Genomic evaluation using QTL information

Vincent Ducrocq, Pascal Croiseau, Aurélia Baur, Romain Saintilan, Sebastien Fritz, Didier Boichard

Genomic Evaluation using QTL Information

V. Ducrocq[1], P. Croiseau[1], A. Baur[2], R. Saintilan[2], S. Fritz[2] and D. Boichard[1]
[1] INRA UMR1313 GABI, [2]UNCEIA, UMR1313 GABI, Jouy-en-Josas, France

ABSTRACT: The French dairy cattle genomic evaluation uses an extension of Marker-Assisted BLUP (MA-BLUP) based on a haplotype model with a residual polygenic effect. A two-step approach to select SNP and construct haplotypes linked to QTL was tested by cross-validation for two breeds with moderate (Montbéliarde) and large (Holstein) reference populations. The comparison with other genomic evaluation approaches showed MA-BLUP validation performances which are very similar for Montbéliarde and slightly lower for Holstein. Some simplifying or ad-hoc assumptions explaining these results have been identified and reveal directions for improvement. In particular, a genome-wide strategy to pinpoint the most informative haplotypes linked to QTL is highly desirable. The robustness and computing simplicity of MA-BLUP remain two appealing features for routine genomic evaluations. The use of haplotypes is expected to give more robust and reliable predictions over more distantly related animals.

Keywords: dairy cattle; genomic evaluation; genomic selection; quantitative trait loci

Introduction

In the nineties, several large QTL detection programs in dairy cattle led to the discovery of many QTL for various economic traits (see Khatkar et al. (2004), for a review). In France, based on the results of such a program (Boichard et al. (2003)), a rather unique large scale marker assisted selection program called MAS1 was implemented between 2001 and 2008: more than 70,000 animals were genotyped for 45 microsatellites to enhance EBV of young bulls computed using a marker-assisted BLUP (MA-BLUP) evaluation. MAS1 was used as a pre-selection tool of young bulls before progeny testing (Guillaume et al. (2008)).

QTL detected rather inaccurately with such a reduced number of microsatellites explained only a small part of the total genetic variability. With the possibility to genotype thousands of SNP, associations between markers and QTL are more easily detected. Increasing the number of QTL in a MAS approach raises the part of the total genetic variability due to QTL but leads to a larger fraction of false QTL due to lack of power of the experimental designs.

Genomic selection (GS) as proposed by Meuwissen et al. (2001) assumes that most QTL have a small effect and that a precise estimation of each QTL effect is not relevant, as long as their sum is a good predictor of the breeding value of selection candidates. With GS, a formal QTL detection step is not needed and inclusion of false QTL is not an issue. GS proved to be very successful when reference populations used to establish prediction equations were large enough. Nevertheless, it seems counterintuitive to completely ignore any strong prior knowledge on where large QTL are located, e.g., major genes such as DGAT1 or GRH.

Among QTL detection approaches, Linkage Disequilibrium and Linkage Analysis (LDLA – Meuwissen and Goddard (2000)) combines the robustness of linkage analysis and the high resolution of LD analysis and is well suited for grand-daughter designs. However, with high density genotypes, it faces problems of high false discovery rate, inaccurate QTL location and large overestimation of the proportion of genetic variance explained by each QTL. The main reason is that each QTL position is tested independently, one at a time, ignoring potential LD with other QTL. Some of these limitations can be at least partly circumvented (e.g., see Jonas et al., this congress) but in fact, most genomic evaluation where SNP effects are explicitly estimated altogether can be used for QTL detection (van den Berg et al. (2013)). It is known that many of the QTL detected then are potentially false positive but this does not preclude overall genetic predictions to be precise. This favorable feature may be the consequence of average estimated effects of false QTL close to 0 and of SNP altogether contributing to some sort of relationship measure between animals.

