Genomics 110 (2018) 143–148

Contents lists available at ScienceDirect

Genomics

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

Diversity of copy number variation in a worldwide population of sheep

Liu Yang[a,b], Lingyang Xu[b,c], Yang Zhou[c,d], Mei Liu[c,d], Lei Wang[e], James W. Kijas[f], Hongping Zhang[a], Li Li[a,*], George E. Liu[c,**]

[a] Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu 611130, China
[b] Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China
[c] Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705, USA
[d] College of Animal Science and Technology, Northwest A & F University, Shaanxi Key Laboratory of Agricultural Molecular Biology, Yangling, Shaanxi 712100, China
[e] CapitalBio Technology, Beijing 101111, China
[f] CSIRO Agriculture, Brisbane 4065, QLD, Australia

ARTICLE INFO

Keywords:
Copy number variation
Diversity
Evolution
Selection
Sheep

ABSTRACT

Copy number variation (CNV) represents a major source of genomic variation. We investigated the diversity of CNV distribution using SNP array data collected from a comprehensive collection of geographically dispersed sheep breeds. We identified 24,558 putative CNVs, which can be merged into 619 CNV regions, spanning 197 Mb of total length and corresponding to ~6.9% of the sheep genome. Our results reveal a population differentiation in CNV between different geographical areas, including Africa, America, Asia, Southwestern Asia, Central Europe, Northern Europe and Southwestern Europe. We observed clear distinctions in CNV prevalence between diverse groups, possibly reflecting the population history of different sheep breeds. We sought to determine the gene content of CNV, and found several important CNV-overlapping genes (*BTG3*, *PTGS1* and *PSPH*) which were involved in fetal muscle development, prostaglandin (PG) synthesis, and bone color. Our study generates a comprehensive CNV map, which may contribute to genome annotation in sheep.

1. Background

Sheep (*Ovis aries*) were domesticated in the Fertile Crescent approximately 9000 years ago [1,2], and have since become an important farm animal. The phenotypic and genetic variability that exists between sheep breeds is due to adaptation and artificial selection for animal production including meat, wool and milk [3]. Recently, high-throughput single nucleotide polymorphism (SNP) arrays have facilitated population genetics studies to improve our understanding of the genetic mechanism underlying complex economic and adaptive traits in domesticated animals [4,5]. Notably, the availability of dense SNP datasets has dramatically improved our understandings about genetic diversity, population history admixture, selection signatures, and other features in local and worldwide sheep populations [6–11].

On the other hand, our knowledge of genome function and evolution is still limited to SNP which, although widely used in genome research in farm animals, is just one type of common genomic variations. Copy number variation (CNV) is widely dispersed in mammalian genomes and also independently contributes to phenotypic diversity and disease susceptibility [12–14]. Recently, many studies have proposed that CNV can be used to study population genetics and show lineage-specific selection signatures in human [15–18], zebrafish [19], stickleback [20], and cattle [21]. However, for sheep, previous studies mostly focused on CNV discovery, which relied on using comparative genomic hybridization (CGH) and SNP arrays [22–25]. Thus, the features of sheep CNV at the population level are not well understood. Additionally, besides SNP, CNV is a separated type of important genomic variants. Due to CNV's less known linkage disequilibrium (LD) patterns, CNV-based population genomics results could offer additional new insights for functional and evolutionary studies in sheep. Therefore, the investigation of population genetics based on CNV could facilitate our understanding of evolution and selection aspects of the sheep genome.

In this study, we used the Sheep HapMap dataset to investigate CNV in the worldwide sheep populations. We utilized PennCNV to detect CNV using Illumina Ovine SNP50 genotyping data, and performed CNV-based population genetics and selection sweep analysis. With a large-scale sheep CNV map, our results provide comprehensive CNV information in sheep, and provide potential candidate variations for further exploration on the roles of CNV underlying important traits and

* Corresponding author.
** Corresponding author at: Animal Genomics and Improvement Laboratory, USDA-ARS, Building 306, Room 111, BARC-East, Beltsville, MD 20705, USA.
E-mail addresses: zhp@sicau.edu.cn (H. Zhang), lily@sicau.edu.cn (L. Li), George.Liu@ars.usda.gov (G.E. Liu).

http://dx.doi.org/10.1016/j.ygeno.2017.09.005
Received 31 March 2017; Received in revised form 24 August 2017; Accepted 11 September 2017
Available online 14 September 2017

Table 1
Summary of CNV and CNVR identified in 2111 samples derived from the sheep HapMap populations. Samples were divided into African (AFR), American (AME), Asian (ASI), Central European (CEU), Northern European (NEU), Southwestern Asian (SAS), and Southwestern European (SEU) groups. CNVR in this table represent non-redundant CNVR counts after merging both gain and loss CNVs identified within each group.

