

RESEARCH ARTICLE

Open Access



Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits

Joaquim Manoel da Silva^{1,2}, Poliana Fernanda Giachetto³, Luiz Otávio da Silva⁴, Leandro Carrijo Cintra³, Samuel Rezende Paiva^{5,6,7}, Michel Eduardo Beleza Yamagishi³ and Alexandre Rodrigues Caetano^{6,7*}

Abstract

Background: Copy number variations (CNVs) have been shown to account for substantial portions of observed genomic variation and have been associated with qualitative and quantitative traits and the onset of disease in a number of species. Information from high-resolution studies to detect, characterize and estimate population-specific variant frequencies will facilitate the incorporation of CNVs in genomic studies to identify genes affecting traits of importance.

Results: Genome-wide CNVs were detected in high-density single nucleotide polymorphism (SNP) genotyping data from 1,717 Nelore (*Bos indicus*) cattle, and in NGS data from eight key ancestral bulls. A total of 68,007 and 12,786 distinct CNVs were observed, respectively. Cross-comparisons of results obtained for the eight resequenced animals revealed that 92 % of the CNVs were observed in both datasets, while 62 % of all detected CNVs were observed to overlap with previously validated cattle copy number variant regions (CNVRs). Observed CNVs were used for obtaining breed-specific CNV frequencies and identification of CNVRs, which were subsequently used for gene annotation. A total of 688 of the detected CNVRs were observed to overlap with 286 non-redundant QTLs associated with important production traits in cattle. All of 34 CNVs previously reported to be associated with milk production traits in Holsteins were also observed in Nelore cattle. Comparisons of estimated frequencies of these CNVs in the two breeds revealed 14, 13, 6 and 14 regions in high (>20 %), low (<20 %) and divergent (NEL > HOL, NEL < HOL) frequencies, respectively.

Conclusions: Obtained results significantly enriched the bovine CNV map and enabled the identification of variants that are potentially associated with traits under selection in Nelore cattle, particularly in genome regions harboring QTLs affecting production traits.

Keywords: Beef cattle, Copy number variants, High-throughput nucleotide sequencing, SNP genotyping

* Correspondence: alexandre.caetano@embrapa.br

⁶Embrapa Recursos Genéticos e Biotecnologia, Brasília, Distrito Federal, Brazil

⁷CNPq Fellow

Full list of author information is available at the end of the article



Background

Copy number variations (CNVs) have been shown to account for substantial portions of genomic variation in humans. Gains or losses in genomic regions varying from 50 bp to several megabases (Mbp) in size have been estimated to cover 77.97 % of the human genome (<http://dgv.tcag.ca/dgv/app/statistics?ref=GRCh37/hg19>) [1]. CNVs have also been shown to cause changes in transcription levels of specific genes and may be an important source of material for evolutionary mechanisms to act upon [2]. Approximately half of observed human CNVs span regions containing protein-coding genes [1] known to be involved in essential cellular functions, general metabolism and the onset of different diseases [3–9], and which may influence disease susceptibility [10–12]. CNV alterations have also been observed in primary and metastatic cancerous tissues [4, 11, 13–15] and to be associated with various genetic traits [11, 16].

Most reported broad population-oriented studies for CNV detection use at least two main platforms: Comparative Genomic Hybridization (CGH) arrays and SNP genotyping arrays [17–19]. Advantages and disadvantages associated with these platforms have been widely discussed in the literature [20–22]. However, with the advent and rapidly decreasing costs of next generation sequencing (NGS), studying CNVs with sequencing data has also become increasingly feasible [23, 24]. The main advantages of sequencing over genotyping lie in the improved resolution of CNV identification, and particularly in the fact that searches for CNVs are not limited to specific, pre-defined regions. NGS protocols randomly generate reads and therefore close to the entire genome can be sampled with high coverage and resolution, thus promoting higher accuracy in CNV detection and greater precision when estimating breakpoints [24, 25].

Studies to identify and catalogue CNVs have been successfully performed on animals of economic importance, including cattle [26–37], chicken [38, 39], pig [40, 41], sheep [42, 43] and goat [44]. A large number of CNVs were identified in taurine (*Bos taurus*) and zebuine cattle (*Bos indicus*) in regions containing genes known to affect complex traits [17, 18, 26, 29, 31, 32]. The overlap of CNVs reported among animals of different taurine breeds is greater than the overlap between taurine and indicine cattle while, even though analyses were performed with data from a single Nelore (*B. indicus*) sample, zebu cattle were observed to have the largest CNV diversity among studied breeds [26].

The present study is the first to widely and deeply analyze a population of Nelore (Zebu) cattle composed of 1,717 animals that were genotyped at high density (~770 K SNPs). In addition, eight key ancestral bulls were resequenced with minimal coverage of 20×. The goal of this study was to perform a high-resolution

analysis to detect and characterize CNVs in this breed while also estimating breed-specific variant frequencies.

Results and discussion

Genome-wide discovery and distribution of CNVs

A total of 68,007 CNVs representing 54,510 single copy duplications, 1,729 double copy duplications, 11,672 single copy deletions, and 96 double copy deletions (Additional file 1) were detected with the analysis of genotyping data from 1,509 Nelore samples which passed data QC procedures. Figure 1 shows the chromosome distribution of all detected CNVs. A total of 1,411, 515, and 24 CNVs were observed in >1, >2 and >10 % of the samples analyzed, respectively.

The number of SNPs in each detected CNV varied from 20 to 1,420 (94 ± 110). CNV length varied from 20.01 Kb to 7.75 Mbp (320 ± 413 Kbp). Figure 2 (Additional file 1) shows the size distribution of detected CNVs. Observed CNVs larger than the average by one standard deviation or more (733 Kbp) and with a frequency greater than 1 % were rare and far apart ($n = 116$), with a mean frequency of 2.47 %.

CNVR identification

Different methods for condensing overlapping CNVs into Copy Number Variant Regions (CNVRs) have been proposed [45–47]. Nelore CNVRs were identified using JM-CNV [48], which considers CNV length and frequency, removes extremely long or infrequent CNVs from the initial analysis, and resolves observed breakpoint issues [24, 25]. The 68,007 detected CNVs were condensed into 7,319 CNVRs (Fig. 3: Additional file 2), representing a total coverage of 1.56 Gigabases (61.91 %) of the bovine autosomal genome (Additional file 3). A total of 2,306 duplications, 212 deletions, and 4,801 duplications and deletions were observed in the identified CNVRs (Fig. 4a). A high positive correlation between the number of detected CNVRs and the size of bovine chromosomes was observed (0.98, Fig. 3), contrary to what was observed in terms of the number of total CNVs detected (0.34, Fig. 1).

CNVR length varied from 20.1 Kbp to 3.81 Mbp (213 ± 237 Kbp, Fig. 5, Additional file 2). BTA1 was found to have the highest number of CNVRs (459), while BTA27 had the lowest number (119) of CNVRs (Additional file 2). As for the average distance between CNVRs, BTA24 and BTA19 were found to have the greatest (444.8Kbp) and the smallest (323.5Kbp) distances, respectively. A total of 962, 713, and 296 CNVRs showed frequencies >1, >2 and >10 % in the studied samples, respectively.

CNVs in NGS data

LUMPY [49] uses signal depth from observed split-reads and from miss-mapped paired-end reads as evidence to

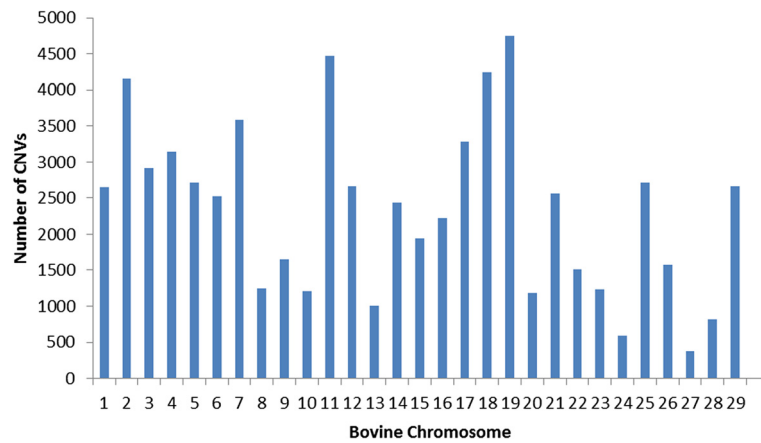


Fig. 1 Chromosome distribution of CNVs detected with high-density SNP genotyping data from Nelore cattle

