CSIRO PUBLISHING

*Animal Production Science*
https://doi.org/10.1071/AN17603

# Genome-wide detection of copy-number variations in local cattle breeds

*Rosalia Di Gerlando*[A,B], *Maria Teresa Sardina*[A], *Marco Tolone*[A], *Anna Maria Sutera*[A], *Salvatore Mastrangelo*[A] *and Baldassare Portolano*[A]

[A]Dipartimento Scienze Agrarie, Alimentari e Forestali, Ed. 4 Ingr. G, Università degli Studi di Palermo, Viale delle Scienze, Palermo 90128, Italy.
[B]Corresponding author. Email: rosalia.digerlando@unipa.it

**Abstract.** The aim of the present study was to identify copy-number variations (CNVs) in Cinisara (CIN) and Modicana (MOD) cattle breeds on the basis of signal intensity (logR ratio) and B allele frequency of each marker, using Illumina's BovineSNP50K Genotyping BeadChip. The CNVs were detected with the PennCNV and SVS 8.7.0 software and were aggregated into CNV regions (CNVRs). PennCNV identified 487 CNVs in CIN that aggregated into 86 CNVRs, and 424 CNVs in MOD that aggregated into 81 CNVRs. SVS identified a total of 207 CNVs in CIN that aggregated into 39 CNVRs, and 181 CNVs in MOD that aggregated into 41 CNVRs. The CNVRs identified with the two softwares contained 29 common CNVRs in CIN and 17 common CNVRs in MOD. Only a small number of CNVRs identified in the present study have been identified elsewhere, probably because of the limitations of the array used. In total, 178 and 208 genes were found within the CNVRs of CIN and MOD respectively. Gene Ontology and KEGG pathway analyses showed that several of these genes are involved in milk production, reproduction and behaviour, the immune response, and resistance/susceptibility to infectious diseases. Our results have provided significant information for the construction of more-complete CNV maps of the bovine genome and offer an important resource for the investigation of genomic changes and traits of interest in the CIN and MOD cattle breeds. Our results will also be valuable for future studies and constitute a preliminary report of the CNV distribution resources in local cattle genomes.

**Additional keywords:** BovineSNP50K BeadChip, Cinisara, genomic variation, Modicana.

Received 5 September 2017, accepted 28 March 2018, published online 13 June 2018

## Introduction

Copy-number variations (CNVs) are classes of polymorphic genomic regions, including deletions, duplications and insertions, in DNA fragments ranging from at least 0.5 kb to several megabases that vary relative to a reference genome (Mills *et al*. 2011; Jiang *et al*. 2013). CNVs represent an important source of genetic variability, and provide genomic structural information that complements single nucleotide polymorphism (SNP) data. They are considered promising markers of some phenotypic variations, environmental adaptability and economically important traits or disease susceptibility in livestock species (Wang *et al*. 2007; Jiang *et al*. 2013; Xu *et al*. 2014). CNVs can be classified as deletions (losses) or duplications (gains) relative to the reference genome. There are several techniques to identify CNVs within a genome, including SNP-genotyping microarrays, next-generation sequencing and comparative genomic hybridisation arrays. Each of these techniques has its advantages and disadvantages. SNP arrays provide a measure of the intensity signal (log R ratio) and B allele frequency (BAF) for each sample. Multiple software have been developed to identify CNVs from SNP arrays: QuantiSNP (Colella *et al*. 2007), PennCNV (Wang *et al*. 2007), SVS (Golden Helix, Inc., Bozeman, MT, USA; www.goldenhelix.com, accessed 27 April 2016), cnvPartition (Illumina Inc., San Diego, CA, USA) and others. Comparative analyses of the algorithms used for CNV identification have been published (Winchester *et al*. 2009; Xu *et al*. 2013). As reported by several authors (Marenne *et al*. 2009; Yavaş *et al*. 2009; Seroussi *et al*. 2010; Ma *et al*. 2017), PennCNV is the most reliable and accurate algorithm in detecting CNVs from Illumina BeadChip data. PennCNV is a tool that incorporates multiple sources of information, such as the signal intensity at each SNP marker, the distance between neighbouring SNPs, and the BAF, and integrates a computational approach by fitting regression models of the GC content to avoid 'genomic waves'. Nevertheless, several authors (Winchester *et al*. 2009; Pinto *et al*. 2011) have recommended the use of at least two algorithms that differ in their performance and impact to identify CNVs.

Copy-number variations have been shown to be associated with complex traits in several species, including chimpanzee (Perry *et al*. 2008), rat (Guryev *et al*. 2008) and mouse (Adams *et al*. 2005), and in livestock species such as cattle (Liu *et al*. 2010; Hou *et al*. 2011), goat (Fontanesi *et al*. 2010), sheep (Fontanesi *et al*. 2011) and horse (Dupuis *et al*. 2013).

The coat colours of horse, pig and sheep are partly determined by CNVs (Clop *et al*. 2012), and milk production (Xu *et al*. 2014), female fertility failure (Kadri *et al*. 2014), osteoporosis (Meyers *et al*. 2010), abortions and stillbirth in cattle (Flisikowski *et al*. 2010) have been shown to be influenced by CNVs. CNVs have been identified in different cattle breeds, including African, Indicine and Taurine breeds (Matukumalli *et al*. 2009), and Bae *et al*. (2010) and Fadista *et al*. (2010) created two CNV maps of the bovine genome using SNP genotyping and CGH arrays. Currently, few studies of genome-wide CNVs have been reported in local cattle breeds, such as Cinisara (CIN) and Modicana (MOD).

