



순종 요크셔 집단의 일당증체량에 대한 GWAS분석

김윤주^{1*}·강호찬¹·김은호¹·유채경²·이재봉³·최태정⁴·임현태^{1,2}

¹경상대학교 응용생명과학부(BK21 plus), ²경상대학교 농업생명과학연구원, ³전북대학교 인수공통전염병연구소, ⁴농촌진흥청 국립축산과학원

Genome-wide association analysis on daily gain of purebred Yorkshire population

Yoon-Joo Kim^{1*}, Ho-Chan Kang¹, Eun-Ho Kim¹, Chae-Kung Yoo², Jae-Bong Lee³, Tae-Jeong Choi⁴, Hyun-Tae Lim^{1,2}

¹Division of Applied Life Science(BK21 plus), Gyeongsang National University, Jinju, 52828, Korea.

²Korea Zoonosis Research Institute(KoZRI), Chonbuk National University, Iksan, 54531, Korea.

³Institute of Agriculture and Life Science, Gyeongsang National University, Jinju, 52828, Korea.

⁴National Institute of Animal Science, Rural Development Administration, Cheonan, 31000, Korea.

Abstract

Growth-related traits (daily gain, back fat thickness, and carcass weight et al.) are complex and economically important in the swine industry. The purpose of this study was to confirm quantitative trait loci (QTL) and the associated positional candidate genes affecting daily gain in pigs. A genome-wide association study (GWAS) was performed using the porcine SNP 60K bead chip. The data used in the study included 1,755 purebred Yorkshire pigs. All experimental animals were genotyped with 44,985 SNPs located throughout the pig autosomes. We identified association between a SNP markers on chromosome 15 (24.8-25.5 Mb). Three positional candidate genes are located on SSC15, *STEAP3*, *C15H2orf76* and *TEX51*. This genes may provide in the selective breeding program after validating its effect on other populations.

Keywords : Average daily gain, GWAS, Single nucleotide polymorphism(SNP), Purebred Yorkshire

서 론

양돈산업에서 돼지의 경제형질은 전통적으로 중요시되어왔고 농가의 수익성과 연결된 중요한 부분이다. 이러한 양돈산업의 수익성을 좌우하는 경제형질에는 육성률, 체중, 사료효율, 일당증체량 및 90kg도달일령 등이 있으며, 이러한 형질은 우수한 종축을 선발하는 기준으로 이용하고 있다. 이 중 돼지의 성장속도는 일당증체량 또는 90kg 도달일령을 사용하여 평가해 왔다. Kim et al.(2000)이 일당증체량은 사료효율과 높은 정의 상관관계를 가지며 비육돈의 증체능력을 나타내는 중요한 형질임을 보고하였고, 일당증체량이 높은 돼지는 낮은 돼지보다 짧은 기간 내 목표체중에 달성할 수 있으므로 사료비 절감이라는 긍정적인 효과를 가져오게 된다(Quan et al., 2018). 또한 2009년 축산과학원 양돈과에서 보고한 내용에 따르면 생산비가 약 60%개선이 되면 농가의 소득이 약 2.8% 향상되며, 이를 통해 성장률이 높은 개체가 생산비 측면에서도 효율적인 것을 확인하였다. 생산비의 효

*Corresponding author: Yoon-Joo Kim

Tel: +82-55-772-1945, E-mail: lilylily3583@gmail.com

Received: 15 June, 2018, Revised: 21 September, 2018, Accepted: 21 September, 2018



© Journal of Animal Breeding and Genomics 2018. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

울성이 증가하게 되면 소득 또한 증가하게 되므로 성장형질은 양돈산업에서 경제적으로 매우 중요한 형질이라고 할 수 있다. 따라서 일당증체량 개선을 위해서는 환경적인 요인도 무시할 수 없지만 유전적인 요인에 의한 손실을 최소화 하기 위한 연구가 필요하다.

