# Genome scan for regions with transmission ratio distortion in cattle

*S. Id-Lahoucine[1], J. Casellas[2], P. A. S. Fonseca[1,3], F. Miglior[1,4], M. Sargolzaei[1,5], L. Brito[1], S. Miller[1,6], J. Chesnais[5], M. Lohuis[5], F. S. Schenkel[1], J. F. Medrano[7] & A. Cánovas[1]*

*[1]University of Guelph, Department of Animal Biosciences, Centre for Genetic Improvement of Livestock, N1G 2W1 Guelph, Ontario, Canada.*
*sidlahou@uoguelph.ca (Corresponding Author)*
*[2]Universitat Autònoma de Barcelona, Departament de Ciència Animal i dels Aliments, 08193 Barcelona, Spain.*
*[3]Universidade Federal de Minas Gerais, Departamento de Biologia Geral, 31270-901 Belo Horizonte, Brazil.*
*[4]Canadian Dairy Network, N1K 1E5 Guelph, Canada.*
*[5]Semex Alliance, N1G 3Z2 Guelph, Canada.*
*[6]Angus Genetics Inc., 64506 St. Joseph, MO, USA.*
*[7]University of California-Davis, Department of Animal Science, 95616 Davis, CA, USA.*

## Summary

Observed deviations from the expected Mendelian inheritance of alleles from heterozygous parents has been reported in a broad range of organisms and it is well known as transmission ratio distortion (TRD). Various biological mechanisms affecting gametes, embryos, foetuses, or even postnatal offspring can produce patterns of TRD. Bayesian models were used to evaluate TRD across the whole genome of Holstein dairy cattle by using either a SNP-by-SNP approach or haplotypes of 2 SNP sliding windows in 79,238 genotyped trios (parents-offspring). From 44,369 autosomal SNP, 942 were significantly detected with TRD that exceeded Bayes factor (BF) ≥ 10 (strong evidence) and 408 with BF ≥ 100 (decisive evidence). The number of SNP with parent-unspecific and parent-specific TRD was 270 and 672 SNP, respectively. After correction by the approximate empirical null distribution of TRD, the number of significant SNP reduced to 669, 462 and 73 with a probability of error of 1%, 0.01% and <0.001%, respectively, which mostly coincided with the genomic regions with high BF. In SNP-by-SNP analyses, the regions with moderate-to-high |TRD| (≥ 0.15) were less polymorphic, providing evidence of TRD selection and exhibiting interesting rare variants. In contrast, more polymorphic regions with moderate-to-high |TRD| were observed by the haplotype analyses. The number of regions detected by haplotype analyses after correction (at <0.001% probability of error) were 775, 348 and 562 for overall, sire- and dam-TRD, respectively. The preliminary functional analysis of detected regions with TRD identified positional genes associated with regulation of embryonic development. In conclusion, the prevalence of TRD was extended across the whole genome and a deeper study of these candidate regions and markers will be required to understand this phenomenon in cattle.

*Keywords: transmission ratio distortion, heterozygous parents, Bayesian analysis, haplotypes, cattle*

## Introduction

Transmission ratio distortion (TRD) is the deviations from the expected Mendelian

inheritance of alleles from heterozygous parents and it has been reported in a broad range of organisms (e.g., Wakasugi, 1974; Hall & Willis, 2005; Casellas *et al*., 2012, 2017). Various biological mechanisms affecting gametes, embryos, fetuses (Huang *et al*., 2013), or even postnatal offspring (Moore, 2006) can produce patterns of TRD, being a consequence of a sort of genetic factors affecting reproduction. Moreover, its link with impaired fertility (Silver, 1989) and differential offspring survival at different stages (Dean *et al*., 2006), emphasises the importance of TRD. Nevertheless, current knowledge about its prevalence and possible causes in livestock species are still scarce. Our research focuses on the study of TRD in the bovine genome, with the aim to identify genomic regions that could contain potential causal mutations, directly affecting reproductive traits. Recently, Casellas *et al*. (2014) developed Bayesian models for the analyses of TRD on biallelic markers and Id-Lahoucine *et al.* (in preparation) derived the Bayesian models of TRD for multiallelic locus, providing a useful and powerful methodology applicable to livestock populations. The objectives of this research were to evaluate the overall (parent-unspecific) and parent-specific TRD across the whole genome in Holstein using SNP-by-SNP and 2-SNP haplotypes TRD methods.