Groups	Sample size	CNV						CNVR			
		Count/ average[a]	Gain/ average	Loss/ average	Length (bp)	Average	SD[b]	Count[c]	Length (bp)	Average	SD[b]
African	117	1470/12.56	61/0.52	1409/12.04	216,493,684	147,274.61	143,971.56	212	53,261,080	251,231.51	285,165.74
American	213	2514/11.80	155/0.73	2359/11.08	346,200,939	137,709.20	115,374.43	211	49,087,317	232,641.31	250,620.72
Asian	206	2574/12.50	152/0.74	2422/11.76	372,547,204	144,734.73	118,862.63	233	55,001,353	236,057.31	253,896.20
Central European	242	2856/11.80	141/0.58	2715/11.22	386,881,910	135,462.85	91,545.60	170	32,014,638	188,321.40	136,481.48
Northern European	496	5763/11.62	245/0.49	5518/11.13	824,741,758	143,109.80	119,099.42	334	86,524,749	259,056.13	319,535.30
Southwestern Asian	136	1953/14.36	73/0.54	1880/13.82	332,389,280	170,194.20	161,479.07	255	70,938,258	278,189.25	265,099.73
Southwestern European	701	7428/10.60	526/0.75	6902/9.85	977,647,582	131,616.53	98,442.08	343	81,019,492	236,208.43	249,716.33

The numbers of Average are divided or normalized by sample counts except that the lengths are average lengths normalized by CNV and CNVR counts.
[a] At sample level, each sample has 12.56 (1470/117) CNVs, 0.52 (61/117) and 12.04 (1409/117) averagely for African.
[b] Standard deviation.
[c] These numbers are non-redundant CNVR counts.

evolutionary adaptation in sheep.

2. Method and materials

2.1. Selecting populations and animals

Data from sixty-eight breeds and 2254 animals are retrieved from the Sheep HapMap dataset which has been previously published [6]. This included 136 animals from African (AFR) breeds, 222 animals from American (AME) breeds, 211 animals from Asian (ASI) breeds, 198 animals from Southwestern Asian (SAS) breeds, 242 animals from Central European (CEU) breeds, 533 animals from Northern European (NEU) breeds, and 712 animals from Southwestern European (SEU) breeds. It is worth noting that for many breeds, individuals were sampled from more than one continent to explore within-breed genetic diversity. The number of animals per population and geographic origin of breed development were described in [6]. All chosen samples had a genotyping success rate of > 99%.

2.2. CNV detection with PennCNV

Signal intensity ratios (log R Ratio: LRR) and allelic frequencies (B allele frequency: BAF) were retrieved using Illumina GenomeStudio1.0 software for each SNP. The population frequency of B allele (PFB) file was calculated based on the BAF of each marker in each population. The sheep GC model file was generated using cal_gc_snp.pl with default settings (http://penncnv.openbioinformatics.org/en/latest/misc/faq/). CNVs were inferred within each individual using PennCNV software based on OARv1.0 [26]. PennCNV quality filters were subsequently applied as follows: We used high quality samples with a standard deviation (SD) of LRR < 0.35 and with the parameter set: BAF drift as 0.01 and waviness factor value between − 0.05 and 0.05, respectively. Appropriate LRR adjustments based on the GC model were incorporated in PennCNV. In addition, we used the program argument: the “lastchr 26” in the “detect” argument for specific CNVs. CNV regions (CNVR) were determined by aggregating overlapping CNVs identified in different animals, as reported previously [27]. For construction of the CNVR map, we classified the status of these CNVR into three categories, ‘Loss’ (CNVR containing deletion), ‘Gain’ (CNVR containing duplication) and ‘Both’ (CNVR containing both deletion and duplication).

2.3. Gene annotation and PANTHER analysis

We retrieved and annotated RefSeq gene overlapping CNV regions, using OARv1.0 from the UCSC Genome Browser. We performed gene ontology (GO) enrichment analysis using PANTHER with a genome-wide gene list based on the cattle (*Bos taurus*) as sheep's list was not present. We only considered terms with gene count > 5 and adjusted *p*-value < 0.05 (Bonferroni correction for multiple testing).

2.4. Population differential analysis

To explore lineage-specific CNVs, we first divided the Sheep HapMap samples into seven groups according to the geography distribution. They included African (AFR), American (AME), Asian (ASI), Southwestern Asian (SAS), Central European (CEU), Northern European (NEU), and Southwestern European (SEU) groups. We then constructed a comparative CNVR map across these seven groups. The frequency of a CNV within each CNVR was measured and utilized as the CNV characteristics for comparison among seven groups. To explore the potential differences involved with selection pressure for CNVs, we estimated the CNV frequency per group and the variance across all seven groups. Based on the frequency across seven groups, Euclidian distances were calculated. Using Ward's method as the linkage criteria, hierarchical cluster analysis was performed using 19 or 39 CNVR at top 5% or top 10% of the variances of frequency, respectively.

3. Results

3.1. Identification of copy number variation in worldwide sheep populations

A total of 52,094 autosomes markers were selected from the OvineSNP50 assay for CNV analysis. Using PennCNV, we performed large-scale CNV explorations in 2254 sheep derived from the Sheep HapMap Project dataset. After quality filtering, we identified 24,588 CNVs across 2111 individuals from 68 breeds with total length 3457 Mb and average length for each individual was 1.6 Mb. Among them, we observed 1470 CNVs in AFR, 2514 in AME, 2574 in ASI, 1953 in SAS, 2856 in CEU, 5763 in NEU, 7428 in SEU groups (see Table 1 for summary statistics of CNV and CNVR).

