Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Single step genome-wide association studies based on genotyping by sequence data reveals novel loci for the litter traits of domestic pigs

Pingxian Wu, Qiang Yang[1], Kai Wang, Jie Zhou, Jideng Ma, Qianzhi Tang, Long Jin, Weihang Xiao, Anan Jiang, Yanzhi Jiang, Li Zhu, Xuewei Li, Guoqing Tang*

Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu 611130, China.

ARTICLE INFO

Keywords:
Swine
Single step GWAS
Genotyping-by-sequencing
Litter traits

ABSTRACT

In this study, data genotyping by sequence (GBS) was used to perform single step GWAS (ssGWAS) to identify SNPs associated with the litter traits in domestic pigs and search for candidate genes in the region of significant SNPs. After quality control, 167,355 high-quality SNPs from 532 pigs were obtained. Phenotypic traits on 2112 gilt litters from 532 pigs were recorded including total number born (TNB), number born alive (NBA), and litter weight born alive (LWB). A single-step genomic BLUP approach (ssGBLUP) was used to implement the genome-wide association analysis at a 5% genome-wide significance level. A total of 8, 23 and 20 significant SNPs were associated with TNB, NBA, and LWB, respectively, and these significant SNPs accounted for 62.78%, 79.75%, and 58.79% of genetic variance. Furthermore, 1 (SSC14: 16314857), 4 (SSC1: 81986236, SSC1: 66599775, SSC1: 161999013, and SSC1: 267883107), and 5 (SSC9: 29030061, SSC2: 32368561, SSC5: 110375350, SSC13: 45619882 and SSC13: 45647829) significant SNPs for TNB, NBA, and LWB were inferred to be novel loci. At SSC1, the AIM1 and FOXO3 genes were found to be associated with NBA; these genes increase ovarian reproductive capacity and follicle number and decrease gonadotropin levels. The genes SLC36A4 and INTU are involved in cell growth, cytogenesis and development were found to be associated with LWB. These significant SNPs can be used as an indication for regions in the *Sus scrofa* genome for variability in litter traits, but further studies are expected to confirm causative mutations.

1. Background

In pigs, the total number born (TNB), number born alive (NBA), and litter weight born alive (LWB) are important reproduction traits that have significant effects on the production efficiency of pigs. In the past few decades, many animal breeders have performed best linear unbiased prediction (BLUP) selection, which led to great genetic improvement in animal production traits. However, these litter traits are complex traits that have low heritability and are controlled by many quantitative trait nucleotides with small effects [1]. Thus, the selecting for them is very hard to accomplish [2].

Until recently, previous studies mainly focused on growth and meat traits, and few researchers devoted themselves to studying litter traits [3,4]. Many significant SNPs and 16,506 pig quantitative trait loci (QTLs) have been identified in the pig genome, but those only include 377 QTLs for reproductive traits (http://www.animalgenome.org/cgi-bin/QTLdb/SS/index). However, a total of 159, 129 and 21 QTLs have been identified for TNB, NBA, and LWB [5], respectively. For pigs, TNB is an indicator of sow reproductive performance [6]. In previous studies, several genes, such as the estrogen receptor gene (ESR1, ESR2) [7,8], catenin alpha-like 1 (CTNNAL1) [9], and transcription factor 12 (TCF12) [9], have been identified to have significant effects on litter size in commercial pig populations, but these candidate genes only explain a small variation in litter size.

The NBA plays an important role in pig production efficiency. To better understand the genetic architecture of farrowing traits, several GWAS have recently been implemented to identify regions associated with NBA [10] in different pig breeds. Many significant SNPs and candidate genes associated with NBA have been detected, for example, catenin alpha-like 1 (CTNNAL1) [9], porcine insulin-like growth factor 2 (IGF2) [11], and retinol-binding protein 4 (RBP4) [11]. Even though many studies have contributed to reproduction traits, few studies have reported the genetic architecture of the LWB. Until recently, a few SNPs and only one candidate gene (MAGE-like 2 (MAGEL2) [12]) have been associated with LWB.

The application of high-throughput genotyping techniques was

* Corresponding author.
E-mail address: tyq003@163.com (G. Tang).
[1] This author contributed equally to this work and should be considered co-first authors.

http://dx.doi.org/10.1016/j.ygeno.2017.09.009
Received 29 August 2017; Received in revised form 18 September 2017; Accepted 19 September 2017

performed to quantify complex traits for better understanding. However, the high cost of high-throughput genotyping was a main critical factor for QTL mapping of the complex traits of animals. Now, a cost-effective technology, which is a high-throughput sequencing-based genotyping approach, has made genotyping possible. One or more restriction enzymes are used to digest the genome into many small fragments [13], and these fragments are sequenced by parallel high-throughput methods, and then bioinformatics pipelines are used to do SNP calling [14–17]; this process is called genotyping by sequencing (GBS). Repetitive genomic regions are avoided using methylation-sensitive restriction enzymes in a genotyping by sequence process. GBS is a highly multiplexed technique based on an Illumina sequencing platform and is suitable for the quantitative study of complex traits in animals [18–20]. The approach provides whole-genome coverage and can be implemented at a low cost with or without reference genome sequences that enable detailed genetic analysis and high power to detect particular variations of animal genes [21]. The GBS marker map is superior to any previous reports, and GBS is a powerful and useful tool in genomic studies. Because of these advantages, this efficient technology is beneficial for studies in pigs, and a large number of SNPs have been identified and used as genetic markers for GWAS.

