Genomics 97 (2011) 158–165

Contents lists available at ScienceDirect

Genomics

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

A first comparative map of copy number variations in the sheep genome

L. Fontanesi [a,*], F. Beretti [a,b], P.L. Martelli [c], M. Colombo [a], S. Dall'Olio [a], M. Occidente [d], B. Portolano [b], R. Casadio [c], D. Matassino [d], V. Russo [a]

[a] DIPROVAL, Sezione di Allevamenti Zootecnici, University of Bologna, Via F.lli Rosselli 107, 42123 Reggio Emilia, Italy
[b] Dept. S.En.Fi.Mi.Zo., Sezione di Produzioni Animali, University of Palermo, Viale delle Scienze – Parco d'Orleans, 90128 Palermo, Italy
[c] Biocomputing Group, Dept. of Experimental and Evolutionary Biology, University of Bologna, Via San Giacomo 9/2, 40126 Bologna, Italy
[d] ConSDABI – Italian Sub National Focal Point FAO AnGR (Mediterranean Biodiversity), 82100 Benevento, Italy

ARTICLE INFO

Article history:
Received 8 September 2010
Accepted 16 November 2010
Available online 24 November 2010

Keywords:
aCGH
Comparative map
Copy number variation
Ovis aries
Ruminants
Sheep breeds

ABSTRACT

We carried out a cross species cattle–sheep array comparative genome hybridization experiment to identify copy number variations (CNVs) in the sheep genome analysing ewes of Italian dairy or dual-purpose breeds (Bagnolese, Comisana, Laticauda, Massese, Sarda, and Valle del Belice) using a tiling oligonucleotide array with ~385,000 probes designed on the bovine genome. We identified 135 CNV regions (CNVRs; 24 reported in more than one animal) covering ~10.5 Mb of the virtual sheep genome referred to the bovine genome (0.398%) with a mean and a median equal to 77.6 and 55.9 kb, respectively. A comparative analysis between the identified sheep CNVRs and those reported in cattle and goat genomes indicated that overlaps between sheep and both other species CNVRs are highly significant ($P<0.0001$), suggesting that several chromosome regions might contain recurrent interspecies CNVRs. Many sheep CNVRs include genes with important biological functions. Further studies are needed to evaluate their functional relevance.

1. Introduction

With the advent of genome wide analysis methods to detect DNA content, such as array comparative genome hybridization (aCGH) [1,2], it has been soon realized that copy number variations (CNVs) (DNA segments ranging from 1 kb to few Mb that present a variable copy number in comparison to a reference genome [3,4]) represent a significant source of genetic variability [5,6]. For example, combining several studies, it appears that about 30% of the human genome is affected by CNVs [7]. CNVs can change gene structure and dosage, can regulate gene expression and function and for these reasons they can potentially have important effects in determining phenotypic differences [8–10]. Many CNVs represent benign polymorphic variants, whereas several others cause or are associated with both Mendelian or complex genetic diseases and traits in humans [11–13]. Following the pioneering studies carried out in humans, genome wide investigations of CNVs have been reported in mouse [14–18], rat [19,20] chimpanzee [21,22], rhesus macaque [23], dog [24,25], pig [26], cattle [27–31], goat [32], and chicken [33] indicating that this variability is widespread in vertebrates as well as in other inferior species [34,35]. Investigations in goat and chimpanzee have been carried out by cross-species aCGH experiments that used the bovine and human genomes as hybridization platforms, considering the closeness between the two ruminant and the two primate species, respectively. Cross-species experiments have been also applied using chicken based arrays to identify CNVs in turkey [36], duck [37], and zebra-finch [38] genomes. Moreover, comparative analyses of the primate (human, chimpanzee and macaque) and ruminant genomes (cattle and goat) indicated that recurrent CNVs occur across close species, suggesting that ancestral segmental duplications or other mechanisms may facilitate CNV formation [21–23,32,39,40].

A few other studies carried out in livestock have shown that CNV affecting genes or gene regions are associated with several phenotypic traits. For example, the *Dominant white* locus in pigs includes alleles determined by duplications of the *KIT* gene [41,42]. CNV in intron 1 of the *SOX5* gene causes the pea-comb phenotype in chicken [43] and the late feathering locus in this avian species includes a partial duplication of the *PRLR* and *SPEF2* genes [44]. CNV affects also the *Agouti* locus in sheep and goats and contributes to the variability of coat colour in these two species [40,45,46].

The sheep genome has been the matter of several refinements starting from the development of genetic, cytogenetic and radiation hybrid maps [47–52]. A virtual genome map based on bovine, dog and human genome assemblies has been recently produced, making use of the homology and synteny conservation between the two ruminant species and the more advanced information available for the compared genomes [53]. Then a first draft of the ovine genome has been produced by the International Sheep Genomics Consortium, but improvements in terms of assembling and annotation are still needed [54]. Furthermore, low and medium density single nucleotide

* Corresponding author. Fax: +39 0522 290523.
E-mail address: luca.fontanesi@unibo.it (L. Fontanesi).

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.11.005

polymorphism (SNP) panels have been recently developed and used in a Sheep HapMap project across a large number of breeds in order to evaluate the extent of variability in this species [54,55]. However, genetic diversity in the sheep genome at the level of CNVs has not been so far evaluated.

Here we applied a cross-species aCGH experiment based on the bovine genome and we obtained an initial comparative map of CNVs of the sheep genome.

2. Results and discussion

2.1. CNVs in the sheep genome

In this study we made use of the high conservation and homology between cattle and sheep genomes [50,53,56] determined by their phylogenetic closeness [57–59] to identify CNVs in sheep applying the same approach we carried out to identify CNVs in the goat genome