To explore the geographic pattern of CNV across the seven groups, we constructed violin plots using the CNV length and the number of SNP within CNV events. This identified patterns relating to CNV length and SNP count across groups (Fig. 1). CNV length generally showed slight difference, while the Central European (CEU) group was notably less variable than the Northern European (NEU) group (Fig. 1A) and the total numbers of SNP covered by each CNV per individual shows similar patterns to those observed for CNV lengths (Fig. 1B).

To compare the frequency of CNV across different groups, we first merged CNVs for each individual into non-redundant CNVR within

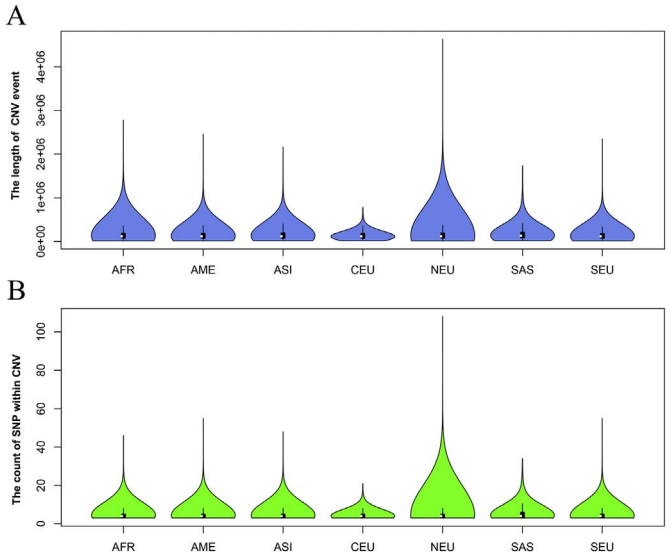

Fig. 1. A. Violin plots showing distribution of the total CNV length per (bp) individual. B. Violin plots showing distribution of the number of SNP covered by CNVs per individual. 24,588 individual CNVs (corresponding to 619 CNVR) were used for these length and count distribution plots. Violin plots combine the density differences depicted in the histograms and the median differences depicted in the box plots into one plot.

groups. This identified 212, 211, 233, 170, 334, 255, 343 CNVR in AFR, AME, ASI, CEU, NEU, SAS and SEU groups, respectively, with the responding CNVR length of 53, 49, 55, 32, 87, 71, 81 Mb (see Table 1 for summary statistics of CNVR).

We observed a large proportion of CNVR shared by multiple groups, which were considered as common variants. However, we also identified several lineage-specific CNVR showing different patterns of distribution among seven sheep groups. Among 619 CNVR, we observed 389 CNVR contained more than one CNV events. These 389 CNVR were subsequently used for CNV differential analysis across seven groups. Notably, we found a varying distribution of CNV counts across sheep autosomes (Fig. 2). Among them, a large amount of CNVs were detected on OAR1, 2, 7, 9 and 10, while there are only a few CNVs on OAR4, 11, 12, 17, 18, 22, 23 and 25.

3.2. Overlap analysis

In order to validate the set of CNVR detected, we compared their genome coordinates to recently published catalogs of ovine CNVR detected in three comparable studies [23,25,28], either separately or in combination (Table 2). Genomic coordinates of these studies were converted between OARv1.0 and OARv3.0 using liftOver [https://genome.ucsc.edu/cgi-bin/hgLiftOver]. The liftOver was carried out at a relaxed threshold (minimum ratio of bases that must remapped ≥ 50%). For example, within the OARv1.0 assembly, we found approximately 46.97 Mb (23.85%) of our CNVR were also covered by a merged CNVR dataset of those three studies. However, on the OARv3.1 assembly, we only observed that about 7.51 Mb (31.62%) were overlapped. This low overlapping length was likely because the liftOver conversion between assemblies was inefficient for DNA segments within copy number variable regions even though liftOver was carried out at a relaxed threshold (minimum ratio of bases that must remapped ≥ 50%). It was also related to the fact that these studies used largely different animals and/or different detection platforms. Although there were two previous studies which explored CNV in sheep, the comparison efforts were not straight forward thus not pursued. The first study [22] used custom tiling array including ~385,000 oligonucleotide probes designed based on the cattle Btau_4.0 assembly. While in the other recent study [24], CNV regions were identified on the UMD3_OA genome build, which was an in-house comparative sheep genome assembly, built using the cattle UMD3.1 assembly to facilitate the design of CGH arrays [24].

3.3. Lineage-specific CNV regions

As described above, to explore the lineage-specific CNVR, we first divided HapMap samples into seven groups according to geographic locations, including Africa, America, Asia, Southwestern Asia, Central Europe, Northern Europe, and Southwestern Europe. We found a total of 619 CNVR with 196 Mb across seven groups and constructed a plot comparing their distribution and frequencies (Fig. 3). Among them, the top five CNVR with high frequency were identified at OAR7 (2 CNVR), OAR10 (2 CNVR), OAR1 (1 CNVR) with their frequencies were 66.24% (chr7:66670941–67222553), 56.68% (chr7:97378846–97650032), 51.37% (chr10:56425004–56572554), 44.27% (chr1:146750090–151381176) and 43.89% (chr10:50801119–51406054), respectively

