Copy number variations within and among diverse cattle breeds

M. Mielczarek¹,², M. Frąszczak¹, E. Nicolazzi³, J. L. Williams⁴ & J. Szyda¹,²

¹Wroclaw University of Environmental and Life Sciences, Department of Genetics, Biostatistic group; Kozuchowska 7, 51-631 Wroclaw, Poland
²National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland
³Council on Dairy Cattle Breeding (CDCB), 4201 Northview Dr, Bowie, MD 20716, the United States
⁴University of Adelaide, School of Animal and Veterinary Sciences, Roseworthy, SA 5371, Australia

Summary

In recent years, the number of studies involving Copy Number Variants in cattle has strongly increased. In the present study, we analysed inter-individual and inter-breed variation of CNVs identified based on whole genome sequences of 146 individuals representing 13 breeds. A highly significant variation among all individuals, as well as among individuals within each breed was observed both in the number of duplications () and in the number of deletions (). We also observed significant differences between breeds for duplications (P=0.01932) and deletions (P=0.01006). The same situation existed for CNV length. The inter-individual variation was significant for both duplications () and deletions () across all individuals and within each breed. Moreover, breed specific variants were identified within each breed, with the largest proportion of breed specific duplications (9.57%) attributable to Fleckvieh. The most frequent breed specific CNVs were mainly located in intragenic regions, however in Simmental, a deletion was reported in a coding sequence of a novel gene ENSBTAG00000000688 on chromosome 18. Brown Swiss, Norwegian Red and Simmental specific deletions were reported in KIT and MC1R genes responsible for different coat colours.

Keywords: breed diversity, cattle, copy number variation, next-generation sequencing

Introduction

Copy Number Variations (CNVs) are deletions or duplications ranging from fifty to millions of base pairs. Because of such considerable lengths, CNVs may potentially exhibit high impact on phenotypes (Bickhart and Liu 2014, Shin et al. 2014, Sasaki et al. 2016). Despite the fact that advances in the Next Generation Sequencing (NGS) provide valuable data sets allowing for identification of CNVs at the whole genome resolution, population-wise CNV studies in cattle are still uncommon. Moreover, the overlap between studies in detected CNVs is very low (Keel et al. 2017), indicating still poor detection accuracy. In this study, we aimed to assess inter-individual and inter-breed variability in CNV numbers and lengths, as well as to identify variants specific to only one breed using a large data set of 146 individuals representing 13 cattle breeds. The CNV identification accuracy was improved by applying
two detection strategies – one based on differences in read depth (Abyzov et al. 2011) and the other on paired-end approach (Ye et al. 2009).

**Material and methods**

**Material**

Whole genome DNA sequences generated with the Illumina HiSeq 2000 Next Generation Sequencing platform were available for 155 bulls representing Brown Swiss (48), Fleckvieh (31), Norwegian Red (26), Guernsey (20), Simmental (16), Parda de la Montaña (4), Pezzata Rossa Italiana (3), Avileña (2), Bruna Italiana (1), Albera (1), Rubia Gallega (1), Toro de Lidia (1) and Pirenaica (1) breed.

**CNV detection and annotation pipeline**

The bioinformatics pipeline comprised: (i) alignment to the UMD 3.1 reference genome using BWA (Li et al. 2009), (ii) detection of CNV using CNVnator (Abyzov et al. 2011) and Pindel (Ye et al. 2009), (iii) validation of CNVs and (iv) the functional annotation of breed specific variants using VEP (McLaren et al. 2010). The genome-average coverage per individual was calculated as where \( n \) denoted the total number of aligned reads, \( r \) - read length in bp (100 bp), and \( d \) - the length of the reference genome (2697.56 Mb). It was used as a criterion for including individuals for further analyses. CNVs outside the length range of 50 bp - 5 Mbp were discarded and a consensus dataset of validated variants was defined based on overlapping results of both programmes. In the next step, breed specific CNVs (i.e. shared by at least two individuals within a breed and absent in other breeds) were subjected to the functional annotation involving: (i) biological processes Gene Ontologies (GO), (ii) Sequence Ontologies (SO), (iii) overlapping QTL from the QTLdb (www.animalgenome.org/).

**Hypothesis testing**

The final set of validated CNVs was subjected to population-wide inferences. Inter-individual and the inter-breed variation of the number of variants was tested separately for duplications and deletions for the most numerous breeds (Brown Swiss, Fleckvieh, Guernsey, Norwegian Red, and Simmental) using the test. Furthermore, the Kurskal-Wallis test was used to test whether the distribution of CNV lengths is the same for all individuals. The same test was applied to check whether the variability of the length of deletions/duplications varies within and also between breeds.

**Results**

**CNV dataset**

After alignment, 9 individuals (7 Norwegian Red, 1 Fleckvieh and 1 Parda la Montaña) were excluded from further analysis due to low genome averaged coverage below 7. The number of variants among the remaining 146 individuals varied between 12 and 11,704 (1,343 ± 1,086) for duplications and between 0 and 3,960 (1,708 ± 700) for deletions. Also, variant lengths varied strongly from 200 bp to 4,992,800 bp (31,018 ± 169,307) for duplications and from 200 bp to 4,536,800 bp (10,836 ± 53,724) for deletions. The average length of
duplications was 13,154.67 (± 26,407.47) in Simmental, 13,905.71 (± 31,887.31) in Guernsey, 18,474.31 (± 87,787.53) in Brown Swiss, 25,468.29 (± 53,082.84) in Norwegian Red, and 76,931.9 (± 37,349.18) in Fleckvieh. The average length of deletions was 7,409.897 (± 43,026.6) in Guernsey, 10,023.01 (± 34,655.82) in Norwegian Red, 10,414.73 (± 29,306.74) in Simmental, 11,750.98 (± 53,277.15) in Brown Swiss, and 12,564.34 (± 79,185.56) in Fleckvieh.

