Rapid Discovery of Mutations Responsible for Sporadic Dominant Genetic Defects in Livestock using Genome Sequence Data: Enhancing the Value of Farm Animals as Model Species

A. Capitan[1,2], P. Michot[1,2], F. Guillaume[1], C. Grohs[1], A. Djari[3], S. Fritz[1,2], S. Barbey[4], P. Otz[5], E. Bourneuf[1,6], D. Esquerré[7,8,9], Y. Gallard[4], C. Klopp[3], D. Boichard[1]

[1] INRA, UMR1313 GABI, 78350 Jouy-en-Josas, France ; [2] UNCEIA, 75012 Paris, France ; [3] INRA, UR875 Biométrie et Intelligence Artificielle, 31320 Castanet-Tolosan, France ; [4] INRA, UE0326, Domaine expérimental du Pin-au-Haras, 61310 Exmes, France, [5] VetAgro Sup, Unité Clinique Rurale, 69210 L'Arbresle, France ; [6] DSV/iRCM/SREIT/LREG, CEA, 78352 Jouy-en-Josas, France ; [7] INRA, UMR1388 Génétique, Physiologie et Systèmes d'Elevage, GeT-PlaGe Genomic Facility, 31326 Castanet-Tolosan, France ; [8] Université de Toulouse INPT ENSAT, UMR1388 Génétique, Physiologie et Systèmes d'Elevage, 31326 Castanet-Tolosan, France ; [9] Université de Toulouse INPT ENVT, UMR1388 Génétique, Physiologie et Systèmes d'Elevage, 31076 Toulouse, France

ABSTRACT: The availability of novel technologies and the development of large data sets of phenotypic records, high density SNP genotyping data and whole-genome sequencing data for genomic selection purpose offers unprecedented opportunity to dissect in record time the genetic architecture of phenotypes in livestock. Here we report the identification of candidate mutations for three dominant genetic defects (bulldog calf syndrome and glass-eyed albino in Holstein and a novel neurocristopathy in Montbéliarde dairy cattle) by sequencing as few as one or two cases and then using whole genome sequencing data from 526 animals from 16 different breeds as controls. The rapid identification of mutations underlying sporadic syndromes which also occur in humans, increases the attractiveness of cattle to become models for such diseases in the post genomic era.

Keywords: whole genome sequencing; sporadic dominant genetic defect; mosaicism; MITF; CHD7; COL2A1.

Introduction

Long before Darwin enunciated the Theory of Evolution, and in multiple societies, humans have taken advantage of their apparent similarities with animals to gain insights into their own anatomy and physiology using them as models. Nowadays laboratory animal models represent an essential resource in all the fields of biology and especially in biomedicine. Apart from ethical or religious considerations, laboratory animal models such as rodents, fruit flies or zebra fish, are often preferable for experimental disease research because of their unlimited supply and ease of manipulation [Simmons, 2008]. Indeed, they enable scientists to control a number of experimental parameters (from environmental factors to genetic background) and provide a sufficient number of subjects to statistically evaluate the results of the experiment. In Genetic research, the development of novel mutagenesis technologies and high throughput genotyping combined with the high reproductive rate of laboratory animals has enabled the characterization of tens of thousands of genes/mutations (mostly with recessive inheritance and affecting coding sequences) in the last 30 years whereas causative mutations for only 556 loci have been identified in domestic animals in the same period of time (http://omia.angis.org.au/home/ accessed 03/10/2014). As it was the case for genetic research in laboratory animals we believe that the availability of novel technologies and the development of large data sets of phenotypic records, high density SNP genotyping data and whole-genome sequencing (WGS) data for genomic selection purpose will revolutionize genetic research in domestic animals. As an example, the use of high density genotyping data for detecting regions displaying a deficit in homozygotes in dairy cattle, as proposed by VanRaden et al. (2011), enables the mapping of deleterious recessive mutations without any phenotypic information. In the same vein, several ongoing projects in bovine aim at discovering deleterious mutations in WGS data and identifying at risk mating in artificial insemination (AI) databases to then phenotype homozygous animal and annotate the function of the affected genes. Both approaches clearly mark a paradigm shift in term of research strategy (from genotype to phenotype versus from phenotype to genotype). Such data sets, combined with the small genetic diversity of bovine breeds, also enable the imputation of the genotypes of hundred thousands of animals for millions of polymorphisms and the successive identification of causative mutations for QTL without a priori on their molecular locations and consequences (Daetwyler et al., 2014). Identifying mutations responsible for spontaneous dominant syndromes using WGS data as a primary source of information represents another challenge and a promising field of research in cattle in a context where the existence of surveillance centers centralizing samples of affected animals combined with the large number of offspring per bull and the end of progeny tests offer the unique opportunity to describe novel dominant syndromes.

