

The application of genome-wide SNP genotyping methods in studies on livestock genomes

Artur Gurgul · Ewelina Semik · Klaudia Pawlina ·
Tomasz Szmatoła · Igor Jasielczuk ·
Monika Bugno-Poniewierska

Received: 22 November 2013 / Revised: 14 January 2014 / Accepted: 4 February 2014 / Published online: 25 February 2014
© Institute of Plant Genetics, Polish Academy of Sciences, Poznan 2014

Abstract Animal genomics is currently undergoing dynamic development, which is driven by the flourishing of high-throughput genome analysis methods. Recently, a large number of animals has been genotyped with the use of whole-genome genotyping assays in the course of genomic selection programmes. The results of such genotyping can also be used for studies on different aspects of livestock genome functioning and diversity. In this article, we review the recent literature concentrating on various aspects of animal genomics, including studies on linkage disequilibrium, runs of homozygosity, selection signatures, copy number variation and genetic differentiation of animal populations. Our work is aimed at providing insight into certain achievements of animal genomics and to arouse interest in basic research on the complexity and structure of the genomes of livestock.

Keywords Copy number variation · Genetic differentiation · Linkage disequilibrium · Runs of homozygosity · Selection signatures

Introduction

Despite the fact that genomics is still a young discipline of knowledge, its achievements are not to be underestimated, especially in the field of human genetics and recognition of the basics of genetic diseases, diagnostics of genetic defects (Ahn et al. 2013) or identification of changes arising in the genetic material of tissues subjected to malignant transformation (Chung et al. 2006). The development of analytical

approaches progressing along with the development of technology and increasing knowledge about sequences of genomes allowed extending the achievements of genomics beyond the framework of experimental applications. Currently, genomics is being heavily engaged in the explanation of complex mechanisms and relationships occurring within a genome in both physiological and pathological conditions and searching for answers that may explain still unknown aspects of its functioning. Following the reduction in costs of genome-wide genetic analysis, genomics has also entered the area of animal genetics, especially in the aspects of biodiversity (Twito et al. 2007; Kijas et al. 2009), animal production (Hayes et al. 2009a; Zhang et al. 2013), susceptibility to diseases (Zhang et al. 2012; Kizilkaya et al. 2013) or identification of genetic factors underlying observed phenotypical traits (Cargill et al. 2008; Ren et al. 2011).

Much has already been said about the use of methods of genomics for assessing the livestock breeding value or identification of genomic regions associated with different production traits (Hayes et al. 2009a; Saatchi et al. 2011; Weber et al. 2012). However, the application of genomics in terms of pure biology, aiming at the identification of the basics of functioning, variability and structure of the genome of these animals, is less popular and often underestimated. Since animal production cannot take place without a series of purely biological processes occurring in cells or tissues, basic research in this area should be within the scope of quantitative geneticists. Moreover, the correct estimation of the breeding value of animals based on molecular markers cannot be performed without extensive knowledge about complexity of the genome.

So far, the most widely used tool in studies on livestock genomes are genotyping microarrays. They allow a relatively quick, reliable and inexpensive determination of genotypes of a large number of single nucleotide polymorphisms (SNPs) being a primary source of genetic variation. The microarrays

A. Gurgul (✉) · E. Semik · K. Pawlina · T. Szmatoła ·
I. Jasielczuk · M. Bugno-Poniewierska
Laboratory of Genomics, National Research Institute of Animal
Production, Krakowska 1, 32-083 Balice, Poland
e-mail: agurgul@izoo.krakow.pl

are designed to describe the genetic variation within a genome of interest in the best possible way, owing to the use of linkage disequilibrium phenomenon (Matukumalli et al. 2009; Kranis et al. 2013). Currently, the most advanced genotyping tools in animal genomics are becoming available for cattle and allow analysis of about 800,000 SNPs in parallel, which means that, in the range of each Mb of genomic sequence, the genetic variation is described by about 300 markers (Rincon et al. 2011). It gives a detailed insight into the genome of the species and allows for a detailed analysis in aspects of its variability, rearrangements and structure.

In view of the growing number of studies on livestock genomics, in this work, we undertook a review of various applications of genomics in terms of population genetics, biodiversity, the structure of the genome and phenomena occurring therein. In this paper, we focused on applications deviating from the prevailing trend of genomic selection, describing the basic research on the complexity of the genome of farm animals.

Studies on linkage disequilibrium (LD)

In the current era of genome-wide association studies, the knowledge of linkage disequilibrium (LD) between markers is important in order to establish the number of markers necessary for genomic selection, efficient association studies and fine mapping of genetic diseases (Pritchard and Przeworski 2001; Espigolan et al. 2013). LD is defined as the non-random association of alleles at two or more loci and is influenced by, inter alia, population history and its evolution (Ardlie et al. 2002; Khatkar et al. 2008).

Studies on LD throughout a genome can be used to reflect population history, breeding systems and patterns of geographic subdivision, while LD in specific genomic regions gives an opportunity to learn more about the history of natural selection, gene conversion, mutations and other factors that cause gene-frequency evolution (Slatkin 2008). In animal populations, these allelic associations are also extremely valuable in localising genes affecting quantitative traits (quantitative trait loci, QTL) and are necessary to detect associations between a QTL and a marker (Pritchard and Przeworski 2001; Du et al. 2007). A local recombination rate is one of the main factors influencing LD. Regions with a low recombination rate, like the Y chromosome, parts of the X chromosome and regions near the centromere in autosomes, are characterised by high LD extent. On the other hand, small LD extent between two loci is typical for regions with a high recombination rate, such as euchromatin and small regions known as recombination hotspots (Jeffreys et al. 2001).

A wide variety of statistics have been proposed to measure LD. D' and r^2 , each with different statistical properties, are two measures most commonly used to evaluate LD between

biallelic markers (Hill and Robertson 1968; Hill 1981; Valdar et al. 2006; Bohmanova et al. 2010). These parameters can vary between 0 (no disequilibrium) and 1 (complete disequilibrium), but their interpretation is slightly different. For biallelic markers, D' takes the value of 1 if at least one allele at each locus is completely associated with an allele at the other locus, in other words, if one or more of the four possible haplotypes are absent. D' values are less than 1 if all four possible haplotypes are present. The extent of LD based on D' is the most useful for representing historical recombination patterns and is very helpful in understanding long-range LD. One disadvantage of this measure is that it tends to be inflated by small-sized samples and in the presence of rare or low-frequency alleles. The other LD measure (r^2) represents the correlation of alleles at two loci and is more useful for predicting the power of association mapping. For a pair of biallelic loci, r^2 is equal to 1 (known as the perfect LD) if only two haplotypes are present within a population. r^2 is a measure less susceptible to an allele frequency fluctuation than D' , but it is not completely independent of it. r^2 appears to be elevated when the average MAF is either too low or too high (Du et al. 2007; Khatkar et al. 2008).