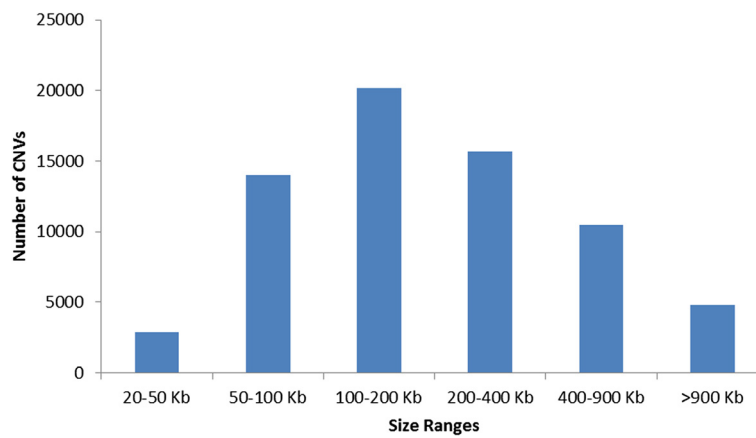


Fig. 2 Size distribution of CNVs detected using Nelore genotyping data

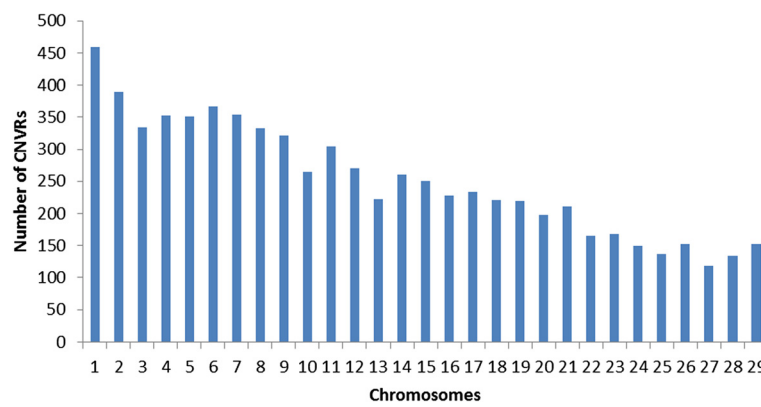


Fig. 3 Distribution of CNVRs detected using Nelore SNP genotyping data across bovine chromosomes

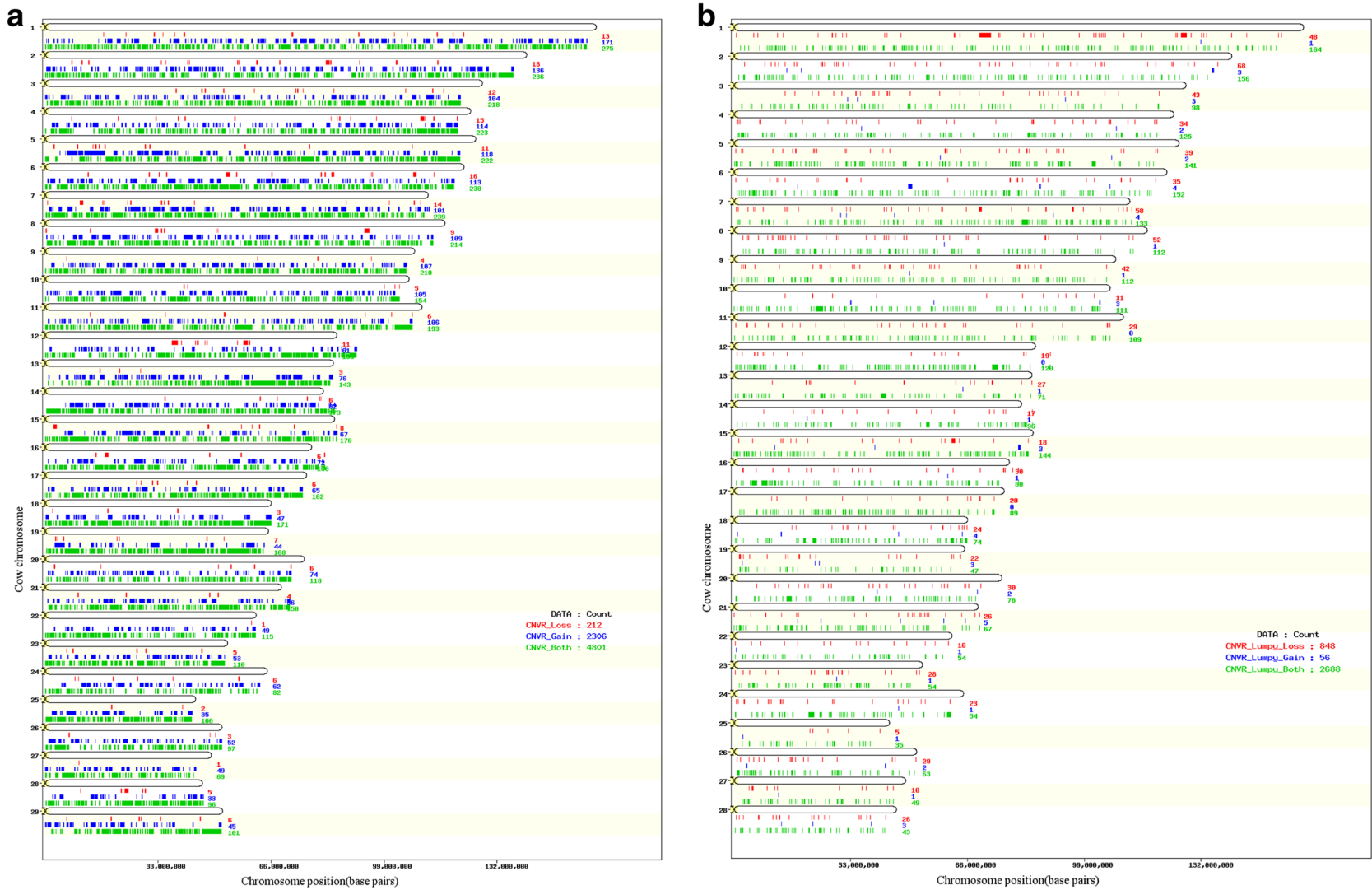


Fig. 4 Distribution of gain, loss and mixed CNRs detected across the Nelore genome (based on UMD3.1). **a** CNRs detected with genotyping data. **b** CNRs <5 Mb detected with NGS data

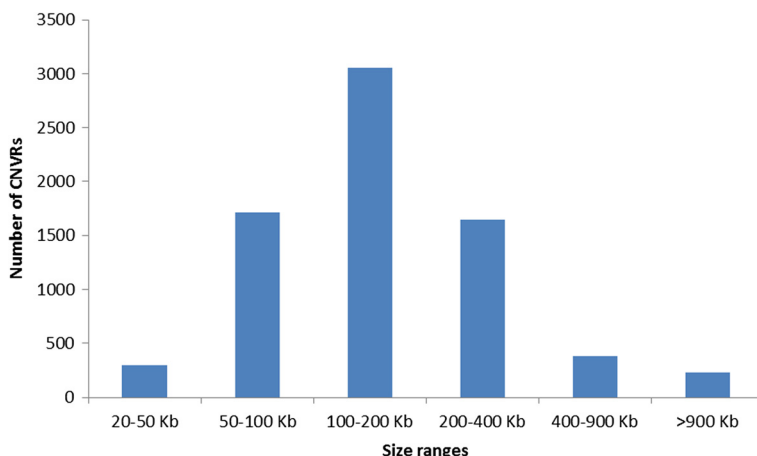


Fig. 5 Size distribution of CNVRs detected using Nelore genotyping data

identify CNVs. A total of 12,786 CNVs distributed non-uniformly (Fig. 6) along the 29 autosomes, representing 999 duplications and 11,787 deletions, with average sizes of 252.8 ± 692.0 Kbp and 22.9 ± 194.2 Kbp, respectively, were detected in NGS data from eight resequenced bulls when both types of evidence were considered (Additional file 4).

Even though the analyzed NGS dataset was exceedingly smaller than the SNP dataset (8 vs 1,509 animals), and represents a reduced sample of the breed’s genetic diversity, LUMPY detected more than ten times the number of CNVs detected with PennCNV, when the same eight animals were considered. Similar results have been reported in other studies [35, 50] and may be attributed to the better resolution of CNV breakpoints which can be obtained from NGS data. Moreover, the CNV ratio of deletions to duplications observed in the results obtained from NGS data (11.80) is more than 56

times larger than the ratio obtained from genotyping data (0.21), suggesting the method is more sensitive in identifying deletions. JM-CNV [48] was used to converge identified CNVs >1,000 bp into CNVRs. The 12,786 detected CNVs were condensed into 3,781 CNVRs, representing a total of 84 duplications, 909 deletions, and 2,788 duplications and deletions (Fig. 4b, Additional file 5). Inevitable ascertainment bias may have influenced obtained results, as the reference bovine genome sequence was derived from a Hereford individual (*Bos taurus*). Future analysis may be used to identify and correct this when a reliable *Bos indicus* reference sequence becomes available.

