Sequence-based association analyses on X chromosome in six dairy cattle breeds

M P Sanchez, C Escouflaire, A Baur, C Hozé, A Capitan, S Fritz, D Boichard

Sequence-based association analyses on X chromosome in six dairy cattle breeds

M.P. Sanchez[1]*, C. Escouflaire[2], A. Baur[2], C. Hozé[2], A. Capitan[2], S. Fritz[2], D. Boichard[1]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]Allice, 75595 Paris, France; *marie-pierre.sanchez@inrae.fr

Abstract

Genomic evaluations in cattle are commonly based on effects of autosomal SNPs, the X chromosome (BTAX) being generally excluded because of its hemizygous state in males. We estimate here the effects of imputed sequence variants of BTAX in six dairy breeds and for 11 complex traits related to milk production, milk composition, mastitis resistance, fertility and stature. We detected QTL in almost all breeds and for almost all traits. They were distributed throughout BTAX and explained between 0.1 and 5.2% of the total phenotypic variance of the trait. In each of these QTL, we identified genes (*GPC3*, *SEPTIN6*, *ZC3H12B*, *COL4A6*, *IDS*, S*LC16A2*, *MIR363*, *HS6T2*, *MAMLD1*, *COL4A6*, *DGKK*, *DDX3X*, *GPC3*, *ERCC6L*, *FRMPD4*, *ENOX2*, *KLHL13*, *HMGB3*, and *CDKL5*) for which the functional link with the studied traits remains to be established. We therefore show here the importance of BTAX in the genetic determinism of complex traits in dairy cattle.

Introduction

Due to its specificity (hemizygous state in males, dosage compensation in females, different genomic relationship matrix), BTAX is often excluded from genetic analyses in cattle. Yet, BTAX is the second largest chromosome of the bovine genome and it is particularly rich in genes. Excluding BTAX from genomic evaluations can thus lead to the loss of a significant part of the genetic variability of traits (Su et al., 2014). To assess the effects of BTAX on dairy traits, we report here results of association analyses from imputed sequences of cows of six French breeds for eleven traits related to milk production, mastitis resistance, fertility and stature.

Materials & Methods

Cows and traits. We analysed 236,496 cows from three national (MON, NOR, HOL) and three regional (ABO, TAR, VOS) breeds with 50k genotypes and phenotypes (Table 1). Traits included were milk (MY), protein (PY) and fat (FY) yields; protein (PC) and fat (FC) contents; somatic cell score (SCS) and clinical mastitis (MAST; except for VOS); calving-first insemination interval (ICFI), heifers' (HCR) and cows' (CCR) conception rate; stature (STAT). Phenotypes used were yield deviations (YD) *i.e.*, phenotypes adjusted for non-genetic effects.

Table 1. Features of populations and sequence variants analysed.

Breed	Abbrev.	# 50k genotypes	# HD genotypes	# WGS animals[2]	# variants after filtering[1]	Mean[1] imputation R^2	Mean[1] MAF
Abondance	ABO	7449	199	9	154,966	0.69	0.18
Tarentaise	TAR	3969	179	12	181,473	0.79	0.19
Vosgienne	VOS	2910	181	4	170,560	0.77	0.20
Montbéliarde	MON	61,881	522	63	186,368	0.81	0.18
Normande	NOR	78,472	526	45	190,280	0.82	0.17
Holstein	HOL	81,815	804	1053	201,554	0.81	0.17

[1] Variants with a MAF ≥ 0.005 and with an imputation $R^2 \geq 0.20$; [2] 2712 multi-breed sequences used for imputation

