Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

Dissection of genomic correlation matrices of US Holsteins using multivariate factor analysis

N.P.P. Macciotta[1], C. Dimauro[1], D.J. Null[2], G. Gaspa[1], M. Cellesi[1] & J.B. Cole[2]

1 Dipartimento di Agraria, Sezione Scienze Zootecniche, Università di Sassari, Sassari, Italy
2 Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland, USA

Keywords
Animal breeding; chromosome; genomic selection.

Correspondence
N.P.P. Macciotta, Dipartimento di Agraria, Sezione Scienze Zootecniche, Università di Sassari, Via de Nicola 9, Sassari 07100, Italy.
Tel: 39 079 229299;
Fax: 39 079 229302;
E-mail: macciott@uniss.it

Received: 24 February 2014;
accepted: 3 July 2014

Summary

The aim of this study was to compare correlation matrices between direct genomic predictions for 31 traits at the genomic and chromosomal levels in US Holstein bulls. Multivariate factor analysis carried out at the genome level identified seven factors associated with conformation, longevity, yield, feet and legs, fat and protein content traits. Some differences were found at the chromosome level; variations in covariance structure on BTA 6, 14, 18 and 20 were interpreted as evidence of segregating QTL for different groups of traits. For example, milk yield and composition tended to join in a single factor on BTA 14, which is known to harbour the *DGAT1* locus that affects these traits. Another example was on BTA 18, where a factor strongly correlated with sire calving ease and conformation traits was identified. It is known that in US Holstein, there is a segregating QTL on BTA18 influencing these traits. Moreover, a possible candidate gene for daughter pregnancy rate was suggested for BTA28. The methodology proposed in this study could be used to identify individual chromosomes, which have covariance structures that differ from the overall (whole genome) covariance structure. Such differences can be difficult to detect when a large number of traits are evaluated, and covariances may be affected by QTL that do not have large allele substitution effects.

Introduction

High-throughput marker platforms are the fundamental tools of the genomic (r)evolution that has caused major changes in dairy cattle breeding over the last 5 years. Cattle are currently genotyped in many countries using SNP chips with different densities (VanRaden *et al.* 2011). Marker data are used both for predicting the genetic merit of individuals and for performing genome-wide association studies aimed at identifying genomic regions that control the expression of traits of economic importance.

Different methods are used to predict genomic estimated breeding values (GEBV), which include direct genomic values (DGV) that are calculated as the sum of genotype*SNP effects on the trait across the whole animal genome, as well as information from conventional genetic evaluation. Direct chromosomal values (DCV) can be computed by summing the genotype*SNP marker effects separately by each chromosome, and the sum of the DCV is the DGV. The DCV may be useful for developing mating plans (Cole & Null 2013). However, they also can be used to compute genomic correlation matrices for individual chromosomes (G_CHR) as well as the whole genome (G_GEN). The G_GEN matrix summarizes relationships between traits averaged across the whole genome, while G_CHR depicts the relationships at a local level.

Genetic relationships between traits are the result of the pleiotropic effects of segregating alleles (Mezey & Houle 2003). Structural differences between

  doi:10.1111/jbg.12113

G_GEN and G_CHR or between different G_CHRs may therefore indicate differences in the genetic mechanisms controlling groups of traits due, for example, to segregating QTLs. For example, Cole *et al.* (2009) reported differences in the correlations between sire calving ease and conformation traits when comparing G_GEN to G_CHR for BTA 18 in US Holsteins. This result confirmed the detection of a segregating QTL in US Holsteins on BTA18 affecting reproductive and type traits, reported also by other authors (Qanbari *et al.* 2011).

A key issue when comparing two correlation matrices is the choice of a suitable methodology for performing the analysis. A matrix has several structural elements that cannot be summarized into a single metric. Moreover, genetic correlation matrices are often singular, with rank equal to the number of genetically independent traits (Hine & Blows 2006). Several approaches to compare G matrices have been proposed, even though none of them seems to be widely accepted (Steppan *et al.* 2002). One of the most popular is the common principal component (CPC) method (Flury 1984). It relies on the assumption that, if two matrices are similar, they share one or more eigenvectors, and similarity is measured as the number of principal components, two matrices have in common. The CPC method relies on principal component analysis, which is a technique mainly used to explain the variance of a system. However, when comparing matrices to find differences in the genetic control of groups of traits, the covariances between variables are of greatest interest.

Multivariate factor analysis (MFA) is a statistical technique, particularly suitable for investigating the correlation structure of complex systems. It has been suggested as a tool for making biologically relevant comparisons among matrices (Houle *et al.* 2002). The basic theoretical assumption of MFA is that the (co) variance of a multivariate system can be partitioned into two portions (Morrison 1976): the first is shared by all variables and it is called communality, and the second is peculiar of each variable and is named uniqueness. As a consequence of (co)variance modelling, each of the n original variables can be represented as a linear combination of p common factors that generates the common covariance between variables plus a residual specific variable (Morrison 1976).

