0.8	7.3±0.9	14.0±1.3
	수	15	6.5±0.5	7.0±0.9	13.5±1.2
	전체	25	6.6±0.6	7.1±0.8	13.7±1.2
2세대	암	21	6.3±0.5	6.4±0.9	12.7±1.0
	수	27	6.3±0.7	6.3±0.6	12.6±1.1
	전체	48	6.3±0.6	6.4±0.6	12.7±1.0
3세대	암	64	6.3±0.6	6.3±0.6	12.5±1.0
	수	40	6.3±0.6	6.4±0.7	12.7±1.1
	전체	104	6.3±0.6	6.3±0.7	12.6±1.0

2. 한국재래돼지의 유두수에 대한 표현형과 유전양상

가. 번식형질중 유두수의 경우 참조축군에서 7번 염색체에서 QTL이 확인되었으며 1세부과제에서 유의성 높은 SNP 또는 유전자가 동정되었다. 이 DNA 마커는 제주재래돼지 집단에 적용하여 marker의 효율성을 평가중에 있으며 현 단계에서는 외형적인 표현형적 자료를 근거로 암컷보다는 수컷의 유두수를 기준으로 강선발하여 교배 및 분만이 진행중에 있다.

나. 한국재래돼지의 성장 특성을 고려하면 늦은 성장으로 인하여 보통 일반돼지는 1세대가 1년 정도 소요되지만 재래돼지의 경우 1.5년 정도 소요됨에 따라 생산두수가 적은 특성이 있으며, 3차년도 과제가 종결되었지만 지속적으로 생산되고 있으며, 현재까지 3세대까지 총 197두가 생산되었다.

- (1) 총 유두수를 살펴보면 기초축군의 전체 평균은 12.2개로 조사되었으며 이후 세대별로 살펴보면 1세대 자손에서 13.7(암, 14.0; 수, 13.5)개, 2세대 자손에서 12.7(암, 12.7; 수, 12.6)개, 3세대 자손은 12.6(암, 12.5; 수, 12.7)개로 표현형적으로 조사되었다(표 1).
- (2) 전체적으로 유두수는 좌우 양측 합계는 적게는 11개에서 많게는 16개로 다양하게 분포하고 있음을 보여주고 있다.
- (3) 개체간 빈도차가 한국재래돼지 집단에서 뚜렷이 확인되었고, 표현형에 의한 선발과 더불어 유전자 분석에 의한 개량의 소지가 많음을 보여주고 있다.
- (4) 그러나 유두수의 유전력이 낮아 개량을 위해서는 유전자 수준에서 검정하는 방법과 능력 검정을 통한 표현형적 자료를 겸하여 번식축 선발에 활용하는 것이 개량의 효과가 좋을 것으로 사료된다.
- (5) 외형적인 유두수의 유전양상 또한 다소 복잡한 양상을 보여주고 있으며, 대다수의 개체들의 표현형적 자료는 총 유두수가 평균 13개 내외이지만 1세대 자손에서 일차적으로 선발된 수컷의 경우 14개 이상의 유두수를 가진 개체가 번식용으로 선발되어 교배가 진행되어 후대를 생산하고 있어 기초축군에 비해 1세대에서 3세대 자손에서 유두수가 증가한 것으로 추정된다.
- (6) 외형적으로 식별되는 유두수 개수 즉 유두수가 14개 이상인 암수를 선발하여 교배를 하더라도 후대에서 생산되는 자손의 유두수 개수가 소폭 증가하는 추세에 있으나 집단내에서 아주 다양하게 측정이 되고 있어 유두수 증가의 목적으로 단순한 개수를 위주로의 표현형적 선발은 어느 정도 한계성이 있어 보이며, 이를 보완할 수 있는 방법으로는 유두수 개수를 조절하는 유전자 또는 표지인자 개발을 통한 우량 종축 선발이 더 중요하게 다루어져야 할 것으로 판단된다.

3. 한국재래돼지의 주령별 체중변화

표 38. 한국재래돼지 집단에서 세대별 성장단계에 따른 체중변화

세대	성별	두수	성장단계별 체중			
			생시	이유시 (4주령)	10주령	20주령
기초 세대	암	8	1.4±0.3	8.2±3.1	21.0±2.6	54.0±6.0
	수	12	1.6±0.5	5.2±0.7	22.3±3.5	58.8±7.9
	전체	20	1.5±0.4	6.4±2.5	22.0±3.3	57.8±7.2
1세 대	암	10	1.2±0.2	5.2±0.9	20.9±3.5	47.4±4.0
	수	15	1.5±0.3	5.9±1.3	25.9±3.4	54.2±11.8
	전체	25	1.4±0.3	5.6±1.2	24.0±4.2	51.6±10.0
2세 대	암	21	1.3±0.2	5.4±1.4	19.4±3.1	47.2±4.6
	수	27	1.3±0.3	5.8±1.6	22.4±5.4	52.1±9.0
	전체	48	1.3±0.3	5.6±1.5	21.1±4.8	49.9±7.8
3세 대	암	64	1.5±0.4	5.4±1.8	19.7±4.2	48.5±7.7
	수	40	1.6±0.4	5.7±2.2	19.8±6.6	49.8±7.3
	전체	104	1.6±0.4	5.5±2.0	19.7±5.2	48.9±7.6

가. 한국재래돼지 세대별 성장 주령별 체중을 분석한 결과 암수 성에 따라 생시체중, 이유시 체중, 10주령 및 20주령 체중에서 유의적인 차이를 나타내지 않았고 각 주령별로 살펴보다도 암수 성간에 차이는 거의 없었고 세대간 차이도 거의 유사한 것으로 나타났다(표 2).