Toward that goal we collected samples from affected animals for five dominant conditions: variant red (VR), Anury (AN), Glass-eyed albino (GEA) and bulldog calf syndrome (BD) in Holstein as well as a novel neurocristopathy (NC) in the Montbéliarde breed. VR is distinct from the traditional recessive red allele of the MC1R gene found in Holstein. It emerged in 1980 with the birth of a red heifer, Surinam Sheik Rosabel-Red, from homozygous black parents. Since then VR has been shown to segregate in Rosabel-Red's progeny in an autosomal dominant manner.

AN refers to the complete absence of caudal vertebrae. AN and Brachyury (partial absence) without

other additional symptoms have been sporadically observed in distinct Holstein pedigrees suggesting a dominant inheritance.

A GEA phenotype similar to that first described by Leipold and Huston (1966) was observed in a pedigree tracing back to a mutant heifer born in 1994. Typical clinical feature includes white coat color, "glass-eyes" characterized by a blue pigmentation of the inner part of the iris and complete depigmentation of the outer part, and, deafness. Some affected individuals may display partial phenotypic reversion consisting in black spots on the tip of the ears and/or heterochromia irides (Fig. 1). BD, a lethal form of chondrodysplasia has been reported in years 1999 and 2000 in a small proportion (1%) of calves of a very widely used French AI sire Igale (Fig. 2). Recessive determinism of the disease was excluded for two reasons: 1) they were no runs of homozygozity in 240 affected calves in the interval to which the defect had been previously mapped (Eggen et al., unpublished data), and 2) thirty two of the dams that produced an affected calf, when mated again with Igale produced 32 unaffected calves, which is extremely unlikely with recessive inheritance ($P<10^{-4}$). The hypothesis that the disease was the result of a dominant mutation, but the sire was a mosaic for this mutation was pursued.

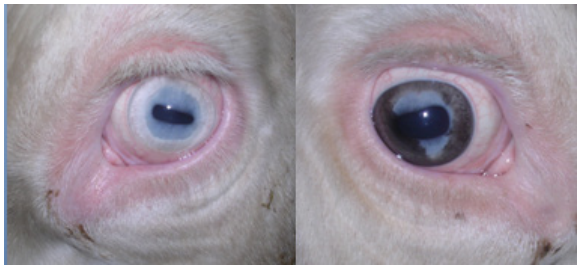

Figure 1. Eye phenotype of GEA cows. Note the typical glass-eyed appearance due to lack of pigmentation of the outer margin of the iris (left) and the partially phenotypic reversion (right) causing heterochromia irides.

Figure 2. BD calf sired by Igale

Neurocristopathy (NC) was observed in approximately half of the progeny of a genomic bull that was a posteriori found to be mildly affected. Phenotypic examination revealed variable expressivity of the syndrome with severely affected individuals displaying cleft lip and/or palate, heart defects, behavioral anomalies resulting from cranial nerves defects, and growth delay whereas mildly affected animals only demonstrated a moderate lack of balance and coordination in the days following their birth and were hardly detectable thereafter (Fig. 3).

Here we report the identification of candidate mutations for three out of these five defects by sequencing as few as one or two cases and then using whole genome sequencing data from 526 animals from 16 different breeds as controls. In addition, we briefly discuss to what extent the study of these defects in cattle can improve our knowledge on the regulation and function of the corresponding genes.