Moreover, Wang et al. proposed new GWAS approaches called GWAS by single-step genomic BLUP (ssGBLUP) and single-step GWAS (ssGWAS). The ssGBLUP procedure provides the most comprehensive information for genomic evaluation. With the ssGWAS approach, we can simultaneously examine all data, such as all genotypes, all phenotype records, and all pedigree information in one step [22]. Recently, many studies have validated this approach and successfully implemented ssGWAS in pigs [23] and other species [24,25]. This approach was efficiently implemented for many models and yielded a greater power and precise estimate values [22].

In this study, we focused on TNB, NBA, and LWB traits in a Yorkshire and Landrace pig population. The main objective of the study is to identify SNPs associated with TNB, NBA, and LWB using ssGWAS in domestic pigs and search for candidate genes in the regions of significant SNPs.

2. Methods

2.1. Animal and phenotype data

A total of 532 multiparous sows from a domestic commercial breeding stock were used in this study, which included 282 Landrace and 250 Yorkshire sows (Table 1). All 532 genotyped sows had phenotype records, the phenotypic data included 2112 reproduction records from 1 to 9 parity data, which were recorded from 2012 to 2015 and included three reproductive traits: TNB, NBA, and LWB. In addition, the dataset contained pedigree information on those 532 female pigs, but these pedigree individuals without genotypes and phenotypes.

2.2. Genotype data

In the study, we extracted genomic DNA from the ear tissue of 532 sows using the surfactant and protease pyrolysis method. Genotyping was performed using GBS [18,20] based on IlluminaHiSeq PE150. SNP markers with a minor frequency (MAF) 0.01, Miss 0.2 and dp3 were excluded from the dataset that was developed in GATK software. After these genomic data quality control measures were complete, a total of 167,355 high-quality SNPs from 532 pigs were used for ssGWAS.

Table 1
Number of sows.

Breed	Number of genotyped sows	Number of sows with phenotypes	Number of litters
Landrace	282	282	934
Yorkshire	250	250	1178
SUM	532	532	2112

2.3. Single-step genome wide association studies (ssGWAS)

To account for the limited breed differences in phenotype, breed was included as a fixed effect in later analyses, and this reduced the possibility that markers could pick up population stratification effects. The following model was used:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{Wp} + \mathbf{Za} + \mathbf{e},$$

where **Y** is the vector of phenotypes, **X** is an incidence matrix of fixed effects, **b** is a vector of fixed effects including breed, year-month, sow farrowing parity effect, **W** is an incidence matrix that related sows to phenotypes, **p** is a vector of permanent environmental effects due to sows, **Z** is a incidence matrix that related animals to phenotypes, **a** is a vector of animal effects, and **e** is residual effect vector, and assumed to be normally distributed, $N(0, \sigma_e^2)$.

In this association analysis, a ssGBLUP approach [26–28] was employed for the GWAS, which was called ssGWAS [22]. With this approach, all genotypes, phenotype records and pedigree information were considered in one step simultaneously. The ssGWAS method modified the inverse of the numerator relationship matrix (A^{-1}) to the inverse of H (H^{-1}) that combines both the genomic and pedigree relationships [26,27]:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

where A^{-1} is the inverse of the numerator relationship matrix, $A_{22}{}^{-1}$ is only the inverse of the pedigree-based relationship matrix for the genotyped animals, and G^{-1} is the inverse of the genomic relationship matrix; G weight markers were obtained by reciprocals of expected marker variance [29]. VanRaden created the G as follows:

$$G = ZDZ'q$$

where Z is an incidence matrix of gene content adjusted for allele frequencies, D is a weight matrix with for SNP (initially D = I), $D_{ii} = \frac{1}{m[2p_i(1-p_i)]}$, which contains the reciprocals of expected marker variance, and q is a normalizing factor, which is derived by ensuring that the average diagonal G is close to A_{22}. The ssGWAS is calculated as follows (Wang et al. [22]): 1. In step 1, to let D = I; 2. Calculate GEBVs for all animals in the dataset using ssGBLUP; 3. Calculate the SNP effect ($\hat{u}$) by GEBVs: $\hat{u} = qDZ'(ZDZ'q)^{-1}\hat{a}$, where $\hat{a}$ is the GEBVs of genotyped animals; 4. Calculate the weight for each SNP: $d_i = \hat{u}_i^2 2p_i(1-p_i)$, where i is the ith SNP; 5. Normalize the weight to keep the total genetic variance constant; 7. Calculate the weighted matrix G; and 8. Loop 2.

The SNP effects were estimated by 3 iterations. The percentage of genetic variance explained by the ith SNP was calculated as follows:

$$\frac{Var(Z_j\hat{u}_j)}{\sigma_a^2} \times 100\%,$$

where $\sigma_a{}^2$ is the total genetic variance, Z_j is a vector of the gene content of the jth SNP for all animals, and $\hat{u}_j$ is the estimated marker effect of the jth SNP.

The significance test of SNP effects was performed using two-sided *t*-test, and the *P* value of each SNP was calculated as follows:

$$p_i = P_t\left(\frac{\hat{u}_i}{\sqrt{\hat{\sigma}_i^2/n}}, n-1\right),$$

where P_t is distribution function of t distribution, $\hat{u}_i$ is ith SNP effect, $\hat{\sigma}_i^2$ is the genetic variance of ith SNP, n is the number of animal with ith SNP.A Bonferroni correction was applied to control false positive association, and the genome significant level was defined as $P < 0.05/N$, where N was the number of SNP loci analyzed. Thus, the significant