Exploring copy number variation in the rabbit (Oryctolagus cuniculus) genome by array comparative genome hybridization

L. Fontanesi a,b,⁎, P.L. Martelli b,c, E. Scotti a, V. Russo a, C. Rogel-Gaillard d, R. Casadio b,c, C. Vernesi e

a Department of Agro-Food Science and Technology, Sezione di Allevamenti Zootecnici, University of Bologna, Viale Fanin 48, 40127 Bologna, Italy
b Centre of Genome Biology, University of Bologna, 40126 Bologna, Italy
c Biocomputing Group, Dept. of Experimental Evolutionary Biology, University of Bologna, Via San Giacomo 9/2, 40126 Bologna, Italy
d INRA, UMR 1313 Génétique Animale et Biologie Intégrative, Domaine de Vilvert, 78352 Jouy en Josas CEDEX, France
e Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, Viote del Monte Bondone, 38040 Trento, Italy

A R T I C L E   I N F O

Article history:
Received 19 January 2012
Accepted 3 July 2012
Available online 16 July 2012

Keywords:
aCGH
Copy number variation
Major Histocompatibility Complex
Oryctolagus cuniculus

A B S T R A C T

The European rabbit (Oryctolagus cuniculus) is relevant in a large spectrum of fields: it is a livestock, a pet, a biomedical model and a biotechnology tool, a wild resource and a pest. The sequencing of the rabbit genome has opened new perspectives to study this lagomorph at the genome level. We herein investigated for the first time the O. cuniculus genome by array comparative genome hybridization (aCGH) and established a first copy number variation (CNV) genome map in this species comprising 155 copy number variation regions (CNVRs; 95 gains, 59 losses, 1 with both gain and loss) covering ~0.3% of the OryCun2.0 version. About 50% of the 155 CNVRs identified spanned 139 different protein coding genes, 110 genes of which were annotated or partially annotated (including Major Histocompatibility Complex genes) with 277 different gene ontology terms. Many rabbit CNVRs might have a functional relevance that should be further investigated.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

The economic importance of the European rabbit (Oryctolagus cuniculus) is probably underestimated: the rabbit is a livestock, it produces meat, fur and wool; it is a pet and a fancy animal with many different breeds/lines with a broad variety of different phenotypic characteristics; it is a biomedical model and a biotechnology tool used to answer many biological questions and to produce biomedical reagents; and it is a wild resource and a pest in several regions in which it has been introduced. Despite the large spectrum of fields in which the rabbit is relevant, relatively few studies, compared to other species, have investigated its genome. In particular, reciprocal chromosome painting between human and rabbit was used to establish homology chromosome maps for these two species [1]. Subsequently, a first whole genome integrated genetic and cytogenetic microsatellite map has been obtained and refined by Chantry-Darmon et al. [2–4]. Then the Broad Institute sequenced the rabbit genome within the Mammalian Genome Project whose aim was to obtain comparative sequencing data useful to annotate the human genome [5]. A preliminary low coverage version (2X) of the rabbit genome has been recently improved and reassembled at ~7X (OryCun2.0 version; http://www.ensembl.org/Oryctolagus_cuniculus/Info/Index). In rabbits only a few reports have characterized the genetic basis of phenotypic traits, like coat colour [6–9] and hair morphology [4,10], or have investigated quantitative trait loci for biomedical and production traits [11–13].

Recently, a large number of studies have demonstrated that copy number variation (CNV), defined as intraspecific gains or losses of ≥ 1 kb of genomic DNA, is quite frequent in the mammalian genomes in which it represents the largest source of variability in terms of affected nucleotides (~0.4–25% of the genome; i.e., [14,15]). Therefore, it is not surprising that CNVs can regulate gene expression and functions (i.e., [16]) and that are involved in many human Mendelian and complex disorders [17] and many other phenotypic traits in other species (i.e., [18-20]). Among domesticated animals, CNV maps have been produced in dog [21,22], cattle [23–25], sheep [26], goat [27], pig [28,29], and chicken [30] using array comparative genome hybridization (aCGH) or high density SNP chips [31,32]. We herein investigated for the first time the O. cuniculus genome by aCGH in order to identify CNVs and establish a first CNV genome map in this species.

