

Article

# Genomic Structural Diversity in Local Goats: Analysis of Copy-Number Variations

Rosalia Di Gerlando <sup>1</sup>, Salvatore Mastrangelo <sup>1</sup>, Angelo Moscarelli <sup>1</sup>, Marco Tolone <sup>1</sup>,  
Anna Maria Sutera <sup>2</sup>, Baldassare Portolano <sup>1</sup> and Maria Teresa Sardina <sup>1,\*</sup>

<sup>1</sup> Dipartimento Scienze Agrarie, Alimentari e Forestali, University of Palermo, 90128 Palermo, Italy; rosalia.digerlando@unipa.it (R.D.G.); salvatore.mastrangelo@unipa.it (S.M.); angelo.moscarelli@unipa.it (A.M.); marco.tolone@unipa.it (M.T.); baldassare.portolano@unipa.it (B.P.)

<sup>2</sup> Dipartimento Scienze Veterinarie, University of Messina, 98168 Messina, Italy; asutera@unime.it

\* Correspondence: mariateresa.sardina@unipa.it

Received: 30 March 2020; Accepted: 12 June 2020; Published: 16 June 2020

**Simple Summary:** Copy-number variations (CNVs) are one of the widely dispersed forms of structural variations in mammalian genomes and are known to be present in genomic regions that regulate important physiological functions. In this study, CNV detection was performed starting from genotypic data of 120 individuals, belonging to four Sicilian dairy goat breeds, genotyped with the Illumina GoatSNP50 BeadChip array. Using PennCNV software, a total of 702 CNVs were identified in 107 individuals. These were merged in 75 CNV regions (CNVRs), i.e., regions containing CNVs overlapped by at least 1 base pair. Functional annotation of the CNVRs allowed the identification of 139 genes/loci within the most frequent CNVRs, which are involved in local adaptation, mild behaviour, immune response, reproduction, and olfactory receptors. This study provides insights into the genomic variations within these Italian goat breeds and should be of value for future studies to identify the relationships between this type of genetic variation and phenotypic traits.

**Abstract:** Copy-number variations (CNVs) are one of the widely dispersed forms of structural variations in mammalian genomes, and are present as deletions, insertions, or duplications. Only few studies have been conducted in goats on CNVs derived from SNP array data, and many local breeds still remain uncharacterized, e.g., the Sicilian goat dairy breeds. In this study, CNV detection was performed, starting from the genotypic data of 120 individuals, belonging to four local breeds (Argentata dell'Etna, Derivata di Siria, Girgentana, and Messinese), genotyped with the Illumina GoatSNP50 BeadChip array. Overall, 702 CNVs were identified in 107 individuals using PennCNV software based on the hidden Markov model algorithm. These were merged in 75 CNV regions (CNVRs), i.e., regions containing CNVs overlapped by at least 1 base pair, while 85 CNVs remained unique. The part of the genome covered by CNV events was 35.21 Mb (1.2% of the goat genome length). Functional annotation of the CNVRs allowed the identification of 139 genes/loci within the most frequent CNVRs that are involved in local adaptations, such as coat colour (*ADAMTS20* and *EDNRA*), mild behaviour (*NR3C2*), immune response (*EXOC3L4* and *TNFAIP2*), reproduction (*GBP1* and *GBP6*), and olfactory receptors (*OR7E24*). This study provides insights into the genomic variations for these Sicilian dairy goat breeds and should be of value for future studies to identify the relationships between this type of genetic variation and phenotypic traits.

**Keywords:** copy number variation; genotyping array; Sicilian goat breeds

## 1. Introduction

Copy-number variations (CNVs) are structural variations widely dispersed in mammalian genomes. Conventionally, CNVs are defined as genomic regions ranging from 50 base pairs (bp) to several Mega bases (Mb), and which are present in the form of deletions, insertions, or duplications [1,2]. CNVs have been shown to contribute to phenotypic diversity in model organisms and are thus associated with genetic diversity as well as important production and disease traits in domesticated livestock species [3–5]. CNVs involving large genomic regions can affect the gene structure and gene dosage, which, in turn, has an impact on the gene expression, influencing in this way phenotypic variability and disease susceptibility [1,6–8].

Several techniques have been developed and used to identify CNVs within the genome, including comparative genomic hybridisation arrays, SNP genotyping arrays, and next-generation sequencing [9,10]. SNP arrays provide the Log R Ratio and B allele frequency, which represent the intensity signal due to the copy number and allelic status of the population [11]. Different algorithms have been used for identification of CNVs from molecular data [12]. For example, the CNV discovery method of the PennCNV software tends to focus on finding long and/or rare CNVs while other methods, like the multivariate one, considering all samples simultaneously, was designed for detecting small and common CNVs [12,13]. However, PennCNV is the most reliable and accurate algorithm in detecting CNVs from Illumina BeadChip data [14–17]. It is also worth pointing out that although the SNP arrays can be used to detect the CNVs, the SNP probes on the chip are neither dense enough nor uniformly distributed to achieve an unbiased high resolution CNV map [18,19].

Previous studies have completed genome-wide CNV analysis in many livestock species, including cattle [19–24], sheep [25–30], pigs [31], and horses [32]. However, in goats, only few studies have been conducted based on SNP array data (e.g., [13,33]).

Since CNVs are known to be present in genomic regions that regulate important physiological functions, they have also been studied for associations with environmental adaptation and economically important traits in livestock species. In sheep, goats, pigs, and cattle, coat colour is partially determined by CNVs [34–37]. In cattle, traits and diseases, such as milk production [5,38,39] and female fertility failure [40], have been shown to be influenced by CNVs. Zhang et al. [41] suggested that CNVs may contribute to different litter size traits in Laoshan dairy goats. Di Gerlando et al. [30] found CNVs associated with milk production and health-related traits in dairy sheep. Qanbari et al. [42] reported duplication of the olfactory receptor family genes within the bovine genome, suggesting their role in environmental adaptability.