대부분의 형질들은 환경적인 요인과 함께 서로 다른 크기의 효과를 가진 유전적 요인에 영향을 받는다는 기본 개념을 바탕으로 표현형에 있어 유전적인 부분과 환경적인 부분이 어떻게 상호작용하며, 그 유전적인 부분의 위치를 알아내고자 하는 양적형질좌위 (Quantitative traits loci)를 탐색하기 위해 현재까지 다양한 방법들이 제시되고 있다. 그 중 Genome-Wide Association Study(GWAS)에 대한 연구결과는 2005년 이후 점점 증가하고 있으며, 이러한 GWAS는 대량의 marker정보를 이용하여 복합적인 표현형 변이를 일으키는 유전적 요인을 확인하는데 사용이 가능하며 지금까지 GWAS를 이용한 연구들이 보고되고 있다 (Sladek et al., 2007; Bergelson and Roux, 2010; Flint and Eskin, 2012). GWAS분석 기법의 발달로 돼지의 표현형 정보를 이용하여 60,000개 이상의 SNP marker를 통계 분석 할 수 있으며, 유전적 다양성과 유전적인 능력 평가의 정확도를 높일 수 있게 되었다(Ramos et al., 2009). 이에 본 연구는 순종 요크셔 집단의 일당증체량과 관련하여 GWAS분석을 수행하고 유의적인 genome 상 위치 및 위치적 후보유전자를 탐색하고자 실시하였다.

재료 및 방법

1. 공시 동물 및 일당증체량 형질 측정

국내 N 종돈장에서 순종 요크셔 집단 1,755두의 모근을 이용하여 추출한 genomic DNA를 제공 받았다. 일당증체량은 순종 요크셔 집단 1,755두의 검정개시체중과 검정종료체중을 측정하고 일당증체량 보정식을 이용하여 계산하였다. 통계분석에 앞서 측정된 형질은 MINITAB 프로그램(Minitab Inc., USA)에서 구현이 되는 Ryan-Joiner 방법(1976)을 이용하여 정규성을 검증하고 이상치를 제거하였다.

2. 데이터 기초분석 및 GWAS

MINITAB 프로그램을 이용하여 통계분석 model에 포함시킬 고정효과와 공변량을 검증하였다. 모근에서 추출된 genomic DNA를 이용하여 Porcine SNP 60K bead chip(Illumina, USA) Ver.1 62,163개, Ver.2 61,565개의 SNP marker를 선발하여 각각 Chip Ver.에 공통으로 포함되어 있는 56,503개 SNP marker data의 quality control은 Plink 프로그램 version 1.07(Purcell et al., 2007)을 이용하였다. Minor allele frequency(MAF) 5%미만, 유전자형분석 오류 10%초과, Hardy-Weinberg평형 오류 P-value 10^{-6} 이하에 해당되는 SNP marker들을 제거하였다. 최종적으로 상염색체상의 44,985개 SNP marker를 선발하였고, 가계보정은 Genome-wide efficient mixed model association(GEMMA, Zhou and Stephens., 2012)프로그램을 이용하였다. SNP marker와 일당증체량과의 관련성 분석은 아래의 선형 회귀 분석식을 이용하였다.

$$y = W\alpha + x\beta + \mu + \varepsilon; \mu \sim \text{MVN}_n(0, \lambda\tau^{-1}K), \varepsilon \sim \text{MVN}_n(0, \tau^{-1}I_n)$$

여기서, y 는 n 개체에 대한 표현형 벡터(or binary disease labels)를 말한다. $W = (w_1, \dots, w_n)$ 는 1의 column을 포함하는 공변량(fixed effects)의 $n \times c$ 효과이다. α 는 intercept를 포함하여 해당 계수에 대응하는 c 벡터이고, x 는 유전자형 marker의 n 벡터이다. β 는 marker효과 크기이고, μ 는 임의 효과(random effect)의 n 벡터이다. ε 는 n 벡터의 오차이다. τ^{-1} 는 잔여 오차의 분산이며, λ 는 두 분산간의 비율이다. K 는 $n \times n$ 연관성 행렬이고 I_n 은 $n \times n$ 의 항등행렬이다. MVN_n 은 n 크기의 다변수 정규분포를 나타낸다. GEMMA는 공통적으로 사용되는 세가지 통계 중 하나를 사용하여 차례대로 각 SNP에 대한 귀무가설 $H_0: \beta = 0$ 의 대체 가설인 $H_1: \beta \neq 0$ 을 사용한다(Wald, likelihood ratio or score). GEMMA는 λ 와 β 의 maximum likelihood estimate(MLE)

또는 restricted maximum likelihood estimate(REML)를 얻고 해당 값을 나타낸다. 또한 GEMMA는 유전자형 또는 “chip heritability”를 입력하여 PVE를 추정한다.