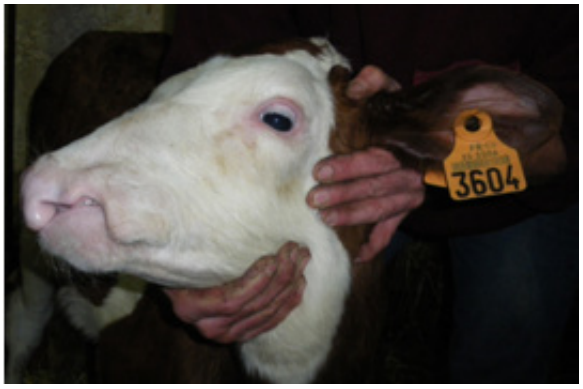

Figure 3. Severe facial phenotype of a NC affected calf

Materials and Methods

Animals. DNA from one VR, one AN, one GEA one NC and two BD affected animals were used for whole genome sequencing purpose. Additional DNA samples were used for validation purpose (see below).

Whole Genome Sequencing, Read Mapping, Variant Calling and Filtering. One paired-end library with a 250-bp insert size was generated per animal using the Illumina TruSeq DNA Sample Prep Kit (for VR, BD and GEA) or (for AN, AASC and NC) the NEXTflex PCR-Free DNA Sequencing Kit (Biooscientific) according to the manufacturer's instruction. Libraries were quantified using the QPCRLibrary Quantification Kit (Agilent) or KAPA Library Quantification Kit (Cliniscience), controlled on a High Sensitivity DNA Chip (Agilent), and sequenced on one or two HiSeq 2000 lanes (Illumina), each with Illumina TruSeq V3 Kit (200 cycles). The 101-bp reads were mapped on the UMD3.1 bovine sequence assembly using BWA (Li and Durbin, 2009). Reads with multiple alignments were removed. PCR duplicates were filtered and variants were called using SAMtools rmdup and pileup options (Li *et al.*, 2009). Only heterozygous variants with a quality score (QUAL) of >=30, a mapping quality (MQ) score of >=30 were conserved. SNP and Indels were then annotated using Ensembl VEP (McLaren et al., 2010). Only frameshift mutations, in-frame insertions or deletions and stop gain or loss variants as well as polymorphisms affecting splice donor or splice acceptor sites and deleterious missense polymorphisms were kept. These were finally filtered for polymorphisms found in whole genome sequencing data from 526 healthy animals consisting of 170 Holstein, 87 Simmental, 54 Angus, 43 Brown Swiss, 27 Jersey, 26 Montbéliarde, 25 Limousin, 22 Normande, 17 Finnish Ayrshire, 16 SwedishRed, 10 Belgian Blue, 9

Table 1. Results of the Whole Genome Sequencing and filtering approach

Bovine syndromes	Coverage	Private heterozygous deleterious mutations	Candidate mutation	Human syndromes (MIM #)
VR	13.4 X	6	-	-
AN	15.3 X	8	-	-
GEA	30.1 X	30	MITF p.R211Del	Tietz syndrome (103500)
BD	9.0 and 12.0 X	23 and 17 (only one in common)	COL2A1 p.G960R	Achondrogenesis type II (200610)
NC	13.4 X	23	CHD7 p.K594AfsX29	CHARGE syndrome (214800)

Guelph Composite, 8 Beef Booster, 8 Charolais, 2 Piedmontese, 1 Hereford and 1 Holstein/Simmental Cross. Most of these genomes (429) are part of the 1000 bull genome project (Daetwyler et al., 2014). Finally, custom scripts were used to further characterize the remaining variants and to compare the affected genes with those listed in the Online Mendelian Inheritance in Man (OMIM; www.omim.org) and Mouse Genome Informatics (MGI; www.informatics.jax.org) databases.

Mapping of the GEA and NC defects. In an attempt to confirm the results of the WGS approach and to map these defects, six GEA cows and five unaffected maternal half sisters were genotyped with the Illumina Bovine 50k Beadchip and phased with their unaffected AI sire and tenth of unaffected paternal half sibs according to the French genomic selection procedure (Boichard et al. 2012). Maternal haplotypes were then investigated for IBD segments shared by the cases. Nine NC subjects were also genotyped with the same chip and phased with their affected sire and 47 paternal half sib of unknown phenotype which had been previously genotyped for genomic selection purpose. Similarly paternal haplotypes were screened for genomic segments shared by all the cases and found at a low frequency in the second group.