During the last century, erosion of livestock genetic resources was observed as a result of the massive replacement of low-productivity local breeds with highly productive ones [43]. In Sicily, dairy goat production represents an important resource for the economy of the hilly and mountainous areas, in which other economic activities are limited. Nowadays, four native dairy goat breeds are reared: Argentata dell'Etna, Derivata di Siria, Girgentana, and Messinese. These breeds present differences in both phenotypic and production traits. Selection breeding schemes are almost absent within Sicilian goat farms. Nevertheless, these breeds are adapted to the harshness of the mountain areas because of their good grazing characteristics, have an aptitude for dairy production, and are resistant to environmental conditions. Their adaptability is also expressed in longevity, resistance to diseases, and good fertility [43]. As reported above, different studies have been conducted on CNV detection in livestock species, but several local breeds still remain uncharacterized since their CNV distribution in the genome has not yet been analysed; for example, the Sicilian goats. Therefore, this study aimed to investigate the CNVs in these four dairy goat breeds by performing a genome-wide screen of autosomes using genomic data generated from the GoatSNP50 BeadChip Genotyping array (Illumina). The genome-wide analysis and characterization of the CNVs could improve our understanding on genetic variations, and could be an important tool for determining the role of CNVs on economic traits and for the development of conservation programs for these local breeds.

## 2. Materials and Methods

### 2.1. Samples and Genotyping

A total of 120 individual blood samples were collected from four Sicilian goat breeds: Argentata dell'Etna ( $n = 24$ ), Derivata di Siria ( $n = 36$ ), Girgentana ( $n = 36$ ), and Messinese ( $n = 24$ ).

The Girgentana goat is one of the ancient Sicilian autochthonous breeds, with long corkscrew horns in both sexes and a cream/white coat colour; the Derivata di Siria breed (also known as Mediterranean Red) is completely red with long ears. The Messinese and Argentata dell'Etna are reared in the mountain areas; the former presents a very different coat colour, e.g., plain, pied or streaked, black, brown, or red with various shades; the latter has a coat colour that has grey shading from light to dark, with silver glints, and a grey skin. All of them are dairy goat breeds and their milk is used for niche products in local markets (<http://eng.agraria.org/>).

All the procedures were approved by "Organismo Preposto al Benessere Animale" (O.P.B.A.) of the University of Palermo in agreement with the recommendations of the European Union (EU) Directive 2010/63/EU to ensure appropriate animal care. About 6 mL of blood were collected from the jugular vein using tubes with EDTA as an anticoagulant. Genomic DNA was extracted from the buffy coats of the nucleated cells using the salting-out method [44].

Samples were genotyped using the Illumina GoatSNP50 BeadChip Genotyping array containing 53,347 SNPs. SNPs were mapped using the goat assembly ARS1 (RefSeq assembly accession: GCF\_001704415.1, [ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/704/415/GCF\\_001704415.1\\_ARS1](ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/704/415/GCF_001704415.1_ARS1)). GenomeStudio v2.0 software (Illumina) was used to create the genotypic and intensity data for further analyses.

### 2.2. Genetic Relationship among Sicilian Goat Breeds

Genotypic data was edited using PLINK 1.7 [45]. Only SNPs located on autosomes were considered. SNPs were filtered according to quality criteria that included a call frequency (proportion of samples with the genotype at each locus)  $\geq 0.98$ , minor allele frequency (MAF)  $\geq 0.05$ , and Hardy–Weinberg equilibrium (HWE)  $p$ -value  $< 0.001$ . SNPs and samples that did not satisfy these quality criteria were excluded. Pair-wise genetic relationships were estimated using the `--cluster` and `--mds-plot` options in PLINK 1.7 [45], and graphically represented by multidimensional scaling (MDS) analysis.

### 2.3. Identification of CNVs and CNVRs

The PennCNV software [46] was used to infer CNVs in our samples. The input file contained the signal intensity ratios (Log R Ratio, LRR) and allelic frequencies (B Allele Frequency, BAF) for 50,619 autosomal SNPs and 120 samples. The CNVs were inferred using the hidden Markov model [46], spanning at least three SNPs, using the population frequency of the B allele (PBF) file calculated on the BAF of each SNP, and considering the GC content within 1 Mb surrounding each SNP for the appropriate LRR adjustments. PennCNV quality filters were applied in order to keep high-quality samples: standard deviation of LRR  $< 0.3$ , BAF drift  $< 0.01$ , and waviness factor value between  $-0.05$  and  $0.05$ . CNV regions (CNVRs) were defined as suggested by Redon et al. [47], i.e., regions containing CNVs overlapped by at least 1 bp, and were identified using the `merge` command of BEDTools software [48]. Finally, CNVRs were conventionally defined as gain, loss, and mixed, i.e., CNVRs containing duplication, deletion, or both duplication and deletion CNV events.

Additionally, a Venn diagram was constructed by an online tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) using the CNVRs identified among the four considered breeds. Moreover, in order to highlight the differences among breeds related to the CNVRs, a matrix was also built assigning "0" to the absence of a CNV in a CNVR (normal state), "1" to a deletion, or "2" to a duplication event. This matrix was used to perform a Principal Components Analysis (PCA) using PAST 3.22 software [49].