By construction, QTL detection as well as GS efficiency strongly depend on LD between markers and QTL. When markers are biallelic (e.g., SNP), LD between neighboring markers is usually fairly small (e.g., $r^2 \sim 0.2$ to 0.25). This is often an overestimation of LD between the marker and any QTL for which the mutated allele is rare. A convenient way to increase LD is to consider haplotypes of markers, making them multiallelic (Edriss et al. (2013)). Then, LD between a rare allele at the QTL and at least one allele of the haplotype may be large. SNP haplotypes have been successfully used in LDLA analysis and GS (e.g., Meuwissen and Goddard (2000), Calus et al. (2009)).

Based on these considerations, it was decided in 2008 to extend the French MAS approach in order to include more QTL regions (300 to 700), traced by haplotypes of 3 to 5 SNP (Boichard et al. (2012)). These regions were defined in an ad-hoc manner using two complementary approaches: 20 to 40 "large" QTL were initially selected using an LDLA approach, the others being chosen grouping SNP selected with an Elastic-Net approach (Croiseau et al. (2011)). In a sense, such a procedure combines interesting features of MAS and GS, and was referred to as MASG. In French conditions, MASG was shown to be at least as efficient as other GS methods (Boichard et al. (2012)).

This paper reconsiders the choice of QTL regions, the comparison with various GS implementations and the pros and cons of a MA-BLUP approach. Data from the two main French dairy breeds were used for illustration.

Material and Methods

Data

For Holstein and Montbéliarde, a reference population was created consisting of all bulls progeny-tested and genotyped with the Illumina BovineSNP50™ Beadchip. All bulls born at least 4 years before the youngest genotyped bull with daughter performances were included in the training population (T). The others composed the validation population (V). All validation bulls were required to have their sire and grand-sires in the training population. This had an impact on the reference population for the Holstein breed for which the genotype of the maternal grand sire of a significant number of young foreign bulls of the Eurogenomics consortium was not available.

The phenotypes used were generally daughter yield deviations (DYD) with their appropriate weight (EDC) transformed into their animal model equivalent. For foreign Holstein bulls, DYD were replaced by deregressed proofs (DRP). Table 1 presents the final distribution of the reference populations for production traits. Slight variations existed for the other traits.

Table 1: composition of the reference populations for production traits

	Holstein	Montbéliarde
Threshold date(1)	1/06/2005	1/10/2004
Training population (T)	13165	1701
Validation population (V)	3391	535

(1) Limit date of birth separating training and validation populations.

For genotyped bulls, a total of 43,801 SNP on the 29 autosomes were kept after pedigree checking and quality control: SNP with a minor allele frequency lower than 0.01, with an invalid position on the UMD3.1 assembly, etc., were discarded. See Baur et al., this congress, for details on an identical procedure applied to French Brown Swiss data. Each genotyped animal was then phased and missing genotypes at each particular SNP were imputed using DagPhase (Druet and George (2010)).

Methodology

In our implementation of MA-BLUP, haplotypes of SNP are used to trace QTL. Ideally, a direct detection of the most informative haplotypes is desirable. Indeed, such strategies have been proposed but they were either not fully implemented into a software (e.g., Calus et al. (2009), Edriss et al. (2013)) or they are currently under test (Croiseau and Fouilloux, this congress). Here a simplified (simplistic?) two-step approach was used: first, SNP influencing the traits of interest were selected using either Elastic Net (EN; Zou and Hastie (2005)) or a Bayesian Sparse Linear Mixed Model (BSLMM, Zhou et al. (2013)).

Elastic Net is a variable selection method which combines Lasso and Ridge Regression into a larger family of models (Zou and Hastie (2005)). It requires the choice of two penalization parameters α and λ. To determine these parameters, a calibration (C) population consisting of the youngest 15% bulls of the training (T) population was created. A grid search was performed to find the values of α and λ leading to the best average prediction in (C) when EN is applied to the (T-C) population. For some traits, EN selected a large number of SNP, incompatible with the later definition of a manageable number of QTL haplotypes for MA-BLUP. Croiseau *et* al (2011) showed that it was possible with very limited loss to choose α and λ such that the total number of SNP selected by EN is below a limit, fixed here at 2500 SNP. The α and λ values were then fixed and applied to the complete training population. Note that in this EN implementation (*gmlnet* package in R), no residual polygenic effect could be included.