Imputation and association analyses. We considered only the X-specific non-pseudo autosomal region (non-PAR) that covers the main part of BTAX (0-133.3 Mbp on the ARS-UCD1.2 reference genome; Johnson et al., 2019). For imputation, males were considered homozygous for all SNPs and pedigree information was not included. HD genotypes of 32,268 SNPs were first imputed from genotypes of 1147 EuroGMD SNPs with FImpute (Sargolzaei et al., 2014) using animals with HD genotypes as a reference (Table 1). Then, 778,576 sequence variants were imputed using a multi-breed population of 2712 animals from the RUN8 reference panel of the 1000 Bull Genomes consortium (Bouwman et al, 2018) with Minimac (Howie et al., 2012). Allele dosages were tested in within breed association analyses using GCTA software (Yang et al., 2011). All phenotypes were measured on females, we therefore applied the following model: $\mathbf{y} = \mathbf{1}\mu + \mathbf{x}b + \mathbf{u} + \mathbf{e}$, where $\mathbf{y}$ is the vector of YD; μ is the overall mean; b is the additive fixed effect of the variant tested; $\mathbf{x}$ is the vector of imputed allele dosages; $\mathbf{u} \sim N(0,\mathbf{G}\sigma^2_u)$ is the vector of random polygenic effects, with $\mathbf{G}$ the genomic relationship matrix based on autosomal 50k SNPs, and σ^2_u is the polygenic variance; $\mathbf{e} \sim N(\mathbf{0},\mathbf{I}\sigma^2_e)$ is the vector of random residual effects. We analysed variants with a MAF $\geq$ 0.005 and with a Minimac imputation $R^2 \geq 0.20$ *i.e.*, from 154,966 to 236,011 variants depending on the breed (Table 1). A unique threshold was used corresponding to 5% significance after Bonferroni correction for ~160,000 variants ($-\log_{10}(P)=6.5$). In each breed and trait, variants with significant effects located less than 10 Mbp apart were grouped together to define QTL. Confidence intervals (CIs) of QTL were then determined based on the positions of variants in the upper third of the peak (van den Berg et al., 2016). The percentage of phenotypic variance (σ^2_p) explained by each QTL was calculated for the variant with the most significant effect as $\%\sigma^2_p = 100.2p(1-p)\alpha^2/\sigma^2_p$, with p the frequency and α the estimated allelic substitution effect.

Results

Association analyses detected QTL in all breeds, except TAR and VOS, and for all traits, except HCR (Table 2). We identified 191 QTL (0 to 10 QTL per breed x trait analysis). For milk traits, we detected QTL in four breeds, the most significant ones being mainly located at the beginning of BTAX for PC and FC and throughout BTAX (~30, 60, 85, 102, 110, 130 Mbp) for MY, PY, and FY (Figure 1). We found a higher number of QTL in national breeds (63, 57, and 63 in MON, NOR, and HOL, respectively) than in regional breeds (8, 0, and 0 in ABO, TAR, and VOS, respectively). The cumulative effects of the QTL per breed explained from 0.1% (ICFI in HOL and CCR in MON) to 5.2% (FY in ABO) of σ^2_p. CIs of the QTL contained 1 to 2193 variants with significant effects. In total 39,429 unique variants were identified.

Table 2. Number of QTL [total % of phenotypic variance explained by the QTL].

Breed[1]	MY	PY	FY	PC	FC	SCS	MAST	ICFI	HCR	CCR	STAT
ABO	1[0.9]	1[0.9]	2[5.2]	2[4.1]	1[1.5]	0	1[0.9]	0	0	0	0
MON	10[1.8]	7[1.4]	7[1.2]	9[2.9]	9[1.5]	3[0.4]	0	8[1.3]	0	1[0.1]	9[1.5]
NOR	8[0.8]	10[1.0]	7[0.6]	7[0.8]	5[0.7]	4[0.4]	0	8[1.2]	0	0	8[3.1]
HOL	8[0.5]	7[0.6]	9[1.0]	10[1.2]	10[1.8]	9[1.1]	0	1[0.1]	0	1[0.2]	8[0.9]

[1] No QTL were found in regional TAR and VOS breeds

Most variants included in CIs were located in intergenic regions (72%) while the other (28%) were in genes or in the vicinity of genes, mainly in introns (20.3%) (Table 3). The best-ranked variants in the peaks *i.e.*, the top 10 and top 1 variants, were more frequently located in genes (37.3% and 47.8%, respectively). Non-intergenic variants ranked first in the peaks were located in 58 different genes. The top 1 variants with the most significant effects ($-\log_{10}(P) \geq 20$) were

located in *GPC3* (FC), *SEPTIN6* (PC), *ZC3H12B* (PC), and *COL4A6* (STAT) in MON; *IDS* (STAT) in NOR; and *SLC16A2* (FC) and *MIR363* (PC) in HOL. For fertility traits (ICFI and CCR), the genic top 1 variants of the QTL, detected in national breeds (MON, NOR or HOL), were located in *HS6T2*, *MAMLD1*, *COL4A6*, *DGKK*, *DDX3X*, *ENSBTAG00000048527*, *GPC3*, *ERCC6L*, and *FRMPD4* genes while for udder health traits (MAST and SCS in ABO, MON, NOR, and HOL), they were located in *ENOX2*, *KLHL13*, *HMGB3*, and *CDKL5*.