CNV and CNVR independent validation and cross-referencing

The importance of comparing CNV detection results with complementary techniques, such as qPCR, FISH,

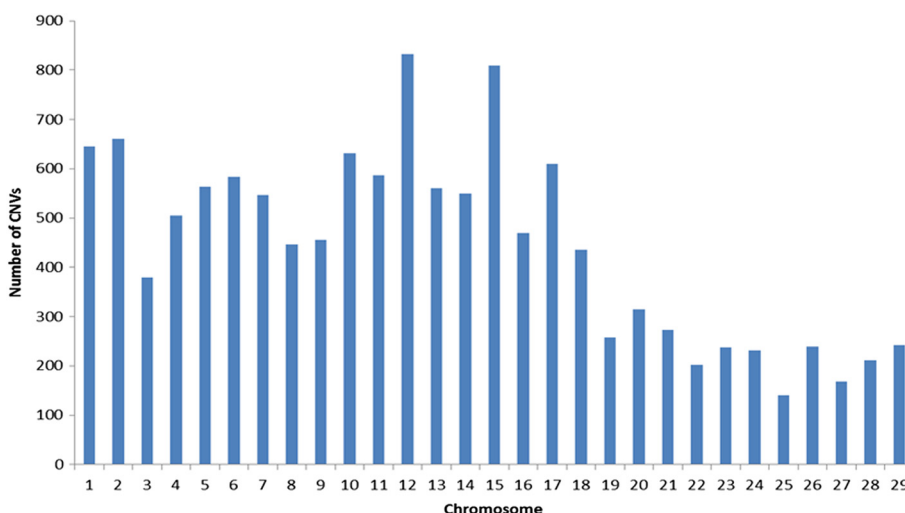


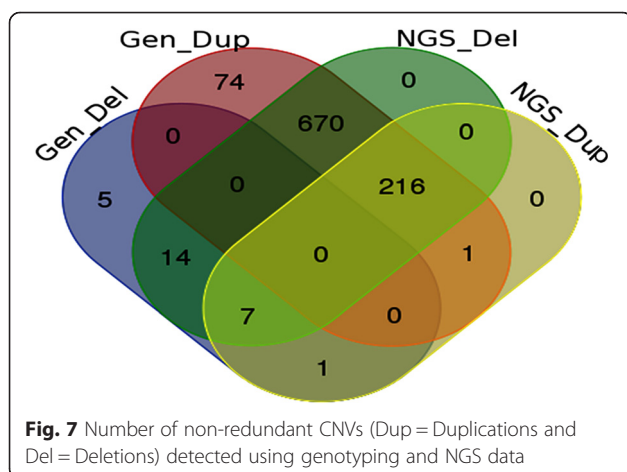
Fig. 6 Chromosome distribution of CNVs detected using Nelore NGS data

Table 1 Summary of CNVs detected using SNP and resequencing data

Animal	# of CNVs detected in SNP data	# of CNVs validated with NGS data	% Validated
BINE_01	11	4	36
BINE_02	87	86	99
BINE_03	27	14	52
BINE_04	22	22	100
BINE_05	93	93	100
BINE_06	729	672	92
BINE_07	9	9	100
BINE_08	10	9	90
Total	988	909	92

CGH arrays, SNP arrays, and sequencing has been extensively reviewed in cattle [35, 51]. Cross-validation of CNVs detected in the genotyping data was performed with NGS data from the eight resequenced animals. A total of 988 CNVs were detected with genotyping data from the eight animals (Additional file 6) and 909 (92 %) of these overlapped with 50 bp or more of at least one of 57,968 CNVs identified with LUMPY using evidence from split-reads and/or miss-mapped paired-end reads - Table 1 (see Additional file 7 for complete list). Further evaluation of the 909 CNVs identified using SNP and NGS data revealed that 173 were identified with all three independent types of evidence (SNP data and signal depth from observed split-reads and from miss-mapped paired-end reads), while 736 were identified with at least two types of evidence (SNP data and observed split-reads or miss-mapped paired-end reads).

A total of 886 of the 988 CNVs (90 %) were observed to contain mixed segments of duplications or deletions considering mostly the NGS data (Fig. 7), which should be considered in future studies as complexity negatively correlates with reproducibility in subsequent CNV



studies with different platforms [52]. The high proportion of observed cross-validated CNVs was contrasted with results reported by previous studies [52, 53]. Observed results show that some CNVs detected with genotyping data overlap with multiple smaller CNVs detected with NGS data (Fig. 8), confirming previous reports [26, 52, 54] which show that NGS offers higher resolution and precision for identification of CNV boundaries.

A total of 68,007 CNVs identified with the SNP dataset were cross-matched with 179 CNVRs previously validated with at least two distinct methods available at DVGarchieve database (<http://www.ebi.ac.uk/dgva/data-download>) and in the literature [26, 27, 30, 33]. A total of 62 % (111) of previously validated cattle CNVRs were found to overlap with CNVs identified in Nelore cattle, considering a minimum of 10 kb of overlap [55, 56]. CNVs with frequencies >1 % were observed in 41 of these previously reported CNVRs in the analyzed Nelore samples (Additional file 8).

Bickhart et al. [26] reported 730 Nelore CNVs from analyses of NGS data from a single animal, considering BTAU4.0 as reference assembly. Conversion of BTAU4.0 to UMD3.1 coordinates using Liftover [57] resulted in 458 CNVs and a total of 295 (64.4 %) of these were found to overlap with one or more of the CNVs currently identified in the NGS data. Observed discrepancies may have resulted from specificities of applied methods as well as sampling bias caused by the extremely reduced sample size used by [26].

CNVRs in regions containing QTLs in cattle

Recent studies [27, 30, 36, 37] revealed CNV variants associated with production traits in dairy and beef cattle. Reported findings suggest that models combining SNP and CNV data could be more powerful at capturing the underlying variation and therefore provide more accurate frameworks to better account for the heritability of complex traits, as the effect of 25 % of identified CNVs could not be accounted for by neighboring SNPs [27].

CNVRs have been detected in genomic regions shown to contain cattle QTLs and have been shown to affect body measurements [17], production traits [37] and parasite resistance [30]. The 7,319 CNVRs detected with genotyping data were compared to the 11,506 regions of the bovine genome reported to contain QTLs (QTL database <http://www.animalgenome.org/cgi-bin/QTLdb/BT/index>). A total of 9.4 % (688/7,319) of the detected CNVRs, which encompass a total of 312Mbp of the bovine autosomal genome, were observed to overlap by >50 % [17] of 286 non-redundant QTLs associated with economically important production traits such as residual feed intake, gestation length, marbling score, fat thickness at the twelfth rib, dry matter intake,

longissimus muscle area, clinical mastitis, and carcass weight (Additional file 9).

All of the 34 CNVs found by [37] to be associated with milk production traits in Holsteins (HOL) were also observed in Nelore (NEL) cattle (Additional file 10). Comparisons of estimated frequencies of these CNVs in the two breeds revealed 14, 13, 6 and 14 regions in high (>20 % in both breeds), low (<20 % in both breeds) and divergent (NEL > HOL, NEL < HOL) frequencies, respectively. Figure 9 shows chromosome positions and frequency differences between Nelore and Holstein cattle at these CNVs.