Genotyping of the candidate causative mutation for BD and confirmation of mosaicism. Ten additional BD calves, their dams and their sire Igale were genotyped by PCR-RFLP. A 341-bp fragment spanning the g.32475732G>A candidate mutation on BTA5 (corresponding to COL2A1 p.G960R) was PCR-amplified with the BD-F (TTCCTGTGTTTTGCTTCTCCT) and BD-R (GGGTCAGGTCAGTGCTTCTC) primers using the Go-Taq Flexi DNA Polymerase (Promega) according to the manufacturer's instructions. The resulting amplicon was subsequently digested with BpmI endonuclease (New England Biolabs) according to the manufacturer's instructions. Whereas digestion of the wild type allele is predicted to produce 166- and 175-bp-fragments, the mutant allele, affecting a BmpI restriction site, is predicted to not be digested. After digestion, products were separated on a 2% agarose gel by electrophoresis and stained with ethidium bromide for visualization. To confirm mosaicism, a second PCR-RFLP analysis was performed. Products were run on a 2% NuSieveTM GTGTM agarose (FMC) gel electrophoresis and IGALE's unrestricted fragment was recovered from agarose plug using Wizard® SV Gel and PCR Clean-up System protocol (Promega). A nested PCR was performed on the purified fragment with BD-F2 (CTGGACCCCCTGGTCCTC) and BD-R primers using the Go-Taq Flexi DNA Polymerase (Promega) according to the manufacturer's instructions. Finally the resulting amplicon was sequenced by Qiagen (Hilden, Germany) using conventional Sanger sequencing.

Results and Discussion

After filtering for variants found in unaffected animals, the numbers of private deleterious SNP and Indels per affected animal were extremely low, ranging from 6 to 30 (Table 1). These results indicate that, thanks to the relatively small effective size of the breeds studied, 170 Holstein genomes, and 26 Montbéliarde + 87 Simmental genomes are enough to capture most of their genetic pool. Interestingly, among Holstein animals, BD#1, BD#2 and GEA which have a three to five fold higher amount of private mutations than VR and AN animals that were born approximately two generations (10 years) before them. As the French Friesian population has undergone progressive introgression of New World Holstein since the 1970s, this could be explained by a slightly higher amount of Friesian genome remaining in these animals.

The small number of private mutations allows the visual examination of each of them, looking for functional candidates. In this process we were not able to find any candidate causative mutation for VR and AN. Of note, no deleterious mutation was found on BTA27, where VR locus is located according to Dreger and Schmutz (2010). The causative mutations for both conditions are most probably caused by SNP and small Indels located in non-coding regions or by structural variations, two hypotheses that have not been investigated here. Considering the reasonable average sequencing depth obtained (13.4 X and 15.3 X) it is less likely that small deleterious mutations have been missed locally in coding regions due to poor sequence coverage. On contrary, we identified one strong functional candidate mutations, supported by a minimum of three different reads, for each of the four other defects.

Bulldog calf syndrome. Remarkably the unique functional candidate mutation for BD was also the only mutation shared by both sequenced calves. This mutation (g.32475742G>A on BTA5; Fig. 4) is predicted to cause the substitution of a glycine residue, which is conserved across vertebrates (Fig 5), by an arginin in alpha-1 chain of type II collagen (COL2A1 p.G960R). COL2A1 mutations have been reported to cause a wide spectrum of skeletal disorders in human including achondrogenesis type II (ACG2; MIM: 200610; Godfrey et al., 1988; Vissing et al.,

1989; Mortler et al., 1995; Bonaventure et al., 1995; Körkkö et al., 2000), which shares strong similarities with the clinical manifestations reported by Agerholm et al. (2004) in Igale's progeny. This substitution, like most of the mutations responsible for ACG2, disrupts the invariant G-x-y structural motif necessary for perfect triple helix formation and, thus, could lead to extensive overmodification, intracellular retention, and reduced secretion of type II collagen, as previously established in the human form of the disease (Vissing et al., 1989). Genotyping by PCR-RFLP of this mutation in 10 additional affected calves showed a perfect association between this mutation and the syndrome, and suggested mosaicism in the Igale germline, given the small proportion of affected calves and the fact that the calves were heterozygous at this position. Finally, Sanger sequencing of a conventional PCR, and of a nested PCR performed after PCR-RFLP, definitively confirmed mosaicism for Igale at this locus (not shown) and causality of the mutation.