BSLMM also combines two popular approaches for genomic selection: a Bayesian Variable Selection Regression and a polygenic model. It can be viewed as an efficient implementation of a BayesCπ approach with a residual polygenic effect and where the pedigree-based relationship matrix is replaced by the genomic one. One important limitation of the current version of the BSLMM software of Zhou et al. (2013) is that all observations are assumed to have the same associated weight, i.e., EDC are ignored. To mimic a QTL detection step similar to the one using EN, only the 2500 SNP with the largest inclusion probability as calculated by BSLMM were retained.

Once SNP were selected by either EN or BSLMM, haplotypes were created: when 3 to 5 selected SNP were included in an interval of less than 2Mb (on average for the 50k chip, 1 Mb contains 18 SNP), they were grouped to form a QTL haplotype (in practice, most haplotypes included much closer SNP). Otherwise the neighboring SNP of a selected SNP was/were added to form a QTL haplotype of at least 3 SNP.

For the MA-BLUP evaluation, the following mixed linear model was applied:

$$y_i = \mu + u_i + \sum_{k=1}^{Nqtl}\left(v_{ik}^{p} + v_{ik}^{m}\right) + e_i \qquad (1)$$

where y_i is the observation (DYD or DRP) of the genotyped animal i, μ is an overall mean, v_{ik}^{p} and v_{ik}^{m} are the random effects of the paternal and maternal alleles of haplotype k which are assumed to be independent, all with the same haplotype variance σ_v^2) ; u_i is the residual polygenic effect and e_i the model residual for animal i and $\mathbf{u} = \{u_i\}$ is distributed as $N(0,\mathbf{A}\,\sigma_u^2)$ where **A** is the relationship matrix. For simplicity, for each trait, σ_u^2 was defined as a proportion (from 10 to 90%, by step of 10%) of the total genetic variance σ_a^2 estimated from previous datasets using an AIREML procedure. The remaining part was attributed to the haplotypes: σ_v^2 was taken equal to $\sigma_a^2 - \sigma_u^2$ divided by the total number of haplotypes. More relevant haplotype variances could have been derived, but for simplicity (and lack of time), they were not considered here. Mixed model equations solutions were obtained in the following way: first, only equations from animals in the reference population were solved iteratively. At convergence, solutions from candidate animals without performances were computed. Reliabilities for all animals were directly derived from the inverse of the coefficient matrix for the reference population.

Analyses

Results from MA-BLUP based on the haplotype lists derived from EN or BSLMM analyses were compared to various SNP-based genomic selection procedures: EN as described above, GBLUP, Bayesian Lasso (Legarra et al. (2009)) and BayesCπ as implemented in the GS3 software (Legarra et al. (2011)). For GBLUP, Bayesian Lasso, Bayes Cπ and both MA-BLUP, a residual polygenic component was also assumed. For its associated variance, different values were considered, from 10% to 90% of the total genetic variance, by step of 10%. Comparison criteria were the weighted correlation $r_{DYD,GEBV}$ between observed DYD (or DRP) and GEBV (which can be regarded as a predicted DYD) and the slope of the regression observed DYD/DRP on GEBV (a slope close to 1 is expected). EDC were used as weights. No correction was performed for the average reliability of DYD/DRP.

Results

Table 2 displays some key figures regarding the MA-BLUP. After the QTL detection step, the number of SNP retained was systematically 2500 with BSLMM (as imposed) and close to 2000 with EN. In half of the traits, the optimal number was less than 2500 (and as low as 205!) in Montbéliarde but always larger than 2500 in Holstein. The average decrease in $r_{DYD,GEBV}$ due to the restriction to 2500 was 0.01 in Montbéliarde and 0.03 in Holstein. Slightly more haplotypes were retained with the EN approach (around 900 per trait) than with BSLMM (close to 1000). The average number of alleles per haplotype was larger in Holstein than in Montbéliarde (e.g., 13.6 *vs* 11.9 with EN)

Table 2: average number (minimum/maximum) of SNP selected and haplotypes formed and average number of alleles per haplotype