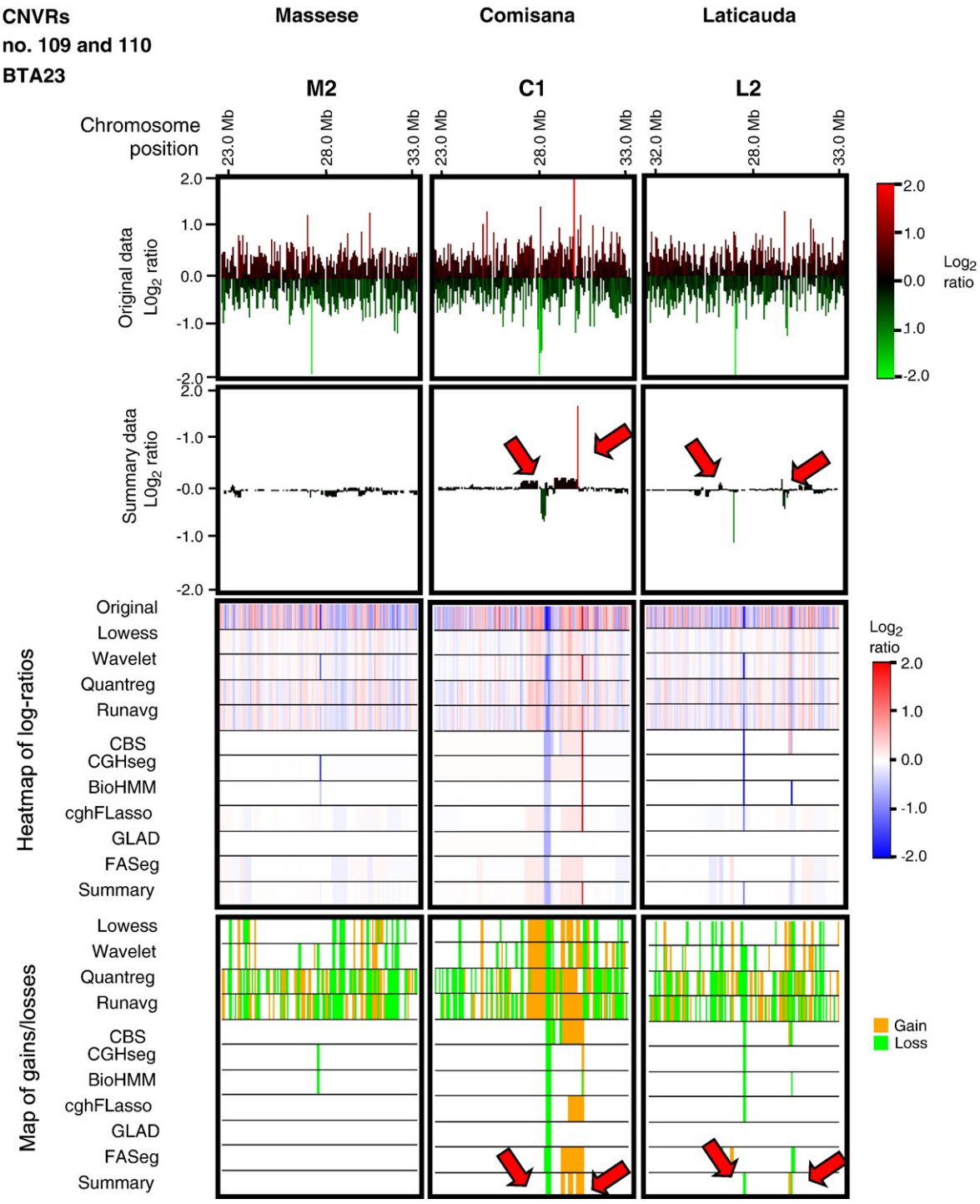

Fig. 1. Examples of aCGH results. Results were reported for three sheep indicated at the top in correspondence of the related images. Images have been reported for i) $\log_2$ ratio plot of original data, ii) $\log_2$ ratio plot of summary data (pointwise averaging of all computed profiles), iii) heatmap of $\log_2$ ratios for original, smoothed/segmented, and summary data, and iv) maps of gains/losses for smoothed/segmented and summary data (gain is indicated in orange, and loss is indicated in green). Red arrows indicate regions of copy gain/loss. Smoothed/segmented data were obtained with several algorithms (Lowess, Wavelet, Quantreg, ruavg, CBS, CGHseg, BioHMM, cghFLasso, GLAD, and FASeg) averaged in the summary data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of CNVs identified in the analysed sheep.

Breeds	No. of sheep	Number of CNVs					CNV average size (kb)
		Mean	Total	Unique	Gain	Loss	
Bagnolese	2	21.5	43	16	22	21	99.2
Comisana	2	20	40	12	24	16	76.8
Laticauda	2	16	32	16	10	22	68.2
Massese	2	13.5	27	14	8	19	52.5
Sarda	2	13.5	27	12	8	19	75.2
Valle del Belice	1	17	17	5	7	10	70.0
Total	11	16.9	186	75	79	107	73.9

[32]. We used a custom tiling array including ~385,000 oligonucleotide probes designed on the Btau_4.0 version of the *Bos taurus* genome assembly [60] and analysed genomic DNA samples of 11 ewes belonging to 6 different Italian dairy or dual-purpose sheep breeds (2 Bagnolese, 2 Comisana, 2 Laticauda, 2 Massese, 2 Sarda, and 1 Valle del Belice) compared to the reference DNA of another Sarda ewe.

Specific criteria were used to call CNVs in this cross-species aCGH experiment as already adopted and validated in the cattle–goat aCGH study [32] (see Materials and methods). CNVs were reported using 10 different algorithms developed for data segmentation and averaging their results as implemented in the CGHweb server [61]. Fig. 1 shows a few examples of CNVs identified in different sheep. Using this conservative approach we reported a total of 186 CNVs (Table 1 and Supplemental file 1). About 61% of the CNVs were observed in only one animal and on average their size was about 73.9 kb (Table 1). The largest mean number of CNVs was identified in the Bagnolese breed (no. 21.5), whereas the lowest mean number was reported for Sarda and Massese breeds (no. 13.5). The lowest number observed for the Sarda sheep was expected as the reference DNA was from another animal of the same breed. This result is in line with what has been reported for mouse, dog, and goat aCGH studies in which breeds/lines of the same subjects used as reference showed the lowest number of CNVs, as low genetic differences among animals is derived by their closeness [15,24,32]. These data can indirectly indicate that CNVs might be useful for population genetic analyses and breed/line characterization and, on the other hand, suggest that most of our CNVs are correctly identified. To further support the latter issue, we validated by semiquantitative multiplex fluorescent PCR (SQF-PCR) three CNVs identified on BTA4, BTA19 and BTA23 (using the cattle as reference chromosomes) (Table S1).

Considering all analysed sheep, we identified on average 16.9 CNVs for each genome. This value is comparable to what we reported in the cattle–goat study (no. = 17.9) [32]. Similar results have been also reported in other aCGH experiments using arrays with ~385,000 oligonucleotide probes that analysed the cattle [30], dog [24], and chicken [33] genomes in which 11.6, 17.2, and 9.6 CNVs were evidenced for each animal in these three species, respectively (Table 2). Two CNVs (not overlapped to any other CNV) were also called in the self-self hybridization providing a rough estimation of the experimental false discovery rate (FDR) among the identified CNVs (FDR = 12%). This is almost the same value we reported in the cattle–goat experiment (FDR = 11%) [32]. However, technical issues, such as heterogeneity of DNA quality among different samples and sequence divergence between the reference genome and the hybridized test and reference DNAs make it difficult to precisely estimate the experimental FDR. In other experiments using homologous DNA hybridizations, estimation of the experimental FDR using the self-self hybridization data ranged from about 3% in dogs [25] to 24% in human [62].

CNV regions (CNVRs) were determined by aggregating overlapping CNVs identified in different animals as previously reported [30,32,62] and considering a conservative approach due to the specificity of our experiment (see Materials and methods and [32] for details). On the whole, we detected 135 CNVRs covering about 10.5 Mb (0.398%) of the virtual sheep genome referred to the bovine genome, version Btau_4.0 (Fig. 2 and Supplemental file 2), considering the 29 autosomes and the X chromosome but not the chrUnAll (unassembled scaffolds) that was not included in the tiling array due to difficulties in interpreting the results [30].

These 135 CNVRs included 75 loss, 59 gain, and one with both events (Supplemental file 2). Of these CNVRs, 18 were found in multiple animals of different breeds, five were observed in multiple animals of both the same and different breeds and one was identified in multiple animals of the same breed only (Laticauda; CNVR no. 112). All CNVRs reported in more than one animal (~18%) were defined as high confidence CNVRs. All other events (no. = 111) were found only in one animal and were considered as second level confidence CNVRs (Supplemental file 2). Mean and median of the 135 CNVRs were equal to 77,621 and 55,876 bp, respectively, ranging from 24,697 (BTA18, CNVR no. 93) to 505,055 bp (BTA12, CNVR no. 72). These values are on the same scale of those reported for the goat genome (mean and median equal to 90,292 and 49,530 bp, respectively [32]). Using a similar aCGH platform Liu et al. [30] reported that in cattle the mean

Table 2
Comparison between the results of this CNV study in sheep and other similar studies using aCGH in mammalian and avian species. Modified from [32].