These two breeds are adapted to the harshness of mountain areas because of their good grazing characteristics, have an excellent aptitude for dairy production, and are resistant to environmental conditions (Mastrangelo *et al*. 2014). However, there is currently no whole-genome CNV map for Sicilian cattle breeds. The aim of the present study was to identify CNVs in the CIN and MOD cattle breeds using the Illumina BovineSNP50K BeadChip v2 (Illumina Inc., San Diego, CA, USA), and to compare the results with previously reported CNVs from other breeds so as to expand the catalogue of CNV regions (CNVRs) in the bovine genome. We hypothesised that in response to the environmental conditions in which these breeds are reared, particularly CIN, they may be characterised by undiscovered CNVs. Therefore, the present study should provide information on genomic variations that have important implications for the development of conservation programs for these local cattle breeds and for future association studies between CNVs and phenotypes.

## Materials and methods

### Sampling and genotyping

In total, 142 individuals from 14 farms were sampled. The procedures involved in animal sample collection were according to the recommendations of European Union (EU) Directive 2010/63/EU. The samples were randomly collected from 71 CIN and 71 MOD individuals. The number of animals sampled per farm ranged from 8 to 10. Genomic DNA was extracted from the blood samples with a salting-out method (Miller *et al*. 1988). The sample DNA was quantified with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), diluted to a final concentration of 50 ng/μL (as required by the Illumina Infinium protocol), and stored at 4°C until use.

Genotyping was performed with the Illumina BovineSNP50K BeadChip v2 with 54609 SNPs. We excluded sexual (ChrX) and unknown (ChrUn) chromosomes from the CNV calls in our analysis because PennCNV assumes that two copies of each SNP occur in the normal copy-number state, which was not likely to be the case within the pseudoautosomal region (Sonstegard *et al*. 2001) or the segmental duplications on chrX. Furthermore, chrUn contains unassigned sequence contigs, so it was not included because of the lack of sequence and SNPs and the uncertainties in SNP mapping. Consequently, the number of SNPs included in the final analyses was reduced to 52 886. The genomic positions of the SNPs on the chromosomes were determined from the bovine UMD3.1 genome sequence assembly.

### CNV and CNVR detection

Two softwares, PennCNV and SVS 8.7.0, based on different algorithms were used for the detection of the CNVs. PennCNV incorporates multiple factors, including the logR ratio (LRR), BAF, the marker distance, and the population frequency of the B allele. The LRR and BAF values for each SNP were obtained with the GenomeStudio 2.0 software (Illumina Inc.). The population frequency of the B allele file was calculated with PennCNV on the basis of the BAF value for each marker in each individual. PennCNV integrates a computational approach by applying a regression model to the GC content to overcome genomic waves. The gcmodel file was generated by calculating the GC content in the 500-kb genomic region on both sides of each SNP and the genomic waves were adjusted using the *-gcmodel* option. Because the bovine genome has 29 autosomal chromosomes, we used an alternative program argument, the *-lastchr 29* in the *-detect* argument. CNVs were also detected using the HMM parameter file. The quality of the final dataset was assessed with the following criteria: a logR ratio standard deviation (LRR_s.d.) of <0.30, BAF drift of <0.01, and waviness factor of >0.05 or <−0.05 for each sample. After quality control, 10 outlier samples were excluded in total. To reduce the possible false CNV calls, we also considered only those CNVs that contained three or more consecutive SNPs.

The copy-number analysis module (CNAM) implemented in the SVS 8.7.0 software was also used for CNV identification. The following options in CNAM were chosen: univariate outlier removal, maximum number of 100 segments per 10 000 markers, minimum markers per Segment 3, 2000 permutations per pair with a *P*-value cutoff of 0.005. Individuals that had −0.05 > waviness factor > 0.05 were also excluded, as suggested by Diskin *et al*. (2008).

The CNVRs were determined by aggregating the overlapping CNVs identified across all samples within each breed (Redon *et al*. 2006). Overlapping was identified with the BEDTools software (Quinlan and Hall 2010). The CNVRs common to PennCNV and SVS were determined by intersecting the datasets and inferring the overlapping CNVRs using the approach described by Wain *et al*. (2009), which identifies CNVRs that fully overlap each other.

### Gene contents and functional annotation

The gene contents of the CNVRs were assessed with Cattle RefSeq in the Genome Data Viewer genome browser at the National Center for Biotechnology Information Database (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/?context = genome&acc = GCF_000003055.3, May 2016). The DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/summary.jsp, May 2016) for gene ontology (GO) analysis and the Kyoto Encyclopedia of Genes and Genomes (KEGG) Database (http://www.genome.jp/kegg/pathway.html, accessed May 2016) for pathway analysis were used. The following options were used for the GO analysis with DAVID: a high classification stringency and a false discovery rate correction. We also performed an enrichment analysis using the cattle quantitative trait locus (QTL) database (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index, accessed May 2016) to identify CNVRs that overlapped QTL regions (QTLRs). We filtered the QTLRs that were >5 Mb, and only those overlapping at least 50% of each CNVR were considered.

## Results and discussion

### CNV and CNVR detection

In the present study, we analysed the CNVs in two local cattle breeds by using two different algorithms.