In the case of genomic matrices, MFA can be carried out separately on G_GEN and G_CHR. Different (co)variance structures can be interpreted as differing genetic relationships between traits at the whole-genome and chromosomal levels. Such an analysis may represent a first step in the identification of differences in genetic architecture among groups of traits. In this work, multivariate factor analysis is used to dissect the structure of different genomic correlation matrices in US Holsteins.

Materials and methods

Direct genomic and chromosomal values for 31 production, functional and conformation traits were calculated for 182 233 Holstein bulls and cows using the SNP effects estimated in May 2012 by the US genomic evaluation system as described in Wiggans *et al.* (2011). Direct genomic values for each chromosome were obtained by summing the effects for only the SNP markers on that chromosome, and all SNP effects were summed to obtain an animal's overall DGV. The traits included in the analysis are listed in Table 1, together with the corresponding means and standard deviations of the DGVs.

Table 1 Means and standard deviations (SD) of direct genomic values (DGV) for the 31 production, fitness, and conformation traits used to construct chromosomal and genomic correlation matrices

Trait	Mean	SD
Milk yield (kg)	222	302
Fat yield (kg)	11.9	11.7
Protein yield (kg)	8.6	8.6
Fat percentage (%)	0.03	0.09
Protein percentage (%)	0.02	0.04
Productive life (d)	1.93	2.22
Net merit ($)	295	224
Somatic cell score	2.87	0.16
Daughter pregnancy rate (%)	−0.07	1.19
Sire calving ease (%)	7.6	1.4
Daughter calving ease (%)	7.3	1.4
Sire stillbirth (%)	7.8	0.78
Daughter stillbirth (%)	7.3	1.3
Final score	1.38	1.07
Stature	1.13	1.21
Strength	0.60	0.93
Dairy form	0.99	1.15
Foot angle	1.12	1.07
Rear legs (side view)	−0.10	0.91
Body depth	0.71	0.99
Rump angle	0.19	0.96
Rump width	0.79	1.01
Fore udder attach	1.41	1.25
Rear udder height	1.70	1.36
Udder depth	1.04	1.14
Udder cleft	0.97	1.10
Front teat placement	0.70	0.99
Teat length	0.02	0.96
Rear legs (rear view)	1.08	1.04
Feet and legs	1.25	1.03
Rear teat placement	0.69	1.06

The G_GEN and G_CHR matrices were then calculated using the DGV for the 31 traits. The suitability of genomic correlation matrices to factor analysis was evaluated using the Kaiser measure of sampling adequacy (MSA). This index compares Pearson's and partial correlations. An empirical threshold of 0.8 is considered as the optimum value to consider a data set suitable for factor analysis (Cerny & Kaiser 1977).

Multivariate factor analysis was then carried out on both G_GEN and the different G_CHRs, separately for each correlation matrix using the maximum likelihood method implemented in the FACTOR procedure of SAS version 9.2 (2008). Factors were rotated using a VARIMAX procedure, and the number of extracted variables was assessed by considering their eigenvalue (only factors with eigenvalue >1 were retained). The interpretation of the extracted factors was assessed by examining the factor loadings, that is, correlations between factors and original variables (in this case, the 31 considered traits). A minimum threshold of 0.60 was assumed for a loading to be considered 'large'. A statistical test was performed to test the salience of each loading, that is, if it was significantly >0.60.

Comparisons were carried out based on the following outputs of MFA: (i) factor pattern, that is, the correlations between extracted common factors and the 31 considered traits; (ii) the variance explained by each extracted factor; and (iii) communalities, that is, the amount of variance of each trait which is explained by the common factors. A popular method for comparing observed (y) and model-predicted (x) values is by the linear regression of y on x. The slope is interpreted as an indicator of bias (it should not be different from 1 if the two variables are equal) and the intercept is related to systematic error (it should not be different from 0). In this analysis, variables considered in the regression were communalities of each original variable. Values referred to the G_GEN were considered as y, whereas corresponding values derived from the different G_CHRs were considered as x, respectively.

Results

Statistics of factors extracted from G_GEN (Table S1) and G_CHROM matrices are reported in Table 2. The Kaiser MSA for G_GEN (0.80) indicates that the partial correlations between the variables are small compared with Pearson's correlations and that the common factor model is appropriate to these data (Morrison 1976; Cerny & Kaiser 1977). The seven extracted factors were able to explain a large part (approximately 0.70) of the variance.