Species	No. of individuals	aCGH platforms	Mean probe spacing (kb)	Total no. of CNVs	Mean no. of CNVs per individual	Total no. of CNVRs	CNVR mean size (kb)	References
Sheep	11	385 k oligo aCGH	6.3	186	16.9	135	77.1	This study
Goat	9	385 k oligo aCGH	6.3	161	17.9	127	90.3	[32]
Cattle	90	385 k oligo aCGH	~6	1041	11.6	177	158.6	[30]
Cattle	20	6.3 million oligo aCGH	0.4	–	–	304	72.0	[31]
Human	270	BAC aCGH[2]	–	–	–	913	228.0	[62]
Human	40	42 million oligo aCGH	0.06	51,997	1300	–	2.9[a]	[13]
Chimpanzee	20	BAC aCGH[b]	–	355	17.8	–	–	[21]
Macaque	9	385 k oligo aCGH	6.5	214	21.4	123	101.2	[23]
Dog	9	385 k oligo aCGH	4.7[c]	155	17.2	60	309.5	[24]
Mouse	21[d]	385 k oligo aCGH	~5	80	2–38[e]	–	271.5	[15]
Mouse	20[d]	2.1 million oligo aCGH	1	10,681	26.4–48.3[e]	3359	64.0	[9]
Rat	3	385 k oligo aCGH	~5[f]	33	11	33	256.0	[19]
Chicken	10	385 k oligo aCGH	2.6	96	9.6	–	166.7	[33]

[a] Median size.
[b] Whole Genome TilePath array comprising 26,574 large insert clones.
[c] Median probe spacing.
[d] No. of strains (2–6 individuals per strains).
[e] Depending on the strain.
[f] Considering nonrepetitive parts of the genome.

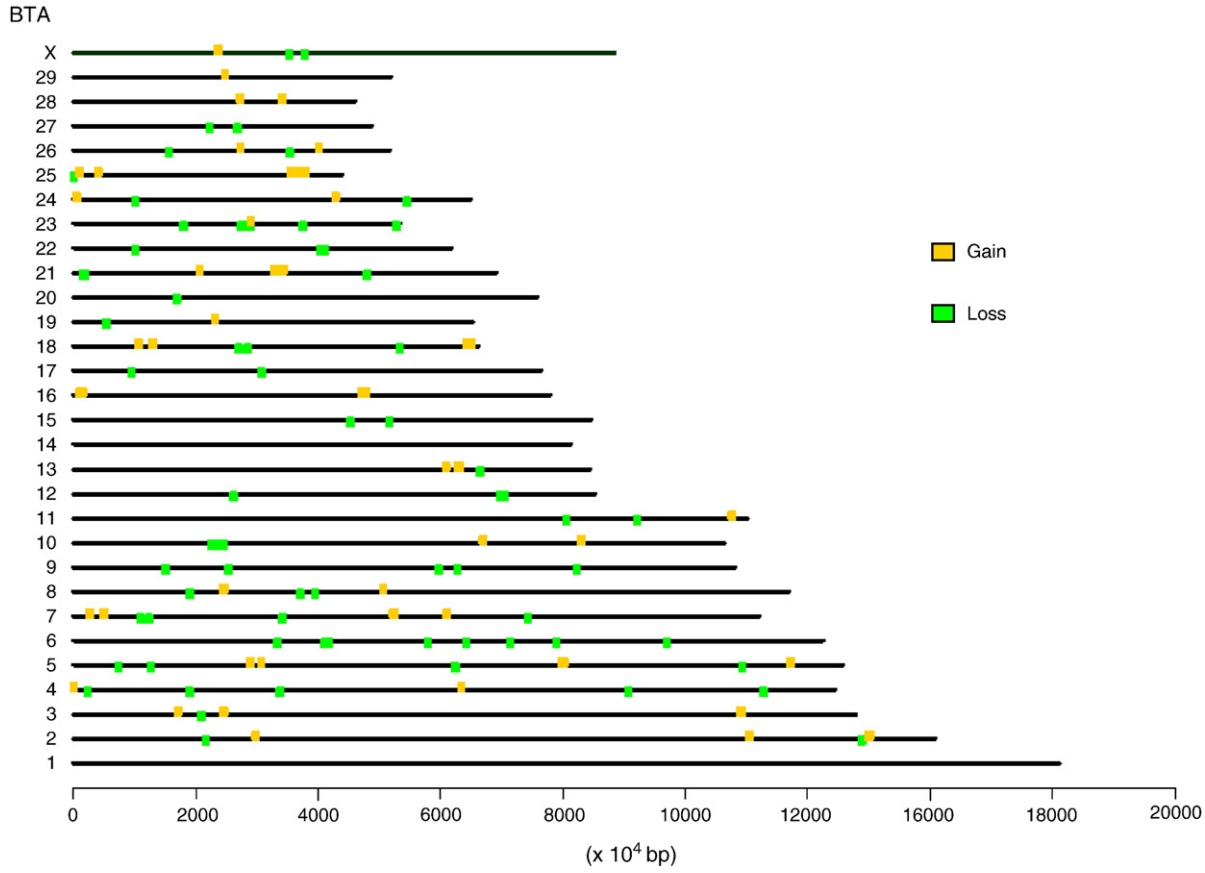

Fig. 2. Map of CNVRs identified in sheep reported on the bovine chromosomes.

and median of the 177 highly confident reported CNVRs were equal to 159 and 89 kb, respectively. These differences between cattle and sheep/goat CNVs identified by aCGH with similar genome resolution might be due to i) the method used to call CNVs between our cross-species experiments and Liu et al. [30] investigation or ii) true differences between species. Another factor might be derived from the extent of analysed animals and their distance/relationship within species that might increase the possibility to detect larger CNVRs.

As in several other aCGH experiments [26,30–32,62], in our sheep data the number of loss events was larger than the number of gain events. This of course may depend on the reference DNA used. However, it seems that the aCGH detection methods might favour the identification of deletions [26,30–32,62] but, to our knowledge, no systematic analysis of this possible bias has been undertaken yet.

2.2. Comparative analysis of CNVRs between sheep and other ruminants

To evaluate if CNVRs we identified in sheep overlap with CNVRs reported in other ruminants, we compared our results with those obtained in five independent cattle or goat experiments [28–32] (Table 3; Supplemental file 2). Three experiments (two in cattle and one in goat) were carried out using aCGH with ~385,000 tiling oligonucleotides [30,32] or including 6.3 million of probes [31]. Two other cattle experiments used the Illumina BovineSNP50 BeadChip containing about 50K SNPs [28,29]. Overlapping between aCGH results obtained in sheep and goat was highly significant ($P<0.0001$). The same significant overlapping results were obtained between sheep CNVRs and cattle CNVRs reported by Bae et al. [29] and Liu et al. [30] ($P<0.0001$). When we compared the merged cattle CNVR set (obtained combining the four cattle experiments), overlapping between sheep and cattle CNVRs was confirmed to be highly significant ($P<0.0001$). Comparing goat and cattle CNVRs, we obtained the same results [32], further suggesting that putative recurrent CNVRs are present across different species of the *Bovidae* family. Significant overlap of CNVRs among different species has been also observed comparing the human with both chimpanzee and rhesus macaque genomes [21–23,39]. These two non-human primate species diverged from the human lineage about 6 and 25 million years ago [63,64]. Together these results suggest that certain genomic regions are prone to recurrent CNV formation and instability in both the primate and the *Artiodactyla* evolutionary lineages. Cattle, sheep, and goat share a common ancestor in the early Miocene about

Table 3
Actual overlaps and P value derived from the permutation tests estimating the significance of the number of overlaps between sheep CNVRs (no. = 135) and different CNVR datasets in goat and cattle.

CNVR dataset/ reference	Method of analysis	No. of CNVRs[a]	No. of actual overlap	P
Goat — Fontanesi et al. [32]	aCGH	127	10	$<10^{-4}$
Cattle — Matukumalli et al. [28]	Illumina BovineSNP50 BeadChip	37	2	0.2
Cattle — Bae et al. [29]	Illumina BovineSNP50 BeadChip	368	17	$<10^{-4}$
Cattle — Liu et al. [30]	aCGH	177	18	$<10^{-4}$
Cattle — Fadista et al. [31]	aCGH	266	7	0.06
Cattle (merged CNVRs)[b]	aCGH + Illumina BovineSNP50 BeadChip	764	33	$<10^{-4}$

[a] No. of CNVRs reported in the corresponding goat or cattle datasets mapped to the Btau_4.0 cattle genome assembly (see Materials and methods).
[b] Cattle CNVRs obtained merging the four cattle dataset reported above [28–31].