Using PennCNV, a total of 487 CNVs was detected in the CIN breed, with an average number of 7.4 per sample and an average length and median size of 147.86 kb and 146.35 kb respectively. In the MOD breed, a total of 424 CNVs was detected, with an average number of 6.4 per sample and an average length and median size of 117.33 kb and 120.56 kb respectively (Fig. 1, File S1–Table S1.1 and File S2–Table S2.1, available as Supplementary material for this paper). By aggregating the overlapping CNVs, a total of 86 (in CIN) and 81 (in MOD) CNVRs (File S1–Table S1.2 and File S2–Table S2.2) was identified, ranging from 50 to ~500 kb. The 86 CNVRs of CIN covered 12.51 Mb, 0.50% of the genomic sequence of the autosomes, and 0.47% of the total genome length; the average number of CNVRs was 4.50 per individual, with an average length and median size of 145.51 kb and 125.93 kb respectively. The 81 CNVRs of MOD covered 11.01 Mb, 0.44% of the genomic sequence of the autosomes, and 0.41% of the total genome length; the average number of CNVRs was 4.25 per individual, with an average length and median size of 136.02 kb and 120.56 kb respectively. In CIN, we found 74 CNVRs with only gains (duplications), nine with only losses (deletions), and three CNVRs with both; 67 CNVRs with a frequency of >2.5% and 19 with a frequency of >5% were detected. The CNVRs with the highest frequencies were located at Chr3:120547501–120647330 (19.7%) and Chr23:34673581–35007295 (18%), whereas the greatest number of genes (i.e. 19 genes) was mapped to one CNVR located at Chr17:74123863–74393620. In MOD, we found 71 CNVRs with gains and 10 with losses; 51 CNVRs had a frequency of >2.5% and 19 had a frequency of >5%. The CNVRs with the highest frequencies were located at Chr17:74123863–74182044 (39.4%) and Chr5: 59364363–59598727 (16.7%), whereas the greatest numbers of genes were mapped to two CNVRs located at Chr17:73944911–74344162 and Chr11:105778702–106019172, which contained 28 and 23 genes respectively.

To assess our results, we compared the 167 CNVRs here reported with those found in other studies using BovineSNP50K BeadChip (Table 1). There was high variability in the total number and length of the CNVRs identified in the different studies and only a small number of CNVRs identified in the present study were reported in other studies, which has also been reported by Bagnato *et al*. (2015). The greatest coincidence of CNVRs was found with the studies of Hou *et al*. (2011, 2012*a*), with 31.1% and 29.9% common CNVRs respectively, and the total lengths of overlapping regions were 7.6 and 7.4 Mb respectively. Thirty-five CNVRs in our study coincided with CNVRs reported by Prinsen *et al*. (2016) and the length of the overlapping regions was 6.6 Mb. In all, 27 of the 39 CNVRs identified by Xu *et al*. (2014), which were associated with one or

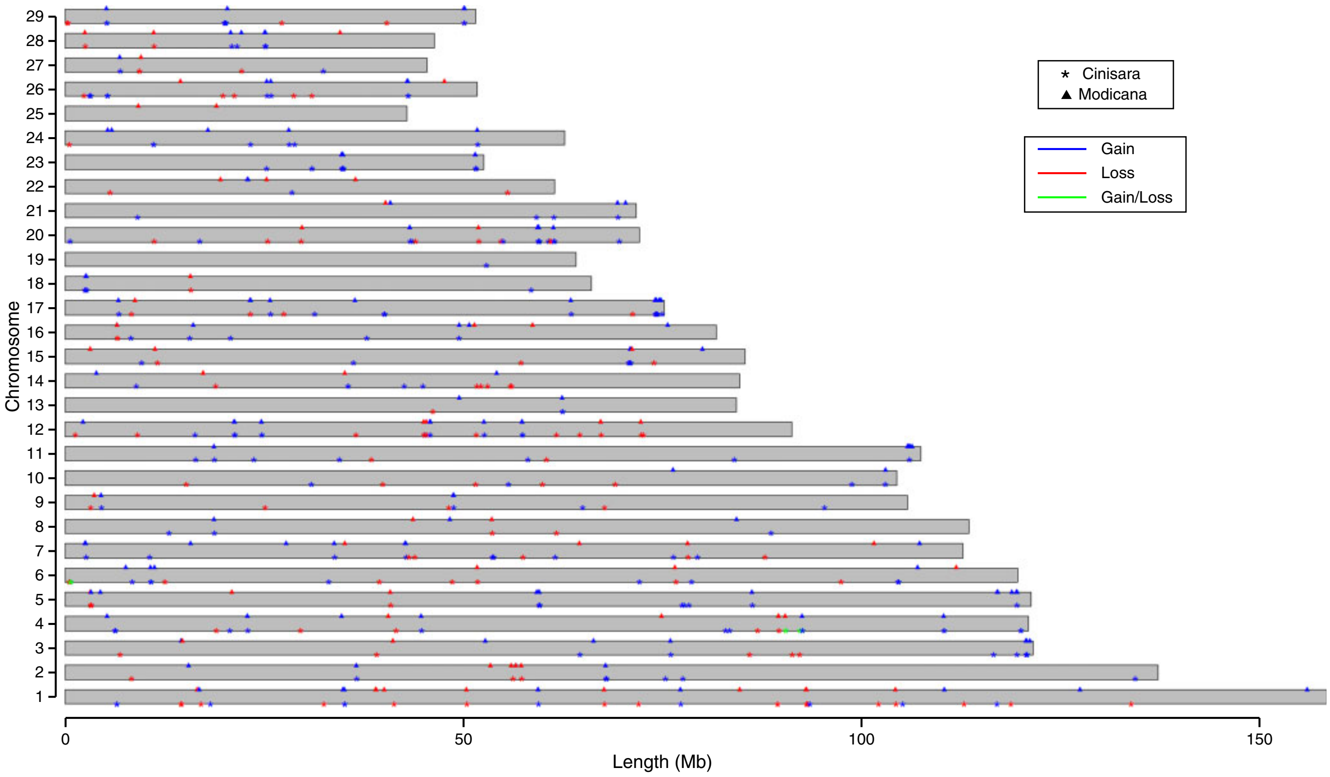

**Fig. 1.** Genomic distribution of copy-number variations (CNVs) in Cinisara and Modicana cattle breeds on the 29 autosomal chromosomes.