Table 2 Statistics of factor extraction

	Factors (n.)	Variance explained	Kaiser MSA
Genome	7	0.69	0.80
BTA1	8	0.69	0.67
BTA2	7	0.60	0.68
BTA3	8	0.67	0.66
BTA4	7	0.66	0.67
BTA5	7	0.80	0.77
BTA6	7	0.69	0.72
BTA7	7	0.67	0.72
BTA8	8	0.72	0.70
BTA9	7	0.68	0.68
BTA10	8	0.73	0.76
BTA11	7	0.69	0.73
BTA12	7	0.61	0.68
BTA13	8	0.68	0.67
BTA14	6	0.67	0.74
BTA15	7	0.58	0.66
BTA16	7	0.68	0.68
BTA17	7	0.65	0.65
BTA18	7	0.76	0.75
BTA19	7	0.70	0.73
BTA20	8	0.69	0.72
BTA21	7	0.63	0.66
BTA22	7	0.67	0.72
BTA23	8	0.69	0.71
BTA24	8	0.71	0.68
BTA25	8	0.77	0.72
BTA26	7	0.77	0.76
BTA27	7	0.62	0.65
BTA28	7	0.68	0.74
BTA29	8	0.71	0.70

Factors extracted from the G_GEN showed a quite readable structure (Table 3), with traits loading onto factors that appear to be functionally related. Each factor had a few large correlations (i.e. significantly larger than 0.60, with $p \leq 0.01$) with considered traits and several rather small loadings. The same conclusions may be drawn if the table is observed across columns: each trait had a large correlation with just one factor and small correlations with the other factors. An exception was represented by fat yield that showed correlations >0.60 with both factors 3 and 6. The first factor (Table 3), explaining approximately 26% of the total variance of the system, was mainly correlated with conformation traits (body size and shape, and udder conformation). The second factor explained approximately half of the variance explained by the first and could be considered as an indicator of longevity, being related to survival traits, SCS, and daughter pregnancy rate. The third factor

Table 3 Factor pattern of the correlation matrix between direct genomic values for 31 production, conformation and functional traits

Trait	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Milk	0.29	0.14	**0.89**	0.02	0.03	−0.17	−0.28
Fat	0.30	0.20	**0.66**	0.06	0.04	**0.65**	0.03
Protein	0.31	0.23	**0.90**	0.05	0.03	−0.03	0.20
Fat percentage	0.04	0.07	−0.20	0.04	0.01	**0.92**	0.33
Protein percentage	0.01	0.14	−0.09	0.05	−0.01	0.29	**0.94**
Net merit	0.36	**0.75**	0.47	0.13	−0.04	0.24	0.07
Productive life	0.22	**0.92**	0.10	0.09	−0.10	0.04	−0.02
Somatic cell score	−0.16	**−0.64**	0.11	−0.09	−0.07	−0.11	0.04
Daughter pregnancy rate	−0.22	**0.71**	−0.30	0.03	0.00	−0.09	0.10
Sire calving ease	0.13	−0.42	−0.16	0.01	0.19	−0.01	−0.05
Daughter calving ease	−0.26	−0.48	−0.17	−0.08	0.03	−0.02	0.00
Sire stillbirth	0.13	−0.33	−0.05	0.00	0.09	0.02	−0.04
Daughter stillbirth	−0.15	−0.40	−0.13	−0.07	−0.01	0.00	0.01
Final score	**0.93**	0.09	0.12	0.23	0.24	0.07	0.01
Stature	**0.72**	−0.17	0.09	0.22	0.46	0.02	0.04
Strength	0.41	−0.12	0.08	0.26	**0.86**	0.04	0.05
Dairy form	**0.75**	−0.29	0.34	0.04	0.00	0.10	−0.06
Foot angle	0.52	0.08	0.05	**0.69**	0.27	0.04	0.06
Rear legs (side view)	0.24	−0.14	0.06	−0.58	−0.13	0.02	0.01
Body depth	0.58	−0.28	0.14	0.20	**0.67**	0.08	0.01
Rump angle	−0.06	0.02	0.11	−0.02	0.08	−0.02	−0.06
Rump width	**0.65**	−0.14	0.11	0.11	0.50	0.04	0.04
Fore udder attachment	**0.85**	0.27	−0.06	0.11	0.17	0.06	−0.01
Rear udder height	**0.88**	0.11	0.15	0.16	0.08	0.06	−0.02
Udder depth	**0.73**	0.34	−0.21	0.08	0.09	0.00	0.03
Udder cleft	**0.81**	0.02	0.09	0.06	0.06	0.01	0.00
Front teat placement	**0.63**	0.16	0.14	−0.03	0.04	0.03	0.02
Teat length	0.00	−0.24	−0.03	0.10	0.24	−0.04	−0.06
Rear legs (rear view)	0.53	0.10	0.07	**0.76**	0.11	0.06	0.04
Feet and legs	**0.65**	0.13	0.07	**0.73**	0.05	0.07	0.05
Rear teat placement	**0.62**	0.01	0.14	−0.04	0.00	0.01	0.01
Variance explained (%)	0.26	0.12	0.09	0.07	0.06	0.05	0.04

Values in bold are significantly higher than 0.60 ($p \leq 0.01$).

was related to yield traits, whereas the fourth showed larger correlation with specific traits of feet and legs. The fifth factor could be interpreted as an indicator of body shape. The final two factors were related to milk composition traits: the sixth is a fat indicator (both for yield and for composition), and the seventh is related to protein content. Such a structure summarizes quite closely the pattern observed in the G_GEN (Table S1), by including in the same factor traits that exhibit large correlations between them.