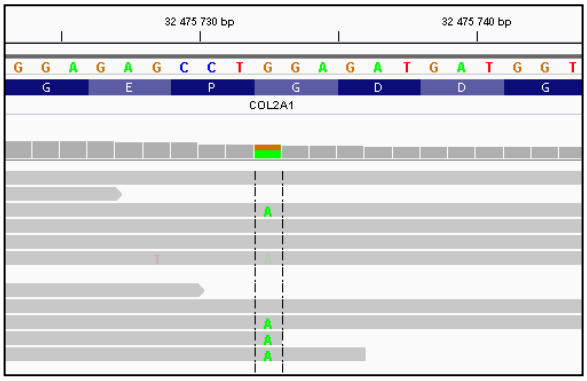

Figure 4. IGV snapshot showing the BTA5 g.32475732G>A mutation of a BD calf

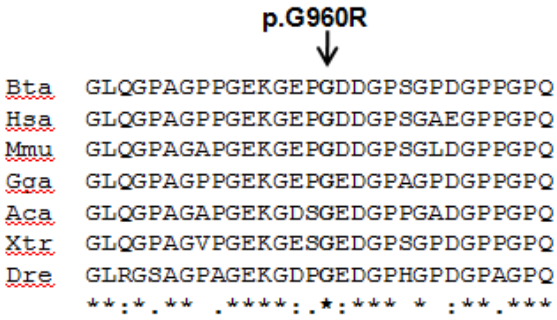

Figure 5. Multispecies alignment of the COL2A1 protein sequence around the p.G960R substitution. Cattle (Bta), human (Hsa), mouse (Mmu), chicken (Gga), anolis (Aca), xenopus (Xtr) and zebrafish (Dre) COL2A1 sequences were retrieved from Ensembl and alignement using CLUSTALW. Note the succession of G-x-y amino acids triplets, typical of triple helical domain of collagen proteins, and the perfect across species conservation of the first G residues.

Glass-eyed albino phenotype. The candidate causative mutation for GEA (BTA22 g. 31746498-31746500del) results in the deletion of an arginin in the basic domain of a functional candidate gene: the microphthalmia-associated transcription factor (MITF p.R211Del; Fig 6). MITF is a basic helix-loop-helix leucine zipper transcription factor which acts as a master regulator of melanocyte development, function and survival (Levy et al., 2006). Of note MITF p.R211Del was also the only private deleterious mutation mapping to a unique 2-Mb segment shared by the maternal genomes of GEA cows. In addition, in a German Fleckvieh cattle pedigree, Philipp et al. (2011) have recently reported the association of GEA phenotype with a p.R210I mutation in the same domain of the same gene. In humans, loss of MITF heterozygozity causes two distinct autosomal dominant syndromes depending on the type and location of the mutation. Non truncating mutations of MITF basic domain have a dominant negative effect (with mutant MITF proteins interfering with the DNA binding domain of wild-type proteins as reported by Takebayashi et al., 1996) resulting most of the time in albinoid-like hypopigmentation of the skin and hair and severe hearing loss (i.e. Tietz syndrome; MIM: 103500) whereas other types of mutations cause patchy depigmentation and uni- or bilateral deafness (i.e. Waardenburg syndrome type 2A; MIM: 193510). Remarkably, the amino-acid sequences of human, bovine and mouse basic domains of MITF are identical and the p.R211Del mutation we identified phenocopies mouse $Mitf^{Mi}$ and human p.R217Del alleles. The latter, is the most frequent MITF mutation observed in human and often occurs de novo due to the existence of a triplet repeat. Interestingly, in a recent study, Léger et al. (2012) reported that p.R217Del not always result in regular Tietz syndrome but rather to a large range of phenotypes with some patient showing Waardenburg-like patches. They also reported that sun-exposed freckles were observed more frequently in Asian populations suggesting possible interaction with modifier loci and highlighted the impossibility to use the mouse as model due the difference of transmission between mouse and human. Indeed $Mitf^{Mi}/Mitf^{+}$ mouse only display slight dilution of the fur in the young which returns to normal in the adult and some, but not all, heterozygotes have a small spot on the head or spots on the belly or tail (Deol, 1970). On contrary the existence of both regular Tiezt phenotype and Tiezt phenotype with heterochromia irides and small black spots in cattle makes GEA a perfect model to investigate the molecular basis of the phenotypic reversion and map modifier loci. Toward that goal we collected semen from a young GEA bull and sampled skin from small black spots and surrounding white skin in partially reverting animals for expression study and histological purpose.