**Table 1. Comparison between copy-number variation regions (CNVRs) detected in previous studies and those detected in the present study**

| Reference | Findings from different studies | | | | Overlapped CNVRs of the present study | | | |
|---|---|---|---|---|---|---|---|---|
| | Method of detection | Total CNVR | Total length (Mb) | Algorithm used | Count | Percentage of count | Total length (Mb) | Percentage of length |
| Bae *et al.* (2010) | Bovine SNP50 | 368 | 63.1 | PennCNV | 17 | 10.2 | 2.8 | 11.9 |
| Hou *et al.* (2011) | Bovine SNP50 | 682 | 139.8 | PennCNV | 52 | 31.1 | 7.6 | 32.3 |
| Hou *et al.* (2012*a*) | Bovine SNP50 | 811 | 141.8 | PennCNV | 50 | 29.9 | 7.4 | 31.5 |
| Seroussi *et al.* (2010) | Bovine SNP50 | 410 | 51.7 | PennCNV | 19 | 11.4 | 3.2 | 13.6 |
| Xu *et al.* (2014) | Bovine SNP50 | 39 | 37.54 | SVS | 27 | 16.2 | 4.2 | 17.9 |
| Jiang *et al.* (2013) | Bovine SNP HD | 367 | 23.8 | PennCNV | 10 | 6.0 | 1.6 | 6.8 |
| Prinsen *et al.* (2016) | Bovine SNP HD | 563 | 57.6 | PennCNV-CNAM_SVS | 35 | 21.0 | 6.6 | 28.1 |
| Present study | Bovine SNP50 | 167 | 23.52 | PennCNV | | | | |

more milk production traits, were detected in our study, and the length of the overlapping regions was 4.2 Mb (Table 1). When we compared our CNVRs with those in the study of Hou *et al.* (2012*a*), which examined the CNVs associated with resistance or susceptibility to gastrointestinal nematodes, we found that many of our CNVRs overlapped CNVs associated with either susceptibility or resistance (File S1–Table S1.3 and File S2–Table S2.6). When we compared our results with those obtained in the aforementioned studies, differences among the identified CNVs were apparent. The use of different algorithms, different technologies, different criteria for various parameters, and differences in the numbers of samples and breeds tested could be responsible for these discrepancies. Moreover, the different environmental conditions in which the cows live, particularly the CIN breed, could also generate unique CNVs. Among the factors that contributed to these discrepancies, the most important is the density of the array used. Several authors (Hou *et al*. 2012*b*; Jiang *et al*. 2013; Salomón-Torres *et al*. 2015) have shown that the use of high-density (HD) SNP arrays provides more precise boundary information for CNV detection. Therefore, some CNVRs were not detected in our study because of the limitations of the array used. Although the BovineSNP50K BeadChip can be used to detect CNVs, the SNP probes on the chip are neither dense enough nor sufficiently uniformly distributed to achieve an unbiased high-resolution cattle CNV map (Jiang *et al*. 2013). The CNVRs identified with HD panels are much shorter than those obtained with BovineSNP50K (Hou *et al*. 2012*a*; Jiang *et al*. 2013). Therefore, when an HD panel is used, more CNVs will be identified throughout the genome (Jiang *et al*. 2013). However, because the CNVRs from different studies do not overlap completely, we infer that a great number of CNVRs are still undiscovered, particularly in local breeds.

In all, 207 CNVs were identified in CIN with CNAM–SVS (File S1–Table S1.9) on the 29 autosomes, and the average number per sample was 3.2, with an average length and median size of 228.92 kb and 150.76 kb respectively. By aggregating the overlapping CNVs, a total of 39 CNVRs (File S1–Table S1.10) was detected, corresponding to 18 gains and 21 losses. These 39 CNVRs covered 7.4 Mb and 0.29% of the genomic autosomal sequence, with an average length and median size of 189.4 kb and 121.0 kb respectively. Using the same algorithm, we found 181 CNVs in MOD on the 29 autosomes (File S2–Table S2.3), and the average number per sample was 2.79, with an average length and median size of 200.45 kb and 159.96 kb respectively. In this breed, 41 CNVRs were identified by aggregating the CNVs (File S2–Table S2.4), corresponding to 11 gains and 30 losses. The 41 CNVRs covered 8.2 Mb and 0.32% of the genomic autosomal sequence, with an average length and median size of 200.58 kb and 163.88 kb respectively. The differences in the numbers of CNVs and CNVRs detected in each breed by the two software were probably attributable to the identification of longer CNVs with the univariate approach of SVS (Bagnato *et al.* 2015).

The overlapping CNVRs identified with the two software led to the identification of only 29 common CNVRs in CIN (File S1–Table S1.11), with a total length of 5.45 Mb, and 17 common CNVRs in MOD (File S2–Table S2.5), with a total length of 2.1 Mb.

Finally, a further analysis of the common CNVRs among breeds detected with both software was conducted. As a first step, we compared the CNVRs in CIN and MOD detected with the PennCNV software, and identified 39 common CNVRs with a total length of 5.19 Mb (File S3–Table S3.1). In the second step, we compared the CNVRs identified in the two breeds with the CNAM–SVS software, and identified 17 common CNVRs, with a total length of 2.31 Mb (File S3–Table S3.2). Finally, we compared the CNVRs common to both breeds and the two software, and identified a small number of common CNVRs, i.e. 11, with a total length of 1.32 Mb (File S3–Table S3.3).

### *CNVR gene contents and functional annotations*

The CNVRs detected with PennCNV contained abundant annotated genes. In total, 178 and 208 genes, completely or partially overlapping the CNVRs, were identified in CIN and MOD respectively, with the Genome Data Viewer genome browser. The 178 genes in CIN and the 208 genes in MOD (File S1–Table S1.3 and File S2–Table S2.6) were found in 62 of 86 CNVRs and 66 of 81 CNVRs identified with PennCNV respectively. Within these CNVRs in CIN, we found 123 protein-coding genes, 29 pseudogenes, 21 noncoding RNA (ncRNA) genes, four tRNA genes and one unknown gene, whereas in MOD, we found 157 protein-coding genes, 22 pseudogenes, 25 ncRNA genes, three tRNA genes and one unknown gene.

