Exploring the rabbit genome to identify single nucleotide polymorphisms useful for association studies

Bertolini F.[1], Schiavo G.[1], Scotti E.[1], Martelli P.L.[2], Casadio R.[2], Fontanesi L.[1]

[1]Department of Agricultural and Food Science. University of Bologna, Italy
[2]Biocomputing Group. University of Bologna, Italy

Corresponding Author: Luca Fontanesi, Dipartimento di Scienze e Tecnologie Agroalimentari. Università di Bologna. Viale Fanin 46, 40127 Bologna, Italy - Tel. +39 051 2096535 - Fax: +39 051 2096355 - Email: luca.fontanesi@unibo.it

ABSTRACT: High throughput genotyping platforms and next generation sequencing are changing the way in which animal genomes are investigated. In this study we applied a next generation sequencing technology (Ion Torrent PGM) to identify single nucleotide polymorphisms (SNPs) in the rabbit genome. Equimolar genomic DNA from different rabbit breeds was pooled and digested with two restriction fragments. Two reduced representation libraries were produced and sequenced on two different Ion Torrent 318 chips. A total of 6343257 reads were aligned to the rabbit genome (oryCun2.0), whereas 621483 reads resulted as "unmapped". Variant calling produced a total of 65630 SNPs with a mapping quality of at least 10Q and covered by at least 4 reads in the point of the detected variation. This study represent the first one that identified a large number of SNPs in the rabbit genome. In addition, the SNPs identified will be useful to design a commercial high throughput genotyping platform that could have an important impact in the study of variability and identification of markers associated with production traits in rabbit populations.

Key words: Rabbit genome, SNP, Ion Torrent, Reduced representation library.

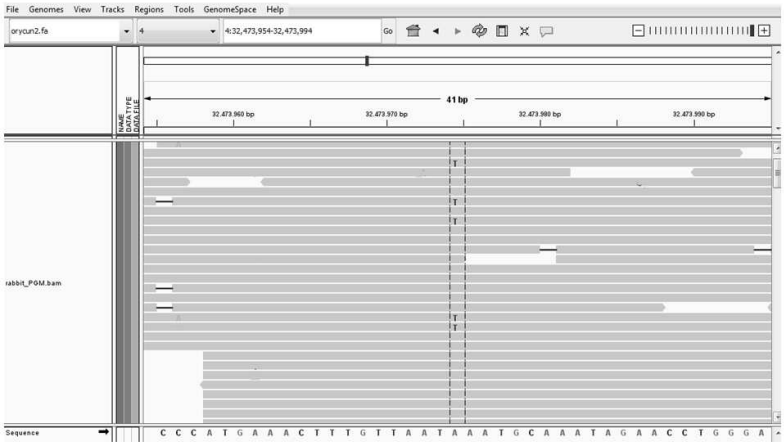
INTRODUCTION – High throughput genotyping platforms are changing the way in which genetic studies and related applications are carried out in livestock. Genome wide association studies and genomic selection approaches have been already applied or under evaluation in cattle, pig and several other species for which thousands or millions of single nucleotide polymorphisms (SNPs) have been identified on their genomes and are available for their use in customized panels or commercial tools. At present, massive information of SNPs in the rabbit genome is not available, yet. A second version of the rabbit genome (oryCun2.0) has been recently assembled and released by the Broad Institute within the mammalian genome project (http://www.broadinstitute.org/scientific-community/science/projects/mammals-models/rabbit/rabbit-genome-project). This version cover about 89% of the predicted genome of this species. Based on this genome assembly several studies have been carried out to identify DNA markers associated with production traits in commercial rabbit lines using a candidate gene approach (e.g. Fontanesi *et al.*, 2011, 2012). Other studies have already identified mutations affecting phenotypic traits in rabbit, like coat colours (e.g. Fontanesi *et al.*, 2010a, 2010b). A quite convenient approach to identify SNPs in one species is to produce Reduced Representation Libraries (RRLs) combined both with traditional Sanger sequencing or, more efficiently, using next generation sequencing technologies (Artshuler *et al.*, 2000). Here we reported the first reduced

representation library sequencing experiment in rabbit with the aim to discover thousands of SNPs.