2. Results and discussion

2.1. Validation of aCGH data

Our previous aCGH studies in sheep and goats were based on a heterologous genome (cattle) used to design the tiling arrays applied to investigate CNV in these two close small ruminant species [26,27], as their genomes were not assembled at that time. In the current study we
used the assembled rabbit genome (OryCun2.0) to analyze CNV in genomic DNA samples coming from the same species. Arrays that contained ~760,000 probes with an average interval of ~3 kb were used to analyze three target DNA samples from rabbits of different lines/breeds (one Commercial white line doe, one Checkered Giant doe and one Champagne d’Argent buck) chosen to maximize possibilities to detect variability, against a reference DNA from an inbred Rhinelander doe. These homologous aCGH data were analyzed using the three better performing algorithms for segmentation analysis implemented in the CGHweb package (see Materials and methods [33]), averaging these results to obtain a suitable log2 ratio threshold. As in this experimental design we included two different controls, we empirically detected the log2 ratio threshold based on them: 1) in the self-self hybridization analysis the threshold should not obtain any CNV call (control of false positive results); 2) in the hybridization having as target genome that of the Champagne d’Argent male (hybridized with a reference female genomic DNA sample), log2 ratio was set up to detect differences of copy number (loss) on more than 98% of the X chromosome (control of false negative results; Fig. 1). Based on these criteria the log2 ratio threshold to call gains and losses was empirically established at 0.4. Applying this threshold we could reduce the problems of false positive and false negative calls even if, in this evaluation, we could not fully control the differences of false positive/negative rates across arrays.

In addition, three CNVRs (see below data about all CNVRs detected in this study), identified on O. cuniculus chromosomes (OCU) 5, 13, and 17 were validated by semiquantitative fluorescent multiplex-PCR (SQF-PCR; Table 1 and Table S1). Two consisted of gains (OCU5 and OCU17) and one of loss (OCU13), as determined in the aCGH experiments. Considering the results obtained for the three animals used in the aCGH analyses and the three investigated CNVRs, all SQF-PCR assays, except one, confirmed results obtained in the aCGH experiments (8 out of 9). Only for the region on OCU13, SQF-PCR in the Commercial white line rabbit did not indicate a loss of copies compared to the reference Rhinelander genomic DNA, as it was obtained from the aCGH experiment (Table S1). Based on our limited validation results, the false positive rate is 1/9 or ~11%. This evaluation is close to the rate of non validated aCGH data reported in cattle using qPCR (~8% [24]), whereas is far lower than the non validated data reported in a pig aCGH study (50%) based on a preliminary assembly of a few porcine chromosomes [27].

SQF-PCR results produced from additional animals not used in the aCGH experiments in order to further investigate these regions indicated

Fig. 1. Results of the aCGH for the X chromosome in the male Champagne d’Argent rabbit. Images have been reported for A) log2 ratio plot of original data, B) log2 ratio plot of summary data (pointwise averaging of all computed profiles), C) heatmap of log2 ratios for original, smoothed/segmented, and summary data, and D) maps of gains/losses for smoothed/segmented and summary data (gain is indicated in orange, loss is indicated in green). Smoothed/segmented data were obtained with three algorithms (CBS, CGHseq, and cghLasso; see also Materials and methods) averaged in the summary data.
that the OCU5 CNVR included both gain and loss of copies, and that gain of copies for the OCU17 CNVR was present in additional rabbits (Table S1).

### 2.2. CNVs in the rabbit genome

Using the stringent and conservative approach we applied in calling CNV from the aCGH experiments, we reported a total of 196 CNVs (Table 2) with an average size of 38.5 kb (ranging from 11.6 kb to 364.4 kb; Supplementary file 1). This is about half the average size observed in the sheep and goat aCGH analyses, that were carried out with about half the number of probes (~385,000) in the chips [26,27], that means that higher is the density of the aCGH platform, higher is the potential resolution power, that, of course, depends also from the criteria used to call CNVs. For example, in this study we considered only CNVs detected by at least 5 consecutive probes. The White commercial rabbit showed the highest number of CNVs (88), most of which were unique (observed only in that animal; 0.75) and consisting of gains (~0.70; Table 2). Champagne d’Argent male showed the lowest number of CNVs (47) that were gains or losses in almost equal proportion (~0.50). Fig. 2 reports a few examples of CNVs identified in the analyzed rabbit using the CGH package.