Detection with[(1)]	Number of	Breed	
		Holstein	Montbéliarde
EN	SNP	2086 (1890-2161)	1853 (205-2500)
	Haplotypes	890 (889-897)	896 (865-927)
	Alleles/hap	13.6 ± 6.9	11.9 ± 5.7
BSLMM	SNP	2500	2500
	Haplotypes	989 (970-1026)	936 (959-1029)
	Alleles/hap	13.9 ± 6.7	11.2 ± 4.8

[(1)] Elastic Net (EN); Bayesian Sparse Linear Mixed Model (BSLMM)

Tables 3 and 4 present $r_{DYD,GEBV}$ and the optimum polygenic variance (as % of the total genetic variance) for three different groups of traits and overall. The overall average absolute deviation of the regression slope from 1 is also displayed. In all scenarios, very few traits led to a regression slope (slightly) larger than 1. Results of Bayesian Lasso are not reported because they were always slightly inferior but always very close to BayesCπ. In Montbéliarde, the overall average correlations were very similar whatever the genomic evaluation, with a maximum difference between approaches of 0.015. A detailed look at each trait reveals that for 12 low heritability traits (6 type traits and 6 functional traits), BayesCπ converged to a situation where no SNP were retained ($\pi \to 1$)) hence to a complete polygenic model. When these pathological cases were excluded, the average $r_{DYD,GEBV}$ correlation increased for all methods, but the improvement varied between methods and as a result, Bayes Cπ was superior by 0.010 to 0.016 to GBLUP and MA-BLUP. For all methods, the optimal proportions of variance to be attributed to the residual polygenic effect were quite large (often around 40-50%). In fact, these values are substantially larger than what is commonly reported in the literature. Regression slopes were closer to 1 for BayesCπ than for GBLUP, the two MA-BLUP being intermediate.

In Holstein, the convergence of BayesCπ to a fully polygenic model was never observed. BSLMM performed poorly for body condition score (a trait associated with a much smaller reference population in France) and for locomotion and stillbirth. However, the MA-BLUP based on the BSLMM SNP lists performed surprisingly much better for those traits. This is illustrated in Figure 1.

Figures 1

GBLUP and BayesCπ gave very similar correlations and these were about 0.03 greater than both EN and BSLMM. The two MA-BLUP approaches gave intermediate results. The increase in correlation with respect to pedigree BLUP was about twice the one observed in Montbéliarde. The optimal proportions of residual polygenic variances were smaller than in the Montbéliarde breed, and larger for MA-BLUP than for the other methods. Regression slopes were closer to 1 than for Montbéliarde, with better results with BayesCπ and GBLUP than with the two MA-BLUP. As for Montbéliarde, the two MA-BLUP approaches gave comparable results with a slight advantage to the situation using EN in the first step.

Discussion

The purpose of this study was to verify whether MA-BLUP with haplotype effects is a valuable alternative to more sophisticated genomic evaluation methods. In previous works leading to the first MASG implementation, reference populations were substantially smaller. A limited number of large QTL were first selected by an LDLA approach where each position was tested independently from the others, with a number of weaknesses when dense markers are used (lower detection power, inaccurate location of neighboring QTL, large overestimation of each QTL contribution to total genetic variance). 'Smaller' QTL were chosen in a separate analysis where SNP forming haplotypes were selected using a GS approach, the Elastic Net in our case. Here, these two components were merged into one: a single GS (EN or BSLMM) approach was used to construct a unique list of haplotypes. Admittedly, the approach is still rather simplistic and suffers from a number of questionable assumptions and short-cuts (leaving room for improvement!). To name just a few, limitations of the software used forced to implement EN without a polygenic component and BSLMM without accounting for the weights of DYD. In both cases, restrictions to a maximum of 2500 SNP retained were imposed. Both the number of retained QTL and the way SNP were grouped into

Table 3: average weighted correlation between GEBV and daughter yield deviation (DYD), average fraction of the genetic variance attributed to the residual polygenic effect and average absolute deviation (slope dev.) from 1 of the regression coefficient of GEBV on DYD for the Montbéliarde breed

