Gene 643 (2018) 1–6

Contents lists available at ScienceDirect

Gene

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

Research paper

Analysis of the SNP loci around transcription start sites related to goat fecundity trait base on whole genome resequencing

Rui-Qian Zhang[a,1], Fang-Nong Lai[b,1], Jun-Jie Wang[b], Hong-Li Zhai[c], Yong Zhao[b], Yu-Jiang Sun[a], Ling-Jiang Min[a], Wei Shen[a,b,*]

[a] College of Animal Science and Technology, Qingdao Agricultural University, Qingdao 266109, China
[b] College of Life Sciences, Qingdao Agricultural University, Qingdao 266109, China
[c] Shandong International Biotechnology Park, Yantai 264670, China

ARTICLE INFO

Keywords:
Goat genome
Whole genome sequencing
TSSs
SNPs

ABSTRACT

Genome flanking regions surrounding transcription start sites (TSSs) are critical for the regulation of gene expression, containing many translational regulatory elements. To investigate whether critical single nucleotide polymorphisms (SNPs) exist around TSSs in the dairy goat genome, we performed high throughput DNA sequencing to compare two dairy goat groups with discrepant litter sizes. After genome mapping, SNP calling, and annotation, we screened the SNPs within 2 kb scales surrounding annotated TSSs in high fecundity (HF) and low fecundity (LF) groups, respectively. We attempted to identify distinct SNPs and motifs near the TSSs in both groups. The SNPs near the TSSs most were consistent; 318 new SNPs were uncovered in the HF group, of which 305 were heterozygote SNPs, 13 were homozygote SNPs, and majority of which were distributed on chromosome 2 and 29. After validation by Sanger sequencing we found that a SNP in CHI16: 27612330 C > A in the PSEN2 gene presented an A/A genotype in the HF group and an A/A or A/C genotype in the LF group. In conclusion, our study provides insightful information into the dairy goat genomic variations surrounding TSSs, which may contribute to enhanced litter size. Based on comparison studies of SNPs exist around transcription start sites between high fecundity group and low fecundity group. Our finding provides insights concerning the goat litter size phenotypic and will promote future goat breeding.

1. Introduction

The goat is one of the first animals to be domesticated, and it continues to play an important role in the agricultural sector, providing milk, meat, and fiber (Luo, 2009). China is home to approximately 21.7% of the global goat herd (Luo, 2009), with dairy goats predominating. Production from the national herd has steadily increased along with the development of the dairy industry, mainly located in the middle Yellow River area including Shaanxi, Shandong, and Henan provinces, where Xinong Saanen, Guanzhong and Laoshan dairy goats are the best performing breeds. Over recent years, long breeding cycles, breeding resource dispersion, and limited understanding of the goat genome have resulted in difficulties regarding genetic marker assisted breeding. In order to satisfy consumer demand, further effective improvements of goat fecundity and production traits are needed. This requires the identification of genomic biomarkers linked to reproductive and production traits.

Fecundity is a complex trait that is regulated by minor polygenes and involves multiple interactions. In recent years, various candidate genes for goat fertility have been reported (An et al., 2015a, 2015b; Lai et al., 2016). For example, the KIT receptor coupled ligand gene (*KITLG*), abundantly expressed in early stage ovarian granulosa cells and involved in primordial follicle activation, contains a number of single nucleotide polymorphisms (SNPs) and exhibits a Hardy-Weinberg disequilibrium in Xinong Saanen and Guanzhong dairy goats, which indicates that specific genome loci undergo selection, mutation, or migration (An et al., 2015a). It has been demonstrated that SNP genotype combinations, for example, g.12654, g.12772, g.12829 and g.23683 SNPs combined with the genotype GG/GG/CC/CC, produce a larger litter size (An et al., 2015a). Furthermore, SNPs existing in the 5′

Abbreviations: TSSs, transcription start sites; SNPs, single nucleotide polymorphisms; HF, high fecundity; LF, low fecundity; KITLG, KIT receptor coupled ligand gene; UTR, untranslated region; GDF9, growth differentiation factor 9; BMP15, bone morphogenetic protein 15; KISS, kisspeptin encoded gene; ZBED6, BED-type containing 6; IGF2, insulin-like growth factors 2; NPY, neuropeptide Y