CNVs were determined by aggregating overlapping CNVs identified in different rabbits using criteria already used in other studies [14,24,26,27]. On the whole we identified 155 CNVRs covering 6.62 Mb (~0.3%) of the OryCun2.0 genome version anchored to the 21 autosomes and the X chromosome (Fig. 3 and Supplementary file 2). The unassembled scaffolds (about 18% of the sequenced genome) were not included in this first investigation due to problems in interpreting the results [24,26]. These 155 CNVRs included 95 gains, 59 losses and one with both gain and loss compared to the reference sample, localized on all chromosomes except on OCU20 (Fig. 3 and Supplementary file 2). Chromosomes with more than ten CNVRs were OCU1, OCU2, OCU3, OCU4 and OCU13 (11, 11, 13, 14, and 14, respectively), whereas OCU5 and OCU4 were the most covered by CNVRs (1.087% and 0.820%, respectively; Table S2). Seventeen CNVRs were identified in more than one rabbit (11%) and were defined as high confidence CNVRs, whereas the remaining 89% were considered as second level confidence CNVRs as previously defined [26]. Mean and median of these 155 CNVRs were 42.7 kb and 23.4 kb, respectively (ranging from ~11.7 to ~346.4 kb). Other aCGH studies in domestic animals used a lower density array (~385,000 probes; [21,24,26,27,30]) or higher density arrays (i.e. [23]) and for these reasons it is possible to have only an indirect comparison between our results and those obtained in other species. For example, the CNVR average size in experiments using about half the number of probes than in our study was almost twice the size that we report here in rabbit, whereas the number of CNV called in each analyzed animal was 2-3 times lower than what we identified in our explorative investigation in O. cuniculus [21,24,26,27,30].

#### 2.3. Annotation of CNVRs

About 50% (78) of the 155 CNVRs identified in the rabbit genome partially or completely spanned 139 different protein coding genes, 110 genes of which were annotated or partially annotated with 277 different Gene Ontology (GO) terms (Supplementary file 3 and data not shown). A few of these categories were significantly or suggestively (from P~0.01 to P~0.1) over-represented in rabbit CNVRs (Table 3). Similarly to what we obtained in sheep, this study in rabbit reported fewer over-represented groups compared to other CNV studies [21,22,27,34]. This might be due to a bias derived by the reference chosen in this experiment and/or by the different number of analyzed animals compared to the other experiments. Even considering these limits, it is particularly interesting to note that antigen binding, antigen processing and presentation were the GO over-represented GO terms of the “Molecular functions” and “Biological Process” categories respectively (Table 3). In the “Cellular component” classification, MHC protein complexes and plasma membrane part were over-represented categories in rabbit CNVRs (Supplementary file 3). The Major Histocompatibility Complex (MHC) of the rabbit appeared affected by copy number polymorphisms. In particular, CNVRs 90 and 93 on OCU12 contained MHC class I and class II genes, respectively (Supplementary file 3). The MHC class I and II genes belong to gene families that include functional genes as well as pseudogenes [35]. The class I molecules are known to present self-antigens to cytotoxic CD8+ T lymphocytes and regulate Natural Killer cell activity and the class II molecules present exogeneous peptides to CD4+ T lymphocytes. Both class I and II genes are highly polymorphic, mostly in the regions encoding the peptide groove. Direct investigations of MHC in other species, including pig and primates, reported similar variability in class I and class II genes [36–38]. In addition, other aCGH studies have already shown that MHC genes in ruminants are affected by CNV [23,24,26,27]. All these studies indicated that CNV is an additional common source of variability in the MHC and differences in copy number might be important to define disease resistance or