Considering the distinct selective pressures Nelore and Holstein cattle have been historically under either naturally (tropical versus temperate climates) or artificially (beef versus milk production), frequency deviations are expected in underlying variant regions controlling traits under selection. CNVR_7294 was observed in 56.93 % of the Nelore samples tested, while a CNV located in the same position at frequency of 2.09 % was reported to be strongly associated with protein percentage in Holsteins (FDR = 5,09E-05 [37]). This genome region harbors QTL controlling carcass weight (QTL 13550), milk fat percentage (QTL 13547) and Milk protein percentage (QTL 13548), and the observed frequencies suggest the CNV may be under positive selection in Nelore while strong negative selection in Holsteins. A similar pattern of frequency divergence can be observed with CNVR_7295. Conversely, CNVR_1557, CNVR_3011 and CNVR_4292, located in regions reported to contain QTL affecting beef production traits, were observed at low frequencies in Nelore cattle (0.07 %) and at high frequencies in Holsteins (60.26, 66.05 and 30.42 %, respectively), suggesting these CNVs may contribute to the underlying variation in traits under divergent selection in

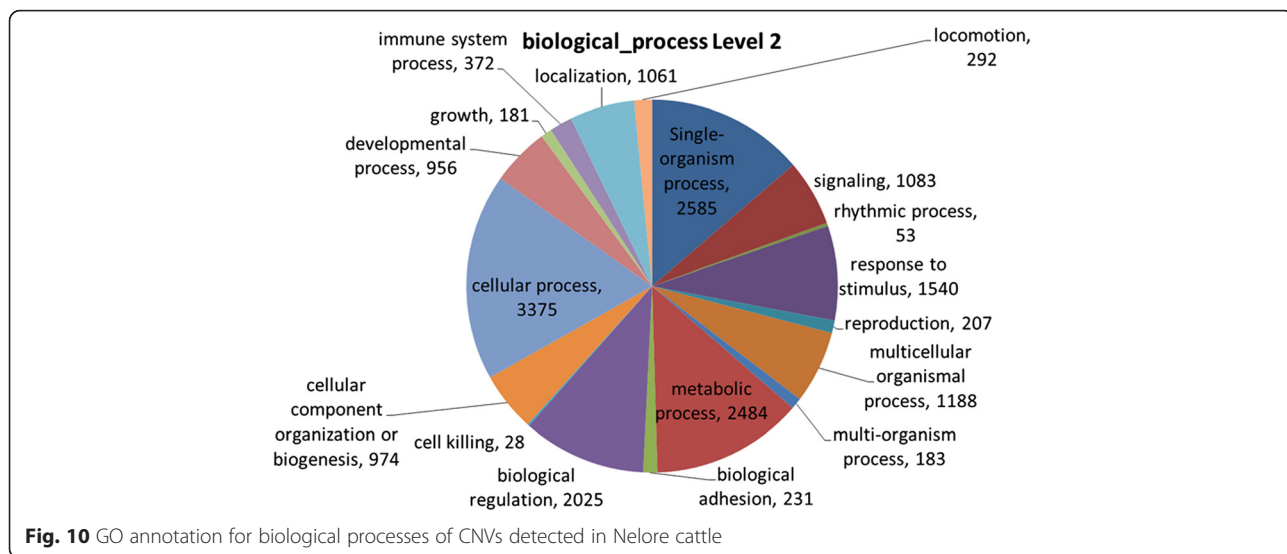
these breeds. These observations suggest that more extensive studies with CNV data from divergent breeds or other population structures could help identify signatures of selection in genome regions containing segmental variations.

Gene ontology and CNVRs

The occurrence of CNVs in genome regions containing functional genes may create opportunities for the emergence of new allelic variants, gene isoforms, and complex mechanisms of gene expression control as a consequence of naturally occurring evolutionary processes. A total of 4,097 CNVRs (55.98 %) are located within genome regions containing 10,399 annotated genes, which can be functionally classified as protein coding (n = 10,070), microRNA (n = 159), snoRNA (n = 148), snRNA (n = 10), miscRNA (n = 8), and rRNA (n = 2).

Automated annotation of these genes with GO terms revealed important categories, including metabolic and cellular processes, biological regulation, response to stimulus, cell signaling, reproduction, and growth (Fig. 10). Many well described contrasting traits between taurine and zebu cattle have been targets of natural selection and production-oriented genetic improvement, and are mediated by genes involved in these biological processes, including reproduction (age of first estrous, fertility, calving interval, etc.) [58], resistance to endo- and ectoparasites [59], heat tolerance [60], disease resistance [61], as well as growth and carcass and meat quality traits [62]. Therefore, further investigation of these regions may unveil important information for understanding underlying mechanisms affecting economically important traits.

Previous studies to identify CNVs in cattle using small numbers of samples from divergent breeds have focused



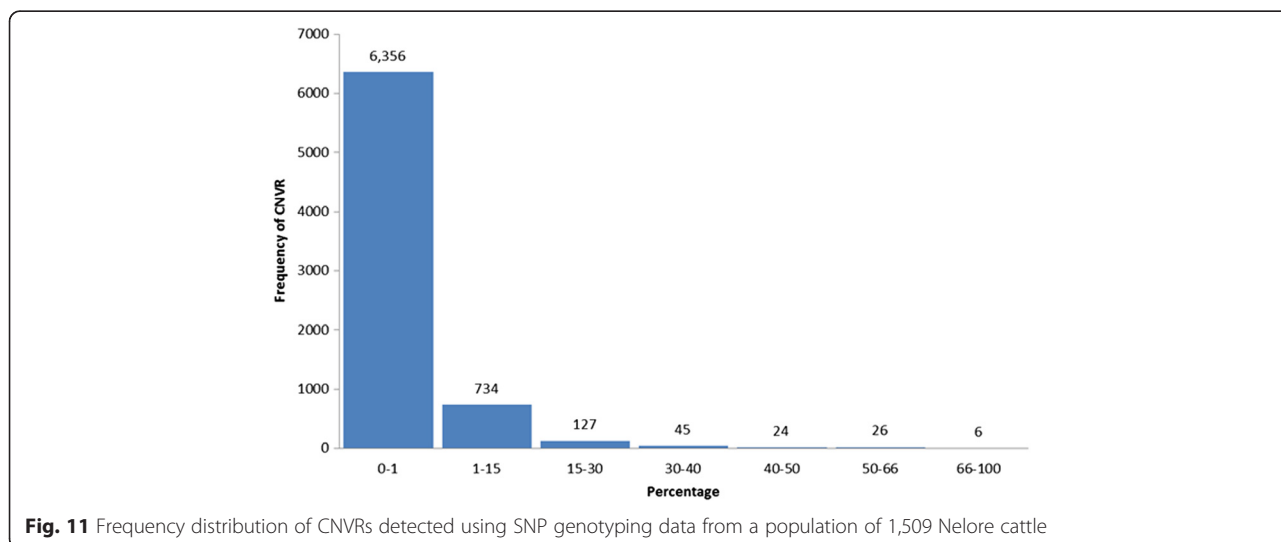
specially on comparisons between breeds [26] and may have provided a comprehensive view of breed-specific CNVs potentially associated with contrasting traits observed among evaluated breeds. Analysis of 1,509 Nelore samples allowed a broad identification of CNVs segregating within the breed in addition to generating population frequency estimates and therefore providing crucial information for inference if observed CNVs may indeed be under selection within the breed. Several previously reported CNVs [26, 63–65] within genome regions containing genes that may control traits of interest for cattle production were observed at extremely low frequencies in the population studied herein (Additional file 10), indicating that these variants may not be positively associated with underlying factors associated with traits under positive selection in the breed.

Sequencing of the bovine reference genome revealed the expansion of the antimicrobial cathelicidin gene, found as a single copy in humans and mice, into a large gene family in cattle [66]. Bickhart et al. [26] reported that one of these cathelicidin genes (CATHL4) was observed to be highly duplicated in the single evaluated Nelore sample. A single copy duplication spanning this gene was observed in both SNP and NGS data but at frequencies <1 %, indicating this particular CNV is not undergoing strong positive selection in the breed (Additional files 2 and 11). Similar divergent results were observed with other genes previously reported to be located in genome regions with CNVs in Nelore cattle and that have been independently shown to affect height (*pleiomorphic adenoma gene 1* - PLAG1), lipid metabolism (*apolipoprotein L3* - APOL3 and *sterol carrier protein 2* - SCP2), transport (*fatty acid binding protein 2* - FABP2, *vesicle associated membrane protein 7* - VAMP7, *lecithin-cholesterol acyltransferase* -

LCAT, and *lecithin-cholesterol acyltransferase* - PCTP), endoparasite resistance (*UL16-binding protein 17* - ULBP17), and oxidative metabolism (*aldehyde oxidase 1* - AOX1) (Additional files 2 and 11).

Genetic imprinting represents a major mechanism of epigenetic regulation of gene expression leading to parent-specific differential expression of a subset of 20 bovine genes (Imprinted Gene Databases - <http://www.geneimprint.com/site/genes-by-species.Bos+taurus> [67]) and DNA sequence polymorphisms in imprinted genes have been shown to affect production traits in cattle [68]. CNVs were observed in regions spanning 11 imprinted genes in Nelore cattle: *mesoderm specific transcript* - MEST (BTA4), *nucleosome assembly protein 1 like 5* - NAP1L5 (BTA6), *insulin like growth factor 2 receptor* - IGF2R (BTA9), *neuronatin* - NNAT (BTA13), *antisense transcript gene of PEG3* - APEG3 (BTA18), *maternally expressed 3* - MEG3 (BTA21), *pleckstrin homology like domain family A member 2* - PHLDA2 (BTA29), *tumor-suppressing subchromosomal transferable fragment 4* - TSSC4 (BTA29), *achaete-scute family bHLH transcription factor 2* - ASCL2 (BTA29), *insulin like growth factor 2* - IGF2 (BTA29), and H19 (BTA29) (Additional file 11).