* Corresponding author at: College of Animal Science and Technology, Qingdao Agricultural University, Qingdao 266109, China.
E-mail address: wshen@qau.edu.cn (W. Shen).
[1] Zhang R.Q. and Lai F.N. are co-first authors.

https://doi.org/10.1016/j.gene.2017.12.002
Received 1 November 2017; Received in revised form 30 November 2017; Accepted 1 December 2017
Available online 05 December 2017
0378-1119/

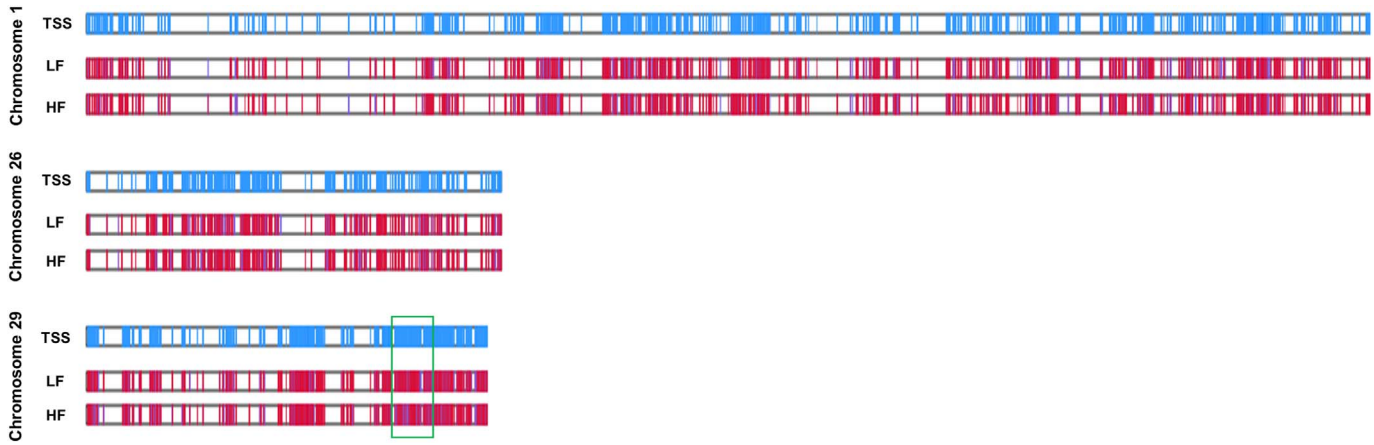

Fig. 1. The schematic diagram of the SNP distribution surrounding TSSs. The diagram shows the TSSs (upper cyan line) and the region surrounding the SNPs in the low fecundity (LF; middle) and high fecundity (HF; lower) groups on chromosome 1, 26, and 29. The red line in the middle and lower row indicates common loci of SNPs; the purple line in the middle and lower row indicate the SNPs unique loci in LF and HF group, respectively. The green box shows SNPs that specifically in LF and HF group, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

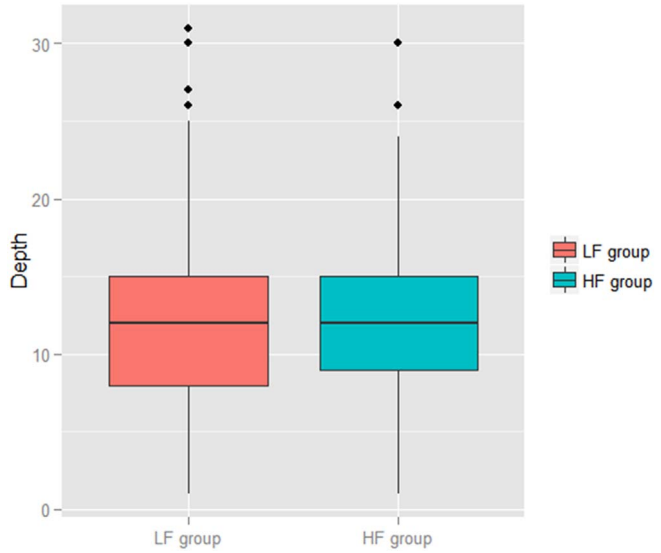

Fig. 2. The box plot shows the sequencing depth of 318 SNPs that are specifically present in the HF group, respectively, in the HF and LF groups. The black spots are outliers.