LD studies have shown that LD in livestock populations is much more extensive than in humans, which can be simply explained by the small effective population and stronger selection that is typical for livestock populations (McRae et al. 2002). A validation work by Khatkar et al. (2008) on Australian Holstein–Friesian cattle suggests that, for the accurate estimation of D' or for any analysis based on the D' matrix (like the construction of LD maps), a sample of 400 or more individuals is required. In contrast, r^2 can be accurately estimated with a smaller sample of 75 individuals. They also reported that LD estimated as r^2 spans over 40 kb and as D' measures over 8.2 Mb. The mean LD among syntenic SNPs measured by r^2 and D' amounted to 0.024 and 0.189, respectively, in the studied Holstein cattle population. Espigolan et al. (2013) investigated LD using 446,986 markers in Nellore cattle, and reported that the average r^2 and D' across the genome were equal to 0.17 and 0.52, respectively. In the study by Bohmanova et al. (2010), $D'=0.72$ and $r^2=0.20$ were observed in North American Holstein cattle between markers distanced by 40–60 kb. Qanbari et al. (2010a) obtained similar results for 810 German Holstein–Friesian cattle genotyped by the Illumina Bovine SNP50K BeadChip. Using a panel of 40,854 SNPs, the authors created a second-generation LD map in this population and presented a mean value of $r^2=0.30\pm 0.32$ in pairwise distances of <25 kb, which dropped to 0.20 ± 0.24 at 50–75 kb. Marques et al. (2008), who analysed 505 SNPs on chromosome 14, estimated LD ($r^2=0.2$) in Holstein cattle using markers separated by less than 100 kb. Similar results were presented by McKay et al. (2007) on the basis of 2,670 SNPs. Using a panel of 54,000 SNPs, Silva

et al. (2010) genotyped 25 Gyr bulls and obtained a mean LD equal to 0.21 (r^2) between adjacent markers.

In the domestic horse, McCue et al. (2012) estimated genome-wide LD within and across different breeds. The authors reported that LD was higher within a breed than across breeds. They also observed that LD declined more rapidly in the Quarter and Mongolian horse than in other studied breeds, with r^2 values dropping below 0.2 within the first 50–100 kb. On the other hand, LD was clearly the highest in the Thoroughbred, where the r^2 value did not drop below 0.2 until 400 kb, and remained higher than in other breeds until approximately 1,200 kb. Similar results were reported by Corbin et al. (2010), who evaluated the extent and distribution of LD in a sample of 817 Thoroughbreds. Using 34,848 autosomal SNP markers, the authors found that the LD was relatively high between closely positioned markers (>0.6 at 5 kb) and extended over long distances, with the average r^2 value maintained above non-syntenic levels for SNPs up to 20 Mb apart.

LD levels between markers have also been studied in the genomes of pig breeds. Du et al. (2007) used 4,500 markers to estimate r^2 in six commercial lines of pigs and observed that, for all pairs of SNPs that are approximately 3 cM apart, the average r^2 was equal to 0.1. Ai et al. (2013) reported that the LD extent across populations is much shorter in Chinese pig breeds than western pigs. With the threshold of $r^2=0.3$, LD extends to 10.5 kb among Chinese pigs and to 125 kb among western breeds. These findings are comparable to a report of Amaral et al. (2008) that was based on the data of 371 SNPs. The authors established that LD extended up to 2 cM in European breeds and up to 0.05 cM in Chinese pigs. Using an SNP panel, Badke et al. (2012) identified the average r^2 between adjacent SNP across all chromosomes for Landrace ($r^2=0.36$), Yorkshire ($r^2=0.39$), Hampshire ($r^2=0.44$) and Duroc ($r^2=0.46$) pigs. The presented values were higher than those reported by Uimari and Tapio (2011), who used the same genotyping platform and obtained average r^2 values of 0.43 and 0.46 for adjacent markers in the Finnish Landrace and Yorkshire populations, respectively.

García-Gómez et al. (2012) presented an analysis of the extent of LD in Spanish Churra sheep using 43,784 SNPs distributed across the autosomal genome. The authors reported that, for SNPs distanced up to 10 kb, the average r^2 was equal to 0.329 and for markers separated by 200–500 kb, the average r^2 was reduced to 0.061. Using the Illumina Ovine SNP50 BeadChip, Miller et al. (2011) examined the extent of genome-wide LD within a population of bighorn sheep (*Ovis canadensis*) and found that high levels of LD persist over 4 Mb. Similar studies were conducted by Usai et al. (2010), who analysed 51,446 SNPs in Sarda rams and showed an average r^2 value of 0.072 for SNPs separated by at least 1,000 kb. These studies showed a substantially lower LD in the sheep when

compared with a wide range of cattle breeds, including dairy and beef cattle (Villa-Angulo et al. 2009).

The differences in the published extent of LD occur because the estimate of LD depends on various factors. Such factors include: the history and structure of an analysed population, a sample size, a marker type (microsatellites or SNPs), a density and distribution of markers, the type of method used for haplotype reconstruction and strictness of SNP filtering (threshold of MAF and Hardy–Weinberg equilibrium). It is important to note that a population characterised by low-range LD will require a higher marker density compared to a population with extensive LD, where fewer markers will be required to obtain the same power to detect association (Meadows et al. 2008).

In summary, LD is an important tool which provides valuable information for selecting SNPs for association and genome selection studies and helps to unravel the recombination history of a population.

Runs of homozygosity

Thanks to the availability of high-density SNP arrays, it is also possible to examine the genome of an animal to identify runs of homozygosity (referred to as ROH). ROHs are contiguous homozygous regions of a DNA sequence where the two haplotypes inherited from parents are identical. This results in a formation of ROHs with different lengths: longer segments represent inbreeding to a recent ancestor and shorter ones are associated with inbreeding from distant generations. To clarify, the length and frequency of ROHs may give information regarding an animal's ancestry and the history of its population (Purfield et al. 2012).

The criteria of ROH identification are, however, still not described precisely, since many authors use different approaches regarding the minimum number of SNPs in ROHs, their length and, also, some of the researchers allow the presence of a small proportion of heterozygote genotypes within ROHs, which may arise as a result of genotyping errors (Ku et al. 2011). From long ROHs, consanguinity may become available to identify. The longer the ROH segments are present in a genome, the higher the chance of recent inbreeding occurring within a pedigree (Kirin et al. 2010). On the other hand, remarkably long ROHs are sometimes present in outbred populations (Gibson et al. 2006). Frequently repeated meiosis and the breaking of chromosomal segments are the reason for long ROHs' decay and creation of short ROHs. Due to the limitations of the pedigree recording process, these short ROHs may not be reflected by the pedigree of an animal (McQuillan et al. 2008).

In human populations, the analysis of ROHs is presented as a tested and valid method of identifying kinship, and may inform about the susceptibility of an individual to recessive

diseases (Gibson et al. 2006; McQuillan et al. 2008; Hildebrandt et al. 2009; Kirin et al. 2010).

ROHs may also be utilised in animal genetics as an estimator of inbreeding levels, which can be used for the assessment of inbreeding depression. In addition, inbreeding estimates obtained conventionally from pedigree data, according to many authors (Ron et al. 1996; Carothers et al. 2006), can be incorrect due to errors and insufficient pedigree depth. These pedigree errors are generated mainly because of an improper recording procedure, mismothering and misidentification of animals. What is more, the results of inbreeding coefficients calculated from pedigree may not reflect the true levels of inbreeding, so the presented approach of ROH utilisation may seem appropriate.

Many authors described high correlations between F_{ROH} (calculated by dividing all of an individual's total length of ROH by the length of the autosomal genome covered by SNPs with the exclusion of centromeres) and inbreeding coefficients. Hamzić (2011) noted that the strongest correlations of F_{ROH} with pedigree inbreeding coefficients were obtained for ROH cut-off lengths of 4 Mb, with a correlation ranging from 0.619 for Norwegian Red up to 0.705 for Tyrol Grey. Purfield et al. 2012 obtained similar results in their study on various cattle populations and presented a strong correlation equal to 0.75. The research of Ferencakovic et al. (2011) corresponded to other authors' results and showed that Austrian Fleckvieh cattle was characterised by a high correlation (0.68) between an inbreeding coefficient calculated from ROHs of lengths greater than 4 Mb and pedigree-based estimates. These results are consistent with the studies conducted on humans.

