Rate of de novo mutation in dairy cattle and potential impact of reproductive technologies

C. Harland[1,4], K. Durkin[4], M. Artesi[4], L. Karim[2], N. Cambisano[2], M. Deckers[2], N. Tamma[2],

E. Mullaart[3], W. Coppieters[2], C. Charlier[4] & M. Georges[4]

[1]Livestock Improvement Corporation, New Zealand
[2]GIGA Genomics platform, University of Liege, Belgium
[3]CRV, Arnhem, The Netherlands
[4]Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège, Belgium

Summary

To study the process of de novo mutation (dnm) in the bovine germ-line, we have sequenced the whole genomes of 743 dairy cattle constituting 131 sire-dam-offspring trios with an average of five grand-offspring each. A first study using five families revealed the common occurrence of somatic and germ-line mosaicism for dnms, pointing towards mutation-prone early cleavage cell divisions. In this extended study, of 131 trios we have identified 7,498 dnms with an overall transition-transversion rate of 1.96, of which 3,413 are mosaic in either the proband or parents' germ-line, confirming the results of our previous study. We detect a significant environmental effect resulting from the use of reproductive technologies, such as in vitro fertilisation, on the rate of dnm in the early embryo.
Keywords: de novo variants, whole genome sequence, cattle, DNA, reproductive technologies

Introduction

De novo mutation is a fundamental biological process, responsible for generating the majority of genetic diversity, that is selected against by natural and artificial selection. We can define germ-line dnms as genetic variants absent in the gametes that formed an individual but present in the gametes it passes onto the next generation. With the advent of next generation sequencing it has become possible to directly detect dnm by the whole genome sequencing of trios consisting of both parents and an offspring. The dnms are then identified as variants that are present in the offspring but absent in the parents. Considerable work has occurred in humans (Conrad et al. 2011; Kong et al. 2012; Palamara et al. 2015; Wong et al. 2016) to characterise the rates and properties of dnm. This has determined that the average human inherits between 70-80 dnms from their parents, with ~80% being inherited from the father. Recently we have applied this approach to dairy cattle showing that the average individual receives 50-60 dnms from its' parents with ~70% from the sire (Harland et al. 2017). A major finding of this work was that in cattle a large proportion of the dnms appear to occur very early in embryo development. A consequence of this is that these early dnms will be mosaic in the soma and germ-line of the individual and will be present in less than 50% of its' DNA (typically detectable between 1-35%), and thus will be inherited by an equivalent proportion of the individuals' offspring (Harland et al. 2017). This differs markedly from dnms that occur late in the development of the germ-line of an individual. Which are consequently restricted to the germ-line and will typically only be inherited by a single offspring.

In dairy cattle embryo based reproductive technologies such as multiple ovulation and embryo transfer (MOET) and in vitro maturation and fertilisation (IVM/IVF) are widely used to increase the number of offspring from important dams. The most commonly used technology is MOET, which involves hormonal induced super ovulation of the dam followed by artificial insemination. The resulting fertilised embryos are then flushed from the cow and collected, before transplantation into surrogate cows. This can typically generate 3-4 successful pregnancies (Galli et al. 2003). The second common method is IVM/IVF, for which immature oocytes are collected. The oocytes then undergo maturation and fertilisation in vitro before freezing for storage or transfer to a surrogate dam (Galli et al. 2003; Humblot et al. 2010). Repeated use of IVM/IVF on a cow can generate up to 70 calves per year (Humblot et al. 2010). These two techniques allow a substantial increase in the number of offspring deriving from a single cow, providing opportunities to accelerate genetic gain. In cattle, approximately 600,000 MOET and 340,000 IVM/IVF embryo transfers occurred in 2010 (Hansen 2014). Due to the associated costs and the resulting focus on high production and genetic worth animals, both techniques are likely to have an oversized contribution to the next generation of elite sires and dams. The widespread use of such technologies and the likely impact they have on such a critical stage as early embryo development, makes them a possible source of environmental effects on events occurring early in embryo development such as dnm. Thus, we have sought to consider this hypothesis as part of a whole genome sequencing experiment focused on dnm in Dutch dairy cattle.

