Small Ruminant Research 128 (2015) 19–26

Contents lists available at ScienceDirect

Small Ruminant Research

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

Genome-wide analysis of copy number variations in Chinese sheep using array comparative genomic hybridization

Cheng-Lin Hou[a,b], Fan-Hua Meng[a], Wei Wang[a], Shen-Yuan Wang[a], Yan-Ping Xing[a], Jun-Wei Cao[a], Kai-Feng Wu[a], Chun-Xia Liu[a], Dong Zhang[a], Yan-Ru Zhang[a,*], Huan-Min Zhou[a,*]

[a] Key Laboratory of Biological Manufacturing, Inner Mongolia Agricultural University, Hohhot 010010, PR China
[b] Institute of Biological Engineering, Baotou Light Industry Vocational Technical College, Baotou 014030, PR China

ARTICLE INFO

Article history:
Received 20 May 2014
Received in revised form 19 April 2015
Accepted 24 April 2015
Available online 8 May 2015

Keywords:
Array comparative genomic hybridization
Copy number variation
Phenotypic variation

ABSTRACT

Copy number variations (CNVs) account for phenotypic variation and disease susceptibility in mammals. The characterization of these CNVs is important for understanding the molecular mechanisms and unraveling the complex genetic architecture of underlying diseases. In this study, we reported a genome-wide CNVs map in Chinese sheep breeds using a custom-designed Roche NimbleGen array comparative genome hybridization (aCGH) platform. We identified a total of 245 CNVs, which were located within 51 CNV Regions (CNVRs) in five different sheep breeds. Of these, 21 (41.18%) involved DNA losses, 23 (45.10%) involved DNA gains and 7 (13.73%) involved both. These CNVRs covered approximately 15.55 Mb of the sheep genome, and this coverage corresponded to 0.60% of the autosomal genome sequence. Moreover, we identified 1726 genes within or overlapping with these 51 CNVRs by functional enrichment analysis, which indicates that they are involved in antigen processing, the immune response, oxygen transport, and heme binding. The results elucidate the genetic variability in the sheep genome and offer genetic foundations for the construction of economically important phenotypes of sheep, such as sheep with high immunity and oxygen binding capability.

Abbreviations: aCGH, array comparative genome hybridization; ASIP, agouti signaling protein; CNV, copy number variation; CNVR, CNV region; Ct, crossing thresholds; DAVID, the database for annotation visualization and integrated discovery; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; LOESS, locally weighted polynomial regression; MHC, major histocompatibility complex; OMIA, online Mendelian inheritance in animals; OR, olfactory receptors; PMEL, premelanosome protein; q-PCR, quantitative real time polymerase chain reaction; QTL, quantitative trait locus; SNP, single nuclotide polymorphism.

* Corresponding authors.
E-mail addresses: houcl@mail.ustc.edu.cn (C.-L. Hou), fhmeng2013@126.com (F.-H. Meng), wangwei13233@163.com (W. Wang), wsy1688521@126.com (S.-Y. Wang), xyping8315@163.com (Y.-P. Xing), caojunwei1970@163.com (J.-W. Cao), wukaifeng@126.com (K.-F. Wu), lcx.imau@163.com (C.-X. Liu), zhangdongsydw@163.com (D. Zhang), yanru1964@hotmail.com (Y.-R. Zhang), huanminzhou@263.net (H.-M. Zhou).