GWAS분석에서 쓰인 SNP marker는 additive model로 분석하였으며, 임계수준(suggestive threshold level)은 Bonferroni adjusted genome-wide suggestive 유의 수준(Bonferroni adjusted threshold : 1/상염색체상의 SNP marker수)으로 하였다. Bonferroni threshold에 대해 각 염색체 별 SNP marker를 확인하고, 이들 marker와 가장 근접한 유전자를 위치상 후보 유전자로 탐색하였다. 후보유전자 탐색은 NCBI dbSNP(<https://www.ncbi.nlm.nih.gov/snp>, map ver. 11.1)을 사용하였고, 유의적으로 확인된 SNP marker의 위치를 기준으로 가장 근접하거나 포함되어 있는 유전자를 위치상 후보유전자로 동정하였다.

동정된 후보 SNP marker와 주변 SNP marker간의 연관성을 보기 위해 R 프로그램의 snp.plotter(Luna & Nicodemus, 2007) package를 이용하여 유의성 있는 SNP로 확인된 marker의 연관 불평등(LD, Linkage disequilibrium)분석을 실시하였다.

결과 및 고찰

순종 요크셔 집단 1,642두의 일당증체량 평균값과 표준편차는 0.660 ± 0.053 kg이고, 최소값 0.491, 최대값 0.854로 나타났다(Table 1).

Illumina사의 Porcine SNP 60K beadchip을 이용하여 생성된 SNP marker중 Plink 프로그램을 통해 $MAF < 0.05$, genotyping error > 0.1 , Hardy Weinberg equilibrium($P\text{-value} \leq 10^{-6}$)에 해당하는 SNP marker를 제거하였다. 그 결과, 상염색체에 위치한 44,985개 SNP marker를 확인할 수 있었고, 이들 SNP marker와 일당증체량 형질 사이의 관련성 분석을 통해 genome상 유의적인 위치를 탐색하였다. SNP marker의 Manhattan plot과 Quantile-quantile(QQ) plot을 작성한 결과(Fig. 1)와 GWAS분석 결과에 대한 Bonferroni임계수준(genome-wide suggestive level)을 이용하여 유의적으로 나타난 SNP marker는 염색체 15번에서 5개가 확인된 결과는 다음과 같다(Table 2). 이들 중 ALGA0108976 marker는 $P = 6.75 \times 10^{-6}$ 유의성으로 가장 높은 유의성을 나타내었다. 5개의 SNP marker는 돼지 15번 염색체의 24,814,161 ~ 25,521,629bp사이에 위치하고 있었으며, NCBI데이터베이스를 통해 각 SNP marker의 위치를 확인하고, 가까이에 위치한 유전자를 탐색하였다. 유의적으로 확인된 5개 SNP marker의 위치적 후보유전자는 STEAP3, C15H2orf76, EX51으로 확인되었다. 그 중 STEAP3(STEAP3 metalloredutase)는 iron수송체로 작용하는 다중 경로 막단백질로 코딩되어 있으며, 인간의 면역 시스템과 관련이 있고 결핍 시 간암과 빈혈을 유발한다고 보고되어 있었으나(Watson et al., 2016; Cadiou et al., 2017), 그 외 C15H2orf76와 EX51유전자의 기능에 대한 보고는 찾아볼 수 없었다.

Fonatanesi et al.(2014)의 Italian Large White종 375두의 일당증체량에 대한 GWAS분석을 이용하여 발표한 연구에 따르면 49,413개의 SNP marker를 통해 127개의 genome-wide significant SNP marker를 동정하였고 11개의 후보유전자를 선정하였다. 이 후보유전자들 중 다수는 돼지뿐만 아니라 사람, 소 등 타가축의 성장형질에도 영향을 미친다고 보고되어 있었다. 이 외에도 Wang et al.(2015)은 Duroc 수컷 돼지 796두의 일당증체량 및 100kg도달일령에 대한 GWAS연구에서 39,436개의 SNP marker를 통해 162개의 genome-wide significant SNP marker를 확인하였고, 14개의 후보유전자를 선정하였다. 그 중 일당증체량과 연관된 후보유전자는 1개(COL27A1)로 보고하였으며, 본 후보유전자는 XXVII형태의 콜라겐을 생성하며 연골의 석회

Table 1. Descriptive statistics of daily gain in a purebred Yorkshire pigs.