Table 3. Functional annotation of all, top 10 or top 1 variants located in CIs of the QTL.

	All variants		Top 10 variants		Top 1 variants	
	Number	%	Number	%	Number	%
Intergenic region	28,168	72.0	1084	62.7	106	58.2
Intronic region	7955	20.3	467	27.0	58	31.9
Upstream region	1471	3.8	87	5.0	8	4.4
Downstream region	1231	3.1	75	4.3	7	3.8
Synonymous variants	87	0.2	8	0.5	2	1.1
3' UTR	80	0.2	3	0.2	0	0
Missense variants	79	0.2	3	0.2	0	0
5'UTR	33	0.1	0	0.0	0	0
Other exonic variants	28	0.08	3	0.2	1	0.5
Total	39132	100	1730	100	182	100

Discussion

In mammals, the non-PAR region of BTAX differs from autosomes because of its hemizygous state in males (XY). Mechanisms of dosage compensation occur in females (XX) to inactivate one BTAX copy and therefore ensuring an equal expression of genes located on BTAX in both sexes. Little is known about the complex mechanisms of BTAX inactivation in cows but both copies appear to be equally expressed in the mammary gland at the population level suggesting a random inactivation of either copy (Couldrey et al., 2017). Because of its specificities, BTAX is usually excluded from genetic analyses and few GWAS were conducted in cattle (*e.g.*, Fortes et al., 2020). In the present study, we focused on this particular chromosome to assess the effects of its variants imputed at the sequence level on various complex dairy traits (milk production, udder health, fertility, and stature) in cows from six dairy cattle breeds. After adapting the imputation process (males assumed homozygous and no pedigree) and considering paternal and maternal X chromosomes equally expressed at the population level, we applied the model used for autosomes in association analyses. Except for breeds with the lowest number of cows (TAR and VOS) and for one fertility trait (HCR), we identified QTL for all breeds and traits analysed. As expected, due to the number of cows analysed in each breed, the number of QTL was higher in the largest breeds and these QTL had larger effects. A higher number of QTL with more significant effects was also found for the most heritable traits (milk production, milk composition, and stature). Depending on the breed and on the trait, these QTL explained up to 5.2% of σ^2_p. We also highlighted QTL for less heritable traits that may explain up to 1.3% (fertility) and 1.1% (udder health) of σ^2_p in some breeds. In particular, for SCS in MON, NOR, and HOL, we found that 0.4%, 0.4%, and 1.1% of σ^2_p was explained by 3, 4, and 9 QTL located on BTAX, respectively. These values can be compared to the values we estimated in a previous study for the same trait in MON, NOR, and HOL bulls *i.e.*, 6.3%, 0%, and 2.6% for 3, 0, and 2 QTL detected on all the 29 bovine autosomes, respectively (Tribout et al., 2020). As illustrated in Figure 1, QTL detected on BTAX presented large CIs. This phenomenon is probably due to the linkage disequilibrium, more extended in the non-PAR region where no recombination

occurs in males (Zhang et al., 2020). We show here that dozens of candidate genes are located in X-linked QTL. However, functional links between these genes and traits analysed remain to be established. This study shows the importance of BTAX in the genetic determinism of complex traits in dairy cattle and supports the results of Su et al. (2014) who found that markers on BTAX contribute to accuracy of genomic predictions in Holstein breed.