### Table 2

<table>
<thead>
<tr>
<th>Rabbits</th>
<th>Number of CNVs</th>
<th>CNV average size ± s.d. (in kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Uniquea</td>
<td>Gain Loss</td>
</tr>
<tr>
<td>Champagne d’Argent (male)</td>
<td>47b</td>
<td>24 23</td>
</tr>
<tr>
<td>Checkered Giant (female)</td>
<td>61</td>
<td>39 22</td>
</tr>
<tr>
<td>Commercial white line (female)</td>
<td>88</td>
<td>62 26</td>
</tr>
<tr>
<td>Total</td>
<td>196</td>
<td>125 71</td>
</tr>
</tbody>
</table>

a Not overlapping with any other CNV.
b Excluding chromosome X data.

d Results obtained in the SQF-PCR assays.

---

**Table 1**

Validated CNVRs using semiquantitative fluorescent multiplex-PCR (SQF-PCR).

<table>
<thead>
<tr>
<th>CNVR no.</th>
<th>Chromosome (OCU)</th>
<th>OCU coordinates</th>
<th>Target gene symbol (Ensembl entry no.)</th>
<th>Gain/loss in aCGH</th>
<th>Gain/loss in SQF-PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>5</td>
<td>24,035,242–24,061,529</td>
<td>–</td>
<td>Gain</td>
<td>Gain</td>
</tr>
<tr>
<td>112</td>
<td>13</td>
<td>131,093,402–131,135,020</td>
<td>–</td>
<td>Loss</td>
<td>Loss</td>
</tr>
<tr>
<td>129</td>
<td>17</td>
<td>31,127,859–31,165,791</td>
<td>DIL4 (ENSOCUG00000010756), CHAC1 (ENSOCUG00000010762)</td>
<td>Gain</td>
<td>Gain</td>
</tr>
</tbody>
</table>

---

a As listed in Supplementary file 2.
b OCU coordinates on the OryCun2.0 genome version of the corresponding CNVR.
c Results obtained in the aCGH assays.
d Results obtained in the SQF-PCR assays.
e Loss of copies reported for other rabbits not used in the aCGH experiment (see Table S1).
Fig. 2. Examples of three CNVRs in different rabbits and chromosomes (OCU5, OCU13 and OCU17) identified by aCGH. Images have been reported as described in the legend of Fig. 1. Red arrows indicate regions of copy gain or loss.

Fig. 3. Map of rabbit CNVRs identified in this study (reported in the different rabbit chromosomes).
susceptibility in animals. It will be critical to assess whether the CNVs identified in the rabbit MHC affect functional genes.

Among the many other CNVRs with annotated genes, it is interesting to mention CNVR n. 62 on OCU7 that includes an olfactory receptor gene (olfactory receptor, family 2, subfamily A, member 14; OR2A1H) that is affected by CNV also in human [17,39]. Extensive CNV in olfactory receptor gene families has been reported within and among human and mouse populations and might be important to define olfactory inter-individual differences [34,39,40]. Other four CNVRs on OCU13 (n. 102, 104, 107, and 110) encompass genes (dolichyl-phosphate mannosyltransferase polypeptide 3, DPM3; vang-like 1 (van gogh, Drosophila), VANGL1; low density lipoprotein receptor-related protein 8, apolipoprotein e receptor, LR8; and fatty acid amide hydrolase, FAAH) that are responsible of human genetic defects (congenital disorder of glycosylation 1o; caudal regression syndrome, and neural tube defects; susceptibility to myocardial infarction 1; and susceptibility to drug addiction; respectively) as reported in OMIM database (January 2012). Moreover, CNVR n. 85 on OCU11 encompasses the sperm flagellar 2 (SPEF2) gene, also known as KPL2, that in pig an intronic insertion causes the immotile short-tail sperm defect [41].

CNV has been recently identified in the rabbit alpha globin gene. Unfortunately this gene is not correctly assembled in the OryCun2.0 genome version and for this reason we could have missed it from our study even if it is still not clear if this variability is present only in wild rabbit populations or also in domesticated populations [42].