#### 2.4. Gene Enrichment and Functional Annotation

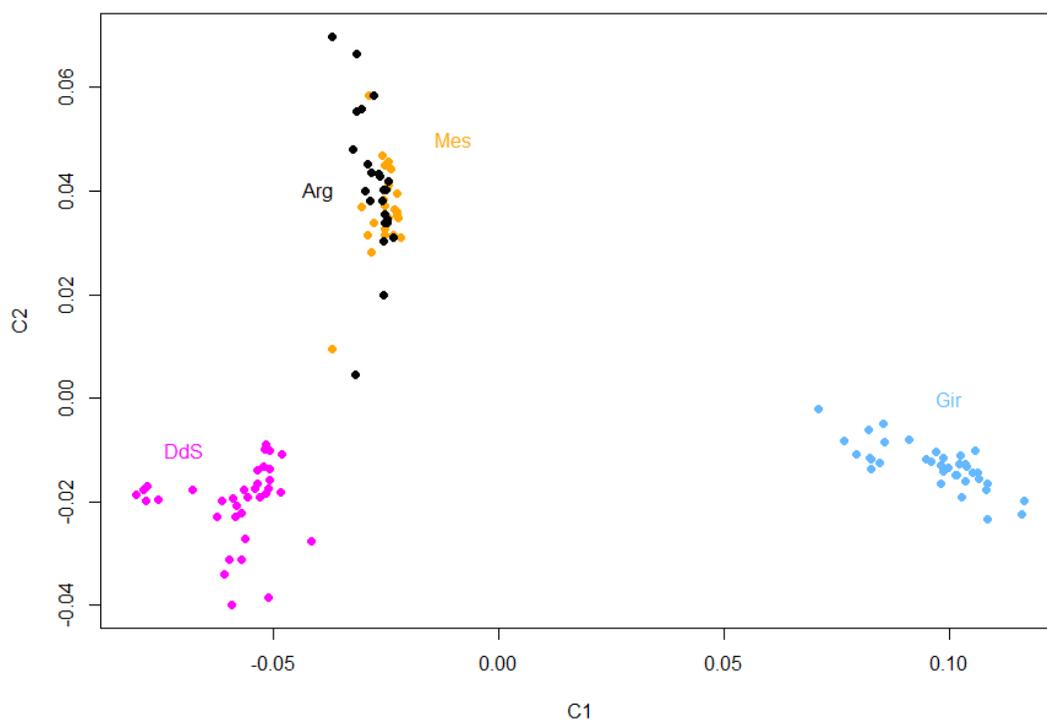
The gene content was assessed for CNVRs with a frequency > 0.10 (i.e., present in at least 11 samples). It was described based on the *Capra hircus* genome assembly ARS1, in the Genome Data Viewer genome browser ([https://www.ncbi.nlm.nih.gov/genome/gdv/browser/?context=genome&acc=GCF\\_001704415.1](https://www.ncbi.nlm.nih.gov/genome/gdv/browser/?context=genome&acc=GCF_001704415.1)). The gene list was further analysed with the PANTHER Classification System (<http://www.pantherdb.org/>) to identify the gene ontology (GO) terms for the molecular function, biological process, cellular component, and biological pathway. It is worth noting that a portion of the genes in the goat genome has not been annotated or has been annotated as LOC, and this could influence the outcome of the GO analysis. For this reason, the GO enrichment analysis was conducted considering bovine genes.

### 3. Results and Discussion

In the present study, we described the CNVs in four goat breeds: Argentata dell'Etna, Derivata di Siria, Girgentana, and Messinese. These are local dairy breeds that are often replaced by highly productive foreign breeds, leading to the progressive abandoning of low-income rural activities [43].

#### 3.1. Genetic Relationship among Sicilian Goat Breeds

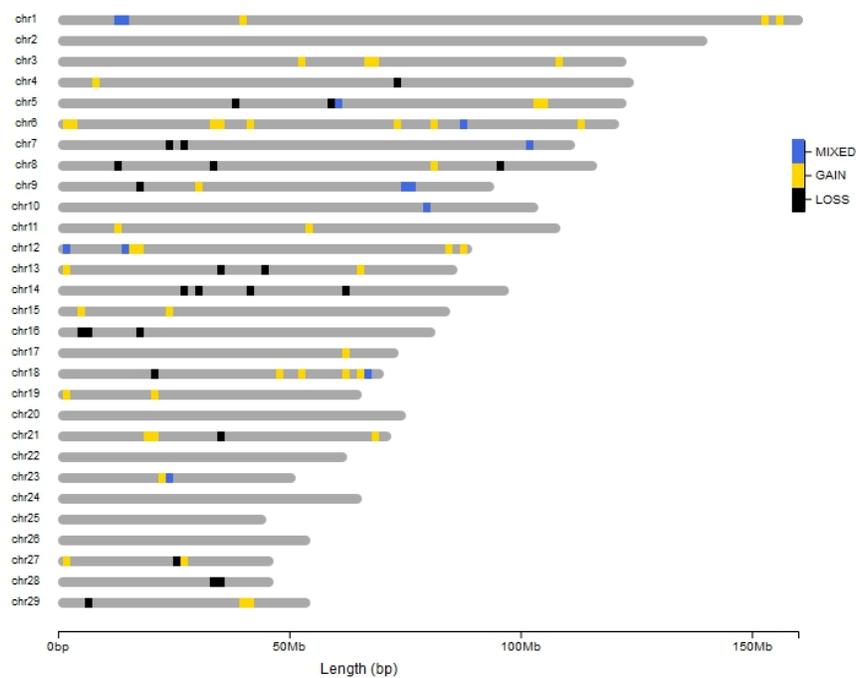
A total of 49,234 SNPs and 120 individuals were used in this analysis. Multidimensional scaling analysis was performed to cluster the animals and to explore the genetic relationship among the Sicilian breeds. The obtained results evidenced that the Girgentana (Gir) and Derivata di Siria (DdS) breeds are distinct groups (Figure 1), as reported in previous studies [43,50,51], while the Argentata dell'Etna (Arg) and Messinese (Mes) individuals are positioned close to each other (Figure 1). The genetic closeness between Arg and Mes could be explained considering their geographic proximity, similarities in environment, breeding practices, and the gene flow among them.



**Figure 1.** MDS plot of the Sicilian goat breeds. Arg = Argentata; DdS = Derivata di Siria; Gir = Girgentana; Mes = Messinese.