MATERIALS AND METHODS – Genomic DNA was extracted from blood or liver of 10 different rabbits belonging to different breeds/lines. Extracted DNA was quantified and pooled at equimolar concentration. Two aliquots of DNA pools of 10 ug each were digested overnight with 50 U of *Hae*III or *Rsa*I (Fermentas) at 37°C. These digested products were separated on agarose gel 0.8% and a band corresponding to the 500-700 bp of length was cut from the gel of the two digested products and the DNA was extracted from them. These procedures produced two reduced representation libraries. Separated DNA was then prepared for the sequencing process carried out with the Ion Torrent PGM sequencer (Rothberg *et al.*, 2011). Briefly, two DNA libraries were created shearing the samples and adding adapters. Selected 100 bp-length fragments were clonally amplified, purified and sequenced using two 318 chips. Sequences obtained from the two runs were aligned with the oryCun 2.0 genome using TMAP aligner of the Ion Torrent suite. Variant calling was performed using the command Samtools mpileup (Li, 2009) and the in silico effect of every single nucleotide mutation was predicted by Variant effect predictor web tool (http://www.ensembl.org/tools.html). Validation of called SNPs was performed in two ways: i) by using the IGV (Integrative Genomics Viewer, Broad Institute) tool and ii) through Sanger sequencing of targeted regions.

RESULTS AND CONCLUSIONS – Combining results of the two high throughput sequencing runs for the two RRLs, a total of 6964750 reads were obtained, with a mean sequence length of 100 bp; 6343257 reads were aligned with the oryCun2.0 genome aff, covering about 15% of the rabbit genome whereas 621483 reads resulted as “unmapped”. Variant calling produced a total of 65630 SNPs with a mapping quality of at least 10Q and covered by at least 4 reads in the point of the detected variation. An example of a SNP identified with IGV is reported in Figure 1.

Figure 1 – Output of the IGV sequence analysis tool with indicated a SNP in the rabbit genome



This work represent the first on that identified a large number of SNPs using a combination of RRLs and Ion Torrent sequencing. This experimental design made it possible to identify a large number of SNPs that have been annotated (data not shown). In particular, 507 SNPs were missense mutations in important coding regions that could be interesting to use in association studies with production traits in this species. In addition, the SNPs identified will be useful to design a commercial high throughput genotyping platform that could have an important impact in the study of variability and identification of markers associated with production traits in rabbit populations.

REFERENCES – **Altshuler**, D., Pollara, V.J., Cowles, C.R., van Etten, W.J., Baldwin, J., Linton, L., Lander, E.S., 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407:513-516. **Fontanesi**, L., Forestier, L., Allain, D., Scotti, E., Beretti, F., Deretz-Picoulet, S, Pecchioli, E., Vernesi, C., Robinson, T.J., Malaney, J.L., Russo, V., Oulmouden, A., 2010a. Characterization of the rabbit agouti signaling protein (ASIP) gene: transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat colour. Genomics 95:166-175. **Fontanesi,** L., Scotti, E., Colombo, M., Beretti, F., Forestier, L., Dall'Olio, S., Deretz, S., Russo, V., Allain, D., Oulmouden, A., 2010b. A composite six bp in-frame deletion in the melanocortin 1 receptor (MC1R) gene is associated with the Japanese brindling coat colour in rabbits (Oryctolagus cuniculus). BMC Genet. 11:59. **Fontanesi**, L., Scotti, E., Frabetti, A., Fornasini, D., Picconi, A., Russo, V., 2011. Identification of polymorphisms in the rabbit (Oryctolagus cuniculus) myostatin (MSTN) gene and association analysis with finishing weight in a commercial rabbit population. Anim. Genet. 42:339. **Fontanesi**, L., Mazzoni, G., Bovo, S., Frabetti, A., Fornasini, D., Dall'Olio, S., Russo, V., 2012. Association between a polymorphism in the *IGF2* gene and finishing weight in a commercial rabbit population. Anim. Genet. 43:651-652. **Li**, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics 16: 2078-2079. **Rothberg**, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards M., Hoon, J., Simons, J.F., Marran, D., Myers, J.W., Davidson, J.F., Branting, A., Nobile, J.R., Puc, B.P., Light, D., Clark, T.A., Huber, M., Branciforte, J.T., Stoner, I.B., Cawley, S.E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J.A., Namsaraev, E., McKernan, K.J., Williams, A., Roth, G.T., Bustillo, J., 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348-452.