(1) 10주령과 20주령 체중의 경우, 수컷이 암컷에 비해 성장이 조금 빠른 것으로 나타났으나 유의차는 확인되지 않았다. 그러나 이유시 평균체중의 경우 수컷이 암컷에 비해 체중이 더 무거운 효과가 있었고, 성장 속도에 있어 10주령 체중이 높을수록 20주령 체중이 높아지는 경향이 확인되었는데, 이 결과는 제주재래돼지와 랜드레이스 교배 집단에서도 거의 유사한 결과를 보여주는 것이라 할 수 있다.

(2) 일반적으로 이유시 체중을 근거로 종돈을 선발하는 경향이 많아 이유체중을 고려하여 1차 선발하는 경향이 많은데 그 효과가 없다고 할 수 있다. 이런 결과는 이유시 체중의 경우 모체의 포유능력과 관계가 있어 유즙분비량이 좋은 어미에서 생산된 자돈은 그 만큼 이유시 체중의 증대를 기대할 수 있는 반면, 이유 이후부터는 전적으로 사료에 의존하기 때문에 성장 속도가 개체의 능력에 의해 평가된다고 할 수 있다.

(3) 초기성장에서 수컷이 암컷보다 이유 후 성장률이 더 좋은 것으로 사료되나, 1세대에서 2세대에 대한 주령별 체중에서 암컷보다는 수컷이 다소 성장이 빠른 면을 보이나 3세대 자손에서는 암수간 차이가 거의 확인되지 않았다.

나. 본 연구에서 제주재래돼지의 이유 시기는 4주령(28일령)에 실시하였는데, 이는 4주령 이유

시기가 일반 도입종의 돼지 품종들은 3주령에 이유하더라도 최소 5.0kg 이상임을 감안하면 제주재래돼지의 초기성장 자체가 상당히 지연된다고 할 수 있기 때문에 포유기간을 길게 하여 초기성장 보조하는 개념이다.

(1) 세대별로 살펴보면 기초 축군에서 생산된 자손의 생시체중은 1.5 kg이었으나 1세대와 2세대 자손에서 각각 1.4 kg과 1.1 kg이었고 3세대 자손에서는 1.6 kg이었다. 생시체중이 기초 축군에서 다소 높은 이유는 산차의 효과에 의한 결과로 추정되어 진다. 일반적으로 생시체중은 품종, 유전적 요인 및 산차 등에 의해 영향을 미치는 것으로 알려져 있다. 4주령 이유체중의 경우 품종에 따른 차이로 대변할 수 있지만 동일한 품종 내에서는 가장 큰 요인으로는 모체의 포유능력에 따라 이유자돈의 체중에 지대한 영향을 미친다고 할 수 있다.

다. 모든의 포유능력을 직접 조사하는 것은 힘든 상황이므로 이를 간접적으로 측정하는 방법으로 이유시까지 자돈의 체중변화를 이용하여 모든의 포유능력을 추정하는 경우도 있다.

(1) 생시체중이 월등히 일반돼지에 비해 적더라도 포유능력이 우수할 경우 이유시 체중이 높을 수 있으나, 1주간 더 포유를 하더라도 4주령 이유시 체중이 일반돼지보다 낮으므로 그만큼 제주재래돼지의 포유능력이 떨어짐을 예측할 수 있다.

(2) 이는 포유능력이 떨어지므로 태어난 자손의 왜소 또는 포유 중 폐사율을 높여 전체적으로 이유 시까지 육성율이 낮아지는 결과를 낳는다.

라. 제주재래돼지의 포유능력을 높이는 것 또한 향후 중요한 과제가 될 수 있을 것으로 사료된다.

(1) 70일령과 140일령 체중에 있어 평균적으로 각각 22 kg과 52 kg으로 조사되었고, 개량종과 비교할 경우 전체 성장속도가 개량종의 50-60% 수준임을 추정할 수 있었다.

(2) 특히 20주령 체중만 비교할 경우 체중이 적은개체는 40 kg 정도에서 체중이 많은 개체는 78 kg으로 개체간 빈도차가 심함을 보여주고 있어 지금까지 제주재래돼지에 대한 육종 사업이 전무한 상황에서 단순히 종 보존 수준에서 사육된 결과라 할 수 있다.

마. 성장에 가장 크게 영향을 미치는 요인은 근친으로 추정이 되는데 그 이유는 초기 집단이 크기가 작기 때문에 소수의 수컷을 활용하여 생산이 된 것이 가장 큰 요인으로 작용되었을 것으로 사료된다.

(1) 그 결과 생산된 자손은 균일하지 못하고 성장속도에서 개체 간 편차가 심하게 나타나는 것으로 추정할 수 있다. 따라서 능력검정 사업을 통하여 최상위 10%내에서 체형과 체중 등을 근거로 몇 세대 강선발에 의해 번식용 우량 종축을 선발한다면 재래돼지에 대한 개량의 효과도 다소 기대할 수 있을 것으로 예상된다.

바. 따라서 한국재래돼지의 단점인 적은 산자수, 성장속도, 유두수, 체장 등 불량형질을 개선하기 위해서는 체형 개선이 필요한데, 이를 보완하기 위하여 외형적 표현형인 장방향 체형을 강선발하고 있으며, 그와 더불어 분자유전학적 분석기법을 이용하여 유전자 수준에서 정보력이 높은 유전자를 개발하는 과정이 진행 중에 있다.