Materials and Method

Blood or sperm samples were collected for 743 Dutch Holstein Friesian cattle (*bos taurus*), which form 131 three or four generation pedigrees each consisting of at least sire, dam, proband and on an average five grand-offspring, with grand-parents where possible. For four pedigrees, all four grandparents were sequenced, 13 pedigrees have three grandparents, and 23 pedigrees have two grandparents sequenced. DNA was extracted and Illumina NextSeq

550bp whole genome libraries were constructed and sequenced on the Illumina HiSeq 2500 in 100bp paired end mode by the University of Liege, GIGA-Genomic core service. The data was prepared following the GATK best practises protocol (version 3) (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013) after alignment to the BosTau6 reference genome by BWA MEM (Li 2013). The GATK Haplotype caller (v3.4) was run following the GVCF N+1 protocol and was utilised to identify variants. DenovoPedFilter (https://github.com/aeonsim/denovoPedFilter) was utilised to identify dnms within the 131 families, determine germ-line of origin via the use of linkage in the grand-offspring and assign the events to a mutational class (Proband Mosaic, Sire Non-Mosaic, Dam Non-Mosaic, Sire Mosaic or Dam Mosaic).

Results

Using a suite of public and custom-made programs (cfr. M&M), complemented by visual inspection of candidate dnm in IGV, we identified a total of 7,498 dnm in the 131 probands. Following Harland et al. (2017) we classified the 7,498 dnms in to 562 proband-mosaic (PM) mutations (having occurred early in the embryo development of the proband), 2,795 sire-non-mosaic (SNM) (having occurred late in the germ-line of the sire), 1,522 sire-mosaic (SM) (having occurred early in the embryo-development of the sire), 1,090 dam-non-mosaic (DNM) (having occurred late in the germ-line of the dam), 1,329 dam-mosaic (DM) (having occurred early in the germ-line of the dam).

Of the 7,498 dnms, there were 13 in splice sites, 35 in UTR, 70 were exonic, 644 within 5kb of a gene, 2,028 in introns and the remaining were intergenic (Variant effect predictor (McLaren et al. 2016), Ensembl gene annotation, build 90). Six thousand and 487 of the dnms were single nucleotide substitutions, 254 tandem mutations, and 757 INDELs.

The numbers of observed dnms of the five classes were adjusted using trio-specific estimates of genome coverage and sensitivity (raw data and correction factors in Suppl. Table 1). From these, we estimated that, in cattle, (i) sperm cells of a 6 year old bull carry on average 41.1 dnms of which 35.3% are detectably mosaic in DNA of the sire, (ii) oocytes carry on average 22.4 dnms of which 54.9% are detectably mosaic in DNA of the dam, (iii) 7 dnms (9.9%) detected in the DNA of male proband, occurred during his development, (iv) 4.1 dnms (5.8%) detected in DNA of a female proband occurred during her development (Table 1).

The numbers of SNM+SM and DNM+DM mutations were used to estimate the sex-averaged bovine dnm rate at 1.21×10^{-8} per base pair per generation (95% CI: $1.15\text{-}1.3 \times 10^{-8}$) (i.e. very similar to human and chimpanzee), with a male-to-female ratio of 1.8:1 (i.e. considerably lower than the 4:1 or 5:1 for human and chimpanzee). Note that cattle breed at a younger age and for a limited time-span, which may mitigate the age-effects observed in primates. It is noteworthy that the male to female ratio for the late occurring SNM/DNM is 2.56 ($p < 1 \times 10^{-5}$), while the SM/DM ratio is 1.15 ($p = 1.7 \times 10^{-4}$).

Table 1: Number of dnms and supporting statistics for the five classes of dnm with PM split by proband gender. Values have been adjusted for estimated coverage of the genome and sensitivity (Suppl. Table 1). Average: mean number of dnms after correction per proband, Range: range of observed dnms in the 131 pedigrees after correction. CV: coefficient of variation for each class of dnm. >1 gamete: the number of individuals with multiple gametes in the population. Repeatability: estimated repeatability of the number of dnms observed in individuals with multiple gametes. p-value REML: significance of the individual animal effect in the REML analysis.