## Material and methods

### Transmission ratio distortion models

The models developed by Casellas *et al.* (2014) parametrized the probability of inheritance ($p$) for a given allele from a heterozygous parent ($A/B$) by including the effect of overall TRD ($\alpha$) or sire- ($\alpha_s$) and dam-specific TRD ($\alpha_d$):

$p(A) = 1 - p(B) = 0.5 + \alpha$ and $p(B) = 1 - p(A) = 0.5 - \alpha$,
(1)
$p_i(A) = 1 - p_i(B) = 0.5 + \alpha_i$ and $p_i(B) = 1 - p_i(A) = 0.5 - \alpha_i$ $i = [s \text{ U } d]$
(2)

where $\alpha$ was the TRD parameter with values defined within the parametric space between -0.5 and 0.5. Under a Bayesian implementation, conditional posterior probabilities of TRD parameters were defined as: $\propto$

$p(\alpha|\boldsymbol{y}) \propto p(\boldsymbol{y}|\alpha)p(\alpha)$
(3)
$p(\alpha_s,\alpha_d|\boldsymbol{y}) \propto p(\boldsymbol{y}|\alpha_s,\alpha_d)p(\alpha_s)p(\alpha_d)$,
(4)

where $\boldsymbol{y}$ was the column vector of genotypes of the offspring generation.

For haplotype analyses, Id-Lahoucine *et al.* (in preparation) assumed the same parametrization but estimating different $\alpha_{jk}$ for each combination of two alleles ($j$ and $k$). The implementation of models in the multiallelic locus approach were: $\propto$

$p(\alpha_{12},\alpha_{13},\ldots\ \alpha_{(n-1)n}|\boldsymbol{y}) \propto p(\boldsymbol{y}|(\alpha_{12},\alpha_{13},..\ \alpha_{(n-1)n})p(\alpha_{12})p(\alpha_{13})\ldots\ p(\alpha_{(n-1)n})$ (5)
$p(a_{s12},\alpha_{s13},\ldots\ \alpha_{s(n-1)n}|\boldsymbol{y}) \propto p(\boldsymbol{y}|(\alpha_{s12},\alpha_{d12},\alpha_{s13},\alpha_{d13},..\ \alpha_{s(n-1)n},\alpha_{d(n-1)n})p(\alpha_{s12})p(\alpha_{d12})p(\alpha_{s13})$
$p(\alpha_{d13})\ldots\ p(\alpha_{s(n-1)n})p(\alpha_{d(n-1)n})$,
(6)

where $\alpha_{jk}$ was the overall TRD and $\alpha_{sjk}$ and $\alpha_{djk}$ were sire- and dam-specific TRD parameters for the specific combination of haplotypes *j* and *k*, and *n* was the number of alleles for the specific locus. Flat priors between -0.5 and 0.5 were assumed for TRD parameters. A multinomial distribution was used to determine the likelihood of the data. Inference on $\alpha$, $\alpha_s$ and $\alpha_d$ were made on the marginal posterior distributions by Markov Chain sampling with Metropolis-Hastings algorithm (Hastings, 1970). The statistical relevance of each TRD parameter was tested by a Bayes factor (BF) approach. A unique Monte Carlo Markov chain of 10,000 iterations was run for each analysis and the first 1,000 iterations were discarded as burn-in, which is enough to perform the analyses in the purpose of optimizing the computational time as suggested by Id-Lahoucine *et al.* (in preparation).