http://dx.doi.org/10.1016/j.smallrumres.2015.04.014
0921-4488/

1. Introduction

Copy number variation (CNV) is a type of genomic structural variation that results from non-allelic homologous recombination or non-homologous end-joining in different organisms (Lupski and Stankiewicz, 2005; Kim et al., 2008; Hastings et al., 2009) and is defined as a DNA segment ranging from 50 bases up to several megabases (Mb) compared to a reference genome (Feuk et al., 2006). CNVs arise from DNA deletion, duplication, insertion, and rearrangement. Because most CNVs contain gene-coding regions and regulatory elements, they may play an important role in gene expression and regulation. Thus, CNVs are regarded as the main contributors to phenotypic diversity and evolutionary adaptation in animals. Many CNVs represent benign polymorphic variants, and several CNVs are associated with Mendelian and complex genetic disorders in humans (Zhang et al., 2009; Conrad et al., 2010; Stankiewicz and Lupski, 2010). Single-nucleotide polymorphism (SNPs) are thought to be the major source of genetic variation between individuals (International HapMap Consortium, 2005), but copy number variation has now been confirmed to have a higher mutation rate than SNPs (Redon et al., 2006; Lupski, 2007; Zogopoulos et al., 2007; Perry et al., 2008; Shaikh et al., 2009). Moreover, current research shows that CNVs affect approximately 30% of the human genome (Database of Genomic Variants, 2010). To date, genome-wide investigations of CNVs have been reported in humans (International HapMap Consortium, 2005; Feuk et al., 2006; Redon et al., 2006; De Smith et al., 2007; Wong et al., 2007; Zogopoulos et al., 2007; Kim et al., 2008; Perry et al., 2008; Zhang et al., 2009; Conrad et al., 2010; Stankiewicz and Lupski, 2010), mice (Adams et al., 2005; Egan et al., 2007; Graubert et al., 2007), rat (Guryev et al., 2008), dogs (Chen et al., 2009; Nicholas et al., 2009; Berglund et al., 2012), pigs (Fadista et al., 2008), cattle (Liu et al., 2008; Matukumalli et al., 2009; Fadista et al., 2010; Seroussi et al., 2010; Bickhart et al., 2012), chickens (Wang et al., 2010), goats (Fontanesi et al., 2010), sheep (Fontanesi et al., 2011; Liu et al., 2013), and horses (Doan et al., 2012), and these studies indicated that the genetic variability is widespread in vertebrates and other inferior species (Cheeseman et al., 2009; Maydan et al., 2010).

Currently, several methods are available to identify CNVs. The commonly used methods are fluorescence in situ hybridization (FISH) and SNP array. FISH suffers from several disadvantages, such low resolution (5–10 Mbp) and labor-intensity, while SNP arrays yield a signal-to-noise ratio array-based comparative genome hybridization (aCGH) has emerged as a powerful tool for detecting gene copy number variants with increased mapping resolution. Because the aCGH technique is based on the co-hybridization of sample and control genomic DNAs, it can minimize irrelevant noises. The first investigation of the sheep genome via aCGH identified 135 CNV regions using a cross-species cattle-sheep aCGH platform. However, the low homology between cattle probes and sheep DNA might reduce the number of detectable CNVRs and further restrict the employment of this platform to scan genetic variations in the sheep genome (Clop et al., 2012). To overcome these problems, we first report the customized Roche NimbleGen aCGH platform containing sheep-specific probes, which can efficiently and exactly obtain the detailed CNV map in the sheep genome. This platform can scan the CNVs with high resolution and accuracy, as well as provide an in-depth understanding of the productive traits construction and evolutionary adaptation of sheep.

2. Materials and methods

2.1. Sample preparation

All study protocols for the collection of the tissue samples from experimental individuals were reviewed and approved by the Agricultural Hall of China Inner Mongolia.

Five sheep were used in this study: two Mongolian sheep (one from Inner Mongolia, China, and another from the Republic of Mongolia), one Chinese Kazakh sheep from Xinjiang, China, one Chinese Tibetan sheep from Qinghai, China, and one Chinese Hu sheep from Jiangsu, China. All animals were females. The genomic DNA was extracted from the blood using the Axygen Genomic DNA Extraction Kit and purified with the Wizard® Genomic DNA Purification Kit according to the manufacturer's instructions. The blood was collected from the jugular vein of each sheep after being injected with tranquillizers. The DNA concentrations were measured using a NanoDrop 2000 (Thermo Fisher Scientific).