Trait	N ¹⁾	Mean	Standard Deviation	MIN	MAX	Effect	
						Fixed	Covariates
DG ²⁾ (kg)	1,642	0.660	0.053	0.491	0.854	Sex, Group, Farm, Litter	End weight

¹⁾N : Number of animal, ²⁾DG : Daily gain

화에 중요한 역할로 돼지의 성장 특성에 직접 또는 간접적으로 밀접한 관련이 있다고 보고하였다(Hjorten et al., 2007). 뿐만 아니라 Meng et al.(2017)이 순종 요크셔 집단 562두의 일당증체량을 측정하여 보고한 연구는 중국의 Geneseek neogen porcine SNP 80K beadchip을 사용하여 MAF < 0.03, genotyping error > 0.9, Hardy Weinberg equilibrium(P -value $\leq 10^{-6}$)을 만족

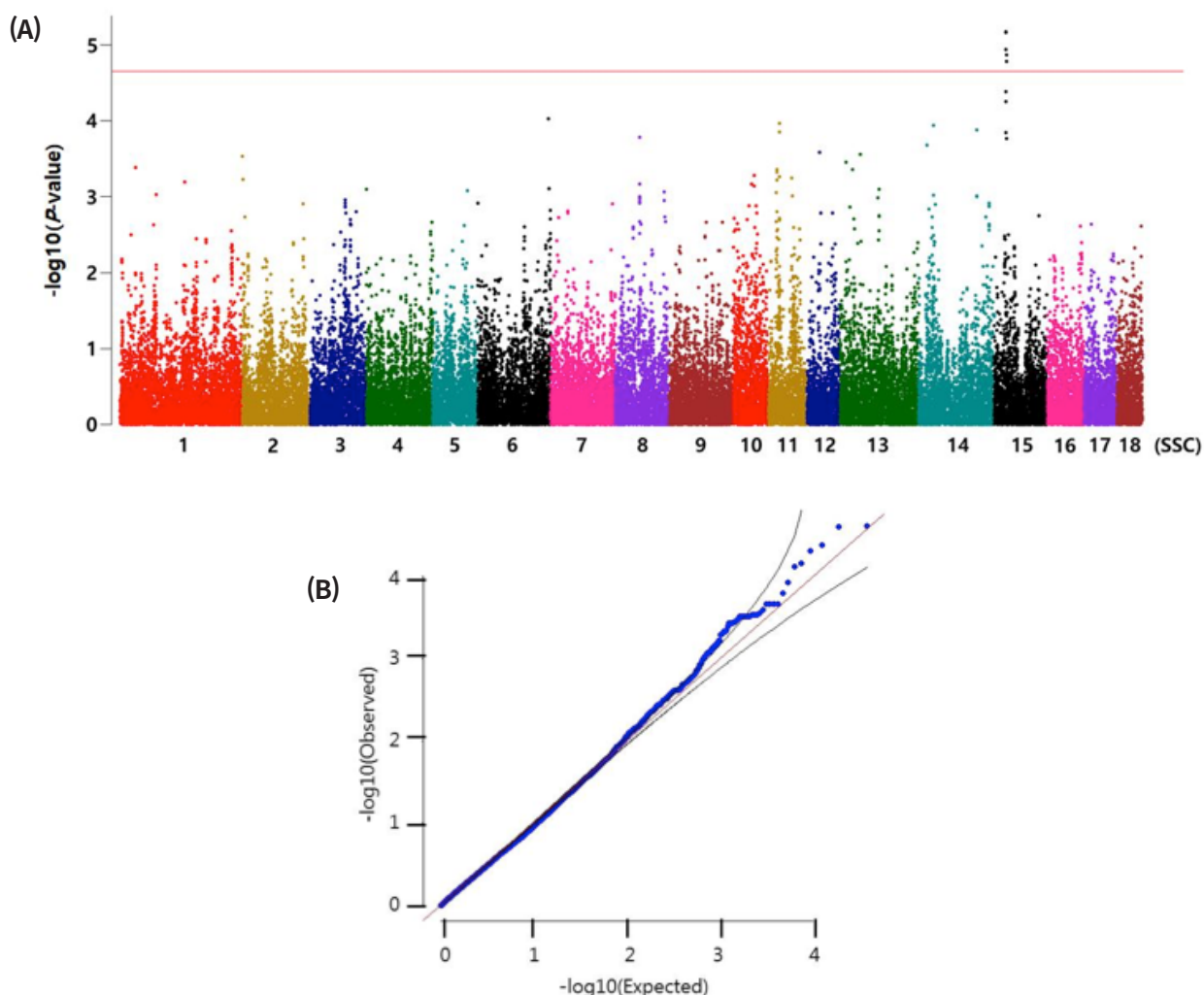


Fig. 1. Manhattan plot(P -value) and Quantile-Quantile plot of genome wide association study of daily gain. SNPs locate their position on each chromosome against on the x-axis and the association signal on the y-axis (shown as $-\log_{10}$ of the P -value). (A) Association between daily gain and 44,985 mapped SNP markers in 18 pig autosomes using additive model. (B) Quantile-Quantile plots of SNP markers.

Table 2. GWAS results of SNP markers for daily gain in purebred Yorkshire.

SSC ¹	SNP	Location(bp) ²	Allele ³	MAF ⁴	P -value	Nearest Gene
15	MARC0097408	24,814,161	C/T	0.165	1.15×10^{-5}	STEAP3
	ALGA0108976	24,828,435	A/G	0.155	6.75×10^{-6}	C15H2orf76
	MARC0084917	24,830,838	G/T	0.155	6.84×10^{-6}	C15H2orf76