The results of GO and KEGG analyses showed that the functions of the proteins encoded by these genes included a wide spectrum of biological processes, cellular components, molecular functions, clustering gene (File S1–Tables S1.4, S1.5, and S1.6, and File S2–Tables S2.7 and S2.8), and pathways (File S1–Table S1.7 and File S2–Table S2.9). In particular, the GO terms for cellular components were most strongly represented by 'plasma membrane' in both CIN, with 18 genes, and in MOD, with 25 genes. The GO terms for molecular functions were most strongly represented by 'olfactory receptor activity', with 11 genes in CIN and 14 genes in MOD. In MOD, 'G-protein-coupled receptor' was represented by 13 and 10 genes in the GO terms for 'molecular functions' and 'biological processes' respectively. The KEGG results showed that 'olfactory transduction' was the most strongly represented pathway in both breeds (11 genes in CIN and 17 in MOD), followed by 'metabolic pathways' (six genes in both CIN and MOD), PI3K–AKT (six genes in CIN and five in MOD), and the prolactin signalling pathway (seven genes in CIN and three in MOD; File S1–Table S1.7 and File S2–Table S2.9). We do not discuss all the genomic regions within the CNVRs in detail, but selected regions that include genes associated with traits peculiar to livestock breeding. For example, the *LOC788322*, *LOC788372*, *LOC788285*, *LOC788258*, *LOC788357*, *OR5AS1*, *OR5 L2*, *LOC788242*, *LOC513384*, *LOC788210* and *OR10A7* genes are involved in the olfactory transduction pathways (from KEGG). In animals, chemosensory receptors are used to find food, detect mates and offspring, recognise territories and avoid danger (Nei *et al*. 2008). There are significant numbers of CNVs in these genes in other species (Nei *et al*. 2008; Qanbari *et al*. 2014). Some interspecific variations in copy number can readily be explained by the adaptation of organisms to different environments. The CNVs observed within populations may be largely neutral, but if a population moves to a new niche, a proportion of these may be used selectively by that population to adapt to the new niche (Nei *et al*. 2008). The *SERPIND1* gene in mammals plays a crucial role in the control of the endopeptidases that mediate important pathways, such as blood coagulation, fibrinolysis, inflammation and complement activation (Rau *et al*. 2007; Silverman *et al*. 2001), some of which are considered the host's first line of defence to hematophagous and blood-dwelling parasites. The *POU1F1* gene encodes a protein reportedly involved in productive life (Huang *et al*. 2008), whereas *PTK2* (Wang *et al*. 2013), *FOXF2* (Capomaccio *et al*. 2015) and *LAMB3* (Gutiérrez-Gil *et al*. 2015) reportedly encode proteins associated with milk production traits. *ZBTB20* is a gene that overlaps the bovine QTL regions associated with milk traits and encodes a transcription factor that has been implicated in hematopoiesis, oncogenesis and the immune response in mammals (dos Santos *et al*. 2017). The *SEC24D* gene encodes a protein associated with various pathways, including in the immune system and transport to the Golgi, and is involved in transcriptional regulation in cattle (Salleh *et al*. 2017). *SDF2 L1*, *MAPK1* and *mir301b* are involved in the processing of a wide range of defensins (Meredith *et al*. 2013). We also detected *MEGF10*, which encodes a protein that regulates the myogenesis of satellite cells in skeletal muscles (Park *et al*. 2014), and *CRIM1*, which is located close to a QTL on Chr11 and encodes an insulin-like growth factor-binding domain (Kolle *et al*. 2000). Other putative candidate genes included *SMAD9*, which encodes a protein with a potentially important role in follicular initiation and development (Xu *et al*. 2015), *GRIK2*, which encodes a protein that plays important physiological roles in maturation and puberty (Widmann *et al*. 2013), *LRBA*, which encodes a protein related to reproductive traits (Taye *et al*. 2017), *H19*, which encodes a protein associated with embryogenesis and fetal growth in livestock species (Zaitoun and Khatib 2006) and *IGF2*, which encodes a protein considered to regulate the postnatal growth and differentiation of the mammary gland (Bagnicka *et al*. 2010). The protein encoded by*LSP1* has a negative regulatory role in leukocyte recruitment to sites of inflammation and in resistance/susceptibility to intestinal nematodes (Hou *et al*. 2012*a*). A biological link to traits such as milk production, reproduction and behaviour, the immune response and resistance/susceptibility to infectious diseases, which are known to be under selection, can be inferred for most genes within CNVRs. CIN and MOD are two local breeds that are extremely well adapted to harsh environments, are resistant to infectious diseases, have good maternal aptitude, and produce high-quality milk. The genes involved in these traits were detected in our study, and are consistent with the phenotypic characteristics of these two breeds.

When we compared the 86 and 81 CNVRs with the reported QTL regions annotated in the cattle QTL database, we found 151 QTLRs in only 51 CNVRs identified in CIN (File S1–Table S1.8), and 167 QTLRs in 49 CNVRs identified in MOD (File S2–Table S2.9). Many of these QTLs have been associated with calving traits, milk fatty acids, the percentage protein in milk, bodyweight and clinical mastitis.

## Conclusions

Although many studies of CNVs are available for specialised cattle breeds, such as Holstein and Brown Swiss, no information has been available until now for local breeds such as CIN and MOD. The present study is the first to use an SNP data analysis to detect CNVs in the CIN and MOD breeds, using two different algorithms. Although we used BovineSNP50K, high-resolution methods should also be used. Several of the genes detected within the identified CNVRs have important roles in adaptation or resistance to diseases. Our results have provided significant information for the construction of a more complete CNV map of the bovine genome and offer an important resource for the investigation of genomic changes and traits of interest in the CIN and MOD cattle breeds. Therefore, our results should be of value for future studies, and constitute a preliminary report on the distribution of CNV resources in local cattle genomes.