**Genotype data and analyses**

All animals that were previously genotyped in Canada and in the United States, as of August 2017, with SNP panels that include 45,187 SNP used in Canada for genomic evaluation purpose, were included in the analyses. A total of 90,400 genotyped animals, including 2,707 sires, 19,089 dams and 79,238 of parents-offspring (without selection of offspring within trios) were used. The genotypes were imputed and phased using FImpute (Sargolzaei *et al*., 2014). The overall TRD and sire- and dam-TRD were evaluated across the 44,369 autosomal SNP. Analyses were performed either by a SNP-by-SNP approach or by a 2-SNP haplotype sliding windows approach.

## Results and discussion

Results across the whole genome showed 942 SNP displaying TRD that exceeded the BF threshold for strong evidence (BF ≥ 10) according to Jeffreys' (1984) scale (Figure 1). Between them, 270 SNP were parent-unspecific TRD and 672 SNP were parent-specific TRD, where 393 and 271 were sire- and dam-TRD, respectively. In addition, 8 markers were detected with sire- and dam-TRD in opposite direction of preference of transmission. According to the approximate empirical null distribution of TRD proposed by Id-Lahoucine *et al.* (in preparation) the number of detected regions reduced to 669, 462 and 73 with a probability of error of 1%, 0.01% and <0.001%, respectively. On the other hand, the number of markers with decisive evidence (BF ≥ 100) was 408, while 154 and 73 SNP were detected with a BF of $10^5$ and $10^{10}$, respectively. In addition, the most of TRD regions that were discarded as random TRD (385 SNP) presented low BF < 100, basically given the low number of informative offspring where the chance to generate TRD randomly is more probable and the obtained BF is usually lees high.

For SNP with moderate-to-high |TRD| (≥ 0.15), a minor allele frequency (MAF) of 0.001, 0.027 and 0.002 were observed for overall, sire- and dam-TRD, respectively. This low frequency is a possible evidence of TRD selection, a phenomenon that can result in the spread of the allele through the population and eventually its fixation (Hurst & Werren, 2001). However, the number of individuals with these minor alleles were still substantial given the size of the dataset and, at the same time, represent interesting rare variants that can be associated with complex disorders. Furthermore, small |TRD| are no less relevant, as ready reported by Id-Lahoucine *et al.* (in preparation), where the maximum |TRD| that can be generated by chance in large population is very small (e.g., ~0.02 in ≥ 40,000 informative offspring). On the other hand, the results from haplotypes analyses showed more interesting regions, where moderate-to-high |TRD| was identified in more polymorphic regions (MAF ~

0.045). Notice that from a population genetics perspective, the maintenance of the disadvantaged allele in the population can be caused by some countervailing force, such as recombination, mutation, genetic drift, and an immunogenetic advantage for survival in later adulthood, regardless of low fertility as suggested by several authors (Labbe *et al*., 2013). Moreover, the alleles with advantage transmission are not always beneficial, but can present deleterious fitness effects and are maintained in balanced polymorphisms (Carvalho & Vaz, 1999), such as in the mouse t-haplotype system and the segregation distorter system in Drosophila. The number of regions detected by haplotype analyses of overall TRD were 2,494, 1,467, 680 and 358 with a BF $\geq 10$, $\geq 100$, $\geq 10^5$ and $\geq 10^{10}$, respectively. The number region with BF $\geq 10$ reduced after correction to 2,441, 1,363 and 775 with a probability of error of 1%, 0.01% and <0.001%, respectively. The number of regions detected was 1,241 and 1,761 for sire- and dam-TRD, respectively; 348 and 562 after correction considering a threshold of <0.001%. The preliminary functional analyses of the detected regions suggested positional candidate genes related with the regulation of embryonic development.