### 3.2. CNV and CNVR Identification among Breeds

Using PennCNV, we were able to identify a total of 702 CNVs (Table S1) in 107 individuals. Among the detected CNVs, 500 were gains and 202 losses, with an average number of 6.5 CNVs per individual and an average length of 184.2 kb per CNV. Similar results (i.e., 6.2 CNVs per individual) were reported by Liu et al. [33], who studied the CNVs in worldwide goat populations, also using the Illumina GoatSNP50 BeadChip array. The highest number of CNVs ( $n = 70$ ) was found on chromosome 10, while only one CNV was identified on chromosomes 2, 20, 24, and 25. No CNV was found on chromosome 22. Merging the overlapping CNVs, 85 CNVs remained unique, i.e., present in only one individual (Table S2), while a total of 75 CNVRs were identified among the breeds (Table S3). The part of the genome covered by CNV events (CNVRs and unique CNVs, together) was 35.21 Mb, which corresponded to 1.2% of the *Capra hircus* genome length. The length of the CNVRs varied from 49.4 kb (CNVR\_21) to 1.98 Mb (CNVR\_16) (Table S3), with an average length of 282.94 kb. All these statistics are consistent with the study of Liu et al. [33]. Among the 75 CNVRs, 42 were gains, 22 were losses, and 11 were mixed (i.e., containing duplication, deletion, or both duplication and deletion CNV events) (Table S3, Figure 2).



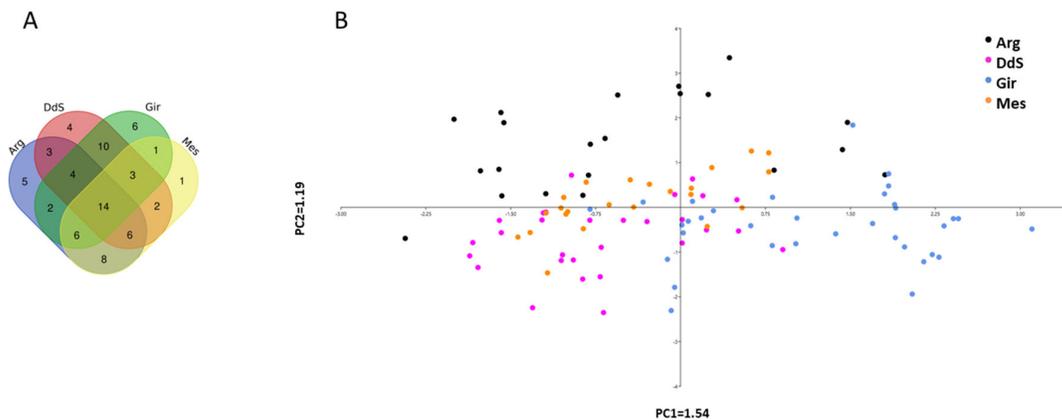
**Figure 2.** Ideogram of the distribution of the copy-number variation regions (CNVRs) in Sicilian goat breeds. Blue, yellow, and black colours represent mixed, gain, and loss events, respectively.

Table 1 includes the CNVR details for each breed. It is possible to note that the Arg breed had the highest number of CNVRs ( $n = 48$ ) and the highest number of gain events ( $n = 40$ ) compared to the other breeds. Mes was the breed with the lowest number of CNVRs ( $n = 41$ ), and DdS was the breed with the lowest number of gain events ( $n = 24$ ).

**Table 1.** CNVR details for each goat breed and overall. Number of animals per breed are reported within brackets. Arg = Argentata; DdS = Derivata di Siria; Gir = Girgentana; Mes = Messinese.

Breed	No CNVRs	Gain	Loss	Mixed
Arg ( <i>n</i> = 21)	48	40	6	2
DdS ( <i>n</i> = 29)	46	24	18	4
Gir ( <i>n</i> = 34)	46	27	17	2
Mes ( <i>n</i> = 23)	41	30	8	3
Overall ( <i>n</i> = 107)	75	42	22	11

The Venn diagram reported in Figure 3A show the number of CNVRs shared among the four breeds (*n* = 14) (Table S4). Moreover, the PCA analysis performed using the matrix of the CNVRs (Table S5; Figure 3B) show two separated groups for Gir and DdS, and closer ones for Arg and Mes. The highest number of shared CNVRs is between Gir and DdS (*n* = 10), followed by the ones shared between Arg and Mes (*n* = 8), which could be the reason for the partial overlap among these breeds presented in the PCA graph.



**Figure 3.** (A) Venn diagram of shared and proper CNVRs of each breed. (B) Principal Component Analysis obtained using the CNVR matrix 0, 1, 2 (0 = normal state; 1 = deletion; 2 = duplication). Arg = Argentata; DdS = Derivata di Siria; Gir = Girgentana; Mes = Messinese.

### 3.3. CNVR Gene Contents and Functional Annotations

Gene enrichment was described for the most representative CNVRs (frequency > 0.10; i.e., present in at least 11 samples) identified among the breeds. A total of 13 CNVRs presented frequency values between 0.103 and 0.430, and contained 139 genes/loci, completely or partially overlapping the CNVRs (Table S6). To further analyse the biological functions of the CNVRs, we conducted a GO analysis. Similar to the studies on CNVs carried out on other livestock species (e.g., [19,23,24,28,30,32]), the results showed that the functions of the proteins encoded by these genes included several biological processes, molecular functions, cellular components, and pathways, as reported in Table 2. Among the most representative CNVRs, many of them included candidate genes associated with important breeding traits.

**Table 2.** The gene ontology (GO) categories (biological process, molecular function, and cellular component) and pathways containing genes overlapping the most frequent CNVRs identified in Sicilian goat breeds.