## Conflicts of interest

The authors declare no conflicts of interest.

## Acknowledgements

This research was financed by PON02_00451_3133441, CUP: B61C12 00076005 funded by MIUR. The authors thank the referees who helped to improve the manuscript.

## References

Adams DJ, Dermitzakis ET, Cox T, Smith J, Davies R, Banerjee R, Bonfield J, Mullikin JC, Chung YJ, Rogers J, Bradley A (2005) Complex haplotypes, copy number polymorphisms and coding variation in two recently divergent mouse strains. *Nature Genetics* **37**, 532–536. doi:10.1038/ng1551

Bae JS, Cheong HS, Kim LH, Gung SN, Park TJ, Chun JY, Kim JY, Pasaje CF, Lee JS, Shin HD (2010) Identification of copy number variations and common deletion polymorphisms in cattle. *BMC Genomics* **11**, 232. doi:10.1186/1471-2164-11-232

Bagnato A, Strillacci MG, Pellegrino L, Schiavini F, Frigo E, Rossoni A, Fontanesi L, Maltecca C, Prinsen RTMM, Dolezal MA (2015) Identification and validation of copy number variants in Italian Brown Swiss Dairy cattle using Illumina Bovine SNP50 BeadChip. *Italian Journal of Animal Science* **14**, 552–558. doi:10.4081/ijas.2015.3900

Bagnicka E, Siadkowska E, Strzałkowska N, Żelazowska B, Flisikowski K, Krzyżewski J, Zwierzchowski L (2010) Association of polymorphisms in exons 2 and 10 of the insulin-like growth factor 2 (IGF2) gene with milk production traits in Polish Holstein–Friesian cattle. *The Journal of Dairy Research* **77**, 37–42. doi:10.1017/S0022029909990197

Capomaccio S, Milanesi M, Bomba L, Cappelli K, Nicolazzi EL, Williams JL, Ajmone-Marsan P, Stefanon B (2015) Searching new signals for production traits through gene based association analysis in three Italian cattle breeds. *Animal Genetics* **46**, 361–370. doi:10.1111/age.12303

Clop A, Vidal O, Amills M (2012) Copy number variation in the genome of domestic animals. *Animal Genetics* **43**, 503–517. doi:10.1111/j.1365-2052.2012.02317.x

Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J (2007) QuantiSNP: an objective Bayes Hidden–Markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Research* **35**, 2013–2025. doi:10.1093/nar/gkm076

Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, Bucan M, Maris JM, Wang K (2008) Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Research* **36**, e126. doi:10.1093/nar/gkn556

dos Santos FC, Peixoto MGCD, de Souza Fonseca PA, Pires MDFÁ, Ventura RV, Rosse IDC, Tomita Bruneli FA, Machado MA, Carvalho MRS (2017) Identification of candidate genes for reactivity in Guzerat (*Bos indicus*) cattle: a genome-wide association Study. *PLoS One* **12**(1), e0169163. doi:10.1371/journal.pone.0169163

Dupuis MC, Zhang Z, Durkin K, Charlier C, Lekeux P, Georges M (2013) Detection of copy number variants in the horse genome and 22 examination of their association with recurrent laryngeal neuropathy. *Animal Genetics* **44**, 206–208. doi:10.1111/j.1365-2052.2012.02373.x

Fadista J, Thomsen B, Holm LE, Bendixen BMC (2010) Copy number variation in the bovine genome. *BMC Genomics* **11**, 284. doi:10.1186/1471-2164-11-284

Flisikowski K, Venhoranta H, Nowacka-Woszuk J, McKay SD, Flyckt A, Taponen J, Schnabel R, Schwarzenbacher H, Szczerbal I, Lohi H, Fries R, Taylor JF, Switonski M, Andersson M (2010) A novel mutation in the maternally imprinted PEG3 domain results in a loss of MIMT1 expression and causes abortions and stillbirths in cattle (*Bos taurus*). *PLoS One* **5**, e15116. doi:10.1371/journal.pone.0015116

Fontanesi L, Martelli PL, Beretti F, Riggio V, Dall'Olio S, Colombo M, Casadio R, Russo V, Portolano B (2010) An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genomics* **11**, 639. doi:10.1186/1471-2164-11-639

Fontanesi L, Beretti F, Martelli PL, Colombo M, Dall'Olio S, Occidente M, Portolano B, Casadio R, Matassino D, Russo V (2011) A first comparative map of copy number variations in the sheep genome. *Genomics* **97**, 158–165. doi:10.1016/j.ygeno.2010.11.005

Guryev V, Saar K, Adamovic T, Verheul M, van Heesch SA, Cook S, Pravenec M, Aitman T, Jacob H, Shull JD, Hubner N, Cuppen E (2008) Distribution and functional impact of DNA copy number variation in the rat. *Nature Genetics* **40**, 538–545. doi:10.1038/ng.141

Gutiérrez-Gil B, Arranz JJ, Wiener P (2015) An interpretive review of selective sweep studies in *Bos taurus* cattle populations: identification of unique and shared selection signals across breeds. *Frontiers in Genetics* **6**, 167. doi:10.3389/fgene.2015.00167

Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim E, Matukumalli LK, Ventura M, Song J, VanRaden PM, Sonstegard TS, Van Tassell CP (2011) Genomic characteristics of cattle copy number variations. *BMC Genomics* **12**, 127. doi:10.1186/1471-2164-12-127