Observed CNV frequencies in regions harboring MEST, NAP1L5, IGF2R, NNAT, APEG3, and MEG3 were very low (<0.2 %). Conversely, CNV frequencies in the region with imprinted genes on BTA29 (49,329,504-50,163,147 bp) were greater than 9 %. The PHLDA2 gene (also known as TSSC3) is located in the aforementioned region of BTA29 and is expressed in the bovine placenta and embryonic tissues during pregnancy [69, 70]. Comparisons of bovine and human polypeptides revealed a strong homology and suggested that PHLDA2 could be involved in the same regulatory pathways in both species [69]. According to Huang et al. [71], proper



PHLDA2 expression is essential for normal embryo development during early development. Additional studies show that PHLDA2 may affect the development of bovine pre-implantation embryos [72]. A single copy duplication in the region containing PHLDA2 was observed in a total of 128 individuals (Additional files 2 and 11) and should be considered in future studies to evaluate the effect of this gene in early embryo development.

Annotation of most frequent CNVRs in Nelore cattle

CNVs with frequencies higher than 1 % were observed in a total of 13 % (962/7,319) of the detected CNVRs (Fig. 11). Six CNVRs were observed to be highly frequent in Nelore, with more than 1,000 CNVs in the analyzed samples and may therefore be associated with underlying factors positively affecting traits under selection in the breed.

BTA2:104,853,165–105,006,347 contains a duplication that was observed in a total of 1,056 individuals. This genome region harbors genes such as *insulin-like growth factor binding protein 2* (IGFBP2) and *short stature homeobox* (SHOX), among others. Studies in humans show that mutations in this gene can lead to short stature and to different pathological conditions such as Turner syndrome (TS), Léri-Weill dyschondrosteosis, and Langer mesomelic dysplasia [73–76]. IGFBP2 has also been shown to be involved in regulating the estrous cycle and early pregnancy in cattle [77].

BTA4:114,375,180–114,638,146 contains 16 annotated genes, as well as microRNA 671, and was found to be duplicated in more than 1,000 animals and one resequenced individual, and to be deleted in four genotyped animals. Studies on humans show that *cyclin-dependent kinase 5* (CDK5), which is located in this region, plays an important role in central nervous system function. It has also been proposed that CDK5 is important in myogenesis, hematopoietic cell differentiation, spermatogenesis, insulin secretion, and lens differentiation [78, 79]. A study with pigs showed that CDK5 is involved in brain development [80].

BTA6:119,154,914–119,384,691 contains *actin binding LIM protein family member 2* (ABLIM2), *actin filament associated protein 1* (AFAP1), *sortilin related VPS10 domain containing receptor 2* (SORCS2), *prosaposin-like 1* (PSAPL1), and *SH3 domain and tetratricopeptide repeats 1* (SH3TC1). This CNVR was found to be duplicated in more than 1,000 animals and deleted in 20 animals. Klimov et al. showed that the ABLIM2 protein is necessary for normal neuron functioning [81]. SORCS2 was identified as a proneurotrophin receptor and is expressed as a single-chain protein that is essential for proBDNF-induced growth cone collapse in developing dopaminergic processes. Deficiency of SORCS2 in mice caused reduced dopamine levels and metabolism, and dopaminergic hyperinnervation of the frontal cortex [82, 83].

BTA19:48,427,331–48,537,167 harbors *angiotensin I converting enzyme* (ACE), *WD40 repeat-containing protein* (WDR68), and *potassium voltage-gated channel subfamily H member 6* (KCNH6). The ACE gene encodes an enzyme involved in catalyzing the conversion of angiotensin I into angiotensin II, which is a potent vasoconstrictor that controls blood pressure and fluid-electrolyte balance. Gauthier et al. (2013) demonstrated that ACE inhibitor-enhanced bradykinin relaxations of bovine coronary arteries occurs through endothelial cell B1 receptor activation and nitric oxide [84].

BTA19:63,507,097–63,735,382 contains *protein kinase C alpha* (PRKCA), *calcium voltage gated channel auxiliary subunit gamma 4* (CACNG4), and *calcium voltage-gated channel auxiliary subunit gamma 5* (CACNG5) genes, as well as the 7SK misc-RNA and was found to be duplicated in more than 1,000 animals and deleted in 11 animals in the population studied. A study on cattle showed that 7SK misc-RNA is located on a central region of the *hexamethylene bisacetamide inducible 1* (BHEXIM1) gene and may play an important role in gene regulation [85]. The authors proposed that this gene affects the latent life cycle of the bovine immunodeficiency virus (BIV),

Table 2 Genome coverage of eight resequenced animals

Animal ID	Length (bp)	# Aligned	# Unaligned	Total reads	mapped reads	Seq. X coverage
BINE_01	2512082506	1343863053	48758778	1392621831	96.50 %	53
BINE_02	2512082506	1301229539	15069073	1316298612	98.86 %	52
BINE_03	2512082506	1587268578	17505393	1604773971	98.91 %	63
BINE_04	2512082506	645452383	40423110	685875493	94.11 %	26
BINE_05	2512082506	554921159	5980155	560901314	98.93 %	22
BINE_06	2512082506	554580777	6327041	560907818	98.87 %	22
BINE_07	2512082506	1399071315	16203693	1415275008	98.86 %	56
BINE_08	2512082506	639537955	7497809	647035764	98.84 %	25

which leads to a lack of clinical signs of the disease in affected animals. This region may be of interest for studies on the clinical diagnosis and prevention of this disease.

Conclusions

This study represents the first comprehensive CNV survey within the Nelore breed (1,717 animals and ~770 K SNPs). Obtained results allowed for direct comparisons of CNV detection results with two distinct platforms (HD SNP genotyping and NG sequencing), and with previous reports from independent studies.

The bovine CNV map was significantly enriched, particularly for the Nelore breed and associated variant frequency estimates enabled the identification of variants potentially associated with traits under selection, particularly in genome regions harboring QTLs affecting production traits.

Obtained results suggest that more extensive studies using CNV data from divergent breeds with differing population structures could help identify signatures of selection using approaches frequently used with SNP data. The study provides important information that may inspire or contribute to future studies on the association between CNVs and production traits important for genetic improvement in cattle.

Methods

Animals

DNA was extracted from commercially available semen samples, and from hair and venous blood samples obtained from animals in production farms, as part of routine animal handling and testing procedures. Tissues were processed with standard commercial kits.

Genotyping and resequencing data

A total of 1,717 Nelore (*Bos indicus*) samples were genotyped with the Illumina Bovine HD Genotyping Bead Chip. DNA was extracted from semen, blood, or hair samples from registered and production animals from commercial farms in Brazil. In addition, DNA from eight unrelated Nelore founding bulls was resequenced using Illumina HiSeq2000 paired-end reads with a minimum coverage of 20× (Table 2) [86].

CNV and CNVR detection in genotyping data

Illumina genotyping data was analyzed with PennCNV [87]. Log R Ratios (LRR), B Allele Frequencies (BAF), distances between neighboring SNPs, and pedigree information were used by the Hidden Markov Model (HMM) algorithm to detect CNVs. Only autosomal SNPs were considered in the analysis. Initial analysis of the dataset with default LRR and BAF cut-off values normally used

in CNV studies on humans [88, 89] resulted in the exclusion of 997 animals (data not shown). Adjusted LRR and BAF cutoff values were derived for analysis of the Nelore dataset based on the observed distributions of these variables in the studied samples. New LRR and BAF cut off values were identified to independently exclude 10 % of the samples. In addition, a GC content correction was performed for each SNP in regions located 500Kb upstream and downstream from each studied SNP [32]. Use of new LRR (<0.4) and BAF (<0.04) cut-off values in conjunction resulted in removal of 208 samples (12 %) from the final dataset. PennCNV default procedures and parameters were subsequently used in the analysis.

Overlapping CNVs were grouped into CNVRs using JM-CNV [48]. CNVs were grouped into closed intervals of whole numbers. This choice made CNVR definition more natural and included the set of intervals whose overlap did not exceed the average size of the CNV set plus one standard deviation. Meanwhile, long and infrequent CNVRs were grouped separately so they would not skew estimated averages and standard deviations.