Various breeds of cattle show different average ROH lengths in their genome. Purfield et al. (2012) showed that the largest mean portion of the genome classified as ROH was identified for Angus and Hereford breeds (198.6 and 198.7 Mb, respectively; approximately 8 % of their genome) and for other breeds, such as Holstein, Holstein–Friesian, Friesian, Limousin and Simmental, it ranged from 80.58 to 93.48 Mb (almost 3.2–3.7 % of their genome). Moreover, the three most homozygous animals had approximately 700 Mb covered by ROHs, which represented nearly a quarter of their genome.

To conclude, the proportion of the genome covered in long ROHs provides a good indication of the inbreeding levels of an animal and may be utilised as a new tool to determine autozygosity that was derived from recent or distant ancestors.

Selection signatures

Animal domestication and modern animal breeding are closely related, with strong artificial selection, which leads to the genetic improvement of animal production traits and fixation

in the population of favourable traits associated with different aspects of animal production (e.g. behaviour, longevity or resistance to disease). Any type of selection (natural or artificial) leads to changes in the frequency of genetic variants associated with a trait under selection. Thanks to the LD across a genome, regions under selection can be detected by the analysis of allele frequency spectra of genome-wide SNPs that reflect the frequency of a selected variant by a physical linkage. The most common approach in the identification of selection signatures is the analysis of differences in allele or haplotype frequencies between populations with different levels of selected traits. In general, most of the computational methods used for the identification of selection signatures are based on comparison of the distribution of allelic frequencies by calculating population genetic statistics that are a function of allelic or genotypic frequencies. For example, F_{ST} (Weir et al. 2005; Wilkinson et al. 2013) and LD (Przeworski 2002; Kim and Nielsen 2004; Ennis 2007) measures have been used. Additionally, specific significance tests for detecting selection signatures have been proposed (Fay and Wu 2000; Kim and Stephan 2002; Voight et al. 2006; Stella et al. 2010) and some of them allow to study selection signatures in single populations (Stella et al. 2010). Other methods, like that proposed by Sabeti et al. (2002) and modified by Qanbari et al. (2010b), the extended haplotype homozygosity test (EHH) identifies loci under selection by an estimation of the age of core haplotypes. It is established by the assessment of decay of core haplotypes association to alleles at various distances from the locus. The method identifies regions with an unusually long range of haplotype and a high frequency in a population (Qanbari et al. 2011).

By using different computational approaches, several studies aiming at the identification of genomic regions under selection in different populations have been performed. Most of them were concerned with cattle as a species most widely subjected to genomic selection, which generates a large amount of data for population genetics. By the analysis of the allele frequency distribution between dairy and beef cattle breeds in Japan, Hosokawa et al. (2012) identified 11 candidate regions associated with different types of production distributed on eight different autosomes. The regions extended over several hundred kb, ranging from 314 kb on BTA13 to 1.8 Mb on BTA26. Within the regions, the authors identified candidate genes, including those previously associated with meat quality and milk yield traits, like *IGF1* or *STAT1*. By using a similar approach, but employing a simulation for significance testing, Hayes et al. (2009b) identified 15 regions of the genome differentially selected in dairy and beef cattle breeds. Most of these regions were located on BTA20 near the locus of *GHR* (growth hormone receptor), a gene with large effects on protein content in milk from dairy cattle (Blott et al. 2003) and on BTA6, in the proximity of the *ABCG2* gene, which harbours a polymorphism affecting milk protein

content (Cohen-Zinder et al. 2005). The analysis of F_{ST} -based genetic diversity in Australian cattle breeds revealed 129 SNPs that have highly divergent F_{ST} values between the studied breeds and Bovine HapMap data (Barendse et al. 2009). The authors identified 12 genomic regions that had additive effects on traits like: residual feed intake, beef yield or intramuscular fatness measured in Australian cattle. The F_{ST} estimate was also used to detect signatures of diversifying selection in 13 porcine breeds. The signatures were found in regions associated with traits related to breed standard criteria, such as coat colour and ear morphology (Wilkinson et al. 2013). By using the parametric composite log likelihood (CLL) of the differences in allelic frequencies between five different cattle breeds selected for milk production, Stella et al. (2010) detected 699 putative selection signatures. The largest CLL was observed on BTA6 and corresponded to the *KIT* gene, which is responsible for the piebald phenotype present in four of the five breeds studied. Moreover, large CLLs were present at the site of the potassium channel-related genes on BTA14, -16 and -25, as well as within integrins (BTA18 and 19) and serine-/arginine-rich splicing factors (BTA20 and 23). By using the EHH, which detects selection by measuring the characteristics of haplotypes within a single population, in Holstein cattle, Qanbari et al. (2010b) identified 12 core haplotypes expected to be under strong positive selection. The haplotypes were associated with a panel of genes, including *EABP3*, *CLPN3*, *SPERT*, *HTR2A5*, *ABCE1*, *BMP4* and *PTGER2*. This panel comprises some interesting candidate genes and QTL, representing a broad range of economically important traits, such as milk yield and composition, as well as reproductive and behavioural traits.

Detection of the regions of the genome which were added to the selection in the breeds' history is also possible by the identification of so-called 'selective sweeps'. This refers to the regions of a genome which show reduction or even elimination of nucleotide variation which arises in the alleles fixation process occurring under strong positive selection. By the analysis of the minor allele frequency of SNPs included in the Bovine SNP50 assay (Illumina) in 14 diverse cattle breeds, Ramey et al. (2013) found 28 genomic regions on 15 different chromosomes, of which 23 were breed-specific and five were shared among two to seven breeds. The regions encompassed several genes which could not be connected with the enrichment of any specific metabolic pathway. Employing a hidden Markov model-based test, which detects selection by studying local variations in the allele frequency spectrum along a genome, within a single population, Boitard and Rocha (2013) revealed, in the Blonde d'Aquitaine breed, three candidate regions under selection on BTA2, -7 and -11. The region on chromosome 2 encompassed *GDF8* gene (myostatin, *MSTN*), a known muscle growth factor inhibitor.

The studies on selection signatures can be an important step in the recognition of biological factors affecting physiology

and production in farm animals. The selected regions may contain or harbour the functional elements responsible for the development of desired traits and, thus, may help to identify the metabolic processes behind selected traits.

Copy number variation

In recent years, much research has been focused on copy number variants (CNVs), which are a type of structural variation of a genome and are considered to be an important source of genetic diversity, constituting approximately 10 % of the human genome (Orozco et al. 2009). They occur when deletions, duplications or insertions of DNA fragments from 1 kbp to 1 Mbp take place (Feuk et al. 2006; Redon et al. 2006). Regions of CNVs may encompass active genes or groups of genes, as well as promoters, enhancers or other functionally important sequences (Henrichsen et al. 2009; Schrider and Hahn 2010). Moreover, CNVs can arise owing to different molecular mechanisms, such as non-allelic homologous recombination (NAHR), non-homologous end joining (NHEJ), replication slippage and retrotransposition. The most common mechanism in humans is NAHR and the least common is retrotransposition (Kidd et al. 2008; Conrad et al. 2010).