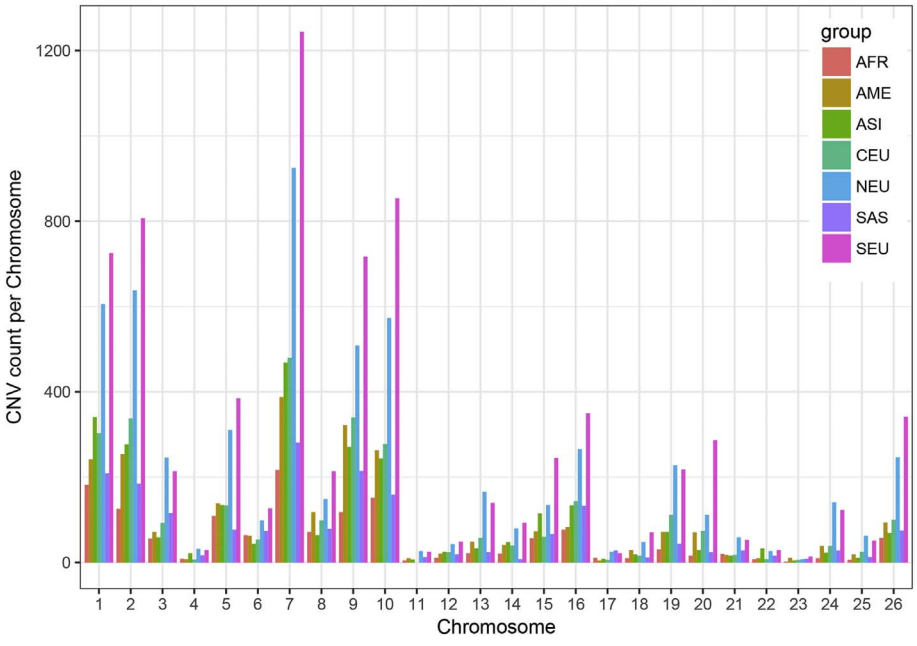

Fig. 2. CNV counts for 26 autosomes across seven groups, including African (AFR), American (AME), Asian (ASI), Central European (CEU), Northern European (NEU), Southwestern Asian (SAS) and Southwestern European (SEU) groups.

Table 2
Comparison with previous sheep CNV studies. We compared our results to recently published catalogs of ovine CNVR detected in three comparable studies, either separately or in combination.

Our study reference	Previous study (reference)	Length (bp)	Current study	Length (bp)	Overlap (bp)	Overlap%	New discovery (bp)
OARv1.0	Liu 2013 (OARv1.0)	60,349,794	Yang et al.	196,936,947	37,386,652	18.98%	159,550,295
OARv1.0	Ma 2015 (OARv1.0)	13,745,861	Yang et al.	196,936,947	12,544,606	6.37%	184,392,341
OARv3.1	Zhu 2016 (OARv3.1)	213,697,605	Yang et al.	23,747,109	4,897,002	20.62%	18,850,107
OARv1.0	Liu 2013, Ma 2015, Zhu 2016	144,060,442	Yang et al.	196,936,947	46,973,839	23.85%	149,963,108
OARv3.1	Liu 2013, Ma 2015, Zhu 2017	222,041,540	Yang et al.	23,747,109	7,508,188	31.62%	16,238,921

(Table S2). Additionally, we also observed one CNVR contained 935 CNV events at 146–151 Mb in OAR1 were enriched in gene *BTG3* and *MIRLET7C*. However, it is noted that this may represent a rare CNV event in the reference animal instead of a common CNV events in the HapMap samples.

To explore the regions which were enriched by CNV in sheep, we first filtered CNVR by only keeping the CNVR identified with 10 or more events in each of seven groups, we obtained 34,47, 44, 51, 75, 48, 72 common CNV in AFR, AME, ASI, CEU, NEU, SAS and SEU groups, respectively (Table S2). Also, we detected 25 hotspot regions on 13 chromosomes across all 7 groups. The top three common hotspots were detected at chr7:66670941–67222553, chr7:97378846–97650032 and chr10:56425004–56572554 and their corresponding total CNV counts were 1399, 1197 and 1085, respectively.

To understand the regions involved in lineage-specific CNV, we investigated and compared CNVR in all groups. In order to measure the CNV diversity, we first estimated the CNV densities (CNVR length in each breed normalized by the sheep genome OARv1.0 size, which is 2,849,359 kb) for each sheep group. We found varying CNV densities across seven groups, for instance, the NEU group had the highest CNV density of 3.04%, while the CEU group had a low CNV density of 1.12%. Also, to explore the potential differences involved with selection pressure for CNVs, we estimated the CNV count per individual and its variance for each CNVR across all seven groups (Table S2). Our analysis revealed the distinguishing pattern of CNV frequency across seven groups. We observed the average frequency for CNVR was almost approximately 0.02 for seven groups, while largest frequencies for CNVR were 0.71 (Line 294 highlighted in red), 0.61, 0.73, 0.79, 0.69, 0.64, and 0.66 (Line 290 highlighted in red) for AFR, AME, ASI, CEU, NEU, SAS and SEU groups, respectively.