CNV detection in NGS data

A previously described strategy for determining high-resolution CNVs in humans [90] was used to identify CNVs in Illumina shotgun data from eight key ancestral Nelore bulls. Paired-end reads were mapped onto the UMD 3.1 assembly using BWA with default parameters [91]. CNVs were detected using LUMPY, a novel CNV discovery framework that uses multiple detection signals including read depth from split reads and mis-mapped paired ends [49] (Additional file 4) for CNV identification. Only autosomal regions were considered in the analysis. Overlapping CNVs >1,000 bp were grouped into CNVRs using JM-CNV [48].

Cross validation of CNVs

CNVs detected with SNP genotyping data were cross-validated using a combination of information derived from eight resequenced Nelore bulls and from published literature, following previously reported strategies [52, 56]. Sequence coordinates from CNVs detected using genotyping methods (Additional file 8) were initially compared to coordinates from 179 CNVRs previously validated in independent studies [26, 27, 30, 92]. Coordinates from CNVs observed with PennCNV and LUMPY were compared using a script written in Python [53] (Additional files 6 and 7). All CNVs >50 bp detected with LUMPY were used in this procedure.

Functional annotation

Automated annotation of genes present within observed CNVs was performed using the scan_region.pl tool from

PennCNV and the annotation file of UMD3.1 assembly [93]. Ensembl Genes 77 database (*Bos taurus* genes UMD3.1) and BioMart were used to annotate observed CNVRs. FASTA sequence files containing annotated gene regions from observed CNVRs were imported into Blast2GO [94, 95] for automatic functional annotation. These files were blasted against the NCBI nr database using default BlastX parameters (e-value threshold 1e-03 and HSP length cut-off of 100). Sequence mapping for Gene Ontology (GO) terms was performed using default parameters (e-value hit filter of 1e-06, annotation cut-off of 55, and GO weight of 5). Annotations were performed using the Annex function of the GO Annotation Toolbox [96]. InterProScan terms were obtained following a previously reported method [97]. In addition, metabolic pathway maps were obtained using the method outlined by the KEGG PATHWAY database [98]. Overlaps between detected CNVs and CNVRs and previously detected QTLs from the Bovine QTL Database [99] were identified with a script in Python (Additional file 9).

Additional files

- Additional file 1:** CNVs detected in Nelore HD SNP genotyping data. (XLSX 48 kb)
- Additional file 2:** CNVRs detected in Nelore HD SNP genotyping data. (XLSX 1160 kb)
- Additional file 3:** Descriptive statistics of CNVRs per individual bovine chromosome. (XLSX 5510 kb)
- Additional file 4:** CNVs detected in Nelore NGS Data. (XLSX 450 kb)
- Additional file 5:** CNVRs detected in Nelore NGS Data. (XLSX 12 kb)
- Additional file 6:** CNVs detected in HD SNP genotyping data from the eight resequenced bulls. (XLSX 648 kb)
- Additional file 7:** Overlap of CNVs detected with both genotyping and NGS data from eight key ancestral bulls. (XLSX 196 kb)
- Additional file 8:** Overlap of all CNVs identified in Nelore cattle with CNVRs currently listed at DVGarchieve database. (XLSX 57 kb)
- Additional file 9:** CNVRs overlapping with previously detected QTLs from the Bovine QTL Database. (XLSX 2973 kb)
- Additional file 10:** CNVs reported by Xu et al. (2014a) to be associated with milk production traits in Holsteins also observed in Nelore cattle. (XLSX 289 kb)
- Additional file 11:** Gene Ontology annotation in CNVRs detected in Nelore HD SNP genotyping data. (XLSX 313 kb)

Acknowledgments

We would like to thank EMBRAPA Multiuser Bioinformatics Lab (Laboratório Multiusuário de Bioinformática da Embrapa) for providing additional computational infrastructure.

Funding

Research supported by Embrapa- Brazilian Agricultural Research Corporation (grants 02.10.06.009.00, 01.11.07.002.06.00), Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant 578592/2008-8) and Fundação de Amparo à Pesquisa do Estado de São Paulo (grant 2012/05002-9). S.R. Paiva and A.R. Caetano are CNPq research fellows.

Availability of data and material

The datasets supporting the conclusions of this article are included within the article and additional files. The NGS resequencing data supporting the

conclusions of this article are available in the NCBI Sequence Read Archive repository (Acc. #SRP068091 - http://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser) and (Acc. #estd227) in Database of Genomic Variants archive.

Authors' contributions

Experimental design: JMS, PFG, MEBY, ARC. Sample selection, collection and processing for NGS: LOCS, SRP, ARC. Genotyping and NGS data analysis: JMS, PFG, LCC, MEBY. Result interpretation: JMS, PFG, MEBY, SRP, ARC. Manuscript preparation: JMS, PFG, MEBY, ARC. All authors read and approved the final manuscript.

Authors' information

Samuel Rezende Paiva and Alexandre Rodrigues Caetano: CNPq Fellow.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Specific approval from an Animal Care and Use Committee was not obtained for this study as analyses were performed with data previously generated from samples previously collected as part of commercial testing procedures.

Author details

¹Faculdade de Ciências Agrárias, Biológicas e Sociais Aplicadas, Universidade do Estado de Mato Grosso (UNEMAT), Av. Prof Dr. Renato Figueiro Varella, CEP 78.690-000 Nova Xavantina, Mato Grosso, Brazil. ²Programa de Pós-Graduação em Genética e Biologia Molecular-Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil. ³Embrapa Informática Agropecuária - Laboratório Multiusuário de Bioinformática (LMB), Campinas, São Paulo, Brazil. ⁴Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brazil. ⁵Embrapa - Secretária de Relações Internacionais, Brasília, Distrito Federal, Brazil. ⁶Embrapa Recursos Genéticos e Biotecnologia, Brasília, Distrito Federal, Brazil. ⁷CNPq Fellow.

Received: 20 February 2016 Accepted: 19 May 2016

Published online: 13 June 2016

References

- Database of Genomic Variants. <http://dgvbeta.tcag.ca/dgv/app/statistics?ref=GRCh37/hg19>. Accessed 18 Sept 2015.
- Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science*. 2008;320(5883):1629–31.
- Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi MY, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004;305(5683):525–8.
- Gupta A, Place M, Goldstein S, Sarkar D, Zhou S, Potamouisis K, Kim J, Flanagan C, Li Y, Newton MA, et al. Single-molecule analysis reveals widespread structural variation in multiple myeloma. *Proc Natl Acad Sci U S A*. 2015;112(25):7689–94.
- Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, et al. A copy number variation morbidity map of developmental delay. *Nat Genet*. 2011;43(9):838–U844.
- Cooper NJ, Shtir CJ, Smyth DJ, Guo H, Swafford AD, Zanda M, Hurles ME, Walker NM, Plagnol V, Cooper JD, et al. Detection and correction of artefacts in estimation of rare copy number variants and analysis of rare deletions in type 1 diabetes. *Hum Mol Genet*. 2015;24(6):1774–90.
- Casey JP, Magalhaes T, Conroy JM, Regan R, Shah N, Anney R, Shields DC, Abrahams BS, Almeida J, Bacchelli E, et al. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum Genet*. 2012;131(4):565–79.
- Almal SH, Padh H. Implications of gene copy-number variation in health and diseases. *J Hum Genet*. 2012;57(1):6–13.
- Moustafa JSE-S, Eleftherohorinou H, de Smith AJ, Andersson-Assarsson JC, Alves AC, Hadjigeorgiou E, Walters RG, Asher JE, Bottolo L, Buxton JL, et al. Novel association approach for variable number tandem repeats (VNTRs) identifies DOCK5 as a susceptibility gene for severe obesity. *Hum Mol Genet*. 2012;21(16):3727–38.

10. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. 2010; 464(7289):704–12.
11. Park RW, Kim T-M, Kasif S, Park PJ. Identification of rare germline copy number variations over-represented in five human cancer types. *Mol Cancer*. 2015;14:25.
12. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 2010; 466(7304):368–72.
13. Ouyang L, Lee J, Park C-K, Mao M, Shi Y, Gong Z, Zheng H, Li Y, Zhao Y, Wang G, et al. Whole-genome sequencing of matched primary and metastatic hepatocellular carcinomas. *BMC Med Genet*. 2014;7:2.
14. Malek SN. The biology and clinical significance of acquired genomic copy number aberrations and recurrent gene mutations in chronic lymphocytic leukemia. *Oncogene*. 2013;32(23):2805–17.
15. Verma M, Khoury MJ, Ioannidis JPA. Opportunities and challenges for selected emerging technologies in cancer epidemiology: mitochondrial, epigenomic, metabolomic, and telomerase profiling. *Cancer Epidemiol Biomark Prev*. 2013;22(2):189–200.
16. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet*. 2015;16(3):172–83.
17. Zhang L, Jia S, Yang M, Xu Y, Li C, Sun J, Huang Y, Lan X, Lei C, Zhou Y, et al. Detection of copy number variations and their effects in Chinese bulls. *BMC Genomics*. 2014;15:480.
18. Cicconardi F, Chillemi G, Tramontano A, Marchitelli C, Valentini A, Ajmone-Marsan P, Nardone A. Massive screening of copy number population-scale variation in *Bos taurus* genome. *BMC Genomics*. 2013;14:124.
19. Xu L, Hou Y, Bickhart DM, Zhou Y, Hay EHA, Song J, Sonstegard TS, Van Tassell CP, Liu GE. Population-genetic properties of differentiated copy number variations in cattle. *Sci Rep*. 2016;6:23161.
20. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, MacDonald JR, Mills R, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol*. 2011;29(6):512–U576.
21. Ionita-Laza I, Rogers AJ, Lange C, Raby BA, Lee C. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics*. 2009;93(1):22–6.
22. Curtis C, Lynch AG, Dunning MJ, Spiteri I, Marioni JC, Hadfield J, Chin S-F, Brenton JD, Tavare S, Caldas C. The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics*. 2009;10:588.
23. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet*. 2009;41(10):1061–U1029.
24. Alkan C, Coe BP, Eichler EE. Applications of next-generation sequencing genome structural variation discovery and genotyping. *Nat Rev Genet*. 2011; 12(5):363–75.
25. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*. 2010;11(10):685–96.
26. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, Schnabe RD, Ventura M, Taylor JF, et al. Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res*. 2012;22(4):778–90.
27. Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim ES, Matukumalli LK, Ventura M, Song J, Vanraden PM, et al. Genomic characteristics of cattle copy number variations. *BMC Genomics*. 2011;12:127.
28. Hou Y, Bickhart DM, Chung H, Hutchison JL, Norman HD, Connor EE, Liu GE. Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. *Funct Integr Genomics*. 2012;12(4):717–23.
29. Hou Y, Bickhart DM, Hvinden ML, Li C, Song J, Boichard DA, Fritz S, Eggen A, Denise S, Wiggans GR, et al. Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. *BMC Genomics*. 2012;13:376.
30. Hou Y, Liu GE, Bickhart DM, Matukumalli LK, Li C, Song J, Gasbarre LC, Van Tassell CP, Sonstegard TS. Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. *Funct Integr Genomics*. 2012;12(1):81–92.
31. Jiang L, Jiang J, Wang J, Ding X, Liu J, Zhang Q. Genome-wide identification of copy number variations in Chinese Holstein. *Plos One*. 2012;7(11):e48732.
32. Jiang L, Jiang J, Yang J, Liu X, Wang J, Wang H, Ding X, Liu J, Zhang Q. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. *BMC Genomics*. 2013;14:33.
33. Liu GE, Van Tassell CP, Sonstegard TS, Li RW, Alexander LJ, Keele JW, Matukumalli LK, Smith TP, Gasbarre LC. Detection of germline and somatic copy number variations in cattle. *Anim Genomics Anim Health*. 2008;132:231–7.
34. Liu GE, Ventura M, Cellamare A, Chen L, Cheng Z, Zhu B, Li C, Song J, Eichler EE. Analysis of recent segmental duplications in the bovine genome. *BMC Genomics*. 2009;10:571.
35. Liu GE, Bickhart DM. Copy number variation in the cattle genome. *Funct Integr Genomics*. 2012;12(4):609–24.
36. Xu L, Hon Y, Bickhart DM, Song J, Van Tassell CP, Sonstegard TS, Liu GE. A genome-wide survey reveals a deletion polymorphism associated with resistance to gastrointestinal nematodes in Angus cattle. *Funct Integr Genomics*. 2014;14(2):333–9.
37. Xu L, Cole JB, Bickhart DM, Hou Y, Song J, Vanraden PM, Sonstegard TS, Van Tassell CP, Liu GE. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. *BMC Genomics*. 2014;15:683.
38. Volker M, Backstrom N, Skinner BM, Langley EJ, Bunzey SK, Ellegren H, Griffin DK. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Res*. 2010; 20(4):503–11.
39. Wang X, Nahashon S, Feaster TK, Bohannon-Stewart A, Adefope N. An initial map of chromosomal segmental copy number variations in the chicken. *BMC Genomics*. 2010;11:351.
40. Fadista J, Nygaard M, Holm L-E, Thomsen B, Bendixen C. A snapshot of CNVs in the Pig genome. *Plos One*. 2008;3(12):e3916.
41. Ramayo-Caldas Y, Castello A, Pena RN, Alves E, Mercade A, Souza CA, Fernandez AI, Perez-Enciso M, Folch JM. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. *BMC Genomics*. 2010; 11:593.
42. Fontanesi L, Beretti F, Martelli PL, Colombo M, Dall'olio S, Occidente M, Portolano B, Casadio R, Matassino D, Russo V. A first comparative map of copy number variations in the sheep genome. *Genomics*. 2011;97(3):158–65.
43. Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F, et al. Analysis of copy number variations in the sheep genome using 50 K SNP BeadChip array. *BMC Genomics*. 2013;14:229.
44. Fontanesi L, Martelli PL, Beretti F, Riggio V, Dall'olio S, Colombo M, Casadio R, Russo V, Portolano B. An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genomics*. 2010;11:639.
45. Butler JL, Locke MEO, Hill KA, Daley M. HD-CNV: hotspot detector for copy number variants. *Bioinformatics*. 2013;29(2):262–3.
46. Kim J-H, Hu H-J, Yim S-H, Bae JS, Kim S-Y, Chung Y-J. CNVRuler: a copy number variation-based case-control association analysis tool. *Bioinformatics*. 2012;28(13):1790–2.
47. Glessner JT, Li J, Hakonarson H. ParseCNV integrative copy number variation association software with quality tracking. *Nucleic Acids Res*. 2013;41(5):e64.
48. Java Merging Copy Number Variants (JM-CNV): A New Algorithm for Identifying Copy Number Variant Regions (CNVR). <https://www.lmb.cnptia.embrapa.br/tools/JMCNV/JMCNVUpload>.
49. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 2014;15(6):R84.
50. Hayes JL, Tzika A, Thygesen H, Berri S, Wood HM, Hewitt S, Pendlebury M, Coates A, Willoughby L, Watson CM, et al. Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. *Genomics*. 2013;102(3):174–81.
51. Xu L, Hou Y, Bickhart D, Song J, Liu G. Comparative analysis of CNV calling algorithms: literature survey and a case study using bovine high-density SNP data. *Microarrays*. 2013;2(3):171.
52. Agam A, Yalcin B, Bhomra A, Cubin M, Webber C, Holmes C, Flint J, Mott R. Elusive copy number variation in the mouse genome. *Plos One*. 2010;5(9):e12839.
53. Zhan B, Fadista J, Thomsen B, Hedegaard J, Panitz F, Bendixen C. Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. *BMC Genomics*. 2011;12:557.
54. Jiang J, Wang J, Wang H, Zhang Y, Kang H, Feng X, Wang J, Yin Z, Bao W, Zhang Q, et al. Global copy number analyses by next generation sequencing provide insight into pig genome variation. *BMC Genomics*. 2014;15:593.