17–23 million years ago [57–59] whereas sheep and goat lineages separated about 6–14 million of years before present [65]. It could be possible that family or lineage specific ancestral segmental duplications are the source of these recurrent CNVRs. As a matter of fact segmental duplications have been shown to significantly overlap with CNVRs in cattle [30,31] as well as in several other species [15,18,23,25]. As soon as high quality assembled sheep and goat genomes are available, it will be interesting to compare segmental duplications across ruminants and evaluate the occurrence of CNVs in these regions.

Considering the bovine chromosomes, sheep CNVRs were not reported only for BTA1 and BTA14 (Fig. 2). BTA5 contained 11 CNVRs (the largest number) as also reported for the goat–cattle aCGH analysis [32]. BTA25 and BTA23 were the most covered by CNVRs (1.44% and 0.94%, respectively), whereas the most enriched in goat were BTA17, BTA10 and BTA18 (1.6%, 1.3% and 1.0% of their length, respectively [32]) and the most enriched in cattle (combining the four mentioned studies) were BTA28 (7.81%), BTA15 (7.35%) and BTA27 (7.22%) (data not shown). Differences of coverage might be mainly due to the number of studies reported in cattle.

2.3. Annotation of sheep CNVRs

Considering the 135 sheep CNVRs, 83 of them (61.5%) partially or completely spanned cattle Ensembl annotated genes (Btau_4.0 version), including 290 Ensembl transcripts representing 185, 160, and 172 gene ontology (GO) categories for molecular function, cellular components and biological processes, and 265 different PANTHER terms (Supplemental file 3 and data not shown). For 188 out of 228 Ensembl cattle gene we retrieved a human orthologous gene (Supplemental file 4). Mutations in only 25 of these genes cause Mendelian disorders in humans and one is associated with longevity, as reported in OMIM database (Supplemental file 4). None of these 228 cattle genes is involved in any reported genetic disease in sheep, goat, or cattle. Nevertheless, it could be expected that a few CNVRs with gain or loss of partial or complete copies of genes might play a role in defining differences among animals and/or breeds [8–12].

Several GO categories are significantly ($P<0.001$ or $P<0.05$) over-represented in sheep CNVRs (Table 4) even if it seems that there are fewer over-represented groups compared to other CNV studies [16,24,25,32,62]. This might be due to differences of the sheep genome or to bias due to the analysed animals and reference. However, it is interesting to note that, in general, lipid metabolism processes and antigen presentation are over-represented GO categories. In dairy sheep, lipid metabolism is a key process for milk production and milk fatty acid composition is an important milk quality trait. A few QTL for milk fat content and composition have been already mapped in sheep even if using, in most cases, different breeds than those we analysed in this CNV study (e.g. [66]). There are overlaps or partial overlaps between these QTL as well as QTL for several other traits (data not shown) and CNVRs even if for the large confidence intervals of these QTL studies it is difficult to evaluate if CNVs are involved in determining variability for these production traits. However, it is also worth to mention that a few regions of the ovine major histocompatibility complex (MHC) are affected by CNVs (Fig. 1). CNVR no. 109 includes a region that is affected by CNV also in goat [32] and cattle (as we derived from the data reported by Liu et al. [30] and Fadista et al. [31]) and harbors MHC class I and non-classical MHC class I genes (Supplemental file 2). Occurrence of CNVs in this region could be important for generation and maintenance of variability in the ruminant MHC class I region together with interlocus recombination as demonstrated in cattle [67,68]. The complete characterization of the sheep MHC may help to elucidate these mechanisms [69]. A large number of studies have reported that polymorphisms in MHC class I genes are associated with disease resistance, with particular effects on resistance to gastrointestinal nematodes infection, in different sheep breeds (e.g. [70]). It will be interesting to evaluate if CNVs in the ovine MHC are associated with resistance to nematode infection and other diseases.

Table 4
Gene ontology (GO) categories significantly overrepresented in sheep CNVRs. Analysis is referred to the GO annotation of the bovine genome (Btau_4.0).

GO level[a]	GO term	GO name	FDR[b]	No. in sheep CNVRs	Expected number
Molecular function	GO:0008289	Lipid binding	<0.001	15	3.1
Biological process	GO:0042157	Lipoprotein metabolic process	<0.001	10	0.7
Biological process	GO:0019882	Antigen processing and presentation	<0.001	10	1.1
Biological process	GO:0006869	Lipid transport	<0.001	10	1.2
Biological process	GO:0010876	Lipid localization	<0.001	10	1.3
Biological process	GO:0002474	Antigen processing and presentation of peptide antigen via MHC class I	<0.05	3	0.1
Cellular component	GO:0042612	MHC class I protein complex	<0.001	9	0.6
Cellular component	GO:0042611	MHC protein complex	<0.001	9	0.8

[a] Analyses are referred to the GO annotation of the bovine genome (Btau_4.0). 185, 172 and 160 transcripts in sheep CNVRs out of 290 are endowed with a GO annotation for molecular function, biological process and cellular component, respectively. 17,077, 16,306 and 16,123 transcripts in the bovine genome of 26,978 are endowed with a GO annotation for molecular function, biological process and cellular component, respectively.
[b] False discovery rate.

3. Conclusions

This cross-species aCGH provided a first comparative map of CNVRs in the sheep genome referred to the cattle genome. As a limited number of sheep and breeds has been analysed, the reported ovine CNVRs represent only a fraction of this kind of variability in the genome of this species. It will be also interesting to investigate sheep of other breeds in order to better evaluate the distribution and extension of CNVs in this species and eventually to evidence differences among breeds and/or breed types. It is also worth to mention that the cross-species analysis, that is based on heterologous hybridization, could have reduced the number of detectable CNVRs due to low homology between cattle probes and sheep DNA for some regions. However, the use of ten algorithms for segmentation analysis of the normalized data might have reduced the possible bias due to differences in hybridization intensity. Further studies are needed to evaluate the effect of differences of homology on the call of CNVs.

Even using a reduced number of analysed sheep, the obtained results indicated that CNVRs are not randomly distributed across sheep, goat, and cattle genomes. Evolutionary conserved mechanisms (probably driven by segmental duplications) might be the causative factors of putative recurrent interspecies CNVs among ruminants. Using this cross-species aCGH designed experiment, it will be also possible to analyse other species and evaluate the extent of CNVR co-occurrence in different lineages of the *Bovidae* family. Several sheep CNVs affect genes with important biological functions and might have important effects on production traits and phenotypic differences among breeds. Further studies are needed to evaluate the functional relevance of these CNVs and the impact of this variability in shaping differences among breeds and their effects on production and disease resistance traits in sheep.