When it comes to their presence in a genome, these variations are common in a range of organisms, not only in humans (Sebat et al. 2004; Conrad et al. 2006; McCarroll et al. 2006; Redon et al. 2006) but also in animals, including mice (Graubert et al. 2007; She et al. 2008), chimpanzees (Perry et al. 2006, 2008), rhesus macaques (Lee et al. 2008), cows (Liu et al. 2010), dogs (Chen et al. 2009; Nicholas et al. 2009), chickens (Griffin et al. 2008), fruit flies (Dopman and Hartl 2007; Emerson et al. 2008), *Caenorhabditis elegans* (Maydan et al. 2010), as well as in plants, such as maize (Springer et al. 2009), *Arabidopsis thaliana* (Ossowski et al. 2008) and even fungi, such as *Saccharomyces cerevisiae* (Carreto et al. 2008). They can also vary between individuals within a species (Schrider and Hahn 2010).

As natural diversity, they can arise de novo in an organism (somatic CNV) or be a result of disruptions in the recombination process in germ cells, which makes them heritable. However, the presence of CNVs in a genome may not be neutral for an organism. Numerous research projects have shown that these variations influence phenotypic features, complex bases of behaviour, susceptibility/resistance to diseases (e.g. autism, autoimmune diseases), as well as the occurrence of genetic disorders in humans (Buckland 2003; Gonzalez et al. 2005; Aitman et al. 2006; Autism Genome Project Consortium 2007; Fanciulli et al. 2007; Yang et al. 2007; Schaschl et al. 2009). There are a couple of mechanisms through which CNVs affect genes and their expression patterns. It can be simply via dosage effect, which may concern a

single gene, a set of adjacent genes (e.g. DiGeorge syndrome, Potocki–Lupski syndrome), as well as allele combinations in the case of complex diseases, particularly those of the central nervous system (Henrichsen et al. 2009). Moreover, CNVs can alter sequences regulating gene expression, like enhancers (McCarroll et al. 2006; Nguyen et al. 2006) or promoters. Such extensive genome rearrangements may lead to the exposure of recessive alleles (when a deletion of the dominant gene takes place) or even to the inactivation of some genes (when a deletion within a gene takes place). Therefore, some diseases may result not from changes of copy numbers of a given CNV, but from a structural alteration in a fragment of a genome, causing a disruption of a metabolic pathway, regardless of the gene dosage (Henrichsen et al. 2009).

Copy number variations can be identified with the use of a wide range of techniques, such as FISH (fluorescent in situ hybridisation), CGH (comparative genomic hybridisation), aCGH (array comparative genomic hybridisation), Southern blotting, PFGE (pulsed-field gel electrophoresis), MAPH (multiplex amplifiable probe hybridisation), MLPA (multiplex ligation-dependent probe amplification), PRT (paralogue ratio test) and qPCR (quantitative polymerase chain reaction). What is more, the cutting-edge methods of analysis as well as advanced computational techniques enable CNV identification at a genome-wide scale using high-throughput genome scan technologies like NGS (next-generation sequencing) or genotyping microarrays (SNP microarrays). To infer copy number changes from a microarray analysis, the combination of the two measures of signal intensities may be used: LRR (log R ratio) and BAF (B allele frequency). A significant deviation from the expected distribution of these parameters implies an incorrect number of copies of a given allele (Wang et al. 2007). When it comes to livestock species, some significant advances have also been made lately. First of all, the construction of low CNV resolution maps for cattle, horse, goat, sheep, pig, dog, chicken, duck and turkey gave us an insight into their genomes and showed that these variations are widespread in these species. Moreover, like in humans, CNVs have been associated with different phenotypes and susceptibility to diseases, as well as developmental disorders, e.g. several pigmentation (white coat in horse, pig and sheep) and morphological (late feathering and pea comb in chicken) traits, osteopetrosis, anhidrotic ectodermal dysplasia, copper toxicosis, intersexuality and cone degeneration (reviewed by Clop et al. 2012).

The first small-scale analysis in cattle was carried out on two Hereford and three Holstein individuals by Liu et al. (2008). It allowed for the identification of 25 CNVs present on 16 autosomes, with a size ranging from 28.7 to 396.8 kb and an average size of around 127.8 kb (Liu et al. 2008). The next step could be taken along with the appearance of the Bovine SNP50 BeadChip, which allowed for the detection of bovine CNVs by high-throughput genotyping of different

breeds. The analysis proved that there were differences in the frequency of CNVs between breeds (African, composite and *Bos indicus* breeds had higher frequency than *Bos taurus* breeds) (Matukumalli et al. 2009). The next studies on bovine CNVs were carried out simultaneously in 2010 by Bae et al. (2010) and Fadista et al. (2010). With the use of the Bovine SNP50 BeadChip and custom aCGH, respectively, they constructed two comprehensive CNV maps. However, the obtained size ranges of CNVRs differed from each other as follows: 50–200 kb (Bae et al. 2010) and 1.7 kb–2 Mb (Fadista et al. 2010). Nonetheless, despite the differences in the size range, in both studies, losses were approximately two to three times more frequent than gains. In 2011, Hou et al. performed research on 539 cows belonging to 21 modern breeds, which enabled them to identify 682 candidate CNVRs that covered 139.9 Mb (i.e. nearly 4.60 % of the bovine genome). Among these 682 CNVRs, there were 370 losses, 216 gains and 96 both (loss and gain in the same region). The chromosomes most rich in CNVs were 1 and 6, as well as pericentromeric and subtelomeric regions of chromosomes. Summarising, these results show that around 50 % of bovine CNVRs may be common to different breeds as well as individuals, albeit when taking into account CNVR frequencies, the existing differences are significant, implying that these structural variations could have participated in the process of breed differentiation (Matukumalli et al. 2009; Liu et al. 2010; Seroussi et al. 2010; Hou et al. 2011). Furthermore, bovine CNVRs may encompass about ~300 and 500 genes (Bae et al. 2010; Fadista et al. 2010; Liu et al. 2010), of which at least 19 are engaged in human diseases. Moreover, CNV regions contain about 110 QTL (Fadista et al. 2010). Overall, with regard to these results, copy number variations may have an impact on traits of economic interest.

The first analyses of CNVs on a genome scale in the horse were performed in 2012 by two teams: Doan et al. with the use of a custom-designed whole-exome tiling array, as well as Dupuis et al. (2013) with the use of the Illumina Equine SNP50 beadchip. Doan et al.'s research was carried out on 16 horses of different breeds (e.g. Andalusian, Vanner, Miniature, Quarter Horse, Shire) and a grey donkey (*Equus asinus*). The number of detected CNVs was 2,368, with size range 197 bp–3.5 Mb and mean size 99.4 kb. Among these CNVs, there were 1,509 gains and 859 losses. A total of 438 CNVs were present in single horses (not shared with the others). When it comes to chromosomal distribution, CNVs were detected on each autosome and the X chromosome; however, some chromosomes (12, 17, 23) were enriched with CNVs (15.1 %, 9.1 %, 8.2 %, respectively). Moreover, the copy variations encompassed 1,707 genes, of which 559 exist as CNVs in humans (Doan et al. 2012). Dupuis et al.' team in 2011 performed a genome-wide association study on 234 cases of horses with recurrent laryngeal neuropathy (RLN) and 228 breed-matched controls (Dupuis et al. 2011). Then,

the data were also used to detect copy number variants and their eventual associations with RLN. In sum, 2,797 CNVs were detected for 477 horses, with an average size of 229 kb. Most of the CNVs (86 %) were observed only in four or fewer horses (i.e. <1 %). None of them were significantly associated with recurrent laryngeal neuropathy (Dupuis et al. 2013).