3.4. CNV differentiation analysis

We next explored the evolution and selection characteristics of copy number variable genes. Based on the CNV count per individual across seven groups, we performed a cluster heatmap analysis using the top

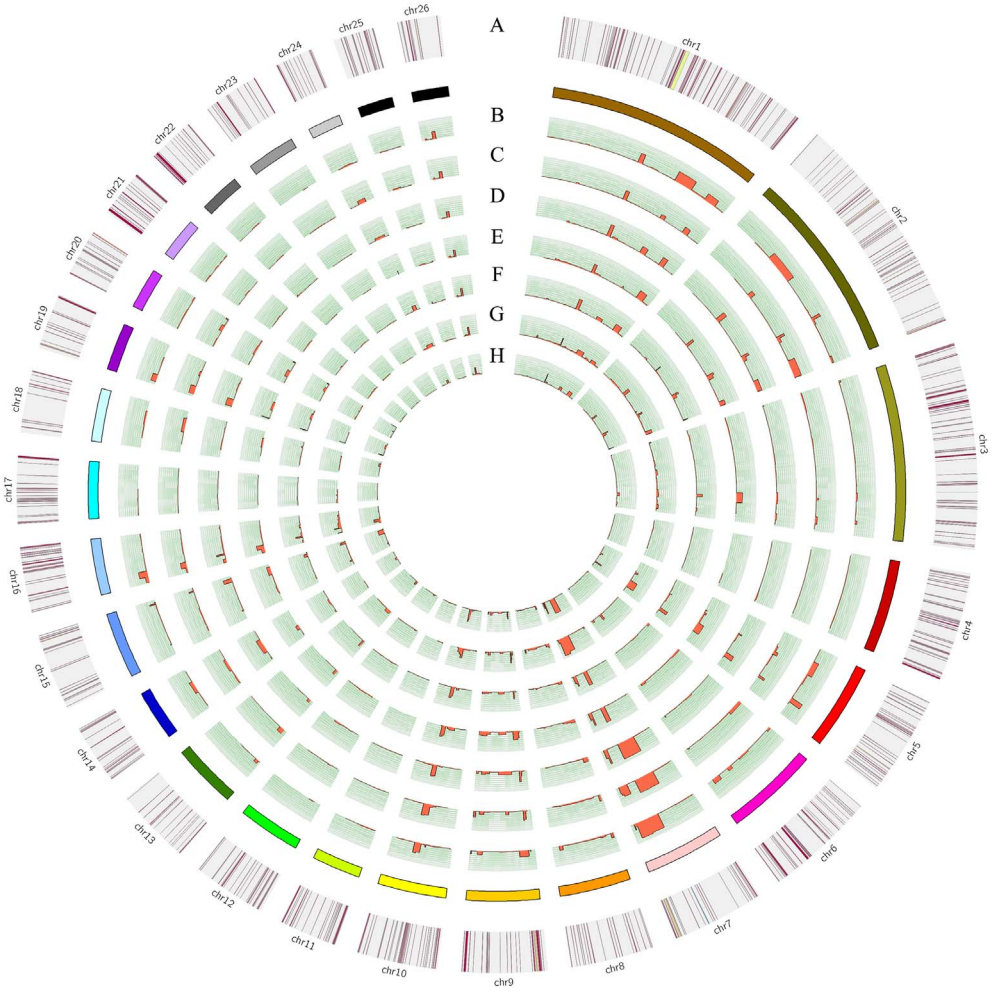

Fig. 3. Genomic distribution and frequencies of CNV regions on autosomes in seven groups from the Sheep HapMap dataset. The tracks from outside to inside are: A. CNVR shared across all seven groups; B–H. The frequency of CNVR was plotted for the African (AFR), American (AME), Asian (ASI), Central European (CEU), Northern European (NEU), Southwestern Asian (SAS) and Southwestern European (SEU) groups, respectively.

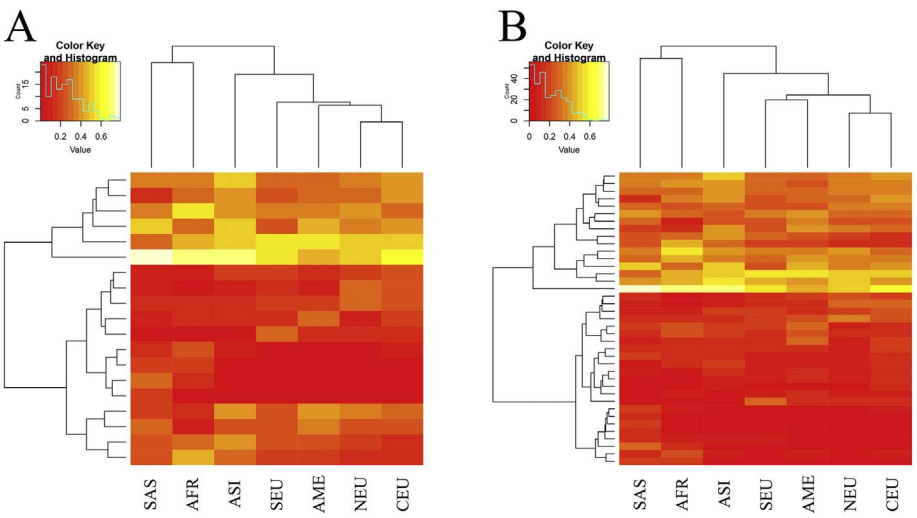

Fig. 4. Heatmap analysis based on hierarchical cluster using top 5% (A) or top 10% (B) of CNVR based on their variances in seven diverse sheep groups, including the African (AFR), American (AME), Asian (ASI), Central European (CEU), Northern European (NEU), Southwestern Asian (SAS), and Southwestern European (SEU) groups. Heatmap analysis based on hierarchical cluster using top 5% or top 10% of CNVR based on their variances. The value in the legend of the heatmap is CNV count per individual (CNV total count/sample size per population in each CNVR).