4. Materials and methods

4.1. Array CGH and identification of CNVs

We analysed CNVs in the sheep genome by means of a cross-species aCGH experiment using the Roche NimbleGen platform (Roche NimbleGen Inc., Madison, WI; http://www.nimblegen.com) based on custom tiling arrays designed on the bovine (*B. taurus*) genome, Btau_4.0 version [60], excluding unassembled scaffolds. Arrays contained ~385,000 probes on a single slide to provide an evenly distributed coverage with an average interval of ~6 kb for the Btau_4.0 genome. Genomic DNA was extracted from blood of 2 Bagnolese, 2 Comisana, 2 Laticauda, 2 Massese, 3 Sarda, and 1 Valle del Belice ewes using the Wizard® Genomic DNA Purification kit (Promega Corporation, Madison, WI). One of the three Sarda DNA sample was used as reference. Reference DNA was labeled with Cy5 and co-hybridized with the other test DNA samples labeled with Cy3 on 11 different arrays. A self hybridization (reference labeled by both Cy5 and Cy3) was carried out in another array. Hybridization and array scanning were performed by Roche NimbleGen as previously described [15]. Data normalization was conducted using the normalize.qsline method from the Bioconductor package in R [15]. Then data were analysed for each hybridization using normalized $\log_2$ ratios using the CGHweb server (http://compbio.med.harvard.edu/CGHweb/ [61]) that includes multiple algorithms. We used the self-self hybridization and the data obtained in a previous goat–cattle cross-species aCGH experiment [32] to define a suitable threshold to apply to the CGHweb calls in order to minimize false positives. Specifically, we retained a predicted CNV if it had at least five consecutive datapoints supporting it (considering an average of probe values inside a smoothing window of five), thus limiting the minimum CNV size to about 30 kb, even if this resolution can vary in different regions depending on the relative distance of the probes that can be different from the averaged ~6 kb. Pointwise averaging of all computed profiles and maps of gains/losses for smoothed/segmented obtained from several algorithms (Lowess, Wavelet, Quantreg, ruavg, CBS, CGHseg, BioHMM, cghFLasso, GLAD, and FASeg) and summary data were generated. All algorithms may generate artefact calls [71,72] and for this reason we have chosen to average their results [61] in order to reduce possible biases that could be systematically introduced by using only one of them in our cattle–sheep heterologous experiment. Summary data were considered to call gain/loss in a chromosome region. Then CNVs were called considering a conservative approach joining regions of at least 5 contiguous probes with CNV signal separated by up to three probes without CNV signal in the same individual (Supplemental file 1). This adjustment was applied in order to overcome possible signal losses or hybridization problems in the cross-species aCGH experiment derived by low homology between cattle designed probes and sheep DNA. CNVRs were reported aggregating overlapping or partially overlapping CNVs in different animals as previously reported [30,32,62] and applying the same criteria for CNVs within individuals (Supplemental file 2). The false discovery rate (FDR) was estimated based on the observation of 2 false positives in the self-self hybridization. A rough estimate of the FDR is the expected number of false positives per array (n. 2) times the number of total arrays divided by the total number of CNVs (n. 186), resulting in an estimated FDR of 12%. This calculation should be considered only an approximation because it does not consider the potential for varying false positive rates across arrays. Based on these criteria the averaged $\log_2$ ratio threshold to call gains and losses [61] was established at 0.175 according to the results previously reported in the cattle–goat experiment in which it was considered a stringent threshold [32]. We used this stringent threshold because we preferred a low false-positive rate even at the expense of having more false negatives in our dataset as determined during validation of selected CNVRs.

4.2. Validation of CNVs

Validation of CNVs was performed by semiquantitative fluorescent multiplex PCR (SQF-PCR) as previously reported [32,40,73] using genomic DNA of the same sheep analysed in the aCGH experiment. Briefly, an internal control region known to have no CNV (*DGAT1* gene fragment) and CNVRs of interest was co-amplified in multiplex PCR under quantitative PCR conditions (with forward primers labeled in 5′ with 6FAM) and the products were separated by capillary electrophoresis using an ABI3100 Avant sequencer (Applied Biosystems, Foster City, CA, USA) [32,40]. Peak heights of regions of interest were normalized against those of the internal controls. Primer pairs for control gene fragments and analysed CNVRs as well as PCR conditions are reported in Table S1. PCR was carried out in a total volume of 10 μL with the following cycling conditions: 5 min at 95 °C; 20–22 amplification cycles of 30 s at 95 °C, 30 s at 59 °C, 30 s at 72 °C; 5 min at 72 °C. Capillary electrophoresis was performed using 1 μL of reaction product, diluted in 10 μL of Hi-Di formamide (Applied Biosystems), and added with 0.1 μL of Rox labeled DNA ladder (500HD Rox, Applied Biosystems). Peak heights were obtained using GeneScan software v. 3.7 (Applied Biosystems). DNA dosages were calculated by comparing the normalized peak height ratios of sheep of interest with the average normalized ratios of the reference Sarda sheep as follows: the peak height of a fragment of interest was divided by the peak height of the internal control; the averaged value obtained from different analyses was divided by the same averaged value obtained for the control sheep DNA. We adopted the theoretical values of 1.5, 2.0, 2.5, and so on for a gain (compared to the reference sheep) of one, two, three or other copies, respectively. A loss of one copy would theoretically result in a value of 0.5. These values should be considered only approximation of the copy number content as the objective was to validate the results obtained with aCGH and not to precisely estimate the number of copies of the analysed DNA fragments. These measures are relative values like the aCGH data and are referred to multiple or reduced copies compared to the unknown number of copies of the reference DNA.

4.3. Bioinformatic and computational analyses

The sheep CNVRs were mapped on the Btau_4.0 version of the bovine genome deposited at the Ensembl database (http://www.ensembl.org/Bos_taurus/Info/Index). To determine whether sheep, goat, and cattle CNVRs occur in orthologous regions more often than expected by chance, we considered the data reported for cattle in four different experiments [28–31] and for goat in another previously reported study [32]. The data reported in the four cattle studies were considered separately or merged. In one of these cattle studies [28], CNVs were reported with reference to the Btau_3.0 version, therefore the LiftOver tool at the UCSC Genome Bioinformatics Site (http://genome.ucsc.edu/cgi-bin/hgLiftOver) was used to map CNVs coordinates on the Btau_4.0 version. In this case, only 45 out of the reported 79 CNVs were successfully re-mapped [32]. Within each experiment, overlapping CNVs were fused to define CNVRs. These procedures ended up with 37, 368, 177, and 266 CNVRs for Matukumalli et al. [28], Bae et al. [29], Liu et al. [30], and Fadista et al. [31] experiments, respectively, for a total of 764 combined cattle CNVRs (the merged cattle list; Supplemental file 2) [32]. The goat genome accounted for 127 CNVRs [32]. The number of overlapping segments between the sheep and all other CNVR sets was computed and the overlap significance was evaluated with a permutation test [23]. For each experiment, 10,000 artificial random rearrangements of the CNVRs were generated and mapped on the Btau_4.0 bovine genome. The CNVR length distribution was preserved in each random rearrangement. In order to evaluate the significance of the overlap between two CNVR sets, we computed the distribution of the number of overlapping segments between one of the CNVR sets and the 10,000

random rearrangements of the other one. The reported P-value is the fraction of random CNVR rearrangements that obtain at least the same number of overlapping segments as the real one.

Sheep CNVRs superimposing with cattle transcripts annotated in the Btau_4.0 version were determined on the basis of the genome coordinates, without imposing a minimum overlap threshold. Gene ontology terms associated with bovine transcripts were downloaded with the Ensembl BioMart retrieval system (http://www.ensembl.org/biomart/index.html) and the complete annotation was obtained by reconstructing the complete list of ancestors of each term in the directed acyclic graph described by the OBO file downloaded from the gene ontology web site on July 2010 (http://www.geneontology.org/). The GOTermFinder tool was adopted for this task (http://search.cpan.org/dist/GO-TermFinder/). We computed the occurrence of each term in the set of transcripts overlapping with sheep CNVRs and we compared it with the occurrence of the same term in the whole bovine genome (Btau_4.0 version). The Fisher exact test was adopted to assess the significance of the overrepresentation of the terms in the set of transcripts overlapping with the sheep CNVRs. The multiple-hypothesis correction [74] was adopted for discriminating the significant terms at different false discovery rates (FDR): 0.001, 0.01 and 0.05.