Despite the improvements in the genome analysis methods, the platforms to discover CNV in domestic animals are not sufficiently precise due to their low resolution, which prevents them from detecting small CNVs. Moreover, the results cannot be easily compared because of technical differences between platforms, and these technical issues can lead to false-negative and -positive results (reviewed by Cantsilieris and White 2013), which is why confirmation with alternative methods is usually required. Furthermore, genomics of livestock species encounter more obstacles when CNV platforms and genome assemblies are not available (e.g. camel, dromedary, alpaca, goat) (Clop et al. 2012). If that is the case, cross-species analyses must be carried out, which may have an impact on their sensitivity (Fontanesi et al. 2010, 2011). However, the application of high-throughput sequencing methods may help to solve these issues owing to their lesser bias (than SNP arrays or aCGH), an ability to identify larger numbers of CNVs during a single experiment and applicability to any species (even without a known genome sequence). Unfortunately, these methods are quite demanding when it comes to computational resources, and the results can also be influenced by technical issues (Alkan et al. 2011).

Hitherto, association studies carried out in domestic animals have concerned mainly Mendelian traits. The next very challenging step in animal genomics will be to identify associations between different CNV genotypes and complex phenotypes such as economic traits (e.g. fatness, milk production) or susceptibility to cancer and infectious diseases, which are important from the point of view of veterinarians and animal breeders (Clop et al. 2012).

Genetic differentiation and breed assignment

The idea of the assignment of individuals to their breed of origin has come from population genetic investigations, such as analysing genetic diversity and structure, evaluating the amount of genetic exchange between populations, identifying immigrants and detecting hidden population structures (Negrini et al. 2009). Genetic markers can be used to identify and verify the origin of individuals when genetic heterogeneity amongst populations is sufficient (Wilkinson et al. 2011). The development of assignment methods would make it possible to allocate animals and animal products to their breed of origin; for example, when requested documentation is lost or when external features of animals cannot be evaluated (Wilkinson et al. 2011; Gurgul et al. 2013). Moreover, the

genetic identification can clear up issues such as, inter alia, the contribution of source populations to mixed fisheries, the identification of migrant individuals, structure and levels of diversity amongst populations, and tracking the trade routes of poached animals (Wilkinson et al. 2011).

SNP chips are highly informative but are relatively costly to produce. Moreover, they are computationally expensive to analyse. Hopefully, there is a possibility to reduce the number of markers by screening according to their information content so as to create reduced panels for population genetic analyses. Several statistical methods can be used to determine which genetic markers contain the most information to discriminate among populations (Wilkinson et al. 2011). Wilkinson et al. (2011) compared marker selection methods (delta, Wright's F_{ST} , Weir and Cockerham's F_{ST} and PCA) for selecting population informative SNP loci. The aim of their study was to determine the lowest number of SNP markers from the Bovine SNP50 BeadChip required for the effective and confident assignment of individual genotypes to European cattle breeds. All of the studied SNP selection methods yielded reduced marker panels capable of breed identification, but the power of assignment varied clearly between analysis methods. The pairwise Wright's F_{ST} subtly outperformed other investigated methods in the individual assignment analysis, but delta, pairwise W&C's F_{ST} and PCA did not perform poorly for assignment success rates (Wilkinson et al. 2011). Gurgul et al. (2013) used 120 SNP markers included in the Bovine SNP50 BeadChip genotyping assay (Illumina), which were recommended for parentage testing and pedigree verification in worldwide cattle populations. The results obtained were not completely satisfying and the authors suggested that the studied markers are not the best tool for breed discrimination, especially with the use of reference populations of small size. It was also suggested that markers' informativeness and the power of discrimination between breeds may be higher for SNPs located in genes responsible for animals' physiological properties (Gurgul et al. 2013). Nishimura et al. (2013), using Wright's F_{ST} values, identified highly differentiated SNPs between Japanese Black and Holstein cattle. Twenty SNPs from the top 100 SNPs with high F_{ST} values (F_{ST} values over 0.61) were selected for primer design, followed by the genotyping of F_1 animals. Of the SNPs, 18 (two SNPs were difficult to genotype and were excluded), located more than 30 Mb apart, were selected for breed assignment and allowed for the correct assignment of all examined samples to JB or to F_1 and Holstein. The authors determined the number of SNPs which should be used for the assignment tests by the examination of an assignment error rate for each number of SNPs used for linear discriminant formula (Nishimura et al. 2013). Several statistical approaches have been developed to enable marker selection with the highest discrimination power between different populations. Nevertheless, the results obtained strongly depend on the differentiation of specific populations,

which strongly influences the power of marker discrimination or their informativeness (Gurgul et al. 2013).

To allocate individuals of unknown breeds to their breed of origin, allocation tests are used. Some of them are implemented in freely available software like GeneClass or Structure, which integrate different algorithms for the assignment of individuals to their breeds or the identification of first-generation migrants and enables calculation of the associated probabilities. Negrini et al. (2009) compared the Bayesian (Rannala and Mountain 1997; Pritchard et al. 2000; Baudouin and Lebrun 2000) and frequency-based methods (Paetkau et al. 1995) implemented in GeneClass 2 and Structure 2.2 software for breed assignment. In the reallocation tests, methods implemented in Structure performed better than those in GeneClass. The percentage of correct assignments accounted for 96 % and 85 %, respectively. However, a higher correct assignment rate in allocating animals treated as unknowns to a reference dataset was shown for methods implemented in the GeneClass software. In the authors' opinion, the results obtained showed that SNPs are suitable markers for the assignment of individuals to reference breeds and the software programs Structure 2.2 and GeneClass 2 can be complementary tools to assess breed integrity (Negrini et al. 2009). Wilkinson et al. (2011) suggested that the method of Rannala and Mountain (1997) is more effective for individual assignment than other methods. However, the authors pointed out that, if the levels of genetic differentiation between reference populations are high, the method of Paetkau et al. (1995) is equally effective. Gurgul et al. (2013) applied the Bayesian (Rannala and Mountain 1997) and frequency-based (Paetkau et al. 1995) methods for allocation tests in their study and found dependence in which worse performance of the Bayesian method for some breeds was compensated by relatively better performance of the frequency-based method of Paetkau et al. (Gurgul et al. 2013).

Even though SNP markers are extensively used in scientific and commercial applications, the methods using SNPs for breed recognition and assignment of individuals are not yet sufficiently developed and tested. However, recent research on the use of SNPs for breed assignment showed promising results and suggested that this kind of studies should be continued (Gurgul et al. 2013).

Summary

In this review, we presented a variety of applications of high-throughput genome analysis methods in studies on livestock and the most up-to-date research performed in this area. The article focuses mainly on the application of genotyping microarrays and gives detailed insight into the most interesting and popular applications of data obtained from the available genotyping platforms. We showed that animal genomics is

currently undergoing dynamic development and provides interesting results, which may find a broader application, e.g. as a model for studies in other species, including humans. A new world of possibilities is currently being opened by next-generation sequencing methods, which allow the study of genomes in one base pair resolution. This will provide a stimulus for further evolution of animal genomics and, in conjunction with present knowledge and achievements of transcriptomics, proteomics and biochemistry, will bring us to the understanding of biological mechanisms shaping economically important traits of farm animals.