	ASGA0084063	25,472,091	A/G	0.121	1.35×10^{-5}	TEX51
	ALGA0111611	25,521,629	T/C	0.128	1.67×10^{-5}	TEX51

¹*Sus scrofa* chromosome

²NCBI annotation Ver. 11.1

³minor/major allele.

⁴minor allele frequency.

시키는 상염색체의 SNP marker를 동정하고 GWAS분석을 실시하였다. 15번 염색체에서 일당증체량과 밀접한 관련이 있는 2개의 후보유전자(*GLI2*, *NYPA2*)를 선정하였고, 그 중 *GLI2* 유전자는 돼지의 후장과 관련이 있었으며, 특히 항문 폐쇄증에 영향을 미친다고 보고하였고(Jin et al., 2013), *NYPA2* 유전자는 혈소판 면역체계와 관련 있다고 보고하였다(Yokoyama et al., 2011). 또한 Jianping et al.(2018)은 Duroc 중 390두를 이용하여 GWAS분석을 통해 1, 5, 6, 9, 13, 16번 염색체에서 일당증체량 형질과 관련된 SNP를 동정하였고, 6개의 후보유전자(*GTF3C5*, *HMGA2*, *NAMPT*, *OLIG1*, *PLEKHG4B*, *ENSSSCG00000031548*)를 선정하였다. 분석 결과 *GTF3C5* 유전자는 돼지 골격근 세포 분화와 밀접한 관련이 있는 유전자로 보고하였고(Hsieh et al., 1999; Lander et al., 2001), *HMGA2* 유전자는 배아 발생 과정에서 고도로 발현된다는 사실이 이전 연구에서 밝혀졌다(Falvo et al., 1995; Reeves et al., 2001; Zhou et al., 1995; Chiappetta et al., 1996; Zhou et al., 1998). 또 인간과 쥐를 이용한 연구에서 *NAMPT* 유전자가 단계별로 전이 효소에 있는 nicotinamide5-phosphoribose-1peropheoptheaenase를 코딩 할 수 있으며, 내장 지방조직에 의해 우선적으로 효소를 분비 할 수 있다고 보고하였다(Imai et al., 2009). 나머지 *OLIG1*와 *PLEKHG4B* 유전자는 인간의 뇌와 신경계의 발달과 관련이 있으며, 신경 영양 인자를 코딩 할 수 있다고 보고하였다(Raffield et al., 2015; Prasad et al., 2017). 따라서 본 연구 결과와 함께 제시된 유전자들은 돼지의 일당증체량을 향상시키는 분자 육종 연구에서 중요한 후보 유전자로 선정될 것으로 생각된다.