Hou Y, Liu GE, Bickhart DM, Matukumalli LK, Li C, Song J, Gasbarre LC, Van Tassell CP, Sonstegard TS (2012*a*) Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. *Functional & Integrative Genomics* **12**, 81–92. doi:10.1007/s10142-011-0252-1

Hou Y, Bickhart DM, Hvinden ML, Li C, Song J, Boichard DA, Fritz S, Eggen A, DeNise S, Wiggans GR, Sonstegard TS, Van Tassell CP, Liu GE (2012*b*) Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. *BMC Genomics* **13**, 376. doi:10.1186/1471-2164-13-376

Huang W, Maltecca C, Khatib H (2008) A proline to histidine mutation in POU1F1 is associated with production traits in dairy cattle. *Animal Genetics* **39**, 554–557. doi:10.1111/j.1365-2052.2008.01749.x

Jiang L, Jiang J, Yang J, Liu X, Wang J, Wang H, Ding X, Liu J, Zhang Q (2013) Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. *BMC Genomics* **14**, 131. doi:10.1186/1471-2164-14-131

Kadri NK, Sahana G, Charlier C, Iso-Touru T, Guldbrandtsen B, Karim L, Nielsen US, Panitz F, Aamand GP, Schulman N, Georges M, Vilkki J, Lund MS, Druet T (2014) A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic red cattle: additional evidence for the common occurrence of balancing selection in livestock. *PLOS Genetics* **10**, e1004049. doi:10.1371/journal.pgen.1004049

Kolle G, Georgas K, Holmes GP, Little MH, Yamada T (2000) CRIM1, a novel gene encoding a cysteine-rich repeat protein, is developmentally regulated and implicated in vertebrate CNS development and organogenesis. *Mechanisms of Development* **90**, 181–193. doi:10.1016/S0925-4773(99)00248-8

Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, Gasbarre LC, Lacalandra G, Li RW, Matukumalli LK, Nonneman D, Regitano LC, Smith TP, Song J, Sonstegard TS, Van Tassell CP, Ventura M, Eichler EE, McDaneld TG, Keele JW (2010) Analysis of copy number variations among diverse cattle breeds. *Genome Research* **20**, 693–703. doi:10.1101/gr.105403.110

Ma Q, Liu X, Pan J, Ma L, Ma Y, He X, Zhao Q, Pu Y, Li Y, Jiang L (2017) Genome-wide detection of copy number variation in Chinese indigenous sheep using an ovine high-density 600 K SNP array. *Scientific Reports* **7**, 912. doi:10.1038/s41598-017-00847-9

Marenne G, Chanock S, Rothman N, Rodríguez-Santiago B, Rico D, Pita G, Pérez-Jurado L, Valencia A, Jacobs K, Pisano DG, Díaz-Uriarte R, Earl J, García-Closas M, Silverman D, Kogevinas M, Génin E, Real FX, Malats N (2009) CNV assessment from Illumina Infinium 1M platform: agreement according to algorithm and source of DNA. *Annals of Human Genetics* **73**, 658–669.

Mastrangelo S, Saura M, Tolone M, Salces-Ortiz J, Di Gerlando R, Bertolini F, Fontanesi L, Sardina MT, Serrano M, Portolano B (2014) The genome-wide structure of two economically important indigenous Sicilian cattle breeds. *Journal of Animal Science* **92**, 4833–4842. doi:10.2527/jas.2014-7898

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, Connell J, Moore SS, Smith TP, Sonstegard TS, Van Tassell CP (2009) Development and 26 characterization of a high density SNP genotyping assay for cattle. *PLoS One* **4**, e5350. doi:10.1371/journal.pone.0005350

Meredith BK, Berry DP, Kearney F, Finlay EK, Fahey AG, Bradley DG, Lynn DJ (2013) A genome-wide association study for somatic cell score using the Illumina high-density bovine bead chip identifies several novel QTL potentially related to mastitis susceptibility. *Frontiers in Genetics* **4**, 229.

Meyers SN, McDaneld TG, Swist SL, Marron BM, Steffen DJ, O'Toole D, O'Connell JR, Beever JE, Sonstegard TS, Smith TPL (2010) A deletion mutation in bovine SLC4A2 is associated with osteoporosis in Red Angus cattle. *BMC Genomics* **11**, 337. doi:10.1186/1471-2164-11-337

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* **16**, 1215. doi:10.1093/nar/16.3.1215

Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, Mu XJ, Nemesh J, Peckham HE, Rausch T, Scally A, Shi X, Stromberg MP, Stütz AM, Urban AE, Walker JA, Wu J, Zhang Y, Zhang ZD, Batzer MA, Ding L, Marth GT, McVean G, Sebat J, Snyder M, Wang J, Ye K, Eichler EE, Gerstein MB, Hurles ME, Lee C, McCarroll SA, Korbel JO 1000 Genomes Project (2011) Mapping copy number variation by population scale genome sequencing. *Nature* **470**, 59–65. doi:10.1038/nature09708

Nei M, Niimura Y, Nozawa M (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nature Reviews Genetics* **9**, 951–963. doi:10.1038/nrg2480

Park SY, Yun Y, Kim MJ, Kim IS (2014) Myogenin is a positive regulator of MEGF10 expression in skeletal muscle. *Biochemical and Biophysical Research Communications* **450**, 1631–1637. doi:10.1016/j.bbrc.2014.07.061

Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurles ME, Tyler-Smith C, Eichler EE, Carter NP, Lee C, Redon R (2008) Copy number variation and evolution in humans and chimpanzees. *Genome Research* **18**, 1698–1710. doi:10.1101/gr.082016.108

Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, Macdonald JR, Mills R, Prasad A, Noonan K, Gribble S, Prigmore E, Donahoe PK, Smith RS, Park JH, Hurles ME, Carter NP, Lee C, Scherer SW, Feuk L (2011) Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nature Biotechnology* **29**, 512–520. doi:10.1038/nbt.1852