4. 한국재래돼지의 유두수 관련 QTL 영역 및 후보 유전자 탐색

가. 번식형질 중 유두수의 경우 QTL 영역 내에서 후보유전자의 SNP를 지속적으로 개발하여 유전자 좌위를 압축하여 표현형 연관 유의성 높은 유전자 또는 DNA marker가 개발될 수 있을 것으로 예측된다.

(1) 현 단계에서 유두수에 직접적으로 영향을 미치는 유전자 개발이 완료가 되지 않아 원인

SNP를 이용한 개체선발은 다소 무리가 있어 보이거나 QTL 영역을 대상으로 IBD mapping 방법 등을 활용하면 원인유전자 주변의 유전자를 선발하여 haplotype 분석에 의한 개체 선발 방법 등은 활용이 가능하다고 할 수 있다.

- (2) 또 다른 방법으로는 표현형을 기준으로 우량 가계를 기준으로 그 가계에서 생산된 개체들을 집중으로 선발하는 방법 등이 있으나 그 효과는 크지 않을 것으로 예상된다. 그 이유는 유두수에 대한 가계내에서 복잡한 유전양상을 보이고, 유전력이 낮으며 특히 표현형적으로 질적형질로 단정할 수 있지만 생산되는 자손을 근거로 표현형을 추정하면 양적형질 개념이 강하다고 할 수 있다.

나. 유전자 수준에서 검토를 하면 제주재래돼지 순종 생산과정에서 유두수는 유전양상이 단순 멘델의 법칙을 벗어나 유전자 침투도가 낮은 것도 선발의 효과를 떨어뜨리는 요인으로 작용할 수 있다. 그 이유는 유두수가 적은 개체 간 교배를 하더라도 일부 개체에서 많은 유두수를 가진 개체가 생산되는 반면 유두수가 많은 개체 간 교배에서도 적은 개체가 생산되기 때문이다. 따라서 유두수 증대에 대한 효과는 해당 가계에 대한 유전양상을 가계도 분석에 의한 1차 검정 후 DNA marker를 이용한 보조수단으로 활용하는 것이 더 효율적인 방법으로 판단되어진다.

5. 한국재래돼지의 갈비뼈 수와 도체장, 도체중의 상관관계

가. 돼지의 생산성을 향상시킬 수 있는 방안 중 또 다른 기준형질로 발생학적으로 태생 시 고정되는 척추뼈의 수의 변화에 따른 도체장 및 도체중의 추이에 대한 상관성을 분석하였다. 척추의 수가 거의 고정되어 있는 대부분의 포유류 종들과는 달리 돼지는 척추뼈, 특히 흉추 (갈비 뼈)의 수가 개체 사이에서 서로 다른 양상을 나타내는 것으로 알려져 있다. 이에 흉곽을 감싸고 있는 갈비뼈 수의 증가가 개체 수준에서 도체장의 길이 및 도체중과 연관되어있을 것으로 가정하였고, 축군에 대한 도축과정에서 개체별 갈비뼈의 수를 계측하였다.

- (1) 조사결과 집단전체 1,041 두에 대한 갈비뼈의 수는 적게는 14개에서 많게는 17개를 나타내었다. 빈도 상으로는 갈비뼈 수가 15개인 동물들이 전체 57.1%로 가장 많았고, 16개 (30.5%), 14개인 동물 (11.0), 마지막으로 갈비뼈 수가 17개인 동물들이 14두(1.3%)로 조사되었다(표 39).

표 39. 돼지 갈비뼈 수의 집단 내 분포 빈도

갈비뼈의 수	개체 수	빈도
14	115	0.110
15	594	0.571
16	318	0.305
17	14	0.013
계	1041	1.000

- (2) 갈비뼈 수의 변화에 따른 도체중의 차이는 표 40에 나타내었다. 갈비뼈의 수는 최소 14, 최대 17까지 나타나지만 17의 경우 빈도 수 낮아 큰 의미가 없다고 하겠다. 전체적으로 갈비

뼈의 수가 증가함에 따라 도체장의 길이가 길어지는 양상을 나타내었다.

- (3) 갈비뼈 수에 따른 도체장의 길이는 유의적인 차이를 나타내었다. 갈비뼈 수가 많아질수록 대략 1 cm 정도 도체장의 길이가 증가함을 확인하였다. 갈비뼈 수의 증가에 따른 도체장의 증가가 돼지 도체에서 갈비뼈가 형성된 지점에서 증가하는 것이라면, 이는 삼겹살이나 갈비 등 돼지 도체에서도 가장 선호되는 경제성이 높은 부위의 양이 증가함을 의미하는 것이나, 갈비뼈 수의 증가와 삼겹살이나 갈비의 양의 증가와의 상관에 대해서는 추가적인 분석이 필요하다고 하겠다. 그림 1은 갈비뼈 수와 도체장과의 상관관계를 나타내는 그래프이다.

표 40. 갈비뼈 수에 따른 도체장의 길이

갈비뼈의 수	도체중 (cm)
14	101.68±6.61 ^a
15	102.36±5.96 ^{ab}
16	103.55±5.39 ^b
17	103.79±3.85 ^{ab}

- (4) 도체장과 도체중의 상관관계를 확인하기 위하여 도체장의 길이에 따른 도체중의 평균을 계산하였다(표 41). 도체장이 90 cm이하인 돼지의 평균체중은 55.33±11.14 kg을 나타내었으며, 빈도가 가장 높았던 101-105 cm의 돼지들은 평균 도체중이 80.75±9.64 kg을 나타내었고, 111 cm 이상인 돼지들은 평균 도체중이 95 kg 이상으로 조사되었다. 그림 2는 도체장의 길이의 증가에 따른 도체중의 변화를 나타낸 것으로, 그래프 상에서 도체장의 길이가 길어질수록 도체중이 점차적으로 증가한다는 것을 확인할 수 있었다.