Class	Average	Range	CV	>1 gamete	Repeatability	p-value REML
PM-M	7.0	0-91	2.2	NA	NA	NA
PM-F	4.1	0-24	1.1	NA	NA	NA
SNM	26.8	8-90	0.4	35	0.15	0.2
SM	14.4	0-41	0.6	35	0.55	$1x10^{-6}$
DNM	10.1	0-48	0.6	33	0	1
DM	12.4	0-69	1.0	33	0.77	$6x10^{-13}$

We first exploited the fact that dnms could be studied in multiple gametes for 39 sires and 48 dams, to estimate the "repeatability" (i.e. is there any evidence for an "animal" effect, whether due to genetics and/or permanent environmental effects?) in the number of dnms of the different classes using a restricted maximum likelihood (REML) model. The repeatability sets the upper limit for the heritability. There was no evidence for a significant sire effect on the number of SNM or dam effect on the number of DNM (Table 1). This was confirmed by the absence of a significant correlation between the number of SNM and DNM mutations between multiple gametes of the same parent (Fig. 1 A and B). There was evidence for a significant individual effect on the number of SM and DM mutations both from the REML and correlation analyses (Table 1 and Fig. 1C and D). It should be noted, however, that a large proportion of SM and DM mutation are shared by sibs, which could unduly inflate these estimates. Work is in progress to control this effect.

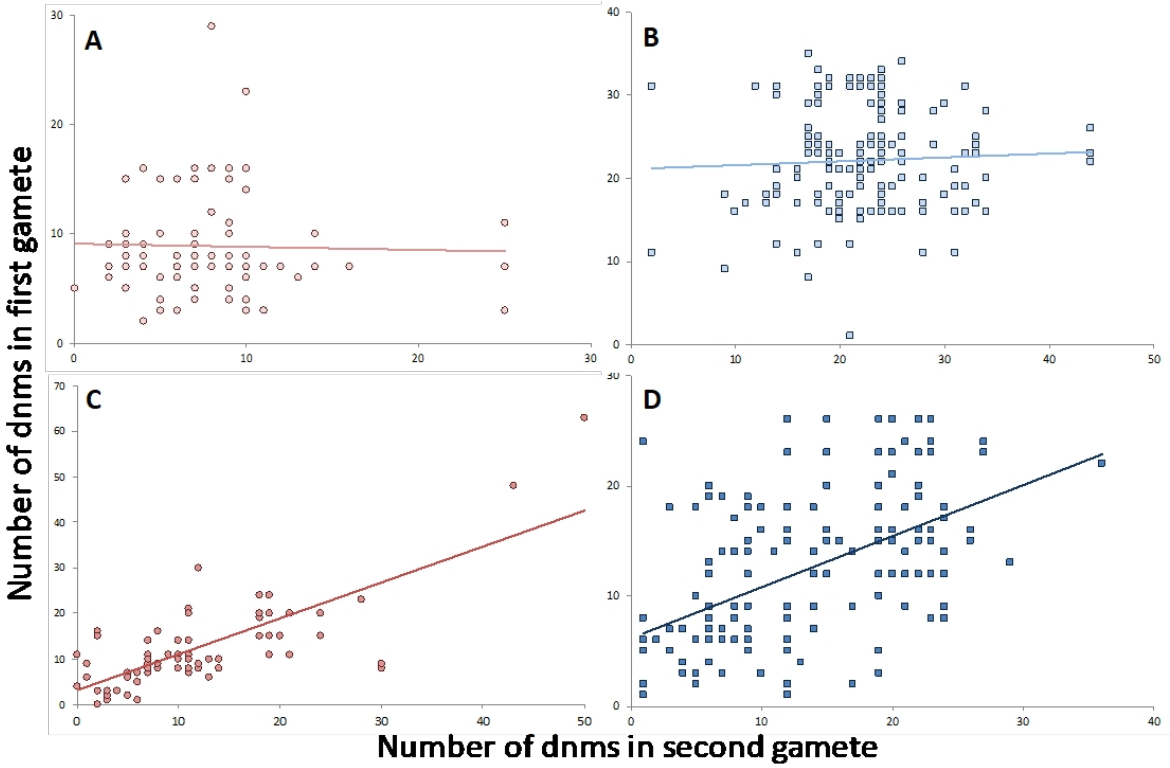

Figure 1: Number of dnms (uncorrected) and line of best fit per dnm class for sires and dams. Numbers of dnms for two gametes from the same parent are shown on the y- and x-axis,

respectively, for (A) DNM, (B) SNM, (C) DM, (D) SM. Red indicates oocytes, blue sperm cells.