Method	Production (5)		Type traits (28)		Functional traits (20)		All traits (43)		
	$r_{DYD,D\hat{Y}D}$	poly (%)	$r_{DYD,D\hat{Y}D}$	poly (%)	$r_{DYD,D\hat{Y}D}$	poly (%)	$r_{DYD,D\hat{Y}D}$	poly (%)	slope dev.
Polygenic	0.365	100	0.438	100	0.362	100	0.412	100	
Elastic Net	0.529	-	0.539	-	0.481	-	0.506	-	0.149
BSLMM	0.553		0.564		0.363		0.516		
GBLUP	0.534	28	0.553	30	0.524	55	0.521	36	0.145
BAYESCpi	0.545	38	0.545	47	0.530	56	0.520	49	0.116
MA-BLUP (EN list)	0.527	38	0.554	37	0.507	40	0.520	39	0.160
MA-BLUP (BSLMM list)	0.515	38	0.552	32	0.501	40	0.518	36	0.148

haplotypes were quite arbitrary. Overcoming these limitations is certainly possible but requires some investment into programming. Indeed, the ultimate and much preferable solution is to develop an approach which directly selects the most suitable haplotypes: a haplotypic extension of BayesCπ (Croiseau and Fouilloux, this congress).

Perhaps more far-fetched, it was assumed that allele effects of a same haplotype were independent, with an equal variance for each haplotype. This is definitely far from true for traits where QTL with large effects are well known, such as DGAT1 for fat content. Again, a better estimation of the contribution of each haplotype to the total variance is certainly feasible (Knürr et al. (2013)).

The use of haplotypes also has some shortcomings. In particular, it requires to phase the SNP in order to construct them. This can be quite time consuming and a potential source of some errors. But the necessity to combine genotypes obtained from a growing number of different chips has transformed imputation into a common practice of which phasing is simply a component. Haplotypes are used to increase LD between markers and any neighboring QTL, but increasing haplotype length quickly leads to a large number of allele effects to estimate. In our case, with haplotypes of 3 to 5 (mostly 5), the average number of alleles was reasonable (11 to 14) but the maximum was 32.Various strategies are possible to reduce this number such as clustering alleles which are relatively similar and likely to carry the same QTL (Calus et al. (2009); Edriss et al. (2013)). Again, this could lead to a better estimation of rare haplotype effects. Another weakness of a haplotypic model for genomic prediction is that new alleles can appear by recombination in the population of candidates, with no equivalent in the reference population. Ignoring these allelic effects or equating them to one or the average of the parental alleles are possible approximations. Perhaps the most promising development to elude some of these complications could come from the use of hidden state models to access ancestral haplotypes as in Kadri et al. (2014)

Based on simulations, haplotypes of 4 to 6 consecutive SNP appeared to give the best prediction results (Guillaume, 2009). This was consistent with a LD between QTL allele and haplotype alleles (r^2) increasing from 0.331 to 0.405 and 0.412 for haplotypes of length 2, 4 and 6 SNP respectively, and an observed accuracy maximum for haplotypes of 4 SNP. Other authors (Calus et al. (2009); Boleckova et al. (2012)) found similar or larger optimal values (up to 10 for Villumsen et al. (2008) but with haplotypes of 5 SNP being nearly as efficient). In our implementation, SNP in a given haplotype were not necessarily consecutive markers from the 50K chip. The impact of this relaxed assumption has not been investigated yet, although the experience of Knürr et al. (2013) using flanking SNP of pre-selected SNP were not very good. The optimal number and length of haplotypes may be linked to the number of independent chromosome segments over the genome, which itself depends on the population structure.

Table 4: average weighted correlation between GEBV and daughter yield deviation (DYD), average fraction of the genetic variance attributed to the residual polygenic effect and average absolute deviation from 1 of the regression coefficient of GEBV on DYD for the Holstein breed