To supplement the functional annotation, PANTHER Molecular Function terms were assigned to all bovine transcripts using the Hidden Markov Model scoring tools of the PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System version 6.1 (http://panther6.ai.sri.com/tools/hmmScoreForm.jsp). Similarly to the GO annotation, the distribution of the PANTHER terms in the set of transcripts overlapping with sheep CNVRs was compared with the occurrence in the whole genome and the significance of the overrepresentation was evaluated with the Fisher exact test adopting the multiple-hypothesis correction. aCGH data have been submitted to the gene expression omnibus (http://www.ncbi.nlm.gov/geo/) under the accession number GSE25122.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2010.11.005.

Acknowledgments

We thank Drs. Charles Lee, Omer Gökçümen (Harvard Medical School, USA), and Giuliano Galimberti (University of Bologna, Italy) for advices on statistical evaluations of sheep–cattle CNVR overlaps. We thank farmers and people who helped in obtaining sheep samples. This work was funded by the Italian MiPAAF SELMOL project, University of Bologna RFO funds, ConsDABI and Assessorato Agricoltura e Foreste of the Regione Siciliana – U.O.B. 108, SOAT n. 69 Aragona (AG), and was associated with the MIUR-PNR 2003 project (FIRB art.8) termed LIBI-Laboratorio Internazionale di BioInformatica.

References

[1] D. Pinkel, R. Segraves, D. Sudar, S. Clark, I. Poole, D. Kowbel, C. Collins, W.L. Kuo, C. Chen, Y. Zhai, S.H. Dairkee, B.M. Ljung, J.W. Gray, D.G. Albertson, High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays, Nat. Genet. 20 (1998) 207–211.
[2] N.P. Carter, Methods and strategies for analyzing copy number variation using DNA microarrays, Nat. Genet. 39 (2007) S16–S21.
[3] A.J. Iafrate, L. Feuk, M.N. Rivera, M.L. Listewnik, P.K. Donahoe, Y. Qi, S.W. Scherer, C. Lee, Detection of large-scale variation in the human genome, Nat. Genet. 36 (2004) 949–951.
[4] J. Sebat, B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Månér, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T.C. Gilliam, B. Trask, N. Patterson, A. Zetterberg, M. Wigler, Large-scale copy number polymorphism in the human genome, Science 305 (2004) 525–528.
[5] L. Feuk, A.R. Carson, S.W. Scherer, Structural variation in the human genome, Nat. Rev. Genet. 7 (2006) 85–97.
[6] J.L. Freeman, G.H. Perry, L. Feuk, R. Redon, S.A. McCarroll, D.M. Altshuler, H. Aburatani, K.W. Jones, C. Tyler-Smith, M.E. Hurles, N.P. Carter, S.W. Scherer, C. Lee, Copy number variation: new insights in genome diversity, Genome Res. 16 (2006) 949–961.
[7] Database of Genomic Variants, http://projects.tcag.ca/variation/. August 2010.
[8] B.E. Stranger, M.S. Forrest, M. Dunning, C.E. Ingle, C. Beazley, N. Thorne, R. Redon, C.P. Bird, A. de Grassi, C. Lee, C. Tyler-Smith, N. Carter, S.W. Scherer, S. Tavaré, P. Deloukas, M.E. Hurles, E.T. Dermitzakis, Relative impact of nucleotide and copy number variation on gene expression phenotypes, Science 315 (2007) 848–853.
[9] P. Cahan, Y. Li, M. Izumi, T.A. Graubert, The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells, Nat. Genet. 41 (2009) 430–437.
[10] L.D. Orozco, S.J. Cokus, A. Ghazalpour, L. Ingram-Drake, S. Wang, A. van Nas, N. Che, J.A. Araujo, M. Pellegrini, A.J. Lusis, Copy number variation influences gene expression and metabolic traits in mice, Hum. Mol. Genet. 18 (2009) 4118–4129.
[11] F. Zhang, W. Gu, M.E. Hurles, J.R. Lupski, Copy number variation in human health, disease, and evolution, Annu. Rev. Genomics Hum. Genet. 10 (2009) 451–481.
[12] P. Stankiewicz, J.R. Lupski, Structural variation in the human genome and its role in disease, Annu. Rev. Med. 61 (2010) 437–455.
[13] D.F. Conrad, D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T.D. Andrews, C. Barnes, P. Campbell, T. Fitzgerald, M. Hu, C.H. Ihm, K. Kristiansson, D.G. Macarthur, J.R. Macdonald, I. Onyiah, A.W. Pang, S. Robson, K. Stirrups, A. Valsesia, K. Walter, J. Wei, Wellcome Trust Case Control Consortium, C. Tyler-Smith, N.P. Carter, C. Lee, S.W. Scherer, M.E. Hurles, Origins and functional impact of copy number variation in the human genome, Nature 464 (2010) 704–712.
[14] D.J. Adams, E.T. Dermitzakis, T. Cox, J. Smith, R. Davies, R.J. Banerjee, J.C. Bonfield, Mullikin, Y.J. Chung, J. Rogers, A. Bradley, Complex haplotypes, copy number polymorphisms and coding variation in two recently divergent mouse strains, Nat. Genet. 37 (2005) 532–536.
[15] T.A. Graubert, P. Cahan, D. Edwin, R.R. Selzer, T.A. Richmond, P.S. Eis, W.D. Shannon, X. Li, H.L. McLeod, J.M. Cheverud, T.J. Ley, A high-resolution map of segmental DNA copy number variation in the mouse genome, PLoS Genet. 3 (2007) e3.
[16] C.M. Egan, S. Sridhar, M. Wigler, I.M. Hall, Recurrent DNA copy number variation in the laboratory mouse, Nat. Genet. 39 (2007) 1384–1389.
[17] G. Cutler, L.A. Marshall, N. Chin, H. Baribault, P.D. Kassner, Significant gene content variation characterizes the genomes of inbred mouse strains, Genome Res. 17 (2007) 743–1754.
[18] X. She, Z. Cheng, S. Zöllner, D.M. Church, E.E. Eichler, Mouse segmental duplication and copy number variation, Nat. Genet. 40 (2008) 909–914.
[19] V. Guryev, K. Saar, T. Adamovic, M. Verheul, S.A. van Heesch, S. Cook, M. Pravenec, T. Aitman, H. Jacob, J.D. Shull, N. Hubner, E. Cuppen, Distribution and functional impact of DNA copy number variation in the rat, Nat. Genet. 40 (2008) 538–545.
[20] F.J. Charchar, M. Kaiser, A.J. Bingham, N. Fotinatos, F. Ahmady, M. Tomaszewski, N.J. Samani, Whole genome survey of copy number variation in the spontaneously hypertensive rat: relationship to quantitative trait loci, gene expression, and blood pressure, Hypertension 55 (2010) 1231–1238.
[21] G.H. Perry, J. Tchinda, S.D. McGrath, J. Zhang, S.R. Picker, A.M. Cáceres, A.J. Iafrate, C. Tyler-Smith, S.W. Scherer, E.E. Eichler, A.C. Stone, C. Lee, Hotspots for copy number variation in chimpanzees and humans, Proc. Natl. Acad. Sci. USA 103 (2006) 8006–8011.
[22] G.H. Perry, F. Yang, T. Marques-Bonet, C. Murphy, T. Fitzgerald, A.S. Lee, C. Hyland, A.C. Stone, M.E. Hurles, C. Tyler-Smith, E.E. Eichler, N.P. Carter, C. Lee, R. Redon, Copy number variation and evolution in humans and chimpanzees, Genome Res. 18 (2008) 1698–1710.
[23] A.S. Lee, M. Gutiérrez-Arcelus, G.H. Perry, E.J. Vallender, W.E. Johnson, G.M. Miller, J.O. Korbel, C. Lee, Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies, Hum. Mol. Genet. 17 (2008) 1127–1136.
[24] W.K. Chen, J.D. Swartz, L.J. Rush, C.E. Alvarez, Mapping DNA structural variation in dogs, Genome Res. 19 (2009) 500–509.
[25] T.J. Nicholas, Z. Cheng, M. Ventura, K. Mealey, E.E. Eichler, J.M. Akey, The genomic architecture of segmental duplications and associated copy number variants in dogs, Genome Res. 19 (2009) 491–499.
[26] J. Fadista, M. Nygaard, L.E. Holm, B. Thomsen, C. Bendixen, A snapshot of CNVs in the pig genome, PLoS ONE 3 (2008) e3916.
[27] G.E. Liu, C.P. Van Tassel, T.S. Sonstegard, R.W. Li, L.J. Alexander, J.W. Keele, L.K. Matukumalli, T.P. Smith, L.C. Gasbarre, Detection of germline and somatic copy number variations in cattle, Dev. Biol. (Basel) 132 (2008) 231–237.
[28] L.K. Matukumalli, C.T. Lawley, R.D. Schnabel, J.F. Taylor, M.F. Allan, M.P. Heaton, J. O'Connell, S.S. Moore, T.P. Smith, T.S. Sonstegard, C.P. Van Tassell, Development and characterization of a high density SNP genotyping assay for cattle, PLoS ONE 4 (2009) e5350.
[29] J.S. Bae, H.S. Cheong, L.H. Kim, S. NamGung, T.J. Park, J.Y. Chun, J.Y. Kim, C.F. Pasaje, J.S. Lee, H.D. Shin, Identification of copy number variations and common deletion polymorphisms in cattle, BMC Genomics 11 (2010) 232.
[30] G.E. Liu, Y. Hou, B. Zhu, M.F. Cardone, L. Jiang, A. Cellamare, A. Mitra, L.J. Alexander, L.L. Coutinho, M.E. Dell'Aquila, L.C. Gasbarre, G. Lacalandra, R.W. Li, L.K. Matukumalli, D. Nonneman, L.C. Regitano, T.P. Smith, J. Song, T.S. Sonstegard, C.P. Van Tassell, M. Ventura, E.E. Eichler, T.G. McDaneld, J.W. Keele, Analysis of copy number variations among diverse cattle breeds, Genome Res. 20 (2010) 693–703.
[31] J. Fadista, B. Thomsen, L.E. Holm, C. Bendixen, Copy number variation in the bovine genome, BMC Genomics 11 (2010) 284.
[32] L. Fontanesi, P.L. Martelli, F. Beretti, V. Riggio, S. Dall'Olio, M. Colombo, R. Casadio, V. Russo, B. Portolano, An initial comparative map of copy number variations in the goat (*Capra hircus*) genome, BMC Genomics 11 (2010) 639.
[33] X. Wang, S. Nahashon, T.K. Feaster, A. Bohannon-Stewart, N. Adefope, An initial map of chromosomal segmental copy number variations in the chicken, BMC Genomics 11 (2010) 351.
[34] I.H. Cheeseman, N. Gomez-Escobar, C.K. Carret, A. Ivens, L.B. Stewart, K.K. Tetteh, D.J. Conway, Gene copy number variation throughout the *Plasmodium falciparum* genome, BMC Genomics 10 (2009) 353.