Prinsen RTMM, Strillacci MG, Schiavini F, Santus E, Rossoni A, Maurer V, Bieber A, Gredler B, Dolezal M, Bagnato A (2016) A genome-wide scan of copy number variants using high-density SNPs in Brown Swiss dairy cattle. *Livestock Science* **191**, 153–160. doi:10.1016/j.livsci.2016.08.006

Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, Nielsen R, Simianer H (2014) Classic selective sweeps revealed by massive sequencing in cattle. *PLOS Genetics* **10**, e1004148. doi:10.1371/journal.pgen.1004148

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842. doi:10.1093/bioinformatics/btq033

Rau JC, Beaulieu LM, Huntington JA, Church FC (2007) Serpins in thrombosis, hemostasis and fibrinolysis. *Journal of Thrombosis and Haemostasis* **5**, 102–115. doi:10.1111/j.1538-7836.2007.02516.x

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME (2006) Global variation in copy number in the human genome. *Nature* **444**, 444–454. doi:10.1038/nature05329

Salleh MS, Mazzoni G, Höglund JK, Olijhoek DW, Lund P, Løvendahl P, Kadarmideen HN (2017) RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high-and low-residual feed intake in Nordic dairy cattle. *BMC Genomics* **18**, 258. doi:10.1186/s12864-017-3622-9

Salomón-Torres R, González-Vizcarra VM, Medina-Basulto GE, Montaño-Gómez MF, Mahadevan P, Yaurima-Basaldúa VH, Villa-Angulo C, Villa-Angulo R (2015) Genome-wide identification of copy number variations in Holstein cattle from Baja California, Mexico, using high-density SNP genotyping arrays. *Genetics and Molecular Research* **14**, 11848–11859. doi:10.4238/2015.October.2.18

Seroussi E, Glick G, Shirak A, Yakobson E, Weller JI, Ezra E, Zeron Y (2010) Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. *BMC Genomics* **11**, 673. doi:10.1186/1471-2164-11-673

Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin PB, Gettins PG, Irving JA, Lomas DA, Luke CJ, Moyer RW, Pemberton PA, Remold-O'Donnell E, Salvesen GS, Travis J, Whisstock JC (2001) The serpins are an expanding superfamily of structurally similar but functionally diverse proteins: evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *The Journal of Biological Chemistry* **276**, 33293–33296. doi:10.1074/jbc.R100016200

Sonstegard TS, Barendse W, Bennett GL, Brockmann GA, Davis S, Droegemuller C, Kalm E, Kappes SM, Kühn C, Li Y, Schwerin M, Taylor J, Thomsen H, Van Tassell CP, Yeh CC (2001) Consensus and comprehensive linkage maps of the bovine sex chromosomes. *Animal Genetics* **32**, 115–117. doi:10.1046/j.1365-2052.2001.0700g.x

Taye M, Lee W, Jeon S, Yoon J, Dessie T, Hanotte O, Mwai OA, Kemp S, Cho S, Jong Oh S, Lee HK, Kim H (2017) Exploring evidence of positive selection signatures in cattle breeds selected for different traits. *Mammalian Genome* **28**, 528–541.

Wain LV, Armour JA, Tobin MD (2009) Genomic copy number variation, human health, and disease. *Lancet* **374**, 340–350. doi:10.1016/S0140-6736(09)60249-X

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M (2007) PennCNV: an integrated hidden Markov model designed for highresolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research* **17**, 1665–1674. doi:10.1101/gr.6861907

Wang H, Jiang L, Liu X, Yang J, Wei J, Xu J, Zhang Q, Liu JF (2013) A post-GWAS replication study confirming the PTK2 gene associated with milk production traits in Chinese Holstein. *PLoS One* **8**(12), e83625. doi:10.1371/journal.pone.0083625

Widmann P, Reverter A, Fortes MRS, Weikard R, Suhre K, Hammon H, Albrecht E, Kuehn C (2013) A systems biology approach using metabolomic data reveals genes and pathways interacting to modulate divergent growth in cattle. *BMC Genomics* **14**, 798. doi:10.1186/1471-2164-14-798

Winchester L, Yau C, Ragoussis J (2009) Comparing CNV detection methods for SNP arrays. *Briefings in Functional Genomics* **8**, 353–366. doi:10.1093/bfgp/elp017

Xu L, Hou Y, Bickhart DM, Song J, Liu GE (2013) Comparative analysis of CNV calling algorithms: literature survey and a case study using bovine high-density SNP data. *Microarrays (Basel, Switzerland)* **2**, 171–185. doi:10.3390/microarrays2030171

Xu L, Cole JB, Bickhart DM, Hou Y, Song J, VanRaden PM, Sonstegard TS, Van Tassell CP, Liu GE (2014) Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. *BMC Genomics* **15**, 683. doi:10.1186/1471-2164-15-683

Xu J, Li J, Wang H, Wang G, Chen J, Huang P, Cheng J, Gan L, Wang Z, Cai Y (2015) A novel SMAD family protein, SMAD9 is involved in follicular initiation and changes egg yield of geese via synonymous mutations in exon1 and intron2. *Molecular Biology Reports* **42**, 289–302. doi:10.1007/s11033-014-3772-7

Yavaş G, Koyuturk M, Ozsoyoglu M, Gould MP, LaFramboise T (2009) An optimization framework for unsupervised identification of rare copy number variation from SNP array data. *Genome Biology* **10**, R119. doi:10.1186/gb-2009-10-10-r119

Zaitoun I, Khatib H (2006) Assessment of genomic imprinting of SLC38A4, NNAT, NAP1L5, and H19 in cattle. *BMC Genetics* **7**, 49. doi:10.1186/1471-2156-7-49