References

- Ahn JW, Bint S, Bergbaum A, Mann K, Hall RP, Ogilvie CM (2013) Array CGH as a first line diagnostic test in place of karyotyping for postnatal referrals—results from four years' clinical application for over 8,700 patients. *Mol Cytogenet* 6:16
- Ai H, Huang L, Ren J (2013) Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. *PLoS One* 8(2):e56001
- Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Robertson-Lowe C, Marshall AJ, Petretto E, Hodges MD, Bhangal G, Patel SG, Sheehan-Rooney K, Duda M, Cook PR, Evans DJ, Domin J, Flint J, Boyel JJ, Pusey CD, Cook HT (2006) Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* 439:851–855
- Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet* 12:363–376
- Amaral AJ, Megens HJ, Crooijmans RP, Heuven HC, Groenen MA (2008) Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* 179:569–579
- Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299–309
- Autism Genome Project Consortium, Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L, Feuk L, Qian C, Bryson SE, Jones MB, Marshall CR, Scherer SW, Vieland VJ, Bartlett C, Mangin LV, Goedken R, Segre A, Pericak-Vance MA, Cuccaro ML, Gilbert JR, Wright HH, Abramson RK, Betancur C, Bourgeron T, Gillberg C, Leboyer M, Buxbaum JD, Davis KL, Hollander E, Silverman JM, Hallmayer J, Lotspeich L, Sutcliffe JS, Haines JL, Folstein SE, Piven J, Wassink TH, Sheffield V, Geschwind DH, Bucan M, Brown WT, Cantor RM, Constantino JN, Gilliam TC, Herbert M, Lajonchere C, Ledbetter DH, Lese-Martin C, Miller J, Nelson S, Samango-Sprouse CA, Spence S, State M, Tanzi RE, Coon H, Dawson G, Devlin B, Estes A, Flodman P, Klei L, McMahon WM, Minshew N, Munson J, Korvatska E, Rodier PM, Schellenberg GD, Smith M, Spence MA, Stodgell C, Tepper PG, Wijsman EM, Yu CE, Rogé B, Mantoulan C, Wittmeyer K, Poustka A, Felder B, Klauck SM, Schuster C, Poustka F, Bölte S, Feineis-Matthews S, Herbrecht E, Schmötzer G, Tsiantis J, Papanikolaou K, Maestrini E, Bacchelli E, Blasi F, Carone S, Toma C, Van Engeland H, de Jonge M, Kemner C, Koop F, Langemeijer M, Hijmans C, Staal WG, Baird G, Bolton PF, Rutter ML, Weisblatt E, Green J, Aldred C, Wilkinson JA, Pickles A, Le Couteur A, Berney T, McConachie H, Bailey AJ, Francis K, Honeyman G, Hutchinson A, Parr JR, Wallace S, Monaco AP, Barnby G, Kobayashi K, Lamb JA, Sousa I, Sykes N, Cook EH,

- Guter SJ, Leventhal BL, Salt J, Lord C, Corsello C, Hus V, Weeks DE, Volkmar F, Tauber M, Fombonne E, Shih A, Meyer KJ (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39:319–328
- Badke YM, Bates RO, Ernst CW, Schwab C, Steibel JP (2012) Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics* 13:24
- Bae JS, Cheong HS, Kim LH, NamGung S, Park TJ, Chun JY, Kim JY, Pasaje CF, Lee JS, Shin HD (2010) Identification of copy number variations and common deletion polymorphisms in cattle. *BMC Genomics* 11:232
- Barendse W, Harrison BE, Bunch RJ, Thomas MB, Turner LB (2009) Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. *BMC Genomics* 10:178
- Baudouin L, Lebrun P (2000) An operational Bayesian approach for the identification of sexually reproduced cross-fertilized populations using molecular markers. *Acta Horticult* 546:81–93
- Blott S, Kim JJ, Moiso S, Schmidt-Küntzel A, Cornet A, Berzi P, Cambisano N, Ford C, Grisart B, Johnson D, Karim L, Simon P, Snell R, Spelman R, Wong J, Vilkki J, Georges M, Farnir F, Coppeters W (2003) Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163:253–266
- Bohmanova J, Sargolzaei M, Schenkel FS (2010) Characteristics of linkage disequilibrium in North American Holsteins. *BMC Genomics* 11:421
- Boitard S, Rocha D (2013) Detection of signatures of selective sweeps in the Blonde d'Aquitaine cattle breed. *Anim Genet* 44:579–583
- Buckland PR (2003) Polymorphically duplicated genes: their relevance to phenotypic variation in humans. *Ann Med* 35:308–315
- Cantsilieris S, White SJ (2013) Correlating multiallelic copy number polymorphisms with disease susceptibility. *Hum Mutat* 34:1–13
- Cargill EJ, Nissing NJ, Grosz MD (2008) Single nucleotide polymorphisms concordant with the horned/pollled trait in Holsteins. *BMC Res Notes* 1:128
- Carothers AD, Rudan I, Kolcic I, Polasek O, Hayward C, Wright AF, Campbell H, Teague P, Hastie ND, Weber JL (2006) Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. *Ann Hum Genet* 70:666–676
- Carreto L, Eiriz MF, Gomes AC, Pereira PM, Schuller D, Santos MAS (2008) Comparative genomics of wild type yeast strains unveils important genome diversity. *BMC Genomics* 9:524
- Chen WK, Swartz JD, Rush LJ, Alvarez CE (2009) Mapping DNA structural variation in dogs. *Genome Res* 19:500–509
- Chung CH, Levy S, Yarbrough WG (2006) Clinical applications of genomics in head and neck cancer. *Head Neck* 28:360–368
- Clop A, Vidal O, Amills M (2012) Copy number variation in the genomes of domestic animals. *Anim Genet* 43:503–517
- Cohen-Zinder M, Seroussi E, Larkin DM, Looor JJ, Everts-van der Wind A, Drackley JH, Band MR, Hernandez AG, Shani M, Lewin HA, Weller JI, Ron M (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res* 15:936–944
- Conrad DF, Andrews TD, Carter NP, Hurler ME, Pritchard JK (2006) A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 38:75–81
- Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J; Wellcome Trust Case Control Consortium, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurler ME (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464:704–712
- Corbin LJ, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams JA (2010) Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim Genet* 41(Suppl 2):8–15
- Doan R, Cohen N, Harrington J, Veazey K, Juras R, Cothran G, McCue ME, Skow L, Dindot SV (2012) Identification of copy number variants in horses. *Genome Res* 22:899–907
- Dopman EB, Hartl DL (2007) A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 104:19920–19925
- Du FX, Clutter AC, Lohuis MM (2007) Characterizing linkage disequilibrium in pig populations. *Int J Biol Sci* 3:166–178
- Dupuis MC, Zhang Z, Druet T, Denoix JM, Charlier C, Lekeux P, Georges M (2011) Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse. *Mamm Genome* 22:613–620
- Dupuis MC, Zhang Z, Durkin K, Charlier C, Lekeux P, Georges M (2013) Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. *Anim Genet* 44:206–208
- Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M (2008) Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* 320:1629–1631
- Ennis S (2007) Linkage disequilibrium as a tool for detecting signatures of natural selection. *Methods Mol Biol* 376:59–70
- Espigolan R, Baldi F, Boligon AA, Souza FRP, Gordo DGM, Tonussi RL, Cardoso DF, Oliveira HN, Tonhati H, Sargolzaei M, Schenkel FS, Carvalheiro R, Ferro JA, Albuquerque LG (2013) Study of whole genome linkage disequilibrium in Nelore cattle. *BMC Genomics* 14:305
- Fadista J, Thomsen B, Holm LE, Bendixen C (2010) Copy number variation in the bovine genome. *BMC Genomics* 11:284
- Fanciulli M, Norsworthy PJ, Petretto E, Dong R, Harper L, Kamesh L, Heward JM, Gough SC, de Smith A, Blakemore AI, Froguel P, Owen CJ, Pearce SH, Teixeira L, Guillevin L, Graham DS, Pusey CD, Cook HT, Vyse TJ, Aitman TJ (2007) FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 39:721–723
- Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413
- Ferenecakovic M, Hamzic E, Gredler B, Curik I, Sölkner J (2011) Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh cattle. *Agric Conspec Sci* 76:325–329
- Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* 7:85–97
- Fontanesi L, Martelli PL, Beretti F, Riggio V, Dall'olio S, Colombo M, Casadio R, Russo V, Portolano B (2010) An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genomics* 11:639
- Fontanesi L, Beretti F, Martelli PL, Colombo M, Dall'olio S, Occidente M, Portolano B, Casadio R, Matassino D, Russo V (2011) A first comparative map of copy number variations in the sheep genome. *Genomics* 97:158–165
- García-Gámez E, Sahana G, Gutiérrez-Gil B, Arranz JJ (2012) Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genet* 13:43
- Gibson J, Morton NE, Collins A (2006) Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet* 15:789–795
- Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'Connell RJ, Agan BK, Ahuja SS, Bologna R, Sen L, Dolan MJ, Ahuja SK (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307:1434–1440
- Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, Shannon WD, Li X, McLeod HL, Cheverud JM, Ley TJ (2007) A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet* 3:e3