Method	Production (5)		Type traits (21)		Functional traits (10)		All traits (36)		
	$r_{DYD,D\hat{Y}D}$	poly (%)	$r_{DYD,D\hat{Y}D}$	poly (%)	$r_{DYD,D\hat{Y}D}$	poly (%)	$r_{DYD,D\hat{Y}D}$	poly (%)	slope dev.
Polygenic (BLUP)	0.472	100	0.477	100	0.477	100	0.451	100	
Elastic Net	0.752	-	0.645	-	0.542	-	0.631	-	0.072
BSLMM	0.788		0.644		0.534	-	0.633		
GBLUP	0.776	12	0.685	24	0.554	21	0.661	22	0.082
BAYESCpi	0.787	18	0.687	25	0.556	22	0.665	23	0.083
MA-BLUP (EN list)	0.763	20	0.667	39	0.541	35	0.645	35	0.106
MA-BLUP (BSLMM list)	0.765	22	0.658	36	0.532	20	0.638	33	0.103

So the optima probably vary across breeds. They are also likely to differ when selected SNP come from a high density chip, e.g., with the aim to apply multi-breed GS.

Despite of all these strong limitations, we showed that in practice, MA-BLUP validation results were very robust to the SNP selection strategy. They were also close to those from standard GS methods and even equivalent in Montbéliarde. Similar conclusions were reported previously (e.g., Boichard et al. (2012)) with yet another SNP selection approach. If we compare the current results with previous works, it is tempting to conclude that the benefit of MA-BLUP are larger when the size of the reference population is limited.

This robustness is quite reassuring. Indeed, the approach presents a number of desirable features:

1) The genetic model (1) is quite simple. In particular, in contrast with most other countries with GS, the resulting predictions have been used in France as GEBV *without* any extra blending. Indeed, whatever the trait considered, the optimal residual polygenic fractions are quite large compared to most other countries. But in fact, these fractions are probably comparable to what is implicitly obtained after blending.

2) The solution of the mixed model equations is fast: the construction of the genomic relationship is avoided and the number of haplotype effects is reduced (usually between 10,000 and 15000 here) compared to a GBLUP involving all SNP. Furthermore, no long chain of simulated values is required.

3) As a result of these two points, model-based genomic reliabilities do not require approximations: they are directly derived from the actual inversion of a coefficient matrix of mixed model equations of manageable size. Also, they are probably less over-evaluated than with more standard GS approaches: Guillaume (2009) showed that a 25% over-estimation of the variance attributed to QTL had a limited influence on $r_{DYD,GEBV}$ (change <1%) but a drastic impact on theoretical reliabilities (change of ~10% or more). The high residual polygenic fractions used are therefore a safeguard against overoptimistic genomic reliabilities.

4) Genomic evaluations of a same animal are relatively stable over time, as long as the QTL list does not change. In contrast, iterative (MCMC) approaches to estimate SNP effects may lead to GEBV variation between runs, just by slight changes of the retained SNP, even though the global accuracy is the same.

5) Because it relies on a stronger LD between haplotype alleles and underlying QTL, the QTL model is expected to be less sensitive to relatedness between reference and candidate populations: the haplotype effects are expected to be better maintained over generations or over groups of more poorly related animals. It was not possible to actually check this here because all validation bulls had both their sire and maternal grand-sire in the reference population.

6) The approach is robust to variation in chip density, as long as true or imputed genotypes are available. For example, haplotypes can consist of groups of SNP selected on the Illumina BovineHD Beadchip™ chip. If alleles of such haplotypes are shared by different breeds, a multibreed genomic evaluation can be implemented. Its accuracy remains to be tested.

7) Identified causal mutations can be considered as a special case of QTL haplotype with only two allele effects. Conversely, large allele effects of a haplotype can be useful to more easily identify causal mutations (Calus et al. (2008), Kadri et al. (2014)), especially relatively recent mutations..

8) As indicated by Legarra and Ducrocq (2012), there is no obstacle to include this approach into a single-step machinery. In matrix notation, model (1) can be written as:

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{u} + \sum_k \mathbf{Z}_k\mathbf{v} + \mathbf{e} \qquad (2)$$

where $\mathbf{Z_k}$ is an incidence matrix relating animals to alleles of the k^{th} haplotype. The **G** matrix introduced to replace the part of **A** corresponding to genotyped animals in single-step mixed model equations is equal to $\mathbf{G}= \kappa_0 \mathbf{A}_{22}+ \kappa_k \sum_k \mathbf{Z}_k\mathbf{Z}_k'$ with κ_0, .. κ_k are fractions of the total genetic variance for alleles of haplotype k. In the iterative implementation of Legarra and Ducrocq (2012), the repeated computation of $\mathbf{Gg_2}$ – where $\mathbf{g_2}$ is the additive genetic value of genotyped animals - does not present any difficulty.