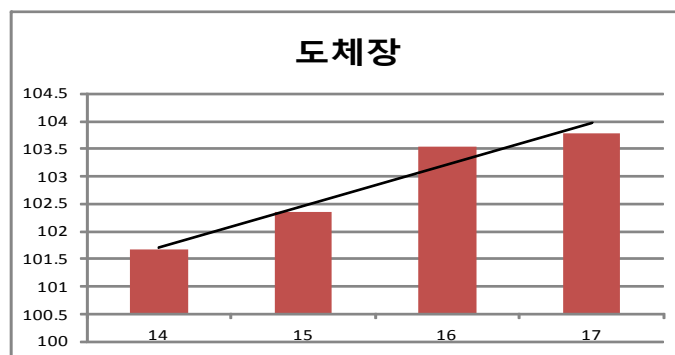


그림 49. 갈비뼈 수의 증가에 따른 도체장의 길이

표 41. 도체장 길이와 도체중간의 상관관계

도체장 (cm)	개체 수	빈도	도체중 (kg)
<90	12	0.013	55.33±11.14
91-95	90	0.097	67.21±11.58
96-100	263	28.3	74.71±8.67
101-105	288	0.310	80.75±9.64
106-110	198	0.213	86.19±10.21
111-115	64	0.069	95.55±10.52
116-120	13	0.014	95.62±9.34
120-125	2	0.002	114.0±8.49
계	930	1.000	

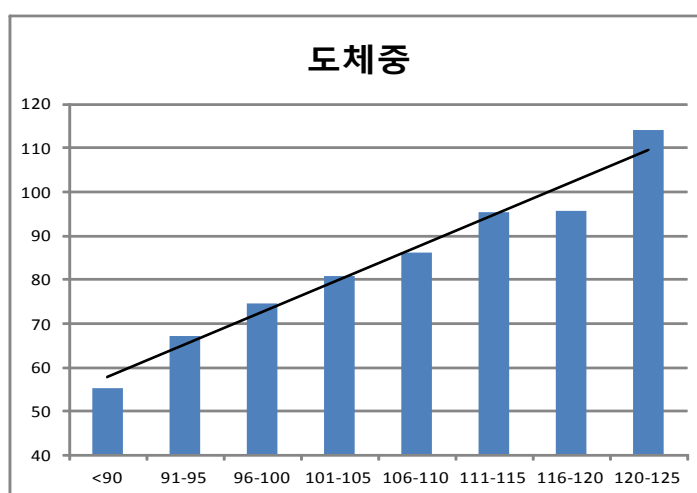


그림 50. 도체장의 길이에 따른 도체중의 변화 추이

나. 이상의 결과에서 도체중, 도체중, 갈비뼈 수의 상호관계를 살펴보면, 갈비뼈 수의 증가는 도체장에 있어 유의적인 상관을 초래하였고, 도체장의 길이는 도체중의 차이를 이끌어낼 수 있음을 확인하였다. 따라서 직접적인 관계가 있는지에 대한 추가적인 조사가 이루어져야 하겠지만, 현재까지의 결과만으로도 갈비뼈 수의 증가가 도체중과도 어느 정도의 상관관계가 있을 것으로 추정할 수 있다. 결과적으로 갈비뼈 수 증가는 일반돼지뿐만 아니라 제주재래돼지 집단에서도 육량과 관련한 주요 개량항목 중 하나로 이용 가능할 것으로 판단된다.

6. 한국재래돼지의 갈비뼈 수 관련 QTL 영역 내 후보유전자형의 효과와 이를 이용한 축군 조성

가. 제 1세부과제에서 발굴한 갈비뼈 수의 관련 후보유전자 중 LTBP2 유전자내에서 확인된 3개의 SNP와 참조축군 도체성적 중 갈비뼈 수의 분포를 확인하였다(표 42). 발견된 3 종의 SNP에서 모두 갈비뼈의 수가 14-17까지의 분포를 나타냄을 확인할 수 있었다.

표 42. 참조축군에서 LTBP2 유전자 내 SNP 유전자형과 갈비뼈 수의 분포

SNP	유전자형	분석두수	평균 및 편차	최소값	최대값
exon 13 199A>G	A/A	18	15.72±0.67	15	17
	A/G	140	15.45±0.63	14	17
	G/G	253	15.04±0.57	14	16
intron 13 18G>A	A/A	118	15.23±0.66	14	17
	A/G	168	15.26±0.66	14	17
	G/G	125	15.13±0.57	14	16
exon 32	A/A	87	14.74±0.52	14	16
	A/C	160	15.25±0.55	15	16
	C/C	78	15.60±0.69	14	17