Of the 31 traits considered, some showed no relationship with the latent factors (Table 3). One group was represented by traits related to calving ease and stillbirth, both for sires and for daughters. Others were morphology measurements of teat, rump and legs. Actually, the salience was related to the communality of variables (Table 4), that is, the amount of variability of each trait that is generated by the common factors. Traits that did not show any relationship with extracted factors were those characterized by the lowest communality (usually lower than 0.30, except for rear leg (side view), which showed loadings closer to the fixed threshold of 0.60). Moreover, these traits did not show large correlations with other traits in the G_GEN (Table S1).

The MFA carried out on single chromosomes showed, as expected, some differences as compared to genome-wide results. The Kaiser MSA (Table 2) was generally lower than the value obtained for the G_GEN. The largest observed values were for BTAs 5, 10 and 26. However, the lowest values (0.65) were not too far from the empirical threshold of 0.80. The total amount of variance explained by the different factors was on average 0.69 ($\pm$0.05), with the lowest and highest values for BTA15 and BTA2, respectively. Moreover, differences between G_GEN and G_CHROM were noted in their distribution across factors. For example, Figure 1 reports the pattern of

Table 4 Communalities of genomic predictions at genome-wide level and statistics of communalities by chromosome

NAME	Whole genome	Average	SD	Maximum	Minimum
Milk	1.00	1.00	0.00	1.00	0.99
Fat	1.00	1.00	0.01	1.00	0.97
Protein	1.00	1.00	0.00	1.00	0.99
Fat percentage	1.00	1.00	0.01	1.00	0.98
Protein percentage	1.00	0.99	0.01	1.00	0.97
Nett merit	0.99	0.96	0.05	1.00	0.79
Productive life	0.92	0.83	0.13	0.98	0.49
Somatic cell score	0.47	0.50	0.14	0.75	0.21
Daughter pregnancy rate	0.67	0.56	0.13	0.82	0.28
Sire calving ease	0.26	0.25	0.14	0.72	0.08
Daughter calving ease	0.33	0.27	0.10	0.53	0.04
Sire stillbirth	0.14	0.24	0.13	0.65	0.07
Daughter stillbirth	0.21	0.28	0.10	0.46	0.08
Final score	1.00	0.93	0.08	1.00	0.58
Stature	0.81	0.67	0.16	0.86	0.09
Strength	1.00	0.81	0.22	1.00	0.05
Dairy form	0.78	0.66	0.19	0.99	0.33
Foot angle	0.83	0.75	0.12	0.92	0.37
Rear legs (side view)	0.43	0.46	0.14	0.71	0.08
Body depth	0.93	0.83	0.21	1.00	0.05
Rump angle	0.03	0.19	0.08	0.37	0.02
Rump width	0.71	0.57	0.16	0.80	0.08
Fore udder attachment	0.84	0.83	0.14	1.00	0.31
Rear udder height	0.85	0.67	0.11	0.81	0.27
Udder depth	0.71	0.73	0.13	0.91	0.37
Udder cleft	0.67	0.66	0.15	0.93	0.30
Front teat placement	0.45	0.60	0.21	1.00	0.28
Teat length	0.13	0.27	0.14	0.56	0.07
Rear legs (rear view)	0.90	0.83	0.12	0.95	0.44
Feet and legs	1.00	0.93	0.13	1.00	0.48
Rear teat placement	0.41	0.64	0.27	0.99	0.19

variance explained by the different factors extracted from both G_GEN and G_CHROM for BTAs 6, 14, 18 and 20. A large reduction in explained variance when moving from the first to the subsequent factors was observed for the G_GEN, with the first factor explaining approximately 2.5 times as much variance as the second factor. While the amount of explained variance decreased with factor number for individual chromosomes, the magnitude was much smaller, especially for BTA 6.

The number of extracted factors by chromosome was very close to that of the G_GEN, ranging from 6 to 8. Their general structure was similar to G_GEN, but specific variations in their pattern have been detected. The communalities of the 31 traits calculated for each chromosome also had similar patterns to the genome-wide matrix (the correlation between communalities calculated from the G_GEN and those averaged by the 29 autosomes was 0.96) (Table 4). However, some traits exhibited large variation in communality among chromosomes. Examples include strength or body weight that ranged from 0.05 (both on BTA1) to 1.00 (on BTA7 and BTA6, respectively). In general, conformation and functional traits were characterized by the largest variation in communality among chromosomes.