이처럼 돼지의 일당증체량과 관련된 GWAS분석이 다양한 결과를 나타내는 이유는 각기 다른 품종, 집단의 구성, 크기, 사육 방

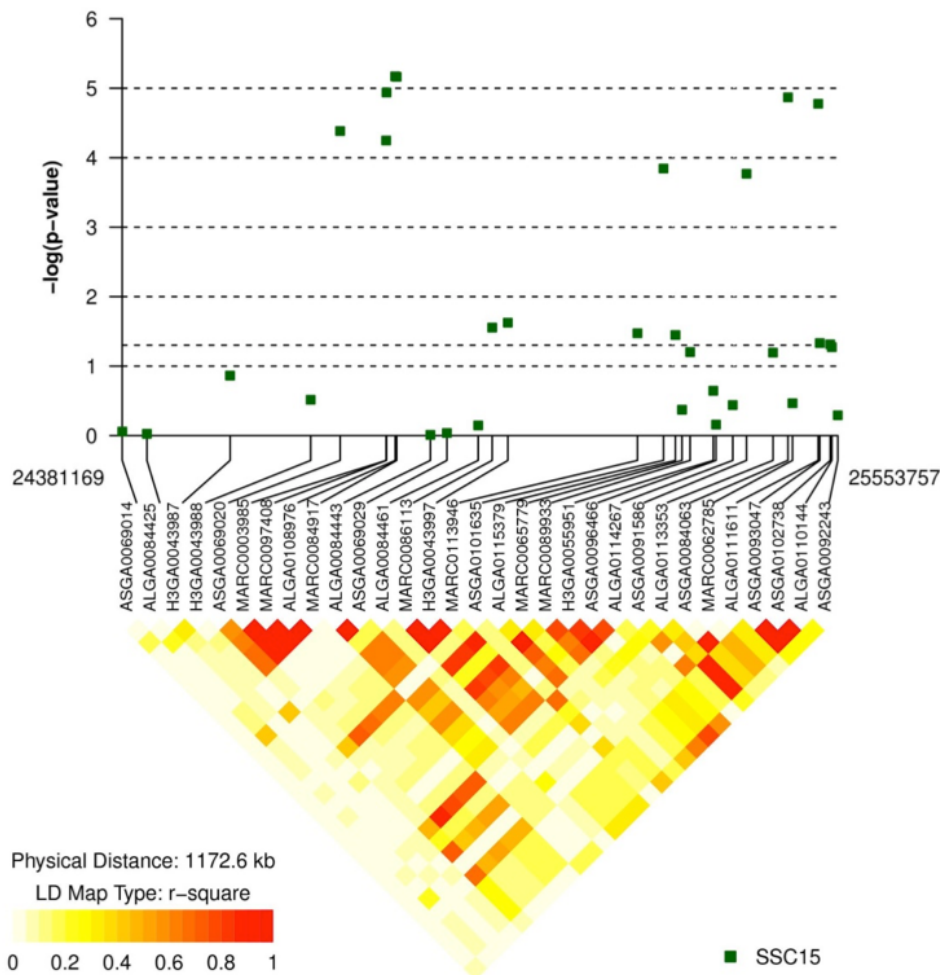


Fig. 2. Association results and LD presented exploiting of purebred Yorkshire data obtained from snp.plotter. SNPs name and their position are locate on x-axis and $-\log_{10}$ of the P -value on y-axis. Dotted lines represent P -value thresholds. It is purebred Yorkshire SSC15 of snp.plotter results.

식 및 다양한 통계적 분석기법 등의 여러 가지 요인들에 의한 차이 일 수 있으며, 또 다른 원인으로는 현재 상용되는 commercial SNP chip이 큰 효과를 발휘하지 못하는 드문 모델들의 영향이 이 같은 차이를 유발했을 것으로 보인다(Monolio et al., 2009; Stringer et al., 2011; Wray et al., 2011).