- (1) 한국재래돼지와 랜드레이스 교배 참조축군에서 갈비뼈수 관련 QTL 영역을 확인한 결과 VRTN에서 LTBP2 유전자 영역으로 전체적으로 비교적 짧은 거리에 있었고, 그 QTL 영역을 좁히기 위하여 15가계 총 1200여두에 대한 IBD mapping을 하였으나, 그중 1개 가계에서만 가계가 확인이 되었는데 이는 당초 참조 축군이 상호교차교배 방법을 이용하였고, 특히 기초축군에서 랜드레이스와 재래돼지를 1:1로 교배하다 보니 전체적으로 많은 가계가 필요하게 되었고 특정형질에 대해선 효과적으로 유전자를 압축하는데는 역교배 방법이 효과적이나 상호교차교배 방법에 의해 생산된 자손을 이용하다 보니 유전자 영역을 좁히는데는 한계가 있었다.
- (2) 갈비뼈수 관련 QTL 영역에서 후보 유전자를 선정하여 RT-PCR에 의한 조기 변이 지역 탐색이 가능하나 갈비뼈의 경우 배발생 초기 단계에 발현되는 특성이 있어 생산된 이후에는 mRNA 서열을 밝히는데는 한계가 있었다. 따라서 QTL 영역을 대상으로 표현형과 가장 연관성이 높을 것으로 사료되는 유전자가 4종이 선정되어 DNA 수준에서 변이체를 검출하였다.
- (3) 특히, LTBP2 exon 32 SNP는 발견된 SNP들 중에서 염기변이에 의한 단백질의 변이가 야기되는 missense mutation으로 확인되었고, AA유전자형에 비해 C염기가 증가할수록 갈비수가 증가하는 양상을 나타내었다. 표현형상에서 CC 유전자형을 갖는 돼지의 갈비뼈 수는 AA 유전자형인 돼지들에 비해 평균 0.86개의 갈비뼈 수가 증가함을 알 수 있었다(표 6).
- (4) LTBP2 유전자에서 발굴된 전체적인 genome 구조와 염기서열상의 변이 영역을 대상으로 표현형과 유전자형과 연관성 분석을 통해 유의성이 높은 SNP를 개발하였다.
 - (가) 발견된 LTBP2의 exon 32 SNP는 제주재래돼지 집단에서도 확인되었다. 특이할만한 결과는 제주재래돼지의 LTBP2 유전자형에서 CC 동형접합체가 발견되지 않는다는 점이고, 이러한 점은 제주재래돼지의 도체에서 갈비뼈 수를 조사했을 때 확인했던 자료와 비교했을 때, 제주재래돼지 집단의 갈비뼈 수가 대부분 14개 또는 15개이며 16개인 경우가

발견되지 않았다는 측면에서 유전자형의 분포와 갈비뼈 수의 분포가 일치하는 면을 확인하였다.

- (나) 한국재래돼지 1세대 자손 집단에 대한 LTBP2 exon 32 유전자형을 분석한 결과 갈비수 증대에 효과가 높은 것으로 추정되는 C/C 유전자형 분포가 기초축군에 비해 1세대 자손에서 보다 빈도가 높게 나타났으며, 특히 기초축군에서 확인되지 않았던 C/C 유전자형까지 발견되어 장방향 체장을 가진 개체 선발이 가능할 것으로 보여주고 있다(표 43).
- (5) 제주재래돼지 집단의 근친도를 낮추기 위해 도입된 재래돼지 집단에 대한 유전자형 분석 결과를 토대로 인위적인 계획교배를 수행하였다. 즉, LTBP2 exon 32의 SNP에 대한 유전자형의 조합을 통해 차후 세대에서 갈비뼈의 수가 많아지는 CC 유전자형의 출현빈도가 높아질 수 있는 교배를 수행하였다.

표 43. 제주재래돼지 집단에서의 LTBP2 exon 32내에 존재하는 SNP의 유전자형 분포

세대	유전자형	분석두수
기초 축군	A/A	14
	A/C	6
	C/C	-
1세대 자손	A/A	12
	A/C	8
	C/C	4

(가) 현재까지는 일부 개체들만 C/C 동형접합체로 확인되나 대다수가 A/A 유전자형으로 향후 1-2년 내에 지속적인 교배를 통해 C/C 유전자형으로 고정할 수 있어, 생식주기가 짧은 돼지의 특성 상 개량의 효과 역시 멀지않은 시점에 가시적인 성과를 거둘 수 있을 것으로 사료된다. 이후 체형이 개선된 제주재래돼지 집단이 조성이 되면 산업화 축군 생산 및 보급으로 고급육 생산에 크게 기여할 수 있을 것으로 예측된다.

(나)



<랜드레이스, 갈비뼈수 16개>



<제주재래돼지, 갈비뼈수 14개>

표 44. 한국재래돼지 갈비뼈 수관련 유전자 접목 효과

제주재래돼지	개체수	갈비뼈 수	비 고
2세대 자손	12	14.5±0.5	갈비뼈수 관련 LTBP2 유전자형 축군내 접목
3세대 자손	20	14.8±0.5	

나. 한국재래돼지 세대별 실험집단의 갈비뼈수

- (1) 기초축군과 1세대 자손에서 갈비뼈 수는 거의 변이가 없을 정도로 14개로 고정이 되어 있었으나 2세대 자손이후 갈비뼈 개수 관련 LTBP2 유전자의 exon32의 C/A 변이를 이용하여 번식축을 선발 및 교배하여 후대축을 생산하였다.
 - (가) 그 결과 2세대 자손에서 평균적으로 14.5개였고 이중 절반은 갈비뼈수가 14개였고, 나머지 절반은 15개였다.
 - (나) 2세대 자손을 번식돈으로 선발하여 3세대 자손을 생산하였을 때 2세대 자손보다 갈비뼈 수 증가 효과가 있었고 평균적으로 14.8개였다.
- (2) 3세대 자손의 갈비뼈수를 세부적으로 살펴보면 모두 20두에 대한 결과로 갈비뼈 수가 14개가 6두, 15개가 13두, 16개가 1두가 확인되어 전반적으로 체형이 장체장형이다. 1세대 자손 까지 15개의 갈비수를 가진 개체는 어렵게 확인이 되었지만 세대가 진행될수록 갈비수가 증가하였고 3세대 자손에서는 16개의 갈비까지 확인되어 유전자 접목에 따른 효과가 크다고 할 수 있다. 추후 한국재래돼지 집단을 대상으로 지속적으로 갈비뼈수에 대한 조사가 도체단계에서 수행되고 있어 현 상황에서는 다소 적은 두수여서 재래돼지 전체를 설명하기에는 다소 무리가 있으나 연구진의 예상과 부합되는 결과를 도출할 수 있어 향후 왜소한 체구와 낮은 생산성이 가장 큰 문제점으로 지적받고 있는 제주재래돼지 품종 자체의 체형을 개선할 수 있는 중요한 계기가 될 것이다. 무엇보다도 중요한 것은 돼지의 체형 개선을 위해 기존의 관행적으로 행해져왔던 타 품종과의 교배육성과정이 없이, 제주재래돼지 집단 내에서 과학적인 근거를 바탕으로 한 체형 개선이 가능해진다는 점이다.