5% or top 10% of CNVR considering their variances (only CNVR contain at least two CNV events were considered here), and we observed the consistent patterns based on CNV count. The clustering based on frequency of CNV events clearly arranged groups according to the population of geographic origin. As expected, AFR and SAS sheep were arranged separately from European groups and American group (Fig. 4). Within Europe, the Northern European group clustered together with the Central European group, while the Southwestern European group was grouped together with the American group. The finding is consistent with the known breed histories as previously reported [6]. However, this clustering result was not superior to the similar results generated based on SNP [6]. This is partially because that SNP's location and status are well defined with high polymorphism, while CNV's true location and status are varying and often are not well defined.

3.5. Copy number variable genes

To understand the potential role that CNV's may play on phenotypic variation, we sought to catalog the gene content spanning CNVR. Overlap analysis using the gene models present in OARv1.0, identified 38 genes that overlapped with a total of 619 CNVR (Table S1), while 31 of them overlapped 389 CNVR with at least 2 CNV events. Based on our PANTHER analysis, the enriched GO terms included molecular function terms (small molecule binding, protein binding); biological process terms (regulation of secretion, regulation of transport, regulation of localization); and cellular component terms (extracellular space, extracellular region) (Table S1). Notably, we found three genes (*BTG3*, *PTGS1* and *PSPH*) imbedded with CNVR displaying differential changes of frequency with large estimated variances ($> 1.0E-3$).

4. Discussion

Our study is first attempt to perform comprehensive study of the CNV population genetic properties for worldwide sheep population across different geographical areas. Using the large-scale data from the Sheep HapMap Project, we identified 619 CNVR with 197 Mb, which occupied around 6.9% of the sheep genome. Compared with previous sheep CNV studies, this study explored several new breeds for their CNV properties for the first time. These newly discovered CNV in specific sheep breeds can be further explored to understand their potential functional and evolutionary features in the sheep genome. The large sheep data offer a valuable platform to explore the population genetic characteristics and evolution selection of CNV. Our results indicated that segregating CNV, just like SNP, express a certain degree of diversity across all breeds. Compared to published studies in other species for human, mouse, and cattle, our results also support the CNV were involved in the differential selective pressure during sheep domestication which are probably related with functional and evolutionary aspects of sheep genome.

Although large number samples have been employed in present study, we only detect a small count of CNV-overlapping genes in sheep genome. The most possible reasons could be related to the design and density of SNP array and/or status of the sheep reference genome, which is incomplete and not well annotated. For example, OvisSNP50 array likely was designed to avoid the regions involved with CNV. Therefore, the detected CNVR in the current study are probably underestimated. It should also be noted that the CNVR identified here are also likely to contain a false positive rate, however this hasn't been explicitly measured. Additionally, CNV discovery using different detection methods in various breeds make it less straight forward to directly compare our study with previous results. Furthermore, previous CNV studies in cattle have suggested using Bovine HD SNPs may provide a comprehensive result of CNVs at a higher resolution and sensitivity, and these analysis results may better facilitate the exploration of population analysis than low density SNPs array [21,32]. Thus, the density of SNPs array is one of important factor for CNV discovery and the subsequently CNV-based population genetic analyses.

The clustering based on matrices estimated from CNV counts clearly arranged groups according to the geographic origin (Fig. 4). Although our study reveals the cluster analysis is consistent with previously reported breed histories [6], the CNV-based population study is still a challenge task due to the difficulty of genotyping them. Most CNV calling algorithms were developed to detect CNV sample by sample, which often led to identified CNV with various coordinates (start and end positions), making it difficult to explore CNV at population levels [33]. Thus, we believe that high-throughput sequencing and new CNV calling algorithms will be needed to further improve the accurate and comprehensive discovery of CNV in animal genomes. CNV-based population genetics studies will be improved with precise CNV mapping *via* high-throughput sequencing and new CNV calling methods, which balance the tradeoffs between individual CNV calling and common CNV calling at population level [34,35].

In the last few years, the discovery of CNVs in farm animal has added more valuable genetics resources. For instance, the *BTG3* gene, named *BTG* Anti-Proliferation Factor 3, belong to a member of the *BTG/Tob* family, which has structurally related proteins that appear to have antiproliferative properties. This encoded protein might play a role in neurogenesis in the central nervous system. Also a previous study found that the gene was related to fetal muscle development [29]. *PTGS1* (Prostaglandin G/H Synthase 1), encoding similar enzymes that catalyze the conversion of arachinodate to prostaglandin. The encoded

protein regulates angiogenesis in endothelial cells, and is inhibited by nonsteroidal anti-inflammatory drugs such as aspirin. *PTGS1* gene has been reported to be associated with prostaglandin (PG) synthetic and finally affected the prolificacy in animals [30]. In addition, we observed one gene named *PSPH*, Phosphoserine Phosphatase, which belongs to a subfamily of the phosphotransferases, and the encoded enzyme is responsible for the third and last step in L-serine formation. Also molecular cloning and characteristics of the *PSPH* showed this gene has been found to be involved in the Black-Boned sheep [31]. This study is one of the first comprehensive CNV analyses with thousand of samples from diverse sheep breeds. The results of CNV identified represent the valuable molecular resources and could be further used for exploring the function and evolutionary aspect of sheep genome. Our better understandings of CNV will further help breeders to design the effective selection strategy and enhance genetics improvement.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2017.09.005.