SNP.plotter 프로그램을 이용하여 유의적으로 확인된 15번 염색체의 5개 SNP marker와 주변 SNP marker간의 LD를 분석하였다(Fig. 2). 분석 결과 유의성 있는 5개의 SNP marker중 MARC0097408, ALGA0108976 및 MARC0084917 (24,814,161 - 24,830,838 bp)이 하나의 블록(block A)을 형성하였고, ASGA0084063과 ALGA0111611(25,472,091 - 25,521,629 bp)이 하나의 블록(block B)을 형성하였다. Block A는 주변 SNP marker(ASGA0069020, MARC0003985, ASGA0101635 및 ASGA0091586)와 강한 LD ($r^2 > 0.6$)를 형성하였으나, block B의 경우 SNP marker ASGA0069020와 강한 LD를 형성하는 것을 확인 할 수 있었다. 또한, Block A의 3개 SNP marker와 Block B의 2개 SNP marker간에는 강한 LD가 형성되지 않는 것을 확인할 수 있었다.

본 연구는 순종 요크셔 집단의 일당증체량 형질에 대한 GWAS분석을 실시하고 Bonferroni임계수준(genome-wide suggestive level)에 따라 유의적인 SNP marker를 확인하였다. 그 결과 15번 염색체에서 5개의 SNP marker를 확인 할 수 있었으며, 이들의 위치를 통해 3개의 위치적 후보유전자를 확인할 수 있었다. 이 3개의 유전자는 돼지의 경제형질과 관련된 중요한 지표로서 다른 품종 및 독립적인 축군에서 유전적 차이를 확인해 볼 필요성이 있다고 여겨지며, 추가적인 기능 연구를 통해 돼지의 일당증체량 형질과 관련된 후보유전자로서 선발될 수 있는 기초자료가 될 것으로 보인다.

사사

본 연구에 참여한 대학원생은 교육과학기술부 BK21Plus 사업의 장학금을 수혜 받았으며, 농촌진흥청 공동연구사업(과제번호: PJ009971032017)의 지원에 의해 이루어진 것이며, 연구비 지원에 감사 드립니다.

References

- Bergelson, J., and F. Roux. 2010. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nat. Rev. Genet.* 11:867-879.
- Cadiou JL, Pichat S, Bondanese VP, Soulard A, Fujii T, Albaredo F, Oger P, Copper transporters are responsible for copper isotopic fractionation in eukaryotic cells. *Sci Rep.* 2017 Mar 17;7:44533.
- Chiappetta G, Avantiaggiato V, Visconti R, et al. High level expression of the HMGI(Y) gene during embryonic development. *Oncogene* 1996;13:2439-46.
- Falvo JV, Thanos D, Maniatis T. Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell* 1995;83:1101-11.
- Flint, j., and E. Eskin. 2012. Genome-wide association studies in mice. *Nat. Rev. Genet.* 13:807-817.
- Fontanesi L, Shiavo G, Galimberti G, Calo DG, Russo V. A genomewide association study for average daily gain in Italian Large White pigs. *J Anim Sci.* 2014 Apr;92(4):1385-94.
- Hjorten R, Hansen U, Underwood RA, Telfer HE, Fernandes RJ, Krakow D, et al. Type XXVII collagen at the transition of cartilage to bone during skeletogenesis. *Bone.* 2007;41(4):535-42.
- Hsieh YJ, Wang Z, Kovelman R, Roeder RG. Cloning and characterization of two evolutionarily conserved subunits (TFIIIC102 and TFIIIC63) of human TFIIIC and their involvement in functional interactions with

TFIIIB and RNA polymerase III. *Mol Cell Biol* 1999;19:4944-52.

Imai S. Nicotinamide phosphoribosyltransferase (*NAMPT*): a link between NAD biology, metabolism, and diseases. *Curr Pharm Des* 2009;15:20-8.

Jianping Q, Roungrong D, Xingwang W, Ming Y, Yang Y, Enqin Z, Ting G, Gengyuan C, Zhenfang W, Dewu L, Jie Y. Genome-wide association study reveals genetic loci and candidate genes for average daily gain in Duroc pigs. *Asian-Australas J Anim Sci*. 2018.

Jin Q, Wang C, Li X, Yu M, Zhao SH. Molecular characterization and genome-wide mutations in porcine anal atresia candidate gene *GLI2*. *Mamm Genome*. 2013 Dec;24(11-12):500-7.