## Conclusions

In conclusion, the prevalence of TRD was extended across the whole genome in cattle. A deeper study of these regions is required to understand better this phenomenon. The preliminary functional analyses of the detected regions suggested positional candidate genes related with the regulation of embryonic development.

## Acknowledgements

This research was funded by Beef Farmers Ontario (BFO-16-04), MITACS accelerate and Semex Alliance. The authors acknowledge the data provided by Canadian Dairy Network.

## List of References

Carvalho, A. B. & S. C. Vaz, 1999. Are Drosophila SR drive chromosomes always balanced? Heredity. 83:221–228.

Casellas, J., J. J. Cañas-Álvarez, A. González-Rodríguez, A. Puig-Oliveras, M. Fina, J. Piedrafita, A. Molina, C. Díaz, J. A. Baró & L. Varona, 2017. Bayesian analysis of parent-specific transmission ratio distortion in seven Spanish beef cattle breeds. Anim. Genet. 48:93-96.

Casellas, J., R. J. Gularte, C. R. Farber, L. Varona, M. Mehrabian, E. E. Schadt, A. J. Lusis, A. D. Attie, B. S. Yandell & J. F. Medrano, 2012. Genome Scans for Transmission Ratio Distortion Regions in Mice. Genetics. 191:247–259.

Casellas, J., A. Manunza, A. Mercader, R. Quintanilla, & M. Amills, 2014. A flexible Bayesian model for testing for transmission ratio distortion. Genetics. 198:1357–1367.

Dean, N.L., J.C. Loredo-Osti, T.M. Fujiwara, K. Morgan, S.L. Tan, A.K. Naumova & A. Ao, 2006. Transmission ratio distortion in the myotonic dystrophy locus in human preimplantation embryos. Eur. J. Hum. Genet. 14:299–306.

Hall, M. C. & J. H. Willis, 2005. Transmission Ratio Distortion in Intraspecific Hybrids of Mimulus guttatus: Implications for Genomic Divergence. Genetics 170:375–386.

Hastings, W.K, 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika. 57:97-109.

Huang, L. O., A. Labbe, & C. Infante-Rivard, 2013. Transmission ratio distortion: a review of

concept and implications for genetic association studies. Human Genetics. 132:245–263.
Hurst, G. D. D., & J. H. Werren, 2001. The role of selfish genetic elements in eukaryotic evolution. Nat. Rev. Genet. 2:597–606.
S. Id-Lahoucine, A. Cánovas, C. Jaton, P. A. S. Fonseca, F. Miglior, S. Miller, M. Sargolzaei, J. P. Chesnais, F. S. Schenkel, J. F. Medrano & J. Casellas, (in preparation). Bayesian analysis of transmission ratio distortion.
Jeffreys, H, 1984. Theory of Probability. Clarendon Press, Oxford.
Labbe, A., L. O. Huang, & C. Infante-Rivard, 2013. Transmission ratio distortion: a neglected phenomenon with many consequences in genetic analysis and population genetics. In: Epigenetics and Complex Traits, A.K. Naumova and C.M.T. Greenwood (editors), Springer, New York. p. 265-285.
Moore, C.S., 2006. Postnatal lethality and cardiac anomalies in the Ts65Dn Down Syndrome mouse model. Mamm Genome. 17:1005-1012.
Sargolzaei, M., J. P. Chesnais & F. S. Schenkel, 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 15:478.
Silver L.M, 1989. Gene dosage effects on transmission ratio distortion and fertility in mice that carry t haplotypes. Genetical Research. 54:221–5.
Wakasugi, N, 1974. A genetically determined incompatibility system between spermatozoa and eggs leading to embryonic death in mice. J. Reprod. Fertil. 41:85–96.