It is important to underline that despite similar prediction accuracies, the various methods compared do not lead to the same rankings. This was illustrated by Liu et al. (2013): the national GEBV of (nearly) the same young bulls differed more between France and three other members of the Eurogenomics consortium (Germany, Netherlands, Nordic Countries) than between these countries. This can be attributed to the use of a QTL model based on haplotypes and a large residual polygenic effect. How the different methods rank in terms of actual efficiency and robustness in the long term remains to be evaluated.

Conclusion

In contrast with most countries with a genomic evaluation in dairy cattle, France has kept an evaluation model involving SNP haplotypes and a large residual polygenic effect. An ad-hoc two-step approach for the selection and construction of these haplotypes generated results which were similar to other commonly used genomic selection approaches when the reference population was of moderate size (Montbéliarde) and slightly lower when the reference population was very large (Holstein). A number of simplifying or ad-hoc assumptions which may penalize MA-BLUP have been identified. Clearly, the direct genome-wide search for the most informative haplotypes linked to QTL offers an interesting perspective for improvement. Meanwhile, the robustness and computing simplicity of MA-BLUP remain very appealing features for routine genomic evaluations. The use of haplotypes is expected to give more robust and reliable predictions over more distant animals.

Acknowledgments

All authors of this work are members of the Mixed Technology Unit “Génétique, Génomique et Gestion de populations bovines” (UMT3G), located at INRA Jouy-en-Josas.

Literature Cited

Boichard, D., Grohs, C., Bourgeois, F., et al. (2003). Genet. Sel. Evol. 35: 77-101.

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

Boleckova J., Christensen, O.F., Sorensen, P. et al. (2012). Czech J. Anim. Sci., 57:1-9.

Calus, M.P.L., Meuwissen, T.H.E, de Roos, A.P.W. et al. (2008). Genetics, 178: 553-561.

Calus, M.P.L, Meuwissen, T.H.E, Windig J.J. et al. (2009) Genet. Sel. Evol. 41:11.

Croiseau, P., Legarra, A., Guillaume, F., et al. (2011) Genet. Res. 93: 409-417.

Druet, T. and Georges M. (2010) Genetics 184: 789–798.

Edriss, V., Fernando R.L., Su G., et al. (2013) Genet. Sel. Evol 45:5

Guillaume, F.(2009) Chapter 5. PhD thesis. AgroParisTech.

Guillaume, F., Fritz, S., Boichard, D. and Druet, T. (2008). J. Dairy Sci. 91: 2520–2522

Kadri, N.K., Sahana, G., Charlier, C. et al. (2014) PLoS Genet 10(1): e1004049. doi:10.1371/journal.pgen.1004049

Khatkar, M.S., Thomson, P.C., Tammen, I. and Raadsmaa, H.W. (2004). Genet. Sel. Evol. 36: 163-190.

Knürr, T.,Strandén, I., Moivula, M. et al. (2013) EAAP – 64th Annual Meeting, Nantes, France 19:454.

Legarra, A., Robert-Granié, C., Croiseau, P. et al. (2011). Genet. Res., 93: 77-87.

Legarra, A. and Ducrocq, V. (2012) J. Dairy Sci. 95:4629–4645

Legarra, A., Ricard, A. and Filangi, O. (2013). GS3 Genomic selection. http://snp.toulouse.inra.fr/~alegarra

Liu, Z., Aamand, G.P., Fritz, S. et al. (2013) Interbull Bulletin, 47:38-42.

Meuwissen, T.H.E., and Goddard, M. E. (2001) Genet. Sel. Evol., 33:605–634

Meuwissen, T.H.E., Goddard, M.E. (2000) Genetics 155: 421–430.

Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001). Genetics 157: 1819-1829.