[35] J.S. Maydan, A. Lorch, M.L. Edgley, S. Flibotte, D.G. Moerman, Copy number variation in the genomes of twelve natural isolates of *Caenorhabditis elegans*, BMC Genomics 11 (2010) 62.
[36] D.K. Griffin, L.B. Robertson, H.G. Tempest, A. Vignal, V. Fillon, R.P. Crooijmans, M.A. Groenen, S. Deryusheva, E. Gaginskaya, W. Carré, D. Waddington, R. Talbot, M. Völker, J.S. Masabanda, D.W. Burt, Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution, BMC Genomics 9 (2008) 168.
[37] B.M. Skinner, L.B. Robertson, H.G. Tempest, E.J. Langley, D. Ioannou, K.E. Fowler, R.P. Crooijmans, A.D. Hall, D.K. Griffin, M. Völker, Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis, BMC Genomics 10 (2009) 357.
[38] M. Völker, N. Backström, B.M. Skinner, E.J. Langley, S.K. Bunzey, H. Ellegren, D.K. Griffin, Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution, Genome Res. 20 (2010) 503–511.
[39] L. Dumas, Y.H. Kim, A. Karimpour-Fard, M. Cox, J. Hopkins, J.R. Pollack, J.M. Sikela, Gene copy number variation spanning 60 million years of human and primate evolution, Genome Res. 17 (2007) 1266–1277.
[40] L. Fontanesi, F. Beretti, V. Riggio, E. Gómez González, S. Dall'Olio, R. Davoli, V. Russo, B. Portolano, Copy number variation and missense mutations of the agouti signaling protein (*ASIP*) gene in goat breeds with different coat colors, Cytogenet. Genome Res. 126 (2009) 333–347.
[41] G. Pielberg, C. Olsson, A.C. Syvänen, L. Andersson, Unexpectedly high allelic diversity at the *KIT* locus causing dominant white color in the domestic pig, Genetics 160 (2002) 305–311.
[42] L. Fontanesi, E. D'Alessandro, E. Scotti, L. Liotta, A. Crovetti, V. Chiofalo, V. Russo, Genetic heterogeneity and selection signature at the *KIT* gene in pigs showing different coat colours and patterns, Anim. Genet. 41 (2010) 478–492.
[43] D. Wright, H. Boije, J.R. Meadows, B. Bed'hom, D. Gourichon, A. Vieaud, M. Tixier-Boichard, C.J. Rubin, F. Imsland, F. Hallböök, L. Andersson, Copy number variation in intron 1 of *SOX5* causes the pea-comb phenotype in chickens, PLoS Genet. 5 (2009) e1000512.
[44] M.G. Elferink, A.A. Vallée, A.P. Jungerius, R.P. Crooijmans, M.A. Groenen, Partial duplication of the *PRLR* and *SPEF2* genes at the late feathering locus in chicken, BMC Genomics 9 (2008) 391.
[45] B.J. Norris, V.A. Whan, A gene duplication affecting expression of the ovine *ASIP* gene is responsible for white and black sheep, Genome Res. 18 (2008) 1282–1293.
[46] L. Fontanesi, S. Dall'Olio, F. Beretti, B. Portolano, V. Russo, Coat colours in the Massese sheep breed are associated with mutations in the agouti signalling protein (*ASIP*) and melanocortin 1 receptor (*MC1R*) genes, Animal 5 (2011) 8–17.
[47] A.M. Crawford, K.G. Dodds, A.J. Ede, C.A. Pierson, G.W. Montgomery, H.G. Garmonsway, A.E. Beattie, K. Davies, J.F. Maddox, S.W. Kappes, R.T. Stone, T.C. Nyen, J.M. Penty, E.A. Lord, J.E. Broom, J. Buitkamp, W. Schwaiger, J.T. Epplen, F.P. Matthew, M.E. Matthew, D.J. Hulme, K.J. Beh, R.A. McGraw, C.W. Beattie, An autosomal genetic linkage map of the sheep genome, Genetics 140 (1995) 703–724.
[48] M.J. de Gortari, B.A. Freking, R.P. Cuthbertson, S.M. Kappes, J.W. Keele, R.T. Stone, K.A. Leymaster, K.G. Dodds, A.M. Crawford, C.W. Beattie, A second-generation linkage map of the sheep genome, Mamm. Genome 9 (1998) 204–209.
[49] J.F. Maddox, K.P. Davies, A.M. Crawford, D.J. Hulme, D. Vaiman, E.P. Cribiu, B.A. Freking, K.J. Beh, N.E. Cockett, N. Kang, C.D. Riffkin, R. Drinkwater, S.S. Moore, K.G. Dodds, J.M. Lumsden, T.C. van Stijn, S.H. Phua, D.L. Adelson, H.R. Burkin, J.E. Broom, J. Buitkamp, L. Cambridge, W.T. Cushwa, E. Gerard, S.M. Galloway, B. Harrison, R.J. Hawken, S. Hiendleder, H.M. Henry, J.F. Medrano, K.A. Paterson, L. Schibler, R.T. Stone, B. van Hest, An enhanced linkage map of the sheep genome comprising more than 1000 loci, Genome Res. 11 (2001) 1275–1289.
[50] T. Goldammer, G.P. Di Meo, G. Lühken, C. Drögemüller, C.H. Wu, J. Kijas, B.P. Dalrymple, F.W. Nicholas, J.F. Maddox, L. Iannuzzi, N.E. Cockett, Molecular cytogenetics and gene mapping in sheep (*Ovis aries*, 2n = 54), Cytogenet. Genome Res. 126 (2009) 63–76.
[51] C.H. Wu, K. Nomura, T. Goldammer, T. Hadfield, J.E. Womack, N.E. Cockett, An ovine whole-genome radiation hybrid panel used to construct an RH map of ovine chromosome 9, Anim. Genet. 38 (2007) 534–536.
[52] P. Laurent, L. Schibler, A. Vaiman, J. Laubier, C. Delcros, G. Cosseddu, D. Vaiman, E.P. Cribiu, M. Yerle, A 12000-rad whole-genome radiation hybrid panel in sheep: application to the study of the ovine chromosome 18 region containing a QTL for scrapie susceptibility, Anim. Genet. 38 (2007) 358–363.