3. Conclusions

The sequencing of the rabbit genome has opened new opportunities to exploit genomic information of this multifaceted species that is important in many basic and applied biological fields. Despite the progress that will be derived by this initiative, many other studies should be carried out to refine this resource, improving its assembly and annotation, and to characterize its variability at different levels. In this study we have explored copy number variability in the rabbit genome by aCGH using the OryCun2.0 genome version to design the tiling arrays, and using strict calling criteria we and reported 155 CNVRs. As in this study we have explored copy number variability in the rabbit genome and for this reason we could have missed it from our study even if it is still not clear if this variability is present only in wild rabbit populations or also in domesticated populations [42].

explain a quote of variability of phenotypic and production traits of different breeds and lines selected for different purposes. In addition, immunological aspects of this species might be influenced, at least in part, by CNV that might also be important to investigate considering the rabbit as a biomedical model and as a biotechnology tool.

4. Materials and methods

4.1. aCGH analyses

Array comparative genome hybridization experiments were based on a customized tiling array designed by Roche NimbleGen Inc. (Madison, WI; http://www.nimblegen.com) on the O. cuniculus genome, OryCun2.0 version (http://www.ensembl.org/Oryctolagus_cuniculus/Info/Index). Arrays contained ~760,000 probes on a single slide to provide an evenly distributed coverage with an average interval of ~3 kb for the OryCun2.0 genome. Rabbit genomic DNA was extracted from blood of one doe of a Commercial white line (CWLO of Grupo Martini [43], one Checked Giant doe (CG1), one Champagne d’Argent buck (CdA), and one Rhinelander (Tricolor: R1) doe using the Wizard® Genomic DNA Purification kit (Promega Corporation, Madison, WI). Reference DNA was from the Rhinelander doe. This DNA was labeled with Cy5 and co-hybridized with the other three test DNA samples labeled with Cy3 on 3 different arrays. A self hybridization (reference DNA labeled by both Cy5 and Cy3) was carried out in another array. Hybridization and array scanning were performed by Roche NimbleGen as previously described [34]. Data normalization was conducted using the normalize.qsline method from the Bioconductor package in R [34]. Then data were analyzed for each hybridization using normalized log2 ratios using the CGHweb R package for Linux (http://compbio.med.harvard.edu/CGHweb/ [33]). Pointwise averaging of all computed profiles and maps of gains/losses for smoothed/segmented obtained from three algorithms implemented in this package (Circular Binary Segmentation [44]; Gaussian Model with Adaptive Penalty [45]; cghFlasso [46]), shown to have better performances ([33]; and data not shown) were generated. We used the self-self hybridization and the chromosome X data of the Champagne d’Argent rabbit used as control region to define a suitable threshold to apply to the CGHweb calls in order to minimize false positive and false negative results. The Champagne d’Argent male is expected to have one copy of the X chromosome compared to the reference Rhinelander doe (loss). We detected loss over at least 98% of the X chromosome in this male/female comparison. Summary data obtained considering the three indicated algorithms were generated to call gain/loss in a chromosome region and to compile a high confidence set of CNVs. Then CNVs were called considering a conservative approach joining regions of at least 5 contiguous probes with CNV signal separated by up to two probes without CNV signal in the same individual. Copy number variation regions (CNVRs) were reported aggregating overlapping or partially overlapping CNVRs in different animals as previously reported [14,24,26,27] and applying the same criteria for CNVRs within individuals.