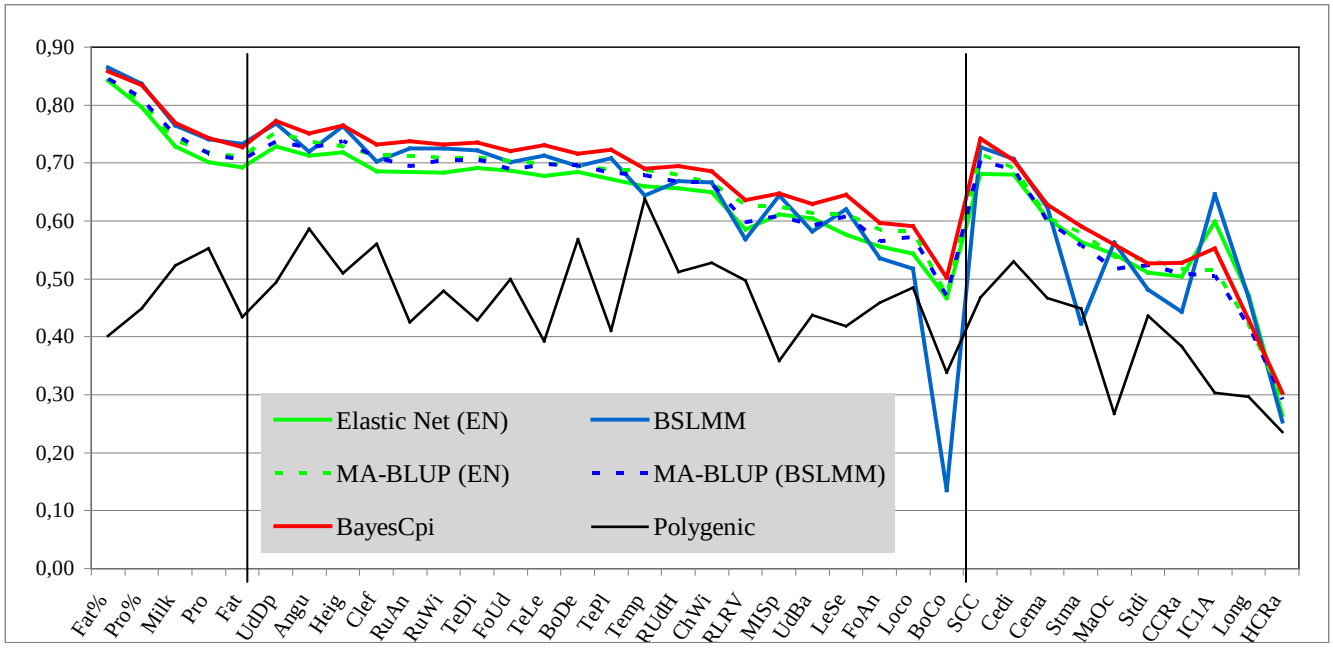
van den Berg, I., Fritz, S. and Boichard, D. (2013) Genet. Sel. Evol. 45:19

Villumsen T.M., Janss, L. and Lund, M.S. (2008). J. Anim. Breed. Genet. 126: 3-13.

Zhou, X., Carbonetto, P. and Stephens, M. (2013). PLoS Genet 9: e1003264. doi:10.1371/journal.pgen.1003264

Zou, H. and Hastie, T. (2005). Royal Stat. Soc. Series B 67:301-320

Figure 1 : weighted correlation between GEBV and DYD per trait, for the Holstein breed



Production traits: Fat%, Protein%, Milk yield, Protein Yield, Fat Yield Type traits: Udder depth, Angularity, Height, Udder Cleft, Rump Width, Teat Distance, Fore Udder, Teat length, Body Depth, Teat Placement, Temperament, Rear Udder Height, Chest Width, Rear leg Rear view, Milking Speed, Udder Balance, Leg Set, Foot Angle Locomotion, Body Condition Functional traits: Somatic cell score, Calving ease direct, Calving ease maternal, Stillbirth maternal, Mastitis occurrence, Stillbirth direct, Cow conception rate, Interval calving first AI, Longevity, heifer Conception rate