제 4장 목표달성도 및 관련분야에의 기여도

1. 연구개발 목표의 달성도

구분	연구개발의 목표	달성도(%)
1차년도 (2010)	○ SNP 대량 분석체계 확립	100
	○ 국산 재래돼지 종축의 맞춤형 통계분석모형 개발	100
	○ 재래돼지 기초축군 선발 및 1세대 자축 형질측정	100
2차년도 (2011)	○ SNP 대량 분석, QTL 확보, 후보 유전자의 변이 확인 및 검증	100
	○ MBV추정 알고리즘 개발 및 전산프로그램 개발	100
	○ 형질관련 marker 적용을 통한 재래돼지 교배 및 2세대 자축 형질측정	100
3차년도 (2012)	○ 형질관련 표지 마커의 개발 및 이를 이용한 선발 실시	100
	○ SNP Marker별 효과추정 및 개체의 유전능력추정	100
	○ 추가 형질 marker 적용을 통한 3세대 자축 생산 및 유전자형 고정축 유지	100
	○ 한국재래돼지를 이용한 산업화 기반 조성	100
최종	○ 한국재래돼지 종돈의 계통확립 기반 조성 ○ 한국재래돼지 종돈의 불량형질(성장 및 포육능력) 개선 및 우량 형질(육질) 보존을 위한 유전자 기반 선발 방법 개발 ○ DNA 표지인자 기반 선발을 통한 한국재래돼지 생산성 극대화 및 보급용 종돈 시범 생산	100

2. 정량적 연구개발 성과

가. 논문성과

게재 연도	논문명	저자			학술지명	Vol.(No.)	국내외 구분	SCI구분
		주저자	교신저자	공동저자				
2011	QTL analysis of white blood cell, platelet and red blood cell-related traits in an F2 intercross between Landrace and Korean native pigs	I. C. Cho, H. B. Park	J. T. Jeon	C. K. Yoo, G. J. Lee, H. T. Lim, J. B. Lee, E. J. Jung, M. S. Ko, J. H. Lee	Animal Genetics	In press	국외	SCI

2011	Maternal Origins of the Jeju Native Pig Inferred from PCR-RFLP Haplotypes and Molecular Phylogeny for Mitochondrial DNA CYTB Gene Sequences	Sang-Hyun Han	In-Cheol Cho	Moon-Suck Ko, Ha-Yeon Jeong, Sung-Soo Lee, Hong-Shik Oh	생명과학회지	21(3): 341~348	국내	비SCI
2011	Characterization of an alternative splicing by a NAGNAG splice acceptor site in the porcine KIT gene	김재환	전진태	종타오, 조인철, 임현태, 유채경, 김병우, 이준현	Genes & Genomics	33: 179-186	국외	SCI
2012	Evaluation of Three Candidate Genes Affecting Fatty Acid Composition in Pigs	Dyah Maharani	이준현	정연국, 조철훈, 정우영, 남기창, 서강석, 이승환	한국축산식품학회	Vol.32, No. 1, pp. 6-12	국내	SCI
2012	Porcine SPP1 gene polymorphism association with phenotypic traits in the Landrace x Jeju (Korea) black pig F2 population	한상현	조인철	신광윤, 이성수, 고문석, 오홍식	Molecular Biology Reports	DOI 10.1007/s11033-012-1606-z (On-line 게재)	국외	SCI
2012	제주재래돼지와 랜드레이스간 교잡종 집단에서 성장형질과 전장의 유전체 정보간 연관성 분석	조충일	이득환	이준호	Journal of Agriculture & Life Science	Vol. 47(2) :75-84	국내	비SCI
2013	The TBC1D21 is a possible candidate gene for teat numbers in pigs	S. Jin	J. H. Lee	J. B. Lee, K. Kang, C. K. Yoo, B. M. Kim, H. B. Park, H. T. Lim, I. C. Cho	Asian-Australian Journal of Animal Science	26(10):1374-1378	국외	SCI
2013	Identification of SNPs Affecting Porcine Carcass Weight with the 60K SNP Chip	Kang, K.	J. H. Lee	D. W. Seo, J. B. Lee, E. J. Jung, H. B. Park, I. C. Cho, H. T. Lim	Korean Journal of Animal Science and Technology	55(4):231-235	국내	비SCI