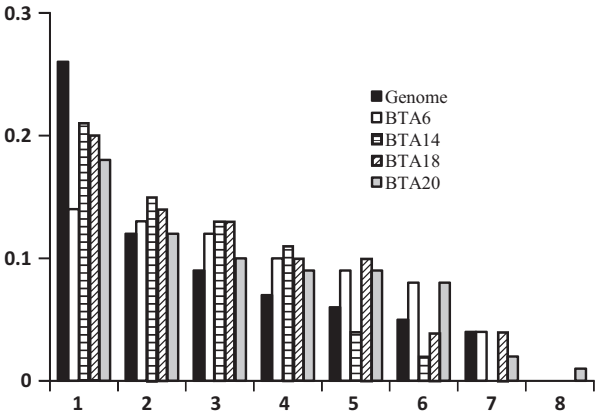

Figure 1 Pattern of explained variance of factors extracted from the genomic and some chromosomal correlation matrices.

Although analyses were performed along the whole genome, to validate the MFA approach, a more detailed examination of results was carried out on four chromosomes known to harbour genes affecting milk production and conformation traits (i.e. BTA 6, 14, 18 and 20) (Grisart *et al.* 2002; Cole *et al.* 2009; Flori *et al.* 2009; Chamberlain *et al.* 2012). Relevant results obtained for other chromosomes are presented in the paper and reported in the supporting information.

The largest extracted factor in terms of explained variance for BTA 6 (Table 5) is similar to the longevity factor of the G_GEN (Table 3), with the exception of a large loading for daughter stillbirth and a loading for daughter calving ease that approaches the threshold of significance. A QTL associated with calving difficulty on this chromosome has been reported for Norwegian Red cattle (Olsen *et al.* 2009), and a genomic region on the same chromosome affecting calving ease in the Piedmontese beef breed has been identified (Bongiorni *et al.* 2012). Some putative candidate

genes related to pelvic morphology, including leucine aminopeptidase (*LAP3)* and ligand-dependent nuclear receptor corepressor-like (*LCORL)*, have been mapped to BTA6 (Flori *et al.* 2009). Large SNP effects on this chromosome have been detected in the US Holstein for daughter pregnancy rate, heifer conception rate and somatic cell score (Cole & VanRaden 2010). Another relevant difference in comparison with the G_GEN could be found on factor 6 (Table 5), which is unfavourably related to milk yield (with a negative sign) and favourably associated with fat and protein percentage. It is widely known that BTA6 harbours several genes involved in milk yield and composition in a group that maps at around 37 Mbp including *FAM13B1*, *SPP1* and *ABCG2* and the casein cluster. As was the case with G_GEN, sire calving traits, rump angle and some teat measures did not load significantly onto any of the extracted factors.

As expected, BTA14 exhibited some variation in comparison with G_GEN as far as milk production traits are concerned (Table 6). The second factor was associated with both yield and composition traits that were associated with different factors (3, 6 and 7) in the genome-wide matrix (Table 3). It is of interest to note that the correlation of fat yield with factor 2 of BTA14 was of a different sign compared with the other yield traits, while it was of the same sign for percentage traits (Table 6). It is known that the *DGAT1* gene maps to this chromosome. The pattern of correlation signs for factor 2 was the same reported for the substitution effects of the K232A mutation on these traits (Grisart *et al.* 2002). It is also of interest to note that protein yield had a correlation slightly lower than the threshold of significance on factor 2, but it showed a large loading on factor 5. Some studies have suggested the existence of a second QTL affecting milk

Table 5 Factor pattern of the correlation matrix between direct chromosomal values for 31 production, conformation and functional traits for BTA6

