A Comparative Analysis of Copy Number Variation of the Sheep and Goat Genomes

L. Fontanesi[*], F. Beretti[*†], V. Riggio[†], S. Dall'Olio[*], M. Occidente[‡], C. Incoronato[‡], P.L. Martelli[§], R. Casadio[§], B. Portolano[†], D. Matassino[‡] and V. Russo[*]

Introduction

Recent studies have shown that copy number variants (CNVs), defined as intraspecific gains or losses of ≥1 kb of genomic DNA (Feuk et al. (2006); Freeman et al. (2006)), are important sources of variability of mammalian genomes as reported in human (e.g. Redon et al. (2006); Shaikh et al. (2009)) and other primates (Perry et al. (2006); Lee et al. (2008)), mouse (Cutler et al. (2007); Graubert, et al. (2007)), rat (Guryev et al. (2008)), dog (Chen et al. (2009); Nicholas et al. (2009)) and pig (Fadista et al.(2008)). Genome-wide discovery and frequency evaluation of CNVs have been possible with the development of high-resolution array comparative genome hybridization (aCGH) and then with data analysis of high-density single nucleotide polymorphism (SNP) platforms (Pinkel et al. (1998); McCarroll et al. (2008)). An advantage of aCGH is that hybridization can be performed using heterologous DNA, i.e. genomic DNA of a different species but close to that used to develop the array, taking advantage from completely sequenced, assembled and annotated genomes. Cross species aCGH experiments have been successfully applied using human array to analyse CNV in chimpanzee (Perry et al. (2006)). Moreover, chicken based aCGH studies have been applied to analyse the turkey and duck genomes (Griffin et al. (2008); Skinner et al. (2009)).
Sheep and goat genomes are not completely sequenced and assembled yet and no systematic studies have been carried out to analyse CNVs in these species. Here, we applied a cross species aCGH study using as reference the cattle genome to identify copy number variation in the sheep and goat genomes.

Material and methods

DNA samples. Blood was collected from 12 sheep of different breeds (3 Sarda, 2 Bagnolese, 2 Comisana, 2 Laticauda, 2 Massese and 1 Valle del Belice) and from 10 goats of different breeds (3 Girgentana, 3 Saanen, 2 Camosciata delle Alpi and 2 Murciano-Granadina). Genomic DNA (gDNA) was extracted using the Promega Wizard® Genomic DNA Purification kit.

[*] DIPROVAL, Sezione di Allevamenti Zootecnici, University of Bologna, 42100 Reggio Emilia, Italy
[†] Dip. S.En.Fi.Mi.Zo., Sezione di Produzioni Animali, University of Palermo, 90128 Palermo, Italy
[‡] ConSDABI – CC (Mediterranean Biodiversity) Italian National Focal Point FAO AnGR , 82100 Benevento, Italy
[§] Biocomputing Group, Dept. of Experimental Evolutionary Biology, University of Bologna, 40126 Bologna, Italy

aCGH platform, chip design and hybridization. For the aCGH experiments we used a Roche NimbleGen bovine whole genome custom tiling array (Roche NimbleGen Inc., Madison, WI) designed on Bta_4.0 genome assembly including part of BTA13 from the University of Maryland (UMD) *Bos taurus* v. 2.0 assembly. The array contained ~385,000 50-mer oligonucleotides with a median probe spacing of ~6 kb). Two separate experiments, one for the analysis of the sheep genome and one for the analysis of the goat genome, were carried out using this customized array. In the sheep experiment the reference gDNA was from one Sarda sheep whereas in the goat experiment the reference gDNA was from one Camosciata delle Alpi goat. The reference gDNAs, labelled with Cy5, were co-hybridized with the other test gDNA samples (labelled with Cy3) on 11 (sheep experiment) or 9 (goat experiment) different arrays. Self hybridizations (reference gDNA labelled by both Cy5 and Cy3) were carried out for the two experiments in two other arrays. Hybridization was performed by Roche NimbleGen as previously described (Graubert et al. (2007)). Slides were scanned and copy numbers were quantified from the dye fluorescence ratios as reported below.

aCGH data analysis and CNV calling. For fluorescence data analysis, Roche NimbleGen conducted the initial data normalization using the normalize.qsline method from the Bioconductor package in R (Graubert et al., 2007). Then we analysed data for each hybridization using normalized $\log_2$ ratios using the CGHweb server (Lai et al., 2008; http://compbio.med.harvard.edu/CGHweb/) that includes multiple algorithms. The average $\log_2$ threshold to call gains and losses was 0.2, with an average of probe values inside a smoothing window of 5. 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. Summary data were considered to call gain/loss in a chromosome region. An empirical false discovery rate (FDR) was estimated based on the observation of false positives in the self-self hybridizations divided by the total number of CNVs identified.

Validation of CNVs. Semiquantitative fluorescence multiplex PCR (SQM-PCR) was carried out as reported by Fontanesi et al. (2009) to validate 5 CNV regions identified in either the sheep or the goat genomes.

Results and discussion

We identified 431 CNVs in the 11 analysed sheep (on average, 39.2 CNVs per sample) with a FDR of 0.08 (based on 3 false positives obtained in the self experiment) and 358 CNVs in the 9 analysed goats (on average, 39.8 CNVs per sample) with a FDR of 0.10 (based on 3 false positives obtained in the self experiment). In other aCGH experiments based on homologous gDNA hybridizations carried out in human and dog, FDR was 0.04 and 0.02, respectively (Hasin et al. (2008); Nicholas et al. (2009)). However, the FDR we observed can be acceptable, as our experiments were based on heterologous gDNA hybridization. A total of 367 and 299 annotated genes in the bovine genome (Bta_4.0 version) were included in CNV regions of the sheep and goat genomes. For the two genomes, the average size of CNVs resulted ~65 kb and ~70 kb, respectively. Table 1 shows a comparison of CNVs

reported in other mammals. Sheep and goat showed the presence of CNVs in the same range of other mammalian species.