Accession Number	Biological Process	Gene Symbol
GO: 0065007	Biological regulation	EDNRA, NR3C2
GO: 0009987	Cellular process	ADAMTS20, EXOC3L4, ERG, KYAT3, LGALS9, TNFAIP2
GO: 0002376	Immune system process	GBP4, GBP5, GBP6
GO: 00051179	Localization	EXOC3L4, KCNJ15, SLC7A3, TNFAIP2
GO: 0008152	Metabolic process	ERG, KYAT3, NR3C2
GO: 0032501	Multicellular organismal process	OR7E24
GO: 0050896	Response to stimulus	GBP4, GBP5, GBP6
Accession Number	Molecular Function	Gene Symbol
GO: 0005488	Binding	ADAMTS20, EXOC3L4, GBP4, GBP5, GBP6, LGALS9, NR3C2, TNFAIP2
GO: 0003824	Catalytic activity	ADAMTS20, ARHGAP10, GBP4, GBP5, GBP6, KYAT3
GO: 0098772	Molecular function regulator	ADAMTS20, ARHGAP10
GO: 0060089	Molecular transducer activity	OR7E24
GO: 0140110	Transcription regulator activity	ERG, NR3C2
GO: 0005215	Transporter activity	KCNJ15, SLC7A3
Accession Number	Cellular Component	Gene Symbol
GO: 0005623	Cell	EXOC3L4, KYAT3, TNFAIP2
GO: 0005576	Extracellular region	ADAMTS20
GO: 0043226	Organelle	ERG
Accession Number	Pathway	Gene Symbol
P00019	Endothelin signalling pathway	EDNRA
P00034	Integrin signalling pathway	ARHGAP10
P00047	PDGF signalling pathway	ARHGAP10, ERG

The CNVR\_11 and CNVR\_54 contained genes previously described as related to coat colour traits and showing CNV in goats [33,52]. The CNVR\_11 partially contained the *ADAMTS20* gene and presented the highest frequency value (0.430). The *ADAMTS20* (ADAM Metalloproteinase with Thrombospondin Type 1 Motif 20) gene is involved in binding (GO:0005488), catalytic activity (GO:0003824), and molecular function regulator (GO:0060089); in the cellular process (GO:0009987) and in extracellular region (GO:0005576) (Table 2). It was reported that this gene had a role in coat colour variation during goat domestication [53] and it was associated with melanocyte development [54]. Moreover, Liu et al. [33] reported for the first time the *ADAMTS20* CNV in goats, with a high frequency (0.4974) within the Eastern Mediterranean (EME) goat breeds. The presence of this gene within the CNVR detected in our breeds confirmed their results, considering that Arg, DdS, and Gir belong to the EME group.

The CNVR\_54 (frequency value 0.280) contain the *EDNRA*, *TMEM184C*, *PRMT9*, *ARHGAP10*, and *NR3C2* genes, which were mapped on a 1 Mb CNV in Boer goats [52]. Among them, the *EDNRA* and *ARHGAP10* genes were present in specific pathways (P00019, P00034, and P00047), and together with *NR3C2* were involved in biological regulation (GO:0065007), metabolic process (GO:008152), and several molecular functions (GO:0005488, GO:0003824, GO:0098772, and GO:0140110). The duplication of the *EDNRA* (endothelin receptor type A) gene was positively associated with the degree of white spotting in goats [52], while the variation in the number of copies of the *NR3C2*

(Nuclear Receptor Subfamily 3 Group C Member 2) gene was related to adaptation traits to harsh and dry environments; moreover, this gene is also involved in the very mild behaviour of the Gir goat breed [55].

Interestingly, among the 30 individuals carrying this CNVR\_54 (Table S4), the ones belonging to the solid-coloured DdS breed were absent, as well as 28 animals belonging to the Gir breed.

The CNVR\_33, CNVR\_65, and CNVR\_7 (frequency values of 0.411, 0.336, and 0.280, respectively) overlapped, with genes related to the immune system and belonging to several GO terms (Table 2). In particular, CNVR\_33 contain the T-cell receptors alpha chain V, which are important components in the adaptive immune system and that were described as the CNVs in Criollo Argentino horses [32] and in Brown Swiss dairy cattle [5]. The CNVR\_65 contain the *EXOC3L4* (exocyst complex component 3 like 4) and *TNFAIP2* (TNF alpha induced protein 2) genes, which are associated with natural antibodies and considered as candidate genes for innate host defence and disease resistance in dairy cows [56]. These genes are involved in the binding function (GO:0005488) and in cellular and localization processes (GO:0009987, GO:0051179). The CNVR\_7 contained several guanylate-binding proteins (*GBP-1*, *GBP-4*, *GBP-5*, *GBP-6*) that are proteins involved in the inflammatory process and host defence against cellular pathogens in mammals (GO:0002376 and GO:0050896) [57]. Moreover, the CNVR\_25 presented the *OR7E24* (olfactory receptor 7E24-like) and putative gustatory receptor genes, which are important to detect and identify a wide range of odours and chemosensory stimuli. It was reported by some authors [19,58,59] that these genes are duplicated within the bovine genome, suggesting their role in local adaptation. All these genes related with the immune system, host defence against pathogens, and olfactory receptors are consistent with the phenotypic characteristics of these breeds, which are well adapted to harsh environments and are resistant to infectious diseases.

Finally, it is interesting to note that the *GBP1*, *GBP6* (on CNVR\_7), and *SLC7A3* (cationic amino acid transporter 3) (on CNVR\_59) genes found by Zhang et al. [51] as CNVs were associated with high fertility and litter size traits in goats. These are noteworthy traits in Sicilian goat breeding farms, where goats often gave twin births and kids are used for meat production.

Although studies on goat CNVs have been previously published [33,43,51], this is the first study on the detection of copy number variations using the Illumina GoatSNP50 BeadChip array in Sicilian local goat breeds. It could be difficult to directly compare the CNV results across different studies. The use of different algorithms, different technologies, different criteria for filtering data, and differences in the numbers of samples and breeds tested could be responsible for the discrepancies. Moreover, all these factors could lead to a lack of overlap among the different studies on the same species. Nevertheless, our results could be considered reliable since the representation of the CNVRs in coat colour traits, immune related genes, and olfactory receptors was reported previously in livestock species [5,19,32,33,43,46]. Moreover, a biological link to traits, such as 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 the CNVRs, and as reported above, the goat breeds involved in this study are well adapted to harsh environments and are resistant to infectious diseases.