Elite dairy cattle are routinely generated by relying on reproductive technologies of increasing sophistication, including (i) artificial insemination (AI), (ii) multiple ovulation and embryo transfer (MOET), and (iii) oocyte pickup, in vitro maturation, in vitro fertilisation (IVM/IVF). Accordingly, 27% of Damona probands were generated by AI, 37% by MOET and 36% by IVF. We tested whether the reproductive technology used to produce an individual (in this case the proband) had an effect on the "early" mutation rate reflected in the number of PM mutations. We observed a significant effect ($p = 1.9x10^{-4}$) for reproductive technologies on the number of PM mutations per proband, with a mean of 1.9 PM dnms for AI, 2.7 for MOET (not significantly different, $p = 0.195$), and 4.5 for IVF (highly significant, $p = 7.35x10^{-5}$) (Fig. 2).

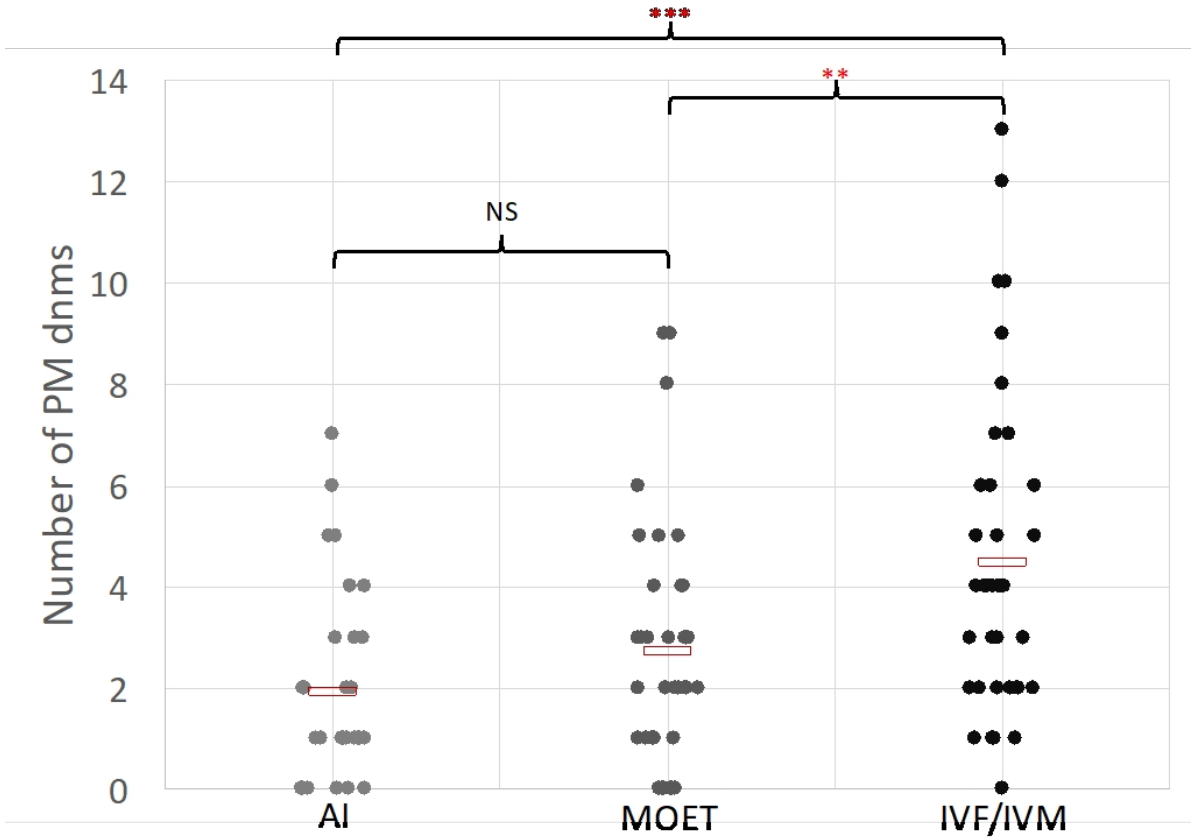

*Figure 2: Distribution of PM dnms by reproductive technology (AI, artificial insemination; MOET, multiple ovulation and embryo transfer; IVF/IVM, in vitro fertilisation and maturation). Red bar indicates the mean for each technology, NS not significant, ** $p < 0.01$, *** $p < 0.001$.*

Discussion

We have utilised 131 pedigrees of dairy cattle to estimate that the dnm rate in Holstein Frisian dairy cattle is $1.2x10^{-8}$ per bp per generation. Thus, any new animal will inherit approximately 41 dnms from its' sire (assuming a 6-year old bull) and 22 from its' dam. Of these dnms at least 25 will have occurred early enough during the development of the parents

that they may be shared by the individual's siblings.