나. 특허성과

번호	출원등록명	출원등록자명	구분	산업재산권 종류	출원등록일
1	교잡종 돼지의 지방산 조성 예측용 SNP 마커	임현대;박희복;이재봉;유채경;정은지;이준현;디아마하라니;정우영	특허출원	특허	2012-04-20
2	돼지 SNP 유전체 정보 활용 유전체 육종가 추정 프로그램	한경대학교 산학협력단;조충일;이득환	특허등록	특허	2012-03-19
3	돼지의 갈비뼈 수 판단용 SNP 마커 및 이의 용도	조인철;임현대;한상현;이준현;고문석;이성수	특허출원	특허	2012-12-31
4	새로운 돼지 유두수 예측용 바이오마커 및 그의 선별방법	임현대, 이재봉, 정은지, 이준현, 진실	특허출원	특허	2013-07-19
5	돼지에서 베이지안 방법을 이용한 분자유종가 (MBV) 추정 프로그램	한경대학교 산학협력단	특허등록	특허	2013-05-24
6	돼지의 대용량 단일염기다형을 이용한 추정 분자유종가 (MBV)의 정확도 계산 프로그램	한경대학교 산학협력단	특허등록	특허	2013-05-24
7	돼지의 갈비뼈 수 판단용 SNP 마커 및 이의 용도	조인철;임현대;한상현;이준현;고문석;이성수	특허등록	특허	2013-08-26

다. 학술발표

번호	발표자	발표제목	발표일시	장소, 국명 출원등록 일
1	S. Jin	The TBC1D21 is the possible candidate gene for the teat number in pigs	2012-06-29	대전, 충남대학교
2	Shil Jin	QTL analysis for teat number in an F2 intercross between Landrace and Korean native pigs	2012-07-16	호주, 케언즈
3	Joonho Lee	Effect of Gene Frequency Deviation to Heterozygosity in Jeju X Landrace Hybrid Populations Using Whole Genomic Data	2013-06-27	제주, 한화리조트

라. 연구성과 활용

(단위 : 건수)

구분		기술실시 (이전)	상품화	정책자료	교육지도	언론홍보	기타
활용건수	목표	1			2	3	
	달성		1 ⁽⁴⁾		2 ⁽¹⁾	4 ⁽²⁾	1 ⁽³⁾

- (1) ⁽¹⁾제주재래흑돼지 산업화 전략 세미나(2011년 10월 20일, 국립축산과학원 난지축산시험장)
- (2) ⁽¹⁾동물 유전체 정보를 활용한 유전능력 평가 기술(2013년 2월 18~22, 한경대학교)
- (3) ⁽²⁾제주도민일보, 제주재래흑돼지 맛 전문가도 인정(2011년 11월 14일)
- (4) ⁽²⁾KBS 뉴스9 TV, 유전자분석 가짜 흑돼지 손쉽게 가려낸다(2012년 3월 26일)
- (5) ⁽²⁾JIBS TV, 가짜 흑돼지 유전자로 확인 (2012년 3월 26일)
- (6) ⁽²⁾KCTV 제주 TV, 흑돼지 유전자로 판별 (2012년 3월 27일)
- (7) ⁽³⁾국립수산물과학원에 유전체자료를 이용한 육종기술 홍보(한경대)
- (8) ⁽⁴⁾제주재래돼지를 활용한 교육질형 흑돼지 분양(2013년 6월)

마. 연구인력 활용/ 양성 성과

- (1) 박사학위 배출 : 5명
- (2) 석사학위 배출 : 2명

3. 연구개발 결과의 관련분야에의 기여도

가. 기술적인 측면

- (1) 한국재래돼지의 개체별 유전능력평가 정확도(신뢰도) 향상
- (2) 세대기간 단축으로 인한 유전적 개량량 상승효과 및 DNA 수준의 조기 선발체계 구축
- (3) 성장 및 포유능력 관련 형질관련 마커의 개발
- (4) 형질이 균일하고 향상된 한국재래돼지 종돈 생산체계 확립
- (5) 한국재래돼지의 국가적 유전자원 활용성 증대 및 국제적인 인지도 향상
- (6) 국제수준의 분석 기술 및 유전자 표지인자에 대한 기술 선점

나. 산업적인 측면

- (1) 제주도에서만 생산되는 흑돈 돈육은 6,000여톤 생산되는 것으로 추정되며, 금전적으로 환산할 경우 년 280억 규모의 시장이 형성되고 있음. 재래돼지는 지금까지 육종사업은 전무한 상태라 개체간 형질차이가 심하고, 전통적인 통계육종과 DNA 마커 도움선발을 병행할 경우 조기에 형질을 고정할 수 있음.
- (2) 기존 한국재래돼지에 비해 성장 및 포유능력이 향상된 종돈 생산 및 보급 가능
- (3) 형질이 균일하고 향상된 종돈생산을 통한 한국재래돼지 이용 흑돼지 산업 활성화 유도
- (4) 생산성 향상을 통한 재래돼지 및 이를 이용한 합성돈 사육 활성화 및 농가이익 증대
- (5) 국산재래돼지의 유전자원확보 및 개량효율로 인한 수입종돈에 대한 의존도 감소
- (6) Genomic selection을 통한 우수한 형질의 국산재래돼지 종돈 확보 및 생산라인 구축

- (7) 국산재래돼지와 외산품종간의 교배를 통한 경제성 높은 신품종 개발
- (8) 국산재래돼지 종돈을 활용한 명품 흑돼지고기 생산 및 특화사업 기반 구축
- (9) 명품 흑돼지고기를 통한 돈육수출 안정화 기반 확립
- (10) 국제 특허 출원 및 산업체 기술이전을 통한 고부가가치 제품 개발