Acknowledgements

We thank members of the Sheep HapMap Consortium for sharing their data. This work was supported in part by AFRI grant no. 2011-67015-30183 and 2013-67015-20951 from USDA NIFA. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The USDA is an equal opportunity provider and employer.

Author contributions

GEL and LX conceived and designed the experiments. LY, LX, YZ, ML, LW, HZ and LL performed *in silico* prediction and computational analyses. RB and JWK retrieved sample information and the SNP genotyping data. LX, GEL, and JWK wrote the paper.

References

[1] B. Chessa, F. Pereira, F. Arnaud, A. Amorim, F. Goyache, I. Mainland, R.R. Kao, J.M. Pemberton, D. Beraldi, M.J. Stear, et al., Revealing the history of sheep domestication using retrovirus integrations, Science 324 (5926) (2009) 532–536.
[2] M.A. Zeder, Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact, Proc. Natl. Acad. Sci. U. S. A. 105 (33) (2008) 11597–11604.
[3] J. Diamond, Evolution, consequences and future of plant and animal domestication, Nature 418 (6898) (2002) 700–707.
[4] M.E. Goddard, B.J. Hayes, Mapping genes for complex traits in domestic animals and their use in breeding programmes, Nat. Rev. Genet. 10 (6) (2009) 381–391.
[5] G.D. Wang, H.B. Xie, M.S. Peng, D. Irwin, Y.P. Zhang, Domestication genomics: evidence from animals, Annu. Rev. Anim. Biosci. 2 (2014) 65–84.
[6] J.W. Kijas, J.A. Lenstra, B. Hayes, S. Boitard, L.R. Porto Neto, M. San Cristobal, B. Servin, R. McCulloch, V. Whan, K. Gietzen, et al., Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection, PLoS Biol. 10 (2) (2012) e1001258.
[7] C. Wei, H. Wang, G. Liu, M. Wu, J. Cao, Z. Liu, R. Liu, F. Zhao, L. Zhang, J. Lu, et al., Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds, BMC Genomics 16 (2015) 194.
[8] J.W. Kijas, D. Townley, B.P. Dalrymple, M.P. Heaton, J.F. Maddox, A. McGrath, P. Wilson, R.G. Ingersoll, R. McCulloch, S. McWilliam, et al., A genome wide survey of SNP variation reveals the genetic structure of sheep breeds, PLoS One 4 (3) (2009) e4668.
[9] H. Wang, L. Zhang, J. Cao, M. Wu, X. Ma, Z. Liu, R. Liu, F. Zhao, C. Wei, L. Du, Genome-wide specific selection in three domestic sheep breeds, PLoS One 10 (6) (2015) e0128688.
[10] K.M. McRae, J.C. McEwan, K.G. Dodds, N.J. Gemmell, Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes, BMC Genomics 15 (2014) 637.
[11] M.I. Fariello, B. Servin, G. Tosser-Klopp, R. Rupp, C. Moreno, M. San Cristobal, S. Boitard, Selection signatures in worldwide sheep populations, PLoS One 9 (8) (2014) e103813.
[12] L. Feuk, A.R. Carson, S.W. Scherer, Structural variation in the human genome, Nat. Rev. Genet. 7 (2) (2006) 85–97.
[13] D.F. Conrad, D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T.D. Andrews, C. Barnes, P. Campbell, et al., Origins and functional impact of copy number variation in the human genome, Nature 464 (7289) (2010) 704–712.
[14] F. Zhang, W. Gu, M.E. Hurles, J.R. Lupski, Copy number variation in human health, disease, and evolution, Annu. Rev. Genomics Hum. Genet. 10 (2009) 451–481.
[15] C.D. Campbell, N. Sampas, A. Tsalenko, P.H. Sudmant, J.M. Kidd, M. Malig, T.H. Vu, L. Vives, P. Tsang, L. Bruhn, et al., Population-genetic properties of differentiated human copy-number polymorphisms, Am. J. Hum. Genet. 88 (3) (2011) 317–332.
[16] M. Kato, T. Kawaguchi, S. Ishikawa, T. Umeda, R. Nakamichi, M.H. Shapero, K.W. Jones, Y. Nakamura, H. Aburatani, T. Tsunoda, Population-genetic nature of copy number variations in the human genome, Hum. Mol. Genet. 19 (5) (2010) 761–773.
[17] A. Narang, P. Jha, D. Kumar, R. Kutum, A.K. Mondal, D. Dash, M. Mukerji, Extensive copy number variations in admixed Indian population of African ancestry: potential involvement in adaptation, Genome Biol. Evol. 6 (12) (2014) 3171–3181.
[18] M. Zarrei, J.R. MacDonald, D. Merico, S.W. Scherer, A copy number variation map of the human genome, Nat. Rev. Genet. 16 (3) (2015) 172–183.