Trait	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Milk	−0.11	0.06	0.02	**0.64**	−0.01	**−0.72**	0.19
Fat	−0.04	0.14	0.03	0.59	−0.09	−0.01	**0.79**
Protein	0.08	0.07	0.09	**0.99**	−0.05	0.01	0.06
Fat percentage	0.08	0.05	0.00	−0.21	−0.08	**0.86**	0.45
Protein percentage	0.19	−0.02	0.04	0.14	−0.04	**0.95**	−0.17
Net merit	**0.88**	0.07	0.28	0.21	0.09	0.15	0.25
Productive life	**0.90**	0.02	0.16	−0.26	0.13	0.16	−0.02
Somatic cell score	**−0.75**	−0.22	−0.19	0.28	−0.01	−0.15	−0.01
Daughter pregnancy rate	**0.70**	−0.09	−0.09	−0.39	0.06	0.22	−0.15
Sire calving ease	−0.29	0.36	0.19	−0.09	0.08	0.11	0.06
Daughter calving ease	−0.57	0.09	−0.09	−0.06	0.02	0.02	0.02
Sire stillbirth	−0.28	0.28	0.16	0.07	0.07	0.08	0.05
Daughter stillbirth	**−0.65**	−0.01	0.13	−0.05	0.03	0.01	0.00
Final score	0.13	0.53	**0.68**	0.20	0.34	0.02	0.13
Stature	0.14	**0.72**	0.30	0.14	0.30	0.02	0.01
Strength	0.00	**0.85**	0.09	−0.09	0.31	−0.03	0.01
Dairy form	−0.34	0.20	0.13	**0.64**	−0.11	−0.12	0.08
Foot angle	0.17	0.40	0.43	−0.07	**0.67**	0.07	−0.04
Rear legs (side view)	0.09	−0.13	0.17	0.21	**−0.65**	0.08	0.02
Body depth	−0.19	**0.93**	0.08	0.23	0.17	−0.07	0.03
Rump angle	0.01	−0.01	−0.39	0.05	0.08	0.02	−0.06
Rump width	0.10	**0.68**	0.30	0.17	0.19	−0.04	0.02
Fore udder attachment	0.34	0.15	**0.86**	−0.04	0.13	0.09	0.06
Rear udder height	−0.07	0.21	0.51	0.22	0.25	−0.12	0.23
Udder depth	0.54	0.07	**0.67**	−0.26	0.12	0.20	0.00
Udder cleft	0.14	0.37	**0.60**	−0.01	0.15	0.00	−0.01
Front teat placement	0.00	0.14	0.51	0.22	0.05	0.02	−0.09
Teat length	−0.27	0.21	−0.14	−0.13	0.12	0.00	−0.02
Rear legs (rear view)	−0.03	0.37	0.18	0.02	**0.86**	−0.05	0.00
Feet and legs	0.09	0.35	0.31	0.06	**0.85**	0.02	−0.04
Rear teat placement	−0.13	0.06	0.44	0.19	0.08	0.00	−0.20
Variance explained (%)	0.14	0.13	0.12	0.10	0.09	0.08	0.04

Values in bold are significantly higher than 0.60 ($p \leq 0.01$).

protein yield and percentage located on BTA14 (Schnabel *et al.* 2005; Cole *et al.* 2011), and it is known that the effect of *DGAT1* on fat and protein is different (Tetens *et al.* 2012).

An additional peculiarity of BTA14 found in this study was the splitting of the factor associated with conformation traits into two latent variables related to udders and feet and legs (the first) and to the size of the animals (the third), respectively (Table 6). The US Holstein population has large marker effects on this chromosome for strength and udder cleft (Cole & VanRaden 2010). An effect of *DGAT1* on rump width and strength has been reported in German Holsteins (Kaupe *et al.* 2007), a QTL related to rump width has been mapped in the US Holstein population (Schnabel *et al.* 2005), and a QTL influencing growth traits has been found in Fleckvieh cattle (Pausch *et al.* 2011).

The results from BTA18 showed relevant variation compared with the genome-wide pattern as far as factor 1 is concerned (Table 7). This variable was strongly correlated with sire calving and conformation traits. As mentioned in the introduction, a QTL affecting sire calving ease and stillbirth and conformation traits was reported in the US (Cole *et al.* 2009) and German (Brand *et al.* 2010) Holstein populations. The maternally imprinted *PG3* domain, a mutation that has recently been associated with the expression of the *MIMT1* protein, affects abortion and stillbirth in Finnish Ayrshire cattle (Flisikowski *et al.* 2010). Cole *et al.* (2014) also have recently reported an association between calf birthweight and a sialic acid-binding immunoglobulin-type lectin that maps on BTA18. This result further supports the role of this putative QTL in influencing body size and shape.

Finally, BTA20 also exhibited some peculiarities in comparison with the G_GEN matrix (Table 8). There was a division of factors related to conformation into one associated with mammary traits (the first) and the second to the animal size (Table 8), which is similar to results observed for BTA14. There was also a

Table 6 Factor pattern of the correlation matrix between direct chromosomal values for 31 production, conformation and functional traits for BTA14