Table 3

<table>
<thead>
<tr>
<th>GO level</th>
<th>GO term</th>
<th>GO name</th>
<th>FDR</th>
<th>No. in rabbit CNVRs</th>
<th>Expected number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular function</td>
<td>GO:0003823</td>
<td>Antigen binding</td>
<td>&lt;0.005</td>
<td>4</td>
<td>0.001</td>
</tr>
<tr>
<td>Biological process</td>
<td>GO:0019882</td>
<td>Antigen processing and presentation</td>
<td>&lt;0.01</td>
<td>5</td>
<td>0.005</td>
</tr>
<tr>
<td>Cellular component</td>
<td>GO:0042611</td>
<td>MHC protein complex</td>
<td>&lt;0.005</td>
<td>5</td>
<td>0.003</td>
</tr>
<tr>
<td>Cellular component</td>
<td>GO:0042612</td>
<td>MHC class I protein complex</td>
<td>&lt;0.05</td>
<td>3</td>
<td>0.002</td>
</tr>
<tr>
<td>Cellular component</td>
<td>GO:0042613</td>
<td>MHC class II protein complex</td>
<td>&lt;0.1</td>
<td>2</td>
<td>0.001</td>
</tr>
<tr>
<td>Cellular component</td>
<td>GO:0044459</td>
<td>Plasma membrane part</td>
<td>&lt;0.1</td>
<td>12</td>
<td>0.08</td>
</tr>
</tbody>
</table>

* Analyses are referred to the GO annotation of the rabbit genome (OryCun2.0): 57, 57, and 59 transcripts in rabbit CNVRs out of 110 are endowed with a GO annotation for molecular function, biological process and cellular component, respectively; 13,916, 11,156, and 10,763 transcripts in the rabbit genome out of 28,188 transcripts are endowed with a GO annotation for molecular function, biological process and cellular component, respectively.

b False discovery rate.
4.2. Validation of CNVs by semiquantitative fluorescent multiplex PCR

Semiquantitative fluorescent multiplex PCR (SQF-PCR) was used to validate several CNVs identified in the aCGH experiments (Table S1). This technique was applied as previously described [19,26,27,47] using genomic DNA of the same rabbits analyzed in the aCGH experiments and genomic DNA of other five rabbits (one Rhinelander, R2; and four Checkered Giant does, CG2-5) from a reference families developed by Fontanesi et al. [48] extracted as described above. Brie

Checkered Giant does, CG2-5 of a reference families developed by and genomic DNA of other

This technique was applied as previously described[19,26,27,47]using

validate several CNVs identi

European Network on Rabbit Genome Biology (Bologna) for technical assistance. We thank the Broad Institute (USA) [16] B.E. Stranger, M.S. Forrest, M. Dunning, C.E. Ingle, C. Beazley, N. Thorne, R. Redon, S.W. Scherer, M.E. Hurles, Origins and functional impact of copy number variation

with the occurrence of the same term in the whole rabbit genome

Ontology web site on November 2011 (http://www.geneontology.org/). We computed the occurrence of each term in

Rabbit CNVRs superimposing with transcripts annotated in the OryCun2.0 version were determined on the basis of the genome coordinates, without imposing a minimum overlap threshold. Gene ontology terms associated with rabbit transcripts were downloaded from the Ensembl BioMart retrieval system (http://www.ensembl.org/)

References

4.3. Bioinformatics analyses

Rabbit CNVRs superimposing with transcripts annotated in the OryCun2.0 version were determined on the basis of the genome coordinates, without imposing a minimum overlap threshold. Gene ontology terms associated with rabbit transcripts were downloaded from the Gene Ontology web site on November 2011 (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 rabbit CNVRs and we compared it with the occurrence of the same term in the whole rabbit genome (OryCun 2.0). The Fisher exact test was adopted to assess the significance of the terms in the set of transcripts overlapping with the rabbit CNVRs. The multiple-hypothesis correction was adopted for discriminating the significant terms at different False Discovery Rates [49]: 0.005, 0.01, 0.05, and 0.1. aCGH data and results have been submitted to the ArrayExpress Archive (http://www.ebi.ac. uk/arrayexpress/), under the accession number E-MEXP-3636.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2012.07.001.

Acknowledgments

We thank Dr. Andrea Frabetti (Rabbit Research Center, Gruppo Martini) for providing samples and Dr. Michela Colombo (University of Bologna) for technical assistance. We thank the Broad Institute (USA) that provided the OryCun2.0 rabbit genome version. This work has been carried out in the framework of the COST Action TD1011 “A Collaborative European Network on Rabbit Genome Biology – RGB-Net” (2011–2014) and has been supported by University of Bologna RFO and Fondazione Edmund Mach Funds.