One significant finding from this dataset is an association between the use of reproductive technologies such as MOET and IVM/IVF and an increase in the number of early mosaic dnms. As we observe a ~2.6x increase in the number of proband mosaic dnms for cattle resulting from embryos generated using IVM/IVF compared to AI. While this may be considered beneficial in its' ability to slightly increase the gain of diversity within the population, it can also accelerate the spread of deleterious dnms into the population.

As mosaic dnms are present at detectable levels (>1%) in parental DNA, they will likely be present in a similar percentage of the gametes produced by the individual and thus inherited by similar proportion of offspring. If a dominant mosaic dnm was present at 5% in the parental DNA it is likely to have little or no effect in that parent, as it is present in only 10% of the cells in the body. However, ~5% of the offspring of the individual will inherit the mutation. They will fully express the associated phenotype, as it will be present in all cells in their body. This is potentially problematic under animal breeding models that are reliant on genomic selection and the use of young sires. Traditionally, AI sires would not have been utilised widely until their daughter proof was complete, allowing for the detection of potential deleterious dominant phenotypes before they reached the general population. However, under the current breeding schemes where young (2 or 3 year old) bulls are widely utilised (10,000s of inseminations), a mosaic dnm effecting 5% of the offspring could result in at least 500 affected individuals. In the case of recessive deleterious dnms the risk comes from the accelerated spread of the dnms through the population while decreasing the chance of their loss to genetic drift.

Considering these potential risks in combination with the decreasing costs of next generation sequencing and with the widespread use of young genomically selected AI sires, it may be worth considering the development sequence based screening for dnms in AI sires. A well implemented screening program could substantially reduce the risk of both dominant and recessive deleterious dnms being introduced into the population. In addition, such a screening program could identify dnms in genes of interest, that would provide additional insights into bovine biology.

List of References

Conrad DF, Keebler JEM, DePristo MA, Lindsay SJ, Zhang Y, Cassals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, et al. 2011. Variation in genome-wide mutation rates within and between human families. *Nat Genet* **43**: 712–714.

DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas M., Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**: 491–498.

Galli C, Duchi R, Crotti G, Turini P, Ponderato N, Colleoni S, Lagutina I, Lazzari G. 2003. Bovine embryo technologies. *Theriogenology* **59**: 599–616.

Hansen PJ. 2014. Current and Future Assisted Reproductive Technologies for Mammalian Farm Animals. In *Current and Future Reproductive Technologies and World Food Production*, *Advances in Experimental Medicine and Biology*, pp. 1–22, Springer, New York, NY https://link.springer.com/chapter/10.1007/978-1-4614-8887-3_1

Harland C, Charlier C, Karim L, Cambisano N, Deckers M, Mullaart E, Coppieters W, Georges M. 2017. Frequency of mosaicism points towards mutation-prone early cleavage cell divisions. *bioRxiv* 079863.

Humblot P, Le Bourhis D, Fritz S, Colleau JJ, Gonzalez C, Guyader Joly C, Malafosse A, Heyman Y, Amigues Y, Tissier M, et al. 2010. Reproductive Technologies and Genomic Selection in Cattle. *Vet Med Int*.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. 2012. Rate of de novo mutations and the importance of father/'s age to disease risk. *Nature* **488**: 471–475.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio*.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–1303.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol* **17**: 122.

Palamara PF, Francioli LC, Wilton PR, Genovese G, Gusev A, Finucane HK, Sankararaman S, Sunyaev SR, de Bakker PIW, Wakeley J, et al. 2015. Leveraging Distant Relatedness to Quantify Human Mutation and Gene-Conversion Rates. *Am J Hum Genet* **97**: 775–789.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al* **11**: 11.10.1-11.10.33.

Wong WSW, Solomon BD, Bodian DL, Kothiyal P, Eley G, Huddleston KC, Baker R, Thach DC, Iyer RK, Vockley JG, et al. 2016. New observations on maternal age effect on germline de novo mutations. *Nat Commun* **7**.