BTA14	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Milk	0.02	**−0.90**	0.01	0.10	0.34	0.27
Fat	0.24	**0.94**	0.09	0.03	0.03	0.20
Protein	−0.06	−0.58	0.02	0.12	**0.80**	0.04
Fat percentage	0.13	**0.98**	0.05	−0.03	−0.15	−0.01
Protein percentage	−0.07	**0.92**	0.01	−0.07	0.06	−0.38
Net merit	0.50	**0.71**	−0.10	−0.39	0.20	0.20
Productive life	0.50	0.19	−0.36	**−0.73**	0.06	0.07
Somatic cell score	−0.40	−0.30	−0.06	0.56	0.31	−0.14
Daughter pregnancy rate	−−0.16	−0.01	−0.25	**−0.62**	−0.03	−0.15
Sire calving ease	−0.14	−0.13	0.19	0.53	−0.17	−0.07
Daughter calving ease	−0.35	−0.02	−0.01	0.35	−0.24	0.00
Sire stillbirth	−0.03	−0.12	0.03	0.43	0.03	0.12
Daughter stillbirth	−0.22	−0.18	0.28	0.36	−0.26	0.27
Final score	**0.89**	0.10	0.42	0.10	0.05	−0.03
Stature	0.28	−0.01	**0.73**	0.18	0.03	0.05
Strength	0.13	0.01	**0.99**	−0.01	0.01	0.03
Dairy form	0.42	0.04	0.29	**0.63**	0.22	0.14
Foot angle	0.53	−0.03	0.45	−0.22	−0.02	0.07
Rear legs (side view)	−0.10	0.23	−0.09	0.49	0.10	−0.09
Body depth	0.21	0.12	**0.89**	0.24	0.05	0.11
Rump angle	−0.39	−0.06	0.03	−0.01	−0.05	−0.01
Rump width	0.27	0.03	**0.81**	0.24	−0.12	0.04
Fore udder attachment	**0.82**	0.19	0.15	−0.15	−0.15	−0.14
Rear udder height	**0.85**	−0.10	0.03	0.01	0.06	−0.05
Udder depth	**0.65**	0.21	−0.07	−0.33	−0.25	−0.16
Udder cleft	**0.71**	−0.02	0.28	0.06	−0.12	0.16
Front teat placement	**0.67**	0.12	0.18	−0.05	0.03	−0.18
Teat length	−0.09	−0.18	0.26	0.10	0.04	0.38
Rear legs (rear view)	0.54	0.09	0.25	−0.29	0.14	0.07
Feet and legs	**0.69**	0.15	0.14	−0.27	0.13	0.13
Rear teat placement	**0.69**	0.00	0.13	0.04	−0.07	0.02
Variance explained (%)	0.21	0.15	0.13	0.11	0.04	0.02

Values in bold are significantly higher than 0.60 ($p \leq 0.01$).

Table 7 Factor pattern of the correlation matrix between direct chromosomal values for 31 production, conformation and functional traits for BTA18

BTA18	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Milk	−0.08	0.04	0.05	0.01	**0.95**	−0.20	−0.20
Fat	0.18	−0.09	−0.01	−0.07	**0.80**	0.55	0.01
Protein	−0.16	0.05	0.13	−0.04	**0.94**	−0.09	0.26
Fat percentage	0.31	−0.15	−0.07	−0.10	−0.10	**0.89**	0.23
Protein percentage	−0.17	0.01	0.18	−0.11	0.05	0.20	**0.94**
Net merit	−0.31	0.13	**0.78**	0.22	0.46	0.09	0.11
Productive life	−0.41	0.13	**0.83**	0.26	0.15	−0.03	0.09
Somatic cell score	0.00	−0.12	**−0.71**	−0.10	0.01	0.00	0.00
Daughter pregnancy rate	−0.22	−0.08	**0.82**	0.25	−0.13	0.00	0.06
Sire calving ease	**0.72**	0.01	−0.41	0.12	−0.04	0.12	−0.01
Daughter calving ease	0.46	−0.09	−0.50	0.09	−0.17	0.08	−0.12
Sire stillbirth	**0.69**	0.11	−0.31	0.19	−0.01	0.17	−0.02
Daughter stillbirth	0.38	0.08	−0.44	−0.07	−0.26	0.08	−0.10
Final score	0.47	**0.69**	0.22	0.46	0.02	0.00	−0.08
Stature	**0.83**	0.21	−0.03	0.19	−0.06	−0.05	−0.11
Strength	**0.96**	0.01	−0.08	0.09	−0.03	0.09	0.02
Dairy form	0.34	0.37	−0.36	0.11	0.28	−0.06	−0.22
Foot angle	0.52	0.32	0.10	**0.67**	−0.10	−0.03	0.00
Rear legs (side view)	−0.03	−0.04	−0.10	−0.46	−0.03	0.08	0.07
Body depth	**0.93**	0.05	−0.23	0.09	−0.01	0.10	−0.06
Rump angle	−0.37	0.05	−0.14	−0.24	0.11	−0.23	−0.05
Rump width	**0.84**	0.21	−0.07	0.17	−0.05	0.05	−0.06
Fore udder attachment	0.30	**0.67**	0.44	0.33	−0.10	0.01	−0.01
Rear udder height	0.04	**0.71**	0.22	0.36	0.07	−0.12	−0.08
Udder depth	0.17	0.51	0.51	0.27	−0.25	−0.08	−0.11
Udder cleft	0.04	**0.85**	0.14	0.18	0.03	−0.09	0.06
Front teat placement	0.01	**0.81**	−0.06	0.04	−0.02	0.00	0.00
Teat length	0.44	−0.27	−0.13	−0.17	0.17	−0.05	−0.10
Rear legs (rear view)	0.29	0.26	0.16	**0.84**	−0.01	0.11	−0.03
Feet and legs	0.13	0.33	0.16	**0.91**	−0.04	0.00	0.01
Rear teat placement	−0.03	**0.84**	−0.17	0.00	0.07	0.00	0.04
Variance explained (%)	0.20	0.14	0.13	0.10	0.10	0.04	0.04