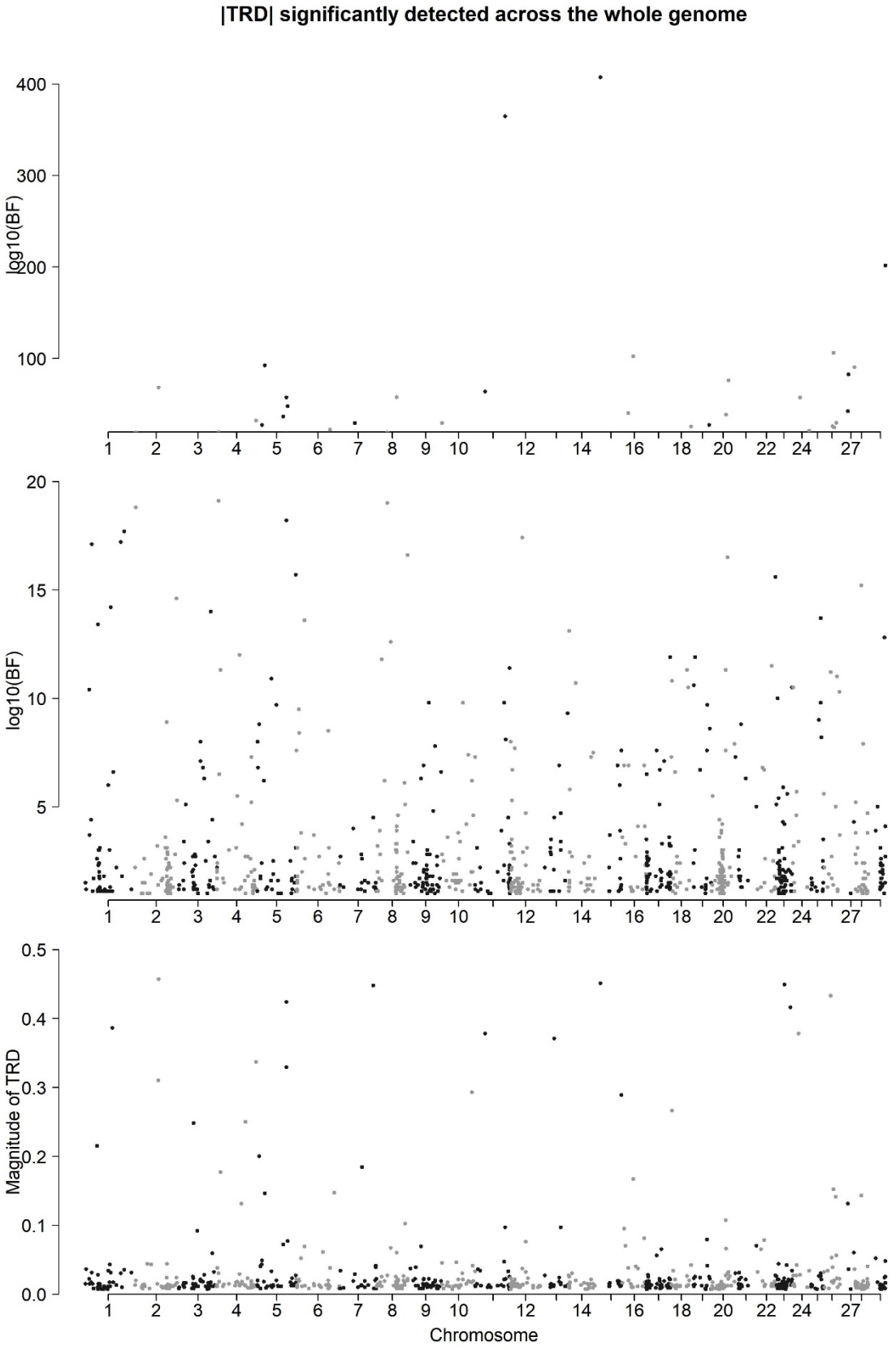


*Figure 1. Byes factor (BF) and magnitude of overall, sire- or dam-TRD of SNP significantly detected across the whole genome of Holstein cattle by SNP-by-SNP analyses.*