55. Berglund J, Nevalainen EM, Molin A-M, Perloski M, Andre C, Zody MC, Sharpe T, Hitte C, Lindblad-Toh K, Lohi H, et al. Novel origins of copy number variation in the dog genome. *Genome Biol.* 2012;13(8):R73.
56. Molin AM, Berglund J, Webster MT, Lindblad-Toh K. Genome-wide copy number variant discovery in dogs using the CanineHD genotyping array. *BMC Genomics.* 2014;15:210.
57. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haussler M, et al. The UCSC genome browser database: 2015 update. *Nucleic Acids Res.* 2015;43(D1):D670–81.
58. Sartori R, Bastos MR, Baruselli PS, Gimenes LU, Ereno RL, Barros CM. Physiological differences and implications to reproductive management of *Bos taurus* and *Bos indicus* cattle in a tropical environment. *Reprod Domest Rumin Vii.* 2010;67:357–75.
59. Piper EK, Jonsson NN, Gondro C, Lew-Tabor AE, Moolhuijzen P, Vance ME, Jackson LA. Immunological profiles of *Bos taurus* and *Bos indicus* cattle infested with the cattle tick, *Rhipicephalus (boophilus) microplus*. *Clin Vaccine Immunol.* 2009;16(7):1074–86.
60. Beatty DT, Barnes A, Taylor E, Pethick D, McCarthy M, Maloney SK. Physiological responses of *Bos taurus* and *Bos indicus* cattle to prolonged, continuous heat and humidity. *J Anim Sci.* 2006;84(4):972–85.
61. Brunelle BW, Greenlee JJ, Seabury CM, Brown II CE, Nicholson EM. Frequencies of polymorphisms associated with BSE resistance differ significantly between *Bos taurus*, *Bos indicus*, and composite cattle. *BMC Vet Res.* 2008;4:36.
62. Bolormaa S, Pryce JE, Kemper KE, Hayes BJ, Zhang Y, Tier B, Barendse W, Reverter A, Goddard ME. Detection of quantitative trait loci in *Bos indicus* and *Bos taurus* cattle using genome-wide association studies. *Genet Sel Evol.* 2013;45:43.
63. Bera A, Singh S, Nagaraj R, Vaidya T. Induction of autophagic cell death in *Leishmania donovani* by antimicrobial peptides. *Mol Biochem Parasitol.* 2003;127(1):23–35.
64. Kulkarni MM, Barbi J, McMaster WR, Gallo RL, Satskar AR, Mcgwire BS. Mammalian antimicrobial peptide influences control of cutaneous *Leishmania* infection. *Cell Microbiol.* 2011;13(6):913–23.
65. Karim L, Takeda H, Lin L, Druet T, Arias JAC, Baurain D, Cambisano N, Davis SR, Farnir F, Grisart B, et al. Variants modulating the expression of a chromosome domain encompassing *PLAG1* influence bovine stature. *Nat Genet.* 2011;43(5):405.
66. Elsik CG, Tellam RL, Worley KC, Gibbs RA, Abatepaulo ARR, Abbey CA, Adelson DL, Aerts J, Ahola V, Alexander L, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science.* 2009;324(5926):522–8.
67. Imprinted Gene Databases. <http://www.geneimprint.com/site/genes-by-species.Bos+taurus>. Accessed 18 Sept 2015.
68. Lawson HA, Cheverud JM, Wolf JB. Genomic imprinting and parent-of-origin effects on complex traits. *Nat Rev Genet.* 2013;14(9):608–17.
69. Guillomot M, Taghouti G, Constant F, Degrelle S, Hue I, Chavatte-Palmer P, Jammes H. Abnormal expression of the imprinted gene *Phlda2* in cloned bovine placenta. *Placenta.* 2010;31(6):482–90.
70. Sikora KM, Magee DA, Berkowicz EW, Lonergan P, Evans ACO, Carter F, Comte A, Waters SM, Machugh DE, Spillane C. *PHLDA2* is an imprinted gene in cattle. *Anim Genet.* 2012;43(5):587–90.
71. Huang W, Yandell BS, Khatib H. Transcriptomic profiling of bovine IVF embryos revealed candidate genes and pathways involved in early embryonic development. *BMC Genomics.* 2010;11:23.
72. Driver AM, Huang W, Kropp J, Penagaricano F, Khatib H. Knockdown of *CDKN1C* (*p57*(*kip2*)) and *PHLDA2* results in developmental changes in bovine Pre-implantation embryos. *Plos One.* 2013;8(7):e69490.
73. Gatta V, Palka C, Chiavaroli V, Franchi S, Cannataro G, Savastano M, Cotroneo AR, Chiarelli F, Mohn A, Stuppia L. Spectrum of phenotypic anomalies in four families with deletion of the *SHOX* enhancer region. *BMC Med Genet.* 2014;15:87.
74. Clement-Jones M, Schiller S, Rao E, Blaschke RJ, Zuniga A, Zeller R, Robson SC, Binder G, Glass I, Strachan T, et al. The short stature homeobox gene *SHOX* is involved in skeletal abnormalities in Turner syndrome. *Hum Mol Genet.* 2000;9(5):695–702.
75. Rappold GA, Fukami M, Niesler B, Schiller S, Zunkeller W, Bettendorf M, Heinrich U, Vlachopapadoupoulou E, Reinehr T, Onigata K, et al. Deletions of the homeobox gene *SHOX* (short stature homeobox) are an important cause of growth failure in children with short stature. *J Clin Endocrinol Metab.* 2002;87(3):1402–6.
76. Shears DJ, Vassal HJ, Goodman FR, Palmer RW, Reardon W, Superti-Furga A, Scambler PJ, Winter RM. Mutation and deletion of the pseudoautosomal gene *SHOX* cause Leri-Weill dyschondrosteosis. *Nat Genet.* 1998;19(1):70–3.
77. McCarthy SD, Roche JF, Forde N. Temporal changes in endometrial gene expression and protein localization of members of the IGF family in cattle: effects of progesterone and pregnancy. *Physiol Genomics.* 2012;44(2):130–40.
78. Yamada M, Saito T, Sato Y, Kawai Y, Sekigawa A, Hamazumi Y, Asada A, Wada M, Doi H, Hisanaga S. *Cdk5-p39* is a labile complex with the similar substrate specificity to *Cdk5-p35*. *J Neurochem.* 2007;102(5):1477–87.
79. Dhavan R, Tsai LH. A decade of *CDK5*. *Nat Rev Mol Cell Biol.* 2001;2(10):749–59.
80. Long H, Zhao S, Lei T, Han J, Yuan J, Qi Y, Yang Z. Cloning and spatio-temporal expression of porcine *CDK5* and *CDK5R1(p35)* genes. *Anim Biotechnol.* 2009;20(3):133–43.
81. Klimov E, Rud'ko O, Rakhmanaliev E, Sulimova G. Genomic organisation and tissue specific expression of *ABLIM2* gene in human, mouse and rat. *Biochim Biophys Acta.* 2005;1730(1):1–9.
82. Glerup S, Olsen D, Vaegter CB, Gustafsen C, Sjoegaard SS, Hermey G, Kjolby M, Molgaard S, Ulrichsen M, Boggild S, et al. *SorCS2* regulates dopaminergic wiring and is processed into an apoptotic Two-chain receptor in peripheral glia. *Neuron.* 2014;82(5):1074–87.
83. Rezgaoui M, Hermey G, Riedel IB, Hampe W, Schaller HC, Hermans-Borgmeyer I. Identification of *SorCS2*, a novel member of the *VPS10* domain containing receptor family, prominently expressed in the developing mouse brain. *Mech Dev.* 2001;100(2):335–8.
84. Gauthier KM, Cepura CJ, Campbell WB. ACE inhibition enhances bradykinin relaxations through nitric oxide and B1 receptor activation in bovine coronary arteries. *Biol Chem.* 2013;394(9):1205–12.
85. Guo H-y, Ma Y-g, Y-M G, Liang Z-b, Ma J, Su Y, Zhang Q-c, Chen Q-m, Tan J. Bovine *HEXIM1* inhibits bovine immunodeficiency virus replication through regulating *BTat*-mediated transactivation. *Vet Res.* 2013;44:44.
86. Jansen S, Aigner B, Pausch H, Wysocki M, Eck S, Benet-Pages A, Graf E, Wieland T, Strom TM, Meitinger T, et al. Assessment of the genomic variation in a cattle population by re-sequencing of key animals at low to medium coverage. *BMC Genomics.* 2013;14:44.
87. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, Hakonarson H, Bucan M. *PennCNV*: an integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 2007;17(11):1665–74.
88. Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, Bucan M, Maris JM, Wang K. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* 2008;36(19):e126.
89. Ballif BC, Hornor SA, Jenkins E, Madan-Khetarpal S, Surti U, Jackson KE, Asamoah A, Brock PL, Gowans GC, Conway RL, et al. Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nat Genet.* 2007;39(9):1071–3.
90. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Eichler EE, et al. Diversity of human copy number variation and multicopy genes. *Science.* 2010;330(6004):641–6.
91. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
92. Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, et al. Analysis of copy number variations among diverse cattle breeds. *Genome Res.* 2010;20(5):693–703.
93. Cow Genome. <http://hgdownload.soe.ucsc.edu/downloads.html#cow>. Accessed 18 Sept 2015.
94. Blast2GO. <https://www.blast2go.com/>. Accessed 18 Sept 2015.
95. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 2008;36(10):3420–35.
96. Myhre S, Tveit H, Mollestad T, Laegreid A. Additional gene ontology structure for improved biological reasoning. *Bioinformatics.* 2006;22(16):2020–7.
97. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. Analysis tool Web services from the EMBL-EBI. *Nucleic Acids Res.* 2013;41(W1):W597–600.
98. Keeg. <http://www.genome.jp/kegg/pathway.html>. Accessed 18 Sept 2015.
99. Animal QTL database. <http://www.animalgenome.org/cgi-bin/QTLdb/BT/index>. Accessed 18 Sept 2015.