## 5. Conclusion

Additional analysis based on sequencing would help confirm our findings to further enhance the power to identify additional CNVs and genes in these goat breeds. Moreover, further studies would allow determining whether the CNVs could influence phenotypic variation, adaptiveness, and susceptibility/resistance to diseases in cases where their presence alter gene dosage, and may serve to improve the conservation and breeding programs for these local breeds.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Table S1: CNVs identified in the Sicilian goat breeds. Table S2: Unique CNVs identified in the Sicilian goat breeds. Table S3: CNVRs identified in the Sicilian goat breeds. Table S4: CNVRs identified for each Sicilian goat breed and

frequency per breed. Table S5: CNVRs matrix used for PCA analysis in the four goat breeds. Table S6: Genes identified within the CNVRs with frequency > 0.10.

**Author Contributions:** Conceptualization, R.D.G. and M.T.S.; methodology, R.D.G., S.M. and M.T.S.; software, R.D.G. and M.T.S.; validation, R.D.G. and M.T.S.; formal analysis, R.D.G. and M.T.S.; investigation, R.D.G., S.M., A.M., M.T., A.M.S. and M.T.S.; resources, B.P.; data curation, R.D.G., S.M. and M.T.S.; writing—original draft preparation, M.T.S.; writing—review and editing, R.D.G., S.M., A.M., M.T., A.M.S. and B.P.; visualization, R.D.G., S.M. and M.T.S.; supervision, B.P. and M.T.S.

**Acknowledgments:** This research was financed by LEO: Livestock Environment Opendata, PSRN 2014-2020 - Sottomisura: 16.2, Project number PRJ-0185, CUP: J84I18000090007.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Feuk, L.; Carson, A.R.; Scherer, S.W. Structural variation in the human genome. *Nat. Rev. Genet.* **2006**, *7*, 85–97.
2. Scherer, S.W.; Lee, C.; Birney, E.; Altshuler, D.M.; Eichler, E.E.; Carter, N.P.; Hurles, M.E.; Feuk, L. Challenges and standards in integrating surveys of structural variation. *Nat. Genet.* **2007**, *39*, S7–S15.
3. Henshall, J.M.; Whan, V.A.; Norris, B.J. Reconstructing CNV genotypes using segregation analysis: Combining pedigree information with CNV assay. *Genet. Sel. Evol.* **2010**, *42*, 34, doi:10.1186/1297-9686-42-34.
4. Pierce, M.D.; Dzama, K.; Muchadeyi, F.C. Genetic diversity of seven cattle breeds inferred using copy number variations. *Front. Genet.* **2018**, *9*, 163, doi.org/10.3389/fgene.2018.00163.
5. Prinsen, R.T.M.M.; Rossoni, A.; Gredler, B.; Bieber, A.; Bagnato, A.; Strillacci, M.G. A genome wide association study between CNVs and quantitative traits in Brown Swiss cattle. *Livest. Sci.* **2017**, *202*, 7–12.
6. Conrad, D.F.; Pinto, D.; Redon, R.; Feuk, L.; Gokcumen, O.; Zhang, Y.; Aerts, J.; Andrews, T.D.; Barnes, C.; Campbell, P.; et al. Origins and functional impact of copy number variation in the human genome. *Nature* **2010**, *464*, 704–712.
7. Zhang, F.; Gu, W.; Hurles, M.E.; Lupski, J.R. Copy number variation in human health, disease, and evolution. *Annu. Rev. Genom. Hum. Genet.* **2009**, *10*, 451–481.
8. Hou, Y.; Bickhart, D.M.; Hvinden, M.L.; Li, C.; Song, J.; Boichard, D.A.; Fritz, S.; Eggen, A.; DeNise, S.; Wiggans, G.R.; et al. Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. *BMC Genom.* **2012**, *13*, 376, doi:10.1186/1471-2164-13-376.
9. Alkan, C.; Coe, B.P.; Eichler, E.E. Applications of next-generation sequencing genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **2011**, *12*, 363–375.
10. Zhang, L.; Jia, S.; Yang, M.; Xu, Y.; Li, C.; Sun, J.; Huang, Y.; Lan, X.; Lei, C.; Zhou, Y.; et al. Detection of copy number variations and their effects in Chinese bulls. *BMC Genom.* **2014**, *15*, 480, doi:10.1186/1471-2164-15-480.
11. Peiffer, D.A.; Le, J.M.; Steemers, F.J.; Chang, W.; Jenniges, T.; Garcia, F.; Haden, K.; Li, J.; Shaw, C.A.; Belmont, J.; et al. High-resolution genomic profiling of chromosomal aberrations using inifinium whole-genome genotyping. *Genome Res.* **2006**, *16*, 1136–1148.
12. Bhanuprakash, V.; Chhotaray, S.; Pruthviraj, D.R.; Rawat, C.; Karthikeyan, A.; Panigrahi, M. Copy number variation in livestock: A mini review. *Vet. World* **2018**, *11*, 535–541.
13. Liu, M.; Woodward-Greene, J.; Kang, X.; Pan, M.G.; Rosen, B.; van Tassell, C.P.; Chen, H.; Liu, G.E. Genome-wide CNV analysis revealed variants associated with growth traits in African indigenous goats. *Genomics* **2020**, *112*, 1477–1480.
14. Marenne, G.; Chanock, S.; Rothman, N.; Rodríguez-Santiago, B.; Rico, D.; Pita, G.; Pérez-Jurado, L.; Valencia, A.; Jacobs, K.; Pisano, D.G.; et al. CNV assessment from illumina inifinium 1M platform: Agreement according to algorithm and source of DNA. *Ann. Hum. Genet.* **2009**, *73*, 658–669.
15. Yavaş, G.; Koyuturk, M.; Ozsoyoglu, M.; Gould, M.P.; LaFramboise, T. An optimization framework for unsupervised identification of rare copy number variation from SNP array data. *Genome Biol.* **2009**, *10*, R119, doi:10.1186/gb-2009-10-10-r119.
16. Seroussi, E.; Glick, G.; Shirak, A.; Yakobson, E.; Weller, J.I.; Ezra, E.; Zeron, Y. Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. *BMC Genomics* **2010**, *11*, 673, doi:10.1186/1471-2164-11-673.