Kim, C. W.; Back, D. H.; Lee, J. K.; Cho, K. K.; Choi, J. S.; Rho, J. M.; Song, Y. M.; Kwon, E. J.; Park, W. S.; Hong, Y. H. identification of DNA markers associated with average daily gain (ADG) in Yorkshire. *Korean Journal of Animal Science* 2000 42(4) : 423-432.

Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921.

Luna A, Nicodemus KK. snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. *Bioinformatics*. 2007 Mar 15;23(6):774-6.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-53.

Meng Q, Wang K, Liu X, Zhou H, Xu L, Wang Z, Fang M. Identification of growth related genes in a Yorkshire purebred pig population by genome-wide association studies. *Asian-Australas J Anim Sci*. 2017 Apr;30(4):462-469.

Prasad A, Teh DBL, Blasiak A, et al. Static magnetic field stimulation enhances oligodendrocyte differentiation and secretion of neurotrophic factors. *Sci Rep* 2017;7:6743.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007 Sep;81(3):559-75.

Quan J, Ding R, Wang, Yang M, Yang Y, Zheng E, Gu T, Cai G, Wu Z, Liu D, Yang J. Genome-wide association study reveals genetic loci and candidate genes for average daily gain in Duroc pigs. *Asian-Australas J Anim Sci*. 2018 Apr;31(4):480-488.

Raffield LM, Cox AJ, Hugenschmidt CE, et al. Heritability and genetic association analysis of neuroimaging measures in the Diabetes Heart Study. *Neurobiol Aging* 2015;36:1602.

Ramos AM, Crooijmans RP, Affara NA, et al. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* 2009;4:e6524.

Reeves R. Molecular biology of HMGA proteins: hubs of nuclear function. *Gene* 2001;277:63-81.

Ryan TA and Joiner BL. 1976. Normal probability plots and tests for normality. Minitab, State College, Pennsylvania.

Sladek, R., G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre, P. Boutin, D. Vincent, A. Belisle, S. Hadjadj, B. Balkau, B. Heude, G. Charpentier, T. J. Hudson, A. Montpetit, A. V. Pshezhetsky, M. Prentki, B. I. Posner, D. J. Balding, D. Meyre, C. Polychronakos, and P. Froguel. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881-885.

Shi Z, Wu D, Tang R, et al. Silencing of HMGA2 promotes apoptosis and inhibits migration and invasion of

prostate cancer cell. *J Biosci* 2016;41:229-36.

Stringer S, Wray NR, Kahn RS, Derks EM. Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. *PLoS One*. 2011;6(11):e27964.

Swine Science Division, National Institute of Animal Science, RDA. 2009.

Wang K, Liu D, Hernandez-Sanchez J, Chen J, Liu C, Wu Z, et al. (2015) Genome Wide Association Analysis Reveals New Production Trait Genes in a Male Duroc Population. *PLoS ONE* 10(9):20139207.

Weston C, Klobusicky J, Weston J, Connor J, Toms SA, Marko NF, Aberrations in the Iron Regulatory Gene Signature Are Associated with Decreased Survival in Diffuse Infiltrating Gliomas. *PLoS One*. 2016 Nov 29;11(11):e0166593.

Wray NR, Purcell SM, Visscher PM. Synthetic Associations Created by Rare Variants Do Not Explain Most GWAS Results. *PLoS Biol*. 2011;9(1).

Yokoyama K, Tezuka T, Kotani M, Nakazawa T, Hoshina N, Shimoda Y, Kakuta K, Watanabe K, Iwakura Y, Yamamoto T. NYAP: a phosphoprotein family that links PI3K to WAVE1 signalling in neurons. *EMBO J*. 2011 Sep 23;30(23):4739-54.

Zho X, Chada K. HMGI family proteins: architectural transcription factors in mammalian development and cancer. *Keio J Med* 1998;47:73-7.

Zhou X, Benson KF, Ashar HR, Chada K. Mutation responsible for the mouse pygmy phenotype in the developmentally re-regulates factor HMGI-C. *Nature* 1995;376:771-4.

Zhou X, Stephens M. Genome-Wide efficient mixed-model analysis for association studies. *Nat Genet*. 2012 Jun 17;44(7):821-4.