flanking region and the 3′ untranslated region (UTR) of *KITLG*, may subsequently regulate *KITLG* expression (An et al., 2015b). In fact, many SNP loci of growth differentiation factor 9 (*GDF9*), bone morphogenetic protein 15 (*BMP15*), and the kisspeptin encoded gene (*KISS*) has been reported to be related with sheep reproduction (Pramod et al., 2013).

In addition, the SNPs located in the untranslated region, which including core promoter regions, CpG islands (CpGs), TATA box elements, and nucleosome-free regions (Valen and Sandelin, 2011), may reflect differences in transcriptional regulation. Research of SNPs around transcription start sites (TSSs) in the same genomic region of humans and chimpanzees showed, a diversity across species and a remarkable periodicity of 146 nucleotides, which suggests the possible involvement of nucleosomes in promoter regions (Higasa and Hayashi, 2006). Then consensus has been reached that specific signals located in TSSs could regulate the initiation of transcription, including protein-specific DNA sequence interaction and advanced chromatin features. While SNPs located in those regions cannot alter the amino acid sequence of peptides or proteins, but might change transcription factor binding capacity. The neuropeptide Y (*NPY*) gene is a key neurotransmitter of the central nervous system that is abundant in the bovine cerebrum, and it has a high diversity of SNPs in the TATA and GC box that serve as regulators of bovine *NPY* gene expression (Bahar and Sweeney, 2008).

The aim of this study was to identify the differential SNPs located nearby TSSs between low fecundity (LF) and high fecundity (HF) groups of Laoshan dairy goat. As a cost-effective alternative of whole-genome resequencing of pools of individuals (Pool-seq), the goat genomic DNA of LF and HF group were homogenized and sequenced, and used the information obtained to represent these groups. Reads were mapped to the reference Yunnan black goat genome, SNPs were called and annotated, which were then counted within 1 kb scales of all annotated TSSs. After validating SNPs using Sanger sequencing, some unique HF group SNPs located nearby TSSs were discovered might contribute to the high fecundity trait and probably participate in the regulation of gene expression.

2. Materials and methods

2.1. Sample collection

High-throughput sequencing data used in this study were collected previously (Lai et al., 2016), however, this study didn't involve the previous research. Sample collection took place as previously reported (Lai et al., 2016). Briefly, ear tissues were used for DNA purification (n = tissue from 34 individuals). Among the total 34 Laoshan dairy goats, 20 individuals with a litter size of one were defined as the LF group, 13 individuals with a litter size of 3 and an individual with a litter size of 4 were considered as the HF group.

Regarding SNP calling and annotation; briefly, DNA purification from each sample took place using a QIAmp DNA Mini Kit (QIAGEN, 51304, Hilden, Germany). Compared to high-coverage sequencing of individuals, pool-seq was cost-effective, and would provide a relatively low deviation of allele frequency and a moderate reliability. It was the optimal strategy to test the overall intraspecific population differentiation around TSS region (Schlötterer et al., 2014). After validating the purity and integrity of genomic DNA using a spectrophotometer (NanoDrop 2000, Thermo, USA) and gel electrophoresis, equimolar quantities of DNA (2 μg/sample) were pooled together.

2.2. Genome DNA resequencing and mapping

After end repair, tail adding, adaptor and PCR amplification, two sequencing libraries were established with insertion sequencing sizes of approximately 0.5 kb and an effective concentration > 2 nM. Sequencing library construction and sequencing conduction were performed by Novogene Bioinformatics Technology Co., Ltd. (Beijing,