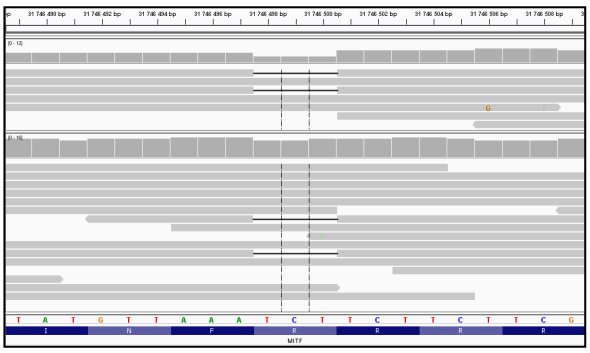

Figure 6. IGV snapshot showing the BTA22 g. 31746498- 31746500del mutation of a GEA cow

Novel neurocristopathy in Montbéliarde. The candidate causative mutation for NC (g. 28085726-28085730del) is a deletion of 5-bp in a AAAAAG/AAAAG repeat motif (Fig. 7) which causes a frame shift, resulting in a truncated CHD7 protein (CHD7 p.K594AfsX29) with a total absence of functional domains. This mutation is also the only private deleterious mutation mapping to a unique 12.2-Mb segment shared by the paternal haplotypes of the nine NC calves. Of note the corresponding haplotype was observed in only 7 out of the 47 animals with unknown phenotype and predicted to be unaffected or mildly affected since they have been genotyped for genomic selection purpose. CHD7 (the chromodomain-helicase-DNA-binding protein 7) is an ATP-dependent chromatin remodeler homologous essential for activation of core components of neural crest transcriptional circuitry, including Sox9, Twist and Slug (Bajpai et al.. 2010). In humans, Loss-of-function mutations in CHD7 cause a variable combination of multiple congenital malformations refered as C.H.A.R.G.E syndrome (MIM: 214800) for Coloboma, Heart disease, Atresia of choanae, Retardation of growth and/or development, Genital Hypoplasia, and Ear abnormalities with or without deafness (Hall, 1979; Hittner et al., 1979 and Pagon et al., 1981). The clinical feature also commonly includes cleft lip and/or palate and nervous system anomalies such as semicircular canal agenesis, cranial nerves defect, arhinencephaly, holoprosencephaly or missing olfactory bulbs (http://omim.org/entry/608892 and http://omim.org/entry/214800). In addition, various other rare symptoms have been described. Patient showing 100% of the major symptoms are rare and several cases of sib pairs or parent/child pairs with the same mutation but different clinical manifestations have been reported (Jongmans et al. 2006; Vuorela et al., 2008; Hughes et al., 2014) suggesting the existence of modifier loci. The large number of the NC bull progeny (1057 calves in total with approximately 50% of cases) represents a unique data set to describe the complete range of clinical feature associated with a unique CHD7 mutation. Of note, clinical examination of 10 NC calves out of the 100 we plan to study have been performed to date and the major feature of CHARGE syndrome expect Genital Hypoplasia and Ear abnormalities have been observed so far. Moreover the existence of two distinct groups of NC animals (severely affected which usually die within their first weeks of life and nearly asymptomatic animals) is very promising for mapping genes partially rescuing CHD7 deficiency.