2.2. Array CGH

We analyzed the CNVs in the sheep genome with an aCGH experiment using the Roche NimbleGen platform based on custom tiling arrays, which is designed for the Sheep (*Ovis aries*) genome; Ovis_aries_1.0. Array contained ~1.4 M probes to provide an evenly distributed coverage with an average interval of 1961 bp for the sheep genome (Additional file 1). The genomic DNA samples were fragmented by sonication. The reference DNA sample of the Dorper sheep was labeled with a fluorescent dye, Cy5, and co-hybridized with the other test DNA samples labeled with Cy3.

The arrays were scanned using a MS200 scanner (NimbleGen) with a 2 μM resolution, and the fluorescence intensity data were obtained with the NimbleScan 2.6 software. For each spot on the array, the $\log_2$-ratios of the Cy3-labeled test sample versus the Cy5 labeled reference sample were computed. Normalization was performed using the q-spline method (Workman et al., 2002), followed by segmentation using the CNV calling algorithm segMNT (Molla, 2007). The segments with a |mean $\log_2$-ratio| ≥ 0.25 and at least 5 consecutive probes were retained and considered CNVs.

2.3. Enrichment analysis

To determine the functional enrichment of CNV genes, the DAVID functional annotation tool (http://david.abcc.ncifcrf.gov/) was used to perform GO classification and the KEGG pathway (Huang et al., 2009). The *P*-value used in the GO and pathway analysis was calculated with a hypergeometric distribution probability formula. The *P*-value reflects the importance of the GO or pathway in the experimental results. Functional annotation terms from the ontologies of "cellular component", "molecular function" and "biological processes" were recorded. Because a small number of genes in the sheep genome have not been annotated, we converted the sheep transcript IDs to orthologous human and mouse gene IDs.

The 790 sheep QTL data were downloaded from the QTL database (http://www.animalgenome). Because all sheep QTLs were only marked on the genetic map, these data needed to be transferred on the physical map. The complete genetic map of all chromosomes of sheep has already been drawn (Maddox et al., 2001). According to the length of the genetic map and physical map corresponding to each chromosome, the conversion proportion of the genetic map and physical map of each chromosome can be obtained. The physical position of each QTL on the chromosome can be converted based on this map. The CNVRs were considered to be overlapping sheep QTLs if they were within 2 Mb of sheep QTLs (Fadista et al., 2010).

2.4. Validation of CNVs genes by qPCR

The CNV genes were validated with real-time quantitative PCR using the Roche LightCycler® 480 Detection System, as previously reported (Fadista et al., 2010). The primers were designed using the Primer Premier 5 software, and a total of 6 pairs of primers plus one pair of control primers for the GAPDH gene were synthesized (Additional file 2). All PCR primers were designed based on the appropriate reference sequence in NCBI. All qPCR reactions were performed using the SYBR Green method. Triplicate wells of reactions (20 μl) contained 10 μl SYBR Green Real-time PCR Master Mix, 20 ng DNA, 0.4 μl of each 10 μM primer and dH_2O. The thermal cycling conditions consisted of 1 cycle at 95 °C for 5 min, followed by 45 cycles at 95 °C for 60 s, 60 °C for 40 s, and 72 °C for 60 s, with fluorescence acquisition at 72 °C in single mode. All PCRs were performed in 96-well clear reaction plates (Roche Applied Science). The resultant cycle thresholds (Ct) were analyzed using the $-\Delta\Delta$Ct method.

3. Results

3.1. CNV discovery

The following sheep breeds were analyzed: Mongolian sheep, Tibetan sheep, Kazakh sheep and Hu sheep. The Mongolian sheep, Tibetan sheep, and Kazakh sheep are the oldest sheep species in China and distributed in the Mongolian plateau, Qinghai-Tibetan Plateau, and Pamir Plateau, where most local Chinese sheep breeds originated. Hu