[19] K.H. Brown, K.P. Dobrinski, A.S. Lee, O. Gokcumen, R.E. Mills, X. Shi, W.W.S. Chong, J.Y.H. Chen, P. Yoo, S. David, Extensive genetic diversity and substructuring among zebrafish strains revealed through copy number variant analysis, Proc. Natl. Acad. Sci. 109 (2) (2012) 529–534.
[20] F.J. Chain, P.G. Feulner, M. Panchal, C. Eizaguirre, I.E. Samonte, M. Kalbe, T.L. Lenz, M. Stoll, E. Bornberg-Bauer, M. Milinski, et al., Extensive copy-number variation of young genes across stickleback populations, PLoS Genet 10 (12) (2014) e1004830.
[21] L. Xu, Y. Hou, D.M. Bickhart, Y. Zhou, H.A. Hay el, J. Song, T.S. Sonstegard, C.P. Van Tassell, Liu GE: Population-genetic properties of differentiated copy number variations in cattle, Sci. Rep. 6 (2016) 23161.
[22] L. Fontanesi, F. Beretti, P.L. Martelli, M. Colombo, S. Dall'olio, M. Occidente, B. Portolano, R. Casadio, D. Matassino, V. Russo, A first comparative map of copy number variations in the sheep genome, Genomics 97 (3) (2011) 158–165.
[23] J. Liu, L. Zhang, L. Xu, H. Ren, J. Lu, X. Zhang, S. Zhang, X. Zhou, C. Wei, F. Zhao, et al., Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array, BMC Genomics 14 (2013) 229.
[24] G.M. Jenkins, M.E. Goddard, M.A. Black, R. Brauning, B. Auvray, K.G. Dodds, J.W. Kijas, N. Cockett, J.C. McEwan, Copy number variants in the sheep genome detected using multiple approaches, BMC Genomics 17 (1) (2016) 441.
[25] C. Zhu, H. Fan, Z. Yuan, S. Hu, X. Ma, J. Xuan, H. Wang, L. Zhang, C. Wei, Q. Zhang, et al., Genome-wide detection of CNVs in Chinese indigenous sheep with different types of tails using ovine high-density 600K SNP arrays, Sci. Rep. 6 (2016) 27822.
[26] K. Wang, M. Li, D. Hadley, R. Liu, J. Glessner, S.F. Grant, H. Hakonarson, M. Bucan, PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data, Genome Res. 17 (11) (2007) 1665–1674.
[27] K. Lee, D.T. Nguyen, M. Choi, S.Y. Cha, J.H. Kim, H. Dadi, H.G. Seo, K. Seo, T. Chun, C. Park, Analysis of cattle olfactory subgenome: the first detail study on the characteristics of the complete olfactory receptor repertoire of a ruminant, BMC Genomics 14 (2013) 596.
[28] Y. Ma, Q. Zhang, Z. Lu, X. Zhao, Y. Zhang, Analysis of copy number variations by SNP50 BeadChip array in Chinese sheep, Genomics 106 (5) (2015) 295–300.
[29] Z. Feng, Z.L. Tang, K. Li, B. Liu, M. Yu, S.H. Zhao, Molecular characterization of the BTG2 and BTG3 genes in fetal muscle development of pigs, Gene 403 (1–2) (2007) 170–177.
[30] H. Zhang, S. Wang, M. Liu, A. Zhang, Z. Wu, Z. Zhang, J. Li, Differential gene expression in the endometrium on gestation day 12 provides insight into sow prolificacy, BMC Genomics 14 (2013) 45.
[31] Q. Li, D. Xi, J. Leng, X. Gou, H. Mao, W. Deng, Molecular cloning and characteristics of the PSPH, snrpa1 and TPM1 genes in black-boned sheep (*Ovis Aries*), Biotechnol. Biotechnol. Equip. 27 (6) (2013) 4231–4235.
[32] Y. Hou, D.M. Bickhart, M.L. Hvinden, C. Li, J. Song, D.A. Boichard, S. Fritz, A. Eggen, S. DeNise, G.R. Wiggans, et al., Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array, BMC Genomics 13 (2012) 376.
[33] L. Xu, Y. Hou, D. Bickhart, J. Song, G. Liu, Comparative analysis of CNV calling algorithms: literature survey and a case study using bovine high-density SNP data, Microarrays 2 (3) (2013) 171.
[34] P.H. Sudmant, S. Mallick, B.J. Nelson, F. Hormozdiari, N. Krumm, J. Huddleston, B.P. Coe, C. Baker, S. Nordenfelt, M. Bamshad, et al., Global diversity, population stratification, and selection of human copy-number variation, Science 349 (6253) (2015) aab3761.
[35] D.M. Bickhart, L. Xu, J.L. Hutchison, J.B. Cole, D.J. Null, S.G. Schroeder, J. Song, J.F. Garcia, T.S. Sonstegard, C.P. Van Tassell, et al., Diversity and population-genetic properties of copy number variations and multicopy genes in cattle, DNA Res. 23 (3) (2016) 253–262.