- Griffin DK, Robertson LB, Tempest HG, Vignal A, Fillon V, Crooijmans RP, Groenen MA, Deryusheva S, Gaginskaya E, Carré W, Waddington D, Talbot R, Völker M, Masabanda JS, Burt DW (2008) Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution. *BMC Genomics* 9:168
- Gurgul A, Rubiś D, Żabek T, Żukowski K, Pawlina K, Semik E, Bugno-Poniewierska M (2013) The evaluation of the usefulness of pedigree verification-dedicated SNPs for breed assignment in three Polish cattle populations. *Mol Biol Rep* 40:6803–6809
- Hamzić E (2011) Levels of inbreeding derived from runs of homozygosity: a comparison of Austrian and Norwegian cattle breeds. Master of Science thesis, University of Natural Resources and Life Sciences, Vienna, Austria
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009a) Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 92:433–443
- Hayes BJ, Chamberlain AJ, Maceachern S, Savin K, McPartlan H, MacLeod I, Sethuraman L, Goddard ME (2009b) A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Anim Genet* 40:176–184
- Henrichsen CN, Chaignat E, Reymond A (2009) Copy number variants, diseases and gene expression. *Hum Mol Genet* 18:R1–R8
- Hildebrandt F, Heeringa SF, Rüschemdorf F, Attanasio M, Nürnberg G, Becker C, Seelow D, Huebner N, Chernin G, Vlangos CN, Zhou W, O’Toole JF, Hoskins BE, Wolf MT, Hinkes BG, Chaib H, Ashraf S, Schoeb DS, Ovunc B, Allen SJ, Vega-Warner V, Wise E, Harville HM, Lyons RH, Washburn J, Macdonald J, Nürnberg P, Otto EA (2009) A systematic approach to mapping recessive disease genes in individuals from outbred populations. *PLoS Genet* 5:e1000353
- Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genet Res* 38:209–216
- Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet* 38:226–231
- Hosokawa D, Ishii A, Yamaji K, Sasazaki S, Oyama K, Mannen H (2012) Identification of divergently selected regions between Japanese Black and Holstein cattle using bovine 50k SNP array. *Anim Sci J* 83:7–13
- Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim ES, Matukumalli LK, Ventura M, Song J, VanRaden PM, Sonstegard TS, Van Tassell CP (2011) Genomic characteristics of cattle copy number variations. *BMC Genomics* 12:127
- Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29:217–222
- Khatkar MS, Nicholas FW, Collins AR, Zenger KR, Cavanagh JAL, Barris W, Schnabel RD, Taylor JF, Raadsma HW (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics* 9:187
- Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tüzün E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M, Saranga D, Brand A, Tao W, Gustafson E, McKernan K, Chen L, Malig M, Smith JD, Korn JM, McCarroll SA, Altshuler DA, Peiffer DA, Dorschner M, Stamatoyannopoulos J, Schwartz D, Nickerson DA, Mullikin JC, Wil RK, Bruhn L, Olson MV, Kaul R, Smith DR, Eichler EE (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* 453:56–64
- Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, Tang D, McEwan J, Cockett N, Oddy VH, Nicholas FW, Raadsma H; International Sheep Genomics Consortium (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4:e4668
- Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics* 167:1513–1524
- Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160:765–777
- Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5:e13996
- Kizilkaya K, Tait RG, Garrick DJ, Fernando RL, Reecy JM (2013) Genome-wide association study of infectious bovine keratoconjunctivitis in Angus cattle. *BMC Genet* 14:23
- Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, Talbot R, Pirani A, Brew F, Kaiser P, Hocking PM, Fife M, Salmon N, Fulton J, Strom TM, Haberer G, Weigend S, Preisinger R, Gholami M, Qanbari S, Simianer H, Watson KA, Woolliams JA, Burt DW (2013) Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14:59
- Ku CS, Naidoo N, Teo SM, Pawitan Y (2011) Regions of homozygosity and their impact on complex diseases and traits. *Hum Genet* 129:1–15
- Lee AS, Gutiérrez-Arcelus M, Perry GH, Vallender EJ, Johnson WE, Miller GM, Korbel JO, Lee C (2008) Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. *Hum Mol Genet* 17:1127–1136
- Liu GE, Van Tassell CP, Sonstegard TS, Li RW, Alexander LJ, Keele JW, Matukumalli LK, Smith TP, Gasbarre LC (2008) Detection of germline and somatic copy number variations in cattle. *Dev Biol (Basel)* 132:231–237
- Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell’Aquila ME, Gasbarre LC, Lacalandra G, Li RW, Matukumalli LK, Nonneman D, Regitano LC, Smith TP, Song J, Sonstegard TS, Van Tassell CP, Ventura M, Eichler EE, McDanel TG, Keele JW (2010) Analysis of copy number variations among diverse cattle breeds. *Genome Res* 20:693–703
- Marques E, Schnabel RD, Stothard P, Kolbehdari D, Wang Z, Taylor JF, Moore SS (2008) High density linkage disequilibrium maps of chromosome 14 in Holstein and Angus cattle. *BMC Genet* 9:45
- Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O’Connell J, Moore SS, Smith TP, Sonstegard TS, Van Tassell CP (2009) Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* 4:e5350
- Maydan JS, Lorch A, Edgley ML, Flibotte S, Moerman DG (2010) Copy number variation in the genomes of twelve natural isolates of *Caenorhabditis elegans*. *BMC Genomics* 11:62
- McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM; International HapMap Consortium (2006) Common deletion polymorphisms in the human genome. *Nat Genet* 38:86–92
- McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, Distl O, Guérin G, Hasegawa T, Hill EW, Leeb T, Lindgren G, Penedo MC, Røed KH, Ryder OA, Swinburne JE, Tozaki T, Valberg SJ, Vaudin M, Lindblad-Toh K, Wade CM, Mickelson JR (2012) A high density SNP array for the domestic horse and extant *Perissodactyla*: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet* 8:e1002451
- McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppeters W, Crews D, Dias Neto E, Gill CA, Gao C, Mannen H, Stothard P, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS (2007) Whole genome linkage disequilibrium maps in cattle. *BMC Genet* 8:74
- McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, Macleod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson

- JF (2008) Runs of homozygosity in European populations. *Am J Hum Genet* 83:359–372
- McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J (2002) Linkage disequilibrium in domestic sheep. *Genetics* 160:1113–1122
- Meadows JR, Chan EK, Kijas JW (2008) Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genet* 9:61
- Miller JM, Poissant J, Kijas JW, Coltman DW; International Sheep Genomics Consortium (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Mol Ecol Resour* 11:314–322
- Negrini R, Nicoloso L, Crepaldi P, Milanesi E, Colli L, Chegiani F, Pariset L, Dunner S, Leveziel H, Williams JL, Ajmone Marsan P (2009) Assessing SNP markers for assigning individuals to cattle populations. *Anim Genet* 40:18–26
- Nguyen DQ, Webber C, Ponting CP (2006) Bias of selection on human copy-number variants. *PLoS Genet* 2:e20
- Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM (2009) The genomic architecture of segmental duplications and associated copy number variants in dogs. *Genome Res* 19:491–499
- Nishimura S, Watanabe T, Ogino A, Shimizu K, Morita M, Sugimoto Y, Takasuga A (2013) Application of highly differentiated SNPs between Japanese Black and Holstein to a breed assignment test between Japanese Black and F₁ (Japanese Black × Holstein) and Holstein. *Anim Sci J* 84:1–7
- Orozco LD, Cokus SJ, Ghazalpour A, Ingram-Drake L, Wang S, van Nas A, Che N, Araujo JA, Pellegrini M, Lusis AJ (2009) Copy number variation influences gene expression and metabolic traits in mice. *Hum Mol Genet* 18:4118–4129
- Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res* 18:2024–2033
- Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Mol Ecol* 4:347–354
- Perry GH, Tchinda J, McGrath SD, Zhang J, Picker SR, Cáceres AM, Iafrate AJ, Tyler-Smith C, Scherer SW, Eichler EE, Stone AC, Lee C (2006) Hotspots for copy number variation in chimpanzees and humans. *Proc Natl Acad Sci U S A* 103:8006–8011
- Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurler ME, Tyler-Smith C, Eichler EE, Carter NP, Lee C, Redon R (2008) Copy number variation and evolution in humans and chimpanzees. *Genome Res* 18:1698–1710
- Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 69:1–14
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Przeworski M (2002) The signature of positive selection at randomly chosen loci. *Genetics* 160:1179–1189
- Purfield DC, Berry DP, McParland S, Bradley DG (2012) Runs of homozygosity and population history in cattle. *BMC Genet* 13:70
- Qanbari S, Pimentel EC, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H (2010a) A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim Genet* 41:377–389
- Qanbari S, Pimentel EC, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H (2010b) The pattern of linkage disequilibrium in German Holstein cattle. *Anim Genet* 41:346–356
- Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, Thaller G, Simianer H (2011) Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics* 12:318
- Ramey HR, Decker JE, McKay SD, Rolf MM, Schnabel RD, Taylor JF (2013) Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* 14:382
- Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci U S A* 94:9197–9201
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacós M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurler ME (2006) Global variation in copy number in the human genome. *Nature* 444:444–454
- Ren J, Mao H, Zhang Z, Xiao S, Ding N, Huang L (2011) A 6-bp deletion in the TYRP1 gene causes the brown colouration phenotype in Chinese indigenous pigs. *Heredity (Edinb)* 106:862–868
- Rincon G, Weber KL, Eenennaam AL, Golden BL, Medrano JF (2011) Hot topic: performance of bovine high-density genotyping platforms in Holsteins and Jerseys. *J Dairy Sci* 94:6116–6121
- Ron M, Blanc Y, Band M, Ezra E, Weller JI (1996) Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. *J Dairy Sci* 79:676–681
- Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, Taxis TM, Chapple RH, Ramey HR, Northcutt SL, Bauck S, Woodward B, Dekkers JC, Fernando RL, Schnabel RD, Garrick DJ, Taylor JF (2011) Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genet Sel Evol* 43:40
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837
- Schaschl H, Aitman TJ, Vyse TJ (2009) Copy number variation in the human genome and its implication in autoimmunity. *Clin Exp Immunol* 156:12–16
- Schrider DR, Hahn MW (2010) Gene copy-number polymorphism in nature. *Proc Biol Sci* 277:3213–3221
- Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. *Science* 305:525–528
- Seroussi E, Glick G, Shirak A, Yakobson E, Weller JI, Ezra E, Zeron Y (2010) Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. *BMC Genomics* 11:673
- She X, Cheng Z, Zöllner S, Church DM, Eichler EE (2008) Mouse segmental duplication and copy number variation. *Nat Genet* 40:909–914
- Silva CR, Neves HHR, Queiroz SA, Sena JAD, Pimentel ECG (2010) Extent of linkage disequilibrium in Brazilian Gyr dairy cattle based on genotypes of AI sires for dense SNP markers. In: *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany, 1–6 August 2010*
- Slatkin M (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 9:477–485
- Springer NM, Ying K, Fu Y, Ji T, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddeloh JA, Nettleton D, Schnable PS (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet* 5:e1000734
- Stella A, Ajmone-Marsan P, Lazzari B, Boettcher P (2010) Identification of selection signatures in cattle breeds selected for dairy production. *Genetics* 185:1451–1461
- Twito T, Weigend S, Blum S, Granevitze Z, Feldman MW, Perl-Treves R, Lavi U, Hillel J (2007) Biodiversity of 20 chicken breeds assessed by SNPs located in gene regions. *Cytogenet Genome Res* 117:319–326
- Uimari P, Tapio M (2011) Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *J Anim Sci* 89:609–614

- Usai MG, Sechi T, Salaris S, Cubeddu T, Roggio T, Casu S, Carta A (2010) Analysis of a representative sample of Sarda breed artificial insemination rams with the OvineSNP50K BeadChip. In: Proceedings of the 37th International Committee for Animal Recording (ICAR) Annual Meeting, Riga, Latvia, May/June 2010
- Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JN, Mott R, Flint J (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* 38:879–887
- Villa-Angulo R, Matukumalli LK, Gill CA, Choi J, Van Tassell CP, Grefenstette JJ (2009) High-resolution haplotype block structure in the cattle genome. *BMC Genet* 10:19
- Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biol* 4:e72
- Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17:1665–1674
- Weber KL, Thallman RM, Keele JW, Snelling WM, Bennett GL, Smith TP, McDanel TG, Allan MF, Van Eenennaam AL, Kuehn LA (2012) Accuracy of genomic breeding values in multibreed beef cattle populations derived from deregressed breeding values and phenotypes. *J Anim Sci* 90:4177–4190
- Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG (2005) Measures of human population structure show heterogeneity among genomic regions. *Genome Res* 15:1468–1476
- Wilkinson S, Wiener P, Archibald AL, Law A, Schnabel RD, McKay SD, Taylor JF, Ogden R (2011) Evaluation of approaches for identifying population informative markers from high density SNP Chips. *BMC Genet* 12:45
- Wilkinson S, Lu ZH, Megens HJ, Archibald AL, Haley C, Jackson IJ, Groenen MA, Crooijmans RP, Ogden R, Wiener P (2013) Signatures of diversifying selection in European pig breeds. *PLoS Genet* 9:e1003453
- Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, Jones KN, Shu Y, Kitzmiller K, Blanchong CA, McBride KL, Higgins GC, Rennebohm RM, Rice RR, Hackshaw KV, Roubey RA, Grossman JM, Tsao BP, Birmingham DJ, Rovin BH, Hebert LA, Yu CY (2007) Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80:1037–1054
- Zhang H, Wang Z, Wang S, Li H (2012) Progress of genome wide association study in domestic animals. *J Anim Sci Biotechnol* 3:26
- Zhang L, Liu J, Zhao F, Ren H, Xu L, Lu J, Zhang S, Zhang X, Wei C, Lu G, Zheng Y, Du L (2013) Genome-wide association studies for growth and meat production traits in sheep. *PLoS One* 8:e66569