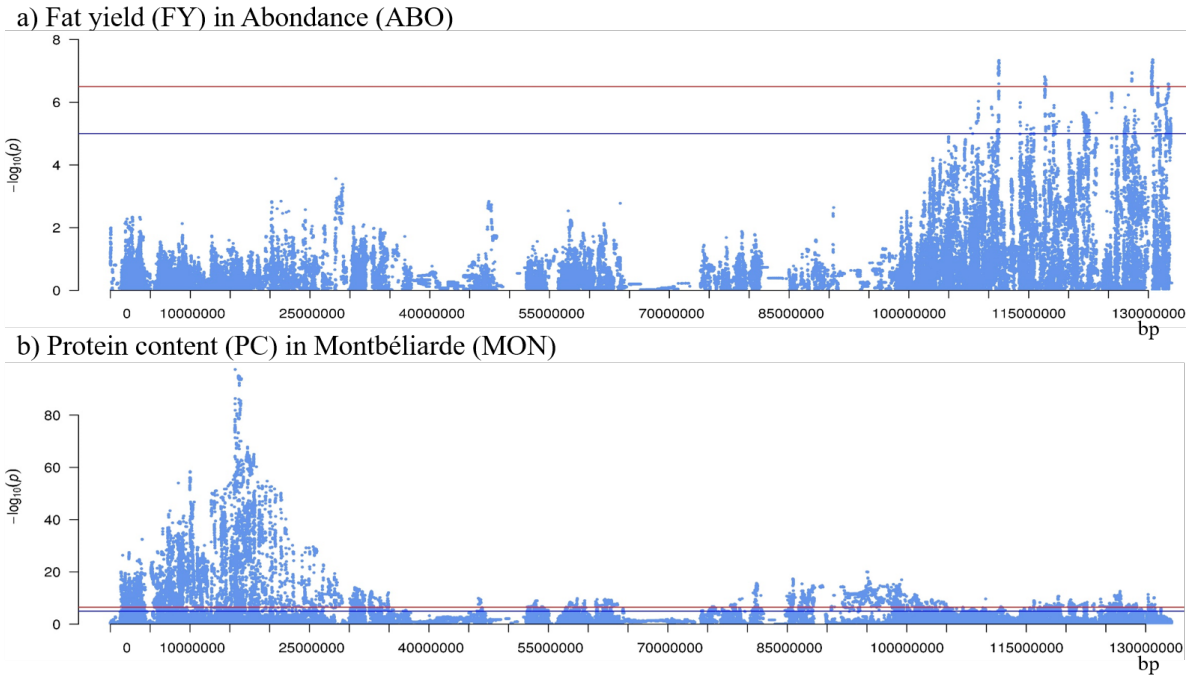

Figure 1. –log10(P) values plotted against the position of variants on X chromosome.

References

Bouwman A. C., Daetwyler H. D., Chamberlain A. J., Ponce C. H., Sargolzaei M. *et al.* (2018) Nat Genet 50(3):362-367. https://doi.org/10.1038/s41588-018-0056-5

Couldrey C., Johnson T., Lopdell T., Zhang I. L., Littlejohn M. D. *et al.* (2017) J Dairy Sci 100:5491-5500. https://doi.org/10.3168/jds.2016-12490

Fortes M., Porto-Neto L. R., Satake N., Nguyen L. T., Freitas A. C. *et al.* (2020) Genet Sel Evol 52:46. https://doi.org/10.1186/s12711-020-00563-5

Johnson T., Keehan M., Harland C., Lopdell T., Spelman R. J. *et al.* (2019) J Dairy Sci 102:1-5. https://doi.org/10.3168/jds.2018-15638

Howie B., Fuchsberger C., Stephens M., Marchini J., and Abecasis G. R. (2012) Nat Genet 44(8):955-959. https://doi.org/10.1038/ng.2354

Sargolzaei M., Chesnais J., and Schenkel F. (2014) BMC Genomics 15(1):478. http://doi.org/10.1186/1471-2164-15-478

Su G., Guldbrandtsen B., Aamand G. P., Stranden I. and Lund M. S. (2014) Genet Sel Evol 46:47. http://doi.org/10.1186/1297-9686-46-47

Tribout T., Croiseau P., Lefebvre R., Barbat A., Boussaha M. *et al.* (2020) Genet Sel Evol 52 :55. https://doi.org/10.1186/s12711-020-00575-1

van den Berg I., Boichard D., Lund M. S. (2016). J Dairy Sci 99(11):8932-8945. https://doi.org/10.3168/jds.2016-11073

Yang J., Lee S., Goddard M., and Visscher P. (2011) Am J Hum Genet 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zhang J., Kadri N. K., Mullaart E., Spelman R., Fritz S. *et al.* (2020) Heredity 125:304-316. https://doi.org/10.1038/s41437-020-0341-9