[53] B.P. Dalrymple, E.F. Kirkness, M. Nefedov, S. McWilliam, A. Ratnakumar, W. Barris, S. Zhao, J. Shetty, J.F. Maddox, M. O'Grady, F. Nicholas, A.M. Crawford, T. Smith, P.J. de Jong, J. McEwan, V.H. Oddy, N.E. Cockett, International Sheep Genomics Consortium, Using comparative genomics to reorder the human genome sequence into a virtual sheep genome, Genome Biol. 8 (2007) R152.
[54] The International Sheep Genomics Consortium, A.L. Archibald, N.E. Cockett, B.P. Dalrymple, T. Faraut, J.W. Kijas, J.F. Maddox, J.C. McEwan, V. Hutton Oddy, H.W. Raadsma, C. Wade, J. Wang, W. Wang, X. Xun, The sheep genome reference sequence: a work in progress, Anim. Genet. 41 (2010) 449–453.
[55] J.W. Kijas, D. Townley, B.P. Dalrymple, M.P. Heaton, J.F. Maddox, A. McGrath, P. Wilson, R.G. Ingersoll, R. McCulloch, S. McWilliam, D. Tang, J. McEwan, N. Cockett, V.H. Oddy, F.W. Nicholas, H. Raadsma, International Sheep Genomics Consortium, A genome wide survey of SNP variation reveals the genetic structure of sheep breeds, PLoS ONE 4 (2009) e4668.
[56] J.W. Kijas, M. Menzies, A. Ingham, Sequence diversity and rates of molecular evolution between sheep and cattle genes, Anim. Genet. 37 (2006) 171–174.
[57] M.W. Allard, M.M. Miyamoto, L. Jarecki, F. Kraus, M.R. Tennant, DNA systematics and evolution of the artiodactyl family *Bovidae*, Proc. Natl. Acad. Sci. USA 89 (1992) 3972–3976.
[58] J. Gatesy, D. Yelon, R. DeSalle, E.S. Vrba, Phylogeny of the *Bovidae* (*Artiodactyla*, *Mammalia*), based on mitochondrial ribosomal DNA sequence, Mol. Biol. Evol. 9 (1992) 433–446.
[59] C.A. Matthee, S.K. Davis, Molecular insights into the evolution of the family *Bovidae*: a nuclear DNA perspective, Mol. Biol. Evol. 18 (2001) 1220–1230.
[60] Bovine Genome Sequencing and Analysis Consortium, The genome sequence of taurine cattle: a window to ruminant biology and evolution, Science 324 (2009) 522–528.
[61] W. Lai, V. Choudhary, P.J. Park, CGHweb: a tool for comparing DNA copy number segmentations from multiple algorithms, Bioinformatics 24 (2008) 1014–1015.
[62] R. Redon, S. Ishikawa, K.R. Fitch, L. Feuk, G.H. Perry, T.D. Andrews, H. Fiegler, M.H. Shapero, A.R. Carson, W. Chen, E.K. Cho, S. Dallaire, J.L. Freeman, J.R. González, M. Gratacòs, J. Huang, D. Kalaitzopoulos, D. Komura, J.R. MacDonald, C.R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M.J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, J. Zhang, L. Armengol, D.F. Conrad, X. Estivill, C. Tyler-Smith, N.P. Carter, H. Aburatani, C. Lee, K.W. Jones, S.W. Scherer, M.E. Hurles, Global variation in copy number in the human genome, Nature 444 (2006) 444–454.
[63] S. Kumar, S.B. Hedges, A molecular timescale for vertebrate evolution, Nature 392 (1998) 917–920.
[64] M.E. Steiper, N.M. Young, T.Y. Sukarna, Genomic data support the hominoid slowdown and an Early Oligocene estimate for the hominoid-cercopithecoid divergence, Proc. Natl. Acad. Sci. USA 101 (2004) 17021–17026.
[65] C. Lalueza-Fox, J. Castresana, L. Sampietro, T. Marquès-Bonet, J.A. Alcover, J. Bertranpetit, Molecular dating of caprines using ancient DNA sequences of *Myotragus balearicus*, an extinct endemic Balearic mammal, BMC Evol. Biol. 5 (2005) 70.
[66] H.W. Raadsma, E. Jonas, D. McGill, M. Hobbs, M.K. Lam, P.C. Thomson, Mapping quantitative trait loci (QTL) in sheep. II. Meta-assembly and identification of novel QTL for milk production traits in sheep, Genet. Sel. Evol. 41 (2009) 45.
[67] E.C. Holmes, A.F.C. Roberts, K.A. Staines, S.A. Ellis, Evolution of MHC class I genes in Cetartiodactyls, Immunogenetics 55 (2003) 193–202.
[68] J. Birch, L. Murphy, N.D. MacHugh, S.A. Ellis, Generation and maintenance of diversity in the cattle MHC class I region, Immunogenetics 58 (2006) 670–679.
[69] J. Gao, K. Liu, H. Liu, H.T. Blair, G. Li, C. Chen, P. Tan, R.Z. Ma, A complete DNA sequence map of the ovine major histocompatibility complex, BMC Genomics 11 (2010) 466.
[70] V.S. Dukkipati, H.T. Blair, D.J. Garrick, A. Murray, Ovar-Mhc—ovine major histocompatibility complex: role in genetic resistance to diseases, N. Z. Vet. J. 54 (2006) 153–160.
[71] W.R. Lai, M.D. Johnson, R. Kucherlapati, P.J. Park, Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data, Bioinformatics 21 (2005) 3763–3770.
[72] H. Willenbrock, J. Fridlyand, A comparison study: applying segmentation to array CGH data for downstream analyses, Bioinformatics 21 (2005) 4084–4091.
[73] C. Le Maréchal, E. Masson, J.M. Chen, F. Morel, P. Ruszniewski, P. Levy, C. Férec, Hereditary pancreatitis caused by triplication of the trypsinogen locus, Nat. Genet. 38 (2006) 1372–1374.
[74] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. B 57 (1995) 289–300.