제 5장 연구개발 성과 및 성과활용 계획

1. 개발기술의 실용화 및 산업화 계획

- (1) 형질관련 마커의 적용을 통한 한국재래돼지 종돈의 선발 및 유지
- (2) 한국재래돼지에서 유전체 육종가를 활용한 교체축 선발 및 도태계획 수립가능
- (3) 제주재래돼지의 성장 및 포육능력에 관한 유전적 배경 구명 및 이를 이용한 특허 출원
- (4) 한국재래돼지 및 합성돈 사육농가에 종돈 보급 및 생산체계 기술이전
- (5) 한국재래돼지 분양 사육 농가의 모니터링 및 피드백을 통한 우수 종돈의 지속적 공급

2. 특허 및 논문 등 지식재산권 확보 계획

- (1) 한국재래돼지 선발 프로그램의 지적 재산권 확보
- (2) 한국재래돼지에서 유전체 육종가를 활용한 교체축 선발과 관련된 논문 게재
- (3) 성장 및 포육능력에 관한 유전적 배경 구명 및 국제 저명 학술지에 논문 게재
- (4) 한국재래돼지 성장 및 포육 능력 개량을 위한 논문 및 지적 재산권 확보
- (5) 한국재래돼지 품종의 등록 및 특허권 획득을 통한 지적 재산권 확보
- (6) 궁극적으로 한국재래돼지의 국제화에 대한 교두보 확보

제 6장 연구개발과정에서 수집한 해외과학기술정보

(Shogo course note 요약 첨부)

1

Genetic analysis using application (BGF90) programs in animal breeding

Shogo Tsuruta
Department of Animal and Dairy Science
College of Agricultural and Environmental Sciences
The University of Georgia
Athens, GA 30602 USA
Email: shogo@uga.edu
February 2013

Contents

0. Fundamental Statistics
1. Analysis of Variance
2. BLUP & BLUE
3. ML and REML
4. Bayesian (MCMC)
5. Practical modeling
6. Genomic selection
7. Application (BGF90) programs for BLUP and variance component estimation
8. References

제 7 장 연구시설 · 장비 현황

제 8장 참고문헌

- Angus. 2013. <http://www.angus.org/>
- Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen and G. Su. 2012. Single-step methods for genomic evaluation in pigs. *Animal*, 6(10):1565-1571.
- CCSI. 2013. <http://www.swineimprovement.com/>
- Falconer, D. S. and Mackay, T. F. C. 1996. *Introduction to Quantitative Genetics*. Longman, Essex, U.K., 4th ed. edition.
- Gianola, D., de los Campos, G., Hill, W. G., Manfredi E., Fernando R. 2009. Additive genetic variability and the Bayesian alphabet. *Genetics*. 183:347-363.
- Han, S. H., K. Y. Shin, S. S. Lee, M. S. Ko, D. K. Jeong, J. T. Jeon, and I. C. Cho. 2008. Effects of ADCYP1R1, FABP3, FABP4, MC4R, MYL2 genotypes on growth traits in F₂ population between Landrace and Jeju Native Black pig. *J. Anim. Sci. & Technol. (Kor.)* 50:621-632.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K. and Goddard, M. E. 2008. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetic Selection. Evolution*. 41-51.
- Hayes, B. J. and Goddard, M. E. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution*. 33:209-229.
- Hayes, B. J. 2008. Course Notes 'QTL Mapping, Mas, and Genomic Selection'.
- Hayes, B. J., Visscher, P. E., McPartlan, H. and Goddard, M. E. 2003. A novel multi-locus measure of linkage disequilibrium and its use to estimate past effective population size. *Genome Research*. 13:635-643.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31(2):423-447
- Illumina. 2013. <http://www.illumina.com/>
- Luo, Z. W. 1998. Linkage disequilibrium in a two-locus model. *Heredity*. 80:198-208.
- Meuwissen, T. H., Hayes, B. J. and Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.
- PIC. 2013. <http://www.picaustralia.com.au/>
- Sargolzaei, M., Schenkel, F. S., Jansen, G. B. and Schaeffer, L. R. 2008. Extent of Linkage Disequilibrium in Holstein Cattle in North America. *J. Dairy Sci.* 91:2106-2117.
- Sargolzaei, M. and Schenkel, F. S. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics*. 25(5):680-681.
- Scheet, P. and Stephens, M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78:629-644.
- Su, G., Guldbbrandtsen, B., Gregersen, V. R. and Lund, M.S. 2010. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein

- population. *J. Dairy Sci.* 93:1175-1183.
- VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F. and Schenkel, F. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J.Dairy Sci.* 92:16-24.
 - VanRaden, P. M. 2007. Genomic measures of relationship and inbreeding. *Interbull Bull.* 37:33-36.
 - VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414-4423.
 - 이중재. 2012. 가계정보와 SNP기반의 유전체 혈연행렬을 이용한 한우의 도체 형질의 유전 능력 추정. 충북대학교 박사학위논문
 - 이준호, 조광현, 조충일, 박경도, 이득환. 2013. 유전체 관계행렬 구성에 따른 Landrace 순종돈의 육종가 비교. *Journal of Animal Science and Technology* 55(3):165-171.
 - 조충일. 2013. 유전체 선발 기술을 활용한 한우 도체형질의 육종가 정확도 및 유전적 개량량 추정에 관한 연구. 한경대학교 박사학위논문

주 의

1. 이 보고서는 농림축산식품부에서 시행한 농림축산식품연구개발사업의 연구 보고서입니다.
2. 이 보고서 내용을 발표할 때에는 반드시 농림축산식품부에서 시행한 농림축산식품연구개발사업의 연구결과임을 밝혀야 합니다.
3. 국가과학기술 기밀유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 아니 됩니다.