17. Ma, Q.; Liu, X.; Pan, J.; Ma, L.; Ma, Y.; He, X.; Zhao, Q.; Pu, Y.; Li, Y.; Jiang, L. Genome-wide detection of copy number variation in Chinese indigenous sheep using an ovine high-density 600 K SNP array. *Sci. Rep.* **2017**, *7*, 912, doi:10.1038/s41598-017-00847-9.
18. Jiang, L.; Jiang, J.; Yang, J.; Liu, X.; Wang, J.; Wang, H.; Ding, X.; Liu, J.; Zhang, Q. Genome-wide detection of copy number variations using highdensity SNP genotyping platforms in Holsteins. *BMC Genom.* **2013**, *14*, 131, doi:10.1186/1471-2164-14-131.
19. di Gerlando, R.; Sardina, M.T.; Tolone, M.; Sutura, A.M.; Mastrangelo, S.; Portolano, B. Genome-wide detection of copy-number variations in local cattle breeds. *Anim. Prod. Sci.* **2019**, *59*, 815, doi:10.1071/an17603.
20. Wang, M.D.; Dzama, K.; Hefer, C.A.; Muchadeyi, F.C. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. *BMC Genom.* **2015**, *16*, 894, doi:10.1186/s12864-015-2122-z.
21. Wu, Y.; Fan, H.; Jing, S.; Xia, J.; Chen, Y.; Zhang, L.; Gao, X.; Li, J.; Gao, H.; Ren, H. A genome-wide scan for copy number variations using high-density single nucleotide polymorphism array in Simmental cattle. *Anim. Genet.* **2015**, *46*, 289–298.
22. da Silva, J.M.; Giachetto, P.F.; da Silva, L.O.; Cintra, L.C.; Paiva, S.R.; Yamagishi, M.E.; Caetano, A.R. Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. *BMC Genom.* **2016**, *17*, 454, doi:10.1186/s12864-016-2752-9.
23. Prinsen, R.T.M.M.; Strillacci, M.G.; Schiavini, F.; Santus, E.; Rossoni, A.; Maurer, V.; Bieber, A.; Gredler, B.; Dolezal, M.; Bagnato, A. A genome-wide scan of copy number variants using high-density SNPs in Brown Swiss dairy cattle. *Livest. Sci.* **2016**, *191*, 153–160.
24. Strillacci, M.G.; Gorla, E.; Cozzi, M.C.; Vevey, M.; Genova, F.; Scienski, K.; Longeri, M.; Bagnato, A. A copy number variant scan in the autochthonous Valdostana Red Pied cattle breed and comparison with specialized dairy populations. *PLoS ONE* **2018**, *13*, e0204669, doi:10.1371/journal.pone.0204669.
25. Liu, J.; Zhang, L.; Xu, L.; Ren, H.; Lu, J.; Zhang, X.; Zhang, S.; Zhou, X.; Wei, C.; Zhao, F.; et al. Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genom.* **2013**, *14*, 229, doi:10.1186/1471-2164-14-229.
26. Ma, Y.; Zhang, Q.; Lu, Z.; Zhao, X.; Zhang, Y. Analysis of copy number variations by SNP50 BeadChip array in Chinese sheep. *Genomics* **2015**, *106*, 295–300.
27. Zhu, C.; Fan, H.; Yuan, Z.; Hu, S.; Ma, X.; Xuan, J.; Wang, H.; Zhang, L.; Wei, C.; Zhang, Q.; 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.* **2016**, *6*, 27822, doi:10.1038/srep27822.
28. Yan, J.; Blair, H.T.; Liu, M.; Li, W.; He, S.; Chen, L.; Dittmer, K.E.; Garrick, D.J.; Biggs, P.J.; Dukkupati, V.S.R. Genome-wide detection of autosomal copy number variants in several sheep breeds using Illumina OvineSNP50 BeadChips. *Small Rumin. Res.* **2017**, *155*, 24–32.
29. Yang, L.; Xu, L.; Zhou, Y.; Liu, M.; Wang, L.; Kijas, J.W.; Zhang, H.; Li, L.; Liu, G.E. Diversity of copy number variation in a worldwide population of sheep. *Genomics* **2018**, *110*, 143–148.
30. di Gerlando, R.; Sutura, A.M.; Mastrangelo, S.; Tolone, M.; Portolano, B.; Sottile, G.; Bagnato, A.; Strillacci, M.G.; Sardina, M.T. Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. *PLoS ONE* **2019**, *14*, e0215204, doi:10.1371/journal.pone.0215204.
31. Wang, Y.; Tang, Z.; Sun, Y.; Wang, H.; Wang, C.; Yu, S.; Liu, J.; Zhang, Y.; Fan, B.; Li, K.; et al. Analysis of genome-wide copy number variations in Chinese indigenous and western pig breeds by 60 K SNP genotyping arrays. *PLoS ONE* **2014**, *9*, e106780, doi:10.1371/journal.pone.0106780.
32. Corbi-Botto, C.M.; Morales-Durand, H.; Zappa, M.E.; Sadaba, S.A.; Peral-García, P.; Giovambattista, G.; Díaz, S. Genomic structural diversity in Criollo Argentino horses: Analysis of copy number variations. *Gene* **2019**, *695*, 26–31.
33. Liu, M.; Zhou, Y.; Rosen, B.D.; Van Tassell, C.P.; Stella, A.; Tosser-Klopp, G.; Rupp, R.; Palière, I.; Colli, L.; Sayre, B., et al. Diversity of copy number variation in the worldwide goat population. *Heredity* **2019**, *122*, 636–646.
34. Pielberg, G.; Olsson, C.; Syvänen, A.-C.; Andersson, L. Unexpectedly high allelic diversity at the KIT locus causing dominant white color in the domestic pig. *Genetics* **2002**, *160*, 305–311.
35. Norris, B.J.; Whan, V.A. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep *Genome Res.* **2008**, *18*, 1282–1293.