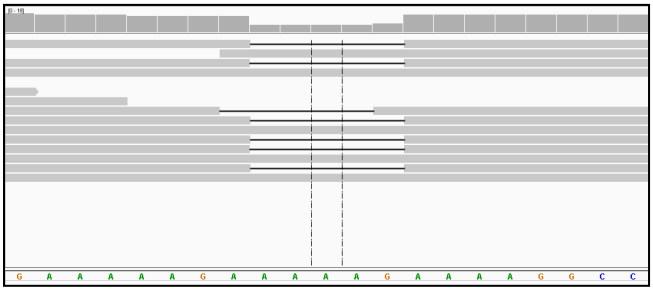

Figure 7. IGV snapshot showing the BTA14 g. 28085726- 28085730del mutation of a NC calf

Conclusion

In conclusion we demonstrate that large Whole Genome Sequencing database combined with the typical structure of livestock populations enable the identification of dominant deleterious mutations by sequencing as few as one or two cases and then using this database as control. The rapid identification of mutations underlying sporadic syndromes which also occur in humans, increases the attractiveness of cattle to become models for such diseases in the post genomic era.

Acknowledgment

The authors are grateful to the members of the 1000 bull genome consortium for the excellent collaboration. We also would like to acknowledge the G. Fayolle (UMOTEST), T. Simon and A. Devillard (EVOLUTION), L. Izard (MIDATEST), B. Giraud (Les Eleveurs du Pays Vert), C. Richard (INRA), G. Belbis (ENVA), P. Costiou (ONIRIS), C. Danchin (IDELE), P. Chabrol, A. Eggen, the French breeders, veterinarians and AI technicians for gently providing samples and phenotypes.

Literature Cited

Agerholm J.S., Arnbjerg J., Andersen O. (2004) J Vet Diagn Invest., 16, 293-298.

Bajpai R., Chen D.A., Rada-Iglesias A., et al. (2010) Nature 463, 958–962.

Boichard D., Guillaume F., Baur A., et al. (2012) Anim Prod Sci, 52, 115–120.

Bonaventure J., Cohen-Solal L., Ritvaniemi P., et al. (1995) Biochem J., 307, 823-830.

Daetwyler H.D., Capitan A., Pausch H. et al. (2014) Nature Genetics, in press.

Deol M.S. (1970) Proc R Soc Lond B Biol Sci. 175, 201-217.

Dreger D.L., Schmutz S.M. (2010) Anim Genet. 41, 109-112.

Godfrey M., Keene D.R., Blank E. et al. (1988) Am J Hum Genet., 43, 894-903.

Hall B.D. (1979). J. Pediat. 95, 395-398.

Hittner H.M., Hirsch N.J., Kreh G.M. et al. (1979). J. Pediat. Ophthal. Strabismus 16, 122-128.

Hughes S.S., Welsh H.I., Safina N.P., et al. (2014) Am J Med Genet A., 164A, 48-53.

Jongmans M.C.J., Admiraal R.J., van der Donk K.P., et al. (2006) J. Med. Genet. 43, 306-314.

Körkkö J., Cohn D.H., Ala-Kokko L., et al. (2000) Am J Med Genet., 92, 95-100.

Léger S., Balguerie X., Goldenberg A., et al. (2012) Eur J Hum Genet., 20, 584-587.

Leipold H.W., Huston K. (1966) J Hered., 57, 179-82.

Levy C., Khaled M., Fisher D.E. (2006). Trends Mol Med., 12, 406-414.

Li H. and Durbin R. (2009) Bioinformatics, 25, 1754–1760.

McLaren W., Pritchard B., Rios D. et al. (2010) Bioinformatics, 26, 2069-2070.

Mortier G.R., Wilkin D.J., Wilcox W.R., et al. (1995) Hum Mol Genet., 4, 285-288.

Pagon R.A., Graham J.M., Zonana J. et al. (1981) J. Pediatr. 99, 223–227.

Philipp U., Lupp B., Mömke S., et al. (2011) PLoS One, 6, e28857.
Simmons D. (2008) Nature Education 1, 70.
Takebayashi K., Chida K., Tsukamoto I., et al. (1996) Mol. Cell. Biol., 16, 1203-1211.
VanRaden P.M., Olson K.M., Null D.J., et al. (2011) J Dairy Sci 94, 6153–6161.
Vissing H., D'Alessio M., Lee B. et al. (1989) J Biol Chem., 264, 18265-18267.
Vuorela P.E., Penttinen M.T., Hietala M.H., et al. (2008) Clin Dysmorphol., 17, 249-253.