Table 1: Comparison of sheep and goat CNVs to those reported in other mammals for similar experiments.

Species	No. of individuals	Methods of analysis	Total no. of CNVs	Mean no. of CNVs per individual	References
Sheep	11	385 k oligo aCGH	431	39.2	This study
Goat	9	385 k oligo aCGH	358	39.8	This study
Human	270	BAC aCGH	1447[α]	70	Redon et al. (2006)
Chimpanzee	20	BAC aCGH	355	17.8	Perry et al. (2006)
Macaque	9	385 k oligo aCGH	123	13.7	Lee et al. (2008)
Dog	9	385 k oligo aCGH	155	17.2	Chen et al. (2009)
Mouse	21	385 k oligo aCGH	80	2-38[β]	Grauben et al. (2007)
Rat	13	Two platforms[γ]	643	63-11[δ]	Guryev et al. (2008)

[α]CNV regions (as defined in Redon et al. (2006)).
[β]Depending on the strain.
[γβ]5 M exon specific oligo arrays (10 animals); 385 k oligo aCGH (3 animals).
[δ]For the two platforms.

A portion of BTA13 from the UMD assembly was included in the customized array, to confirm the presence of a CNVs detected for the goat agouti signaling protein (*ASIP*) gene, which was investigated for its effect on coat colour (Fontanesi et al. (2009)). This gene is correctly assembled in this genome version, whereas it is placed in an unassembled scaffold in the Bta_4.0 version. Results obtained for the goat genome confirmed the presence of CNVs in the BTA13 region, in which the *ASIP* gene is included (Fontanesi et al. (2009)), for Girgentana, Saanen, and Murciano-Granadina goats. These results validated the cross species experiments we carried out. Furthermore, SQM-PCR for 5 genes located in CNV detected by aCGH were used for validation. All targeted gene fragments confirmed the presence of CNVs in the same animals that reported CNVs in the 5 chromosome regions.

Conclusion

This is the first report that analyses CNVs in the whole sheep and goat genomes. For this study we used the bovine genome that is the closest species to the sheep and goat that has an assembled and annotated genome. The two cross-species experiments (sheep *vs* bovine and goat *vs* bovine) were successful to identify CNVs. This variability could be important in determining phenotypic and production differences between and within breeds. Further studies will be carried out to evaluate the identified CNVs from both functional and evolutionary points of view.

Acknowledgments

This study was supported by MiPAAF Selmol, Assessorato Agricoltura e Foreste of the Regione Siciliana – U.O.B. 108, SOAT n. 69 Aragona (AG) and ConSDABI funds.

References

Chen, W.K., Swartz, J.D., Rush, L.J., *et al.* (2009). *Genome Res.,* 19:500–509.

Cutler, G., Marshall, L.A., Chin, N., *et al.* (2007). *Genome Res.,* 17:1743–1754.

Fadista, J., Nygaard, M., Holm, L.E., *et al.* (2008). *PLoS One,* 3:e3916.

Feuk, L., Carson, A.R., and Scherer, S.W. (2006). *Nat. Rev. Genet*., 7: 85–97.

Fontanesi, L., Beretti, F., Riggio, V., *et al.* (2009). *Cytogenet. Genome Res.,* 126:333–347.

Freeman, J.L., Perry, G.H., Feuk, L., *et al.* (2006). *Genome Res.,* 16: 949–961.

Graubert, T.A., Cahan, P., Edwin, D., *et al.* (2007). *PLoS Genet.,* 3:e3.

Griffin, D.K., Robertson, L.B., Tempest, H.G., *et al.* (2008). *BMC Genomics,* 9:168.

Guryev, V., Saar, K., Adamovic, T., *et al.* (2008). *Nat. Genet.,* 40:538–545.

Hasin, Y., Olender, T., Khen, M., *et al.* (2008). *PLoS Genet,* 11: e1000249.

Lai, W., Choudhary, V., and Park, P.J. (2008). *Bioinformatics,* 24:1014–1015.

Lee, A.S., Gutiérrez-Arcelus, M., Perry, G.H., *et al.* (2008). *Hum. Mol. Genet.,* 17:1127–1136.

Nicholas, T.J., Cheng, Z., Ventura, M., *et al.* (2009). *Genome Res.,* 19:491–499.

McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., *et al.* (2008). *Nat. Genet.,* 40:1166–1174.

Perry, G.H., Tchinda, J., McGrath, S.D., *et al.* (2006). *Proc. Natl. Acad. Sci. USA,* 103:8006–8011.

Pinkel, D., Segraves, R., Sudar, D., *et al.* (1998). *Nat. Genet.,* 20:207–211.

Redon, R., Ishikawa, S., Fitch, K.R., *et al.* (2006). *Nature*, 444:444–454.

Shaikh, T.H., Gai, X., Perin, J.C., *et al.* (2009). *Genome Res.,* 19:1682–1690.

Skinner, B.M., Robertson, L.B., Tempest, H.G., *et al.* (2009). *BMC Genomics,* 10:357.