36. Durkin, K.; Coppieters, W.; Drögüller, C.; Ahariz, N.; Cambisano, N.; Druet, T.; Fasquelle, C.; Haile, A.; Horin, P.; Huang, L.; et al. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature* **2012**, *482*, 81–84.
37. Brenig, B.; Beck, J.; Floren, C.; Bornemann-Kolatzki, K.; Wiedemann, I.; Hennecke, S.; Swalve, H.; Schütz, E. Molecular genetics of coat colour variations in white galloway and white park cattle. *Anim. Genet.* **2013**, *44*, 450–453.
38. Xu, L.; Hou, Y.; Bickhart, D.M.; Zhou, Y.; Hay, E.H.A.; Song, J.; Sonstegard, T.S.; Van Tassell, C.P.; Liu, G.E. Population-genetic properties of differentiated copy number variations in cattle. *Sci. Rep.* **2016**, *6*, 23161, doi:10.1038/srep23161.
39. Ma, Y.L.; Wen, Y.F.; Cao, X.K.; Cheng, J.; Huang, Y.Z.; Ma, Y.; Hu, L.Y.; Lei, C.Z.; Qi, X.L.; Cao, H.; et al. Copy number variation (CNV) in the IGF1R gene across four cattle breeds and its association with economic traits. *Arch. Anim. Breed.* **2019**, *62*, 171–179.
40. Kadri, N.K.; Sahana, G.; Charlier, C.; Iso-Touru, T.; Guldbbrandtsen, B.; Karim, L.; Nielsen, U.S.; Panitz, F.; Aamand, G.P.; Schulman, N.; et al. 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 Genet.* **2014**, *10*, e1004049, doi:10.1371/journal.pgen.1004049.
41. Zhang, R.-Q.; Wang, J.-J.; Zhang, T.; Zhai, H.-L.; Shen, W. Copy-number variation in goat genome sequence: A comparative analysis of the different litter size trait groups. *Gene* **2019**, *696*, 40–46.
42. Qanbari, S.; Pausch, H.; Jansen, S.; Somel, M.; Strom, T.M.; Fries, R.; Nielsen, R.; Simianer, H. Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* **2014**, *10*, e1004148, doi:10.1371/journal.pgen.1004148.
43. Mastrangelo, S.; Tolone, M.; Montalbano, M.; Tortorici, L.; Di Gerlando, R.; Sardina, M.T., Portolano, B. Population genetic structure and milk production traits in Girgentana goat breed. *Anim. Prod. Sci.* **2017**, *57*, 430–440.
44. Miller, S.A.; Dykes, D.D.; Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **1988**, *16*, 1215, doi:10.1093/nar/16.3.1215.
45. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575.
46. Wang, K.; Li, M.; Hadley, D.; Liu, R.; Glessner, J.; Grant, S.F.; Hakonarson, H.; Bucan, M. PennCNV: An integrated hidden Markov model designed for high resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **2007**, *17*, 1665–1674.
47. Redon, R.; Ishikawa, S.; Fitch, K.R.; Feuk, L.; Perry, G.H.; Andrews, T.D.; Fiegler, H.; Shapero, M.H.; Carson, A.R.; Chen, W.; et al. Global variation in copy number in the human genome. *Nature* **2006**, *444*, 444–454.
48. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842.
49. Hammer, Ø.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **2001**, *4*, 9.
50. Sardina, M.T.; Ballester, M.; Marmi, J.; Finocchiaro, R.; Van Kaam, J.B.C.; Portolano, B.; Folch, J.M. Phylogenetic analysis of Sicilian goats reveals a new mtDNA lineage. *Anim. Genet.* **2006**, *37*, 376–378.
51. Siwek, M.; Finocchiaro, R.; Curik, I.; Portolano, B. Hierarchical structure of the Sicilian goats revealed by Bayesian analyses of microsatellite information. *Anim. Genet.* **2011**, *42*, 93–95.
52. Menzi, F.; Keller, I.; Reber, I.; Beck, J.; Brenig, B.; Schütz, E.; Leeb, T.; Drögemüller, C. Genomic amplification of the caprine EDNRA locus might lead to a dose dependent loss of pigmentation. *Sci. Rep.* **2016**, *6*, 28438, doi:10.1038/srep28438.
53. Dong, Y.; Zhang, X.; Xie, M.; Arefnezhad, B.; Wang, Z.; Wang, W.; Feng, S.; Huang, G.; Guan, R.; Shen, W.; et al. Reference genome of wild goat (*Capra aegagrus*) and sequencing of goat breeds provide insight into genetic basis of goat domestication. *BMC Genom.* **2015**, *16*, 431.
54. Crepaldi, P.; Nicoloso, L. SNPs in coat colour genes in goats. *Ital. J. Anim. Sci.* **2007**, *6*, 91–93.
55. Fontanesi, L.; Martelli, P.L.; Beretti, F.; Riggio, V.; Dall’Olio, S.; Colombo, M.; Casadio, R.; Russo, V.; Portolano, B. An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genom.* **2010**, *11*, 639.

56. de Klerk, B.; Emam, M.; Thompson-Crispi, K.A.; Sargolzaei, M.; van der Poel, J.J.; Mallard, B.A. A genome-wide association study for natural antibodies measured in blood of Canadian Holstein cows. *BMC Genom.* **2018**, *19*, 624, doi:10.1186/s12864-018-5062-6.
57. Tretina, K.; Park, E.S.; Maminska, A.; MacMicking, J.D. Interferon-induced guanylate-binding proteins: Guardians of host defense in health and disease. *J. Exp. Med.* **2019**, *216*, 482–500.
58. Nei, M.; Niimura, Y.; Nozawa, M. The evolution of animal chemosensory receptor gene repertoires: Roles of chance and necessity. *Nat. Rev. Genet.* **2008**, *9*, 951–963.
59. Mastrangelo, S.; Sardina, M.T.; Tolone, M.; Di Gerlando, R.; Sutera, A.M.; Fontanesi, L., Portolano, B. Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. *Animal* **2018**, *12*, 2480–2488.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).