**Inter-individual and inter-breed variation**

84.85% of all duplications and 77.22% of all deletions were observed in only one bull. The most frequent duplication was common among 117 bulls and the most common deletion was present in 140 bulls. A highly significant variation among all 146 individuals as well as within each breed was observed both in the number of duplications and in the number of deletions (\(P\)). Differences between breeds in the number of duplications (\(P=0.01932\)) and deletions (\(P=0.01006\)) were also significant. Inter-individual variation and within-breed variation of CNV length was significant for duplications and deletions (\(P\)). Therefore, the inter-individual variation within breeds was highly significant. Tests performed within each breed were always highly significant resulting in p-values. Significant differences between breeds in the length of duplications and deletions (\(P\)) were also reported.

**Breed specific CNVs**

The frequency of breed specific variants was lowest in Simmental (1.74% of deletions and 1.31% for deletions), while the most distinct breed in terms of duplications was Brown Swiss (5.00%) and in terms of deletions - Fleckvieh (9.57%). Approximately 1/3 of breed specific duplications and deletions involved genes sequences: 35.98% and 36.10% in Brown Swiss, 36.72% and 31.44% in Fleckvieh, 34.23% and 28.26% in Guernsey, 41.69% and 28.18% in Norwegian Red, 29.05% and 40.62% in Simmental.

In the case of duplications, the most of significantly overrepresented GO terms were found for Simmental and comprised (e.g. natural killer cell mediated cytotoxicity, immunoglobulin production, G-protein coupled receptor signalling pathway, GO terms related to biosynthetic process). Contrarily, no GO term was significantly enriched in Fleckvieh. Moreover, in Guernsey, Norwegian Red and Simmental, duplicated genes were a significantly overrepresented in the same GO term: “chemical stimulus involved in sensory perception of smell”. In the case of genic sequence deletions, the most of significantly overrepresented GO terms were identified only for Norwegian Red (e.g. related to natural killer cell mediated cytotoxicity, cellular response to organic substance, RNA processing).

Ten the most common breed specific duplications and eight deletions were further analysed more precisely. Seven duplications were located in intergenic regions, three belonging to Fleckvieh, Guernsey and Norwegian Red were located in introns or upstream gene regions. In the case of deletions, five were annotated to intergenic regions, two to introns or upstream gene regions and one overlapped with a coding sequence on BTA18 involving a fragment of a transcript of a novel gene ENSBTAG000000000688. This particular deletion was found in five Simmental bulls and overlaps with a deletion from the DGVa database reported by Bickhart et al. (2012).

The most common breed specific CNVs overlapping with QTL representing five phenotypic groups: reproduction, milk, production, exterior, meat and carcass as well as health. In the case of duplications, QTL falling into meat and carcass trait class were found in
all breeds, except Norwegian Red. For the latter breed only two QTL for traits such as calving index and length of productive life were duplicated. Only Fleckvieh and Simmental specific duplications overlapped with QTL related to milk yield. Interestingly, Simmental specific duplication fell into all phenotypic groups, but deletion overlapped only with body weight. QTL related to body weight were also found for breed specific deletions in all analysed breeds. QTL belonging to milk yield as well as meat and carcass classes were found in all breeds specific deletions except Simmental breed.

Discussion and conclusions

In the present study, a highly significant variation among all 146 individuals, as well as among individuals within each breed was observed both in the number of duplications and in the number of deletions. The most of duplications (84.85 %) and deletions (77.22 %) were represented by only one bull in the whole dataset. Similar proportion was also observed by Boussaha et al. (2015), where 61 % of all CNVs were characteristic only for one animal. The inter-individual variation was significant for both duplication and deletion lengths across all individuals and within each breed. Despite this, the significant inter-breed variation was also reported. Moreover, breed specific variants were identified within each breed, with the largest proportion of breed specific duplications (9.57%) attributable to Fleckvieh. The most frequent breed specific CNVs were mainly located in intragenic regions, however in Simmental, a deletion was reported in a coding sequence of a novel gene ENSBTAG00000000688 on chromosome 18. It is probably involved in transcriptional regulation which was documented for human (ZKSCAN4 gene, Li et al. 2007).

CNVs located in the QTLs of milk class and meat and carcass class were identified for all breeds, probably because all of them are dual purposes cattle. Significantly overrepresented GO terms related to chemical stimulus involved in sensory perception of smell in Guernsey, Norwegian Red and Simmental breeds were reported here. Olfactory receptors genes have been already reported to be duplicated within the bovine genome suggesting that they may be under strong selection for newly evolving functions (Qanbari et al. 2014). We have not found any CNV in QTL as well as any significantly enriched gene related to the coat colour. However, we noticed that the part of the KIT gene which explains a considerable proportion of the variation in patterned pigmentation (Hayes et al. 2010) was deleted in Brown Swiss and was present in four remaining breeds having a characteristic spotting phenotype. Moreover, we detected that MC1R receptor, whose loss of function mutations causes red coat colour in different cattle breeds (Qanbari et al. 2014), was partly deleted in Brown Swiss, Norwegian Red and Simmental individuals.

In this study, we showed a high complexity of a CNV landscape in Bos taurus genomes. Significant inter-individual variation, even within particular breeds was reported, however significant differences between breeds also exists.

Acknowledgements

The research was supported by Polish National Science Centre (2014/15/N/NZ9/03914), European Union's Research and Innovation funding programme (FP7) under the Gene2Farm grant (289592) and the Poznan Supercomputing and Networking Centre.

List of References

genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21: 974-984.