Values in bold are significantly higher than 0.60 ($p \leq 0.01$).

factor related to both milk yield and composition (factor 5), and the US population has a strong signal for protein percentage on BTA20 (Cole & VanRaden 2010). A number of SNP associations with milk production traits have also been reported by other groups (Blott *et al.* 2003; Chamberlain *et al.* 2012), and BTA20 harbours some interesting candidate genes for milk production traits, such as the growth hormone receptor (*GHR*; Blott *et al.* 2003) and the prolactin receptor (*PRLR*). Somatic cell score was not included in the factor associated with longevity, and no reports were found in the literature about genomic regions that affect SCS located on this chromosome, but Sodeland *et al.* (2011) did identify a QTL affecting clinical mastitis in Norwegian Red cattle.

The comparisons discussed above were based on visual inspection of factor patterns, evaluating the correspondence of loadings statistically larger than 0.6 between the different factors. However, a more empirical approach may be desirable, particularly as the number of traits continues to grow. Table 9 reports results of regression analyses that compare communalities of different traits estimated by analysing either the whole-genome or chromosomal matrices, respectively. It can clearly be seen that all comparisons differed significantly from expectations; the intercept was always different from zero and the slope from one. Regression models were also used to compare communalities of the G_GEN with those obtained from the G_CHROM of BTA3, which exhibited a factorial pattern similar to the genome wide (data not reported for brevity). In this case, the intercept was not different from zero or the slope from one. The BTA3 results are important because they confirm that intercepts and slopes are consistent with expectations when the whole-genome and chromosome-specific matrices have similar covariance structures.

As far as the other chromosomes are concerned, a difference from genome-wide results was detected on

Table 8 Factor pattern of the correlation matrix between direct chromosomal values for 31 production, conformation and functional traits for BTA20

BTA20	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Milk	−0.16	−0.10	−0.27	−0.04	**−0.76**	0.55	0.12	0.01
Fat	0.13	0.18	−0.22	−0.08	0.31	**0.78**	0.43	−0.04
Protein	0.00	0.25	−0.08	0.00	0.04	**0.95**	−0.19	0.01
Fat percentage	0.23	0.23	0.07	−0.03	**0.90**	0.13	0.23	−0.04
Protein percentage	0.15	0.32	0.23	0.04	**0.84**	0.17	−0.28	0.00
Net merit	0.39	−0.03	0.56	0.26	0.26	0.59	0.13	−0.02
Productive life	0.39	−0.24	**0.80**	0.32	0.04	0.14	0.06	−0.01
Somatic cell score	0.18	0.18	−0.47	−0.26	−0.03	0.06	0.12	−0.03
Daughter pregnancy rate	0.15	−0.23	**0.63**	0.08	0.14	−0.13	−0.03	0.00
Sire calving ease	−0.43	0.25	0.03	−0.03	−0.14	0.19	−0.06	−0.01
Daughter calving ease	−0.07	0.10	0.06	−0.19	−0.43	−0.03	−0.05	0.00
Sire stillbirth	−0.10	0.44	−0.13	0.01	0.07	0.03	−0.05	−0.02
Daughter stillbirth	−0.01	0.17	−0.14	0.06	−0.40	−0.29	−0.04	−0.02
Final score	**0.69**	0.48	0.24	0.40	0.15	0.14	0.06	−0.04
Stature	0.11	**0.82**	0.20	−0.03	0.04	0.09	−0.04	−0.07
Strength	0.10	0.51	0.02	0.20	−0.03	−0.01	−0.05	0.35
Dairy form	0.17	0.53	−0.47	0.09	−0.06	0.27	0.11	−0.13
Foot angle	0.23	0.26	0.30	**0.66**	−0.05	0.12	−0.10	0.01
Rear legs (side view)	0.15	0.38	−0.08	−0.32	0.09	0.14	0.29	−0.07
Body depth	0.14	**0.65**	−0.26	0.18	−0.08	0.08	0.06	0.11
Rump angle	0.11	0.05	−0.17	−0.33	−0.03	0.12	−0.08	−0.01
Rump width	0.14	**0.61**	−0.23	0.20	0.11	−0.03	0.09	−0.01
Fore udder attachment	**0.76**	0.25	0.46	0.19	0.14	0.02	0.06	−0.03
Rear udder height	**0.63**	0.41	0.30	0.32	0.16	0.16	0.07	−0.03
Udder depth	0.52	0.28	**0.66**	0.09	0.21	−0.05	0.04	−0.08
Udder cleft	**0.76**	0.27	0.21	0.21	0.14	−0.03	0.00	−0.02
Front teat placement	**0.98**	0.00	−0.10	0.08	0.10	0.04	−0.02	0.03
Teat length	**−0.67**	0.07	−0.10	0.07	0.01	−0.23	−0.10	−0.02
Rear legs (rear view)	0.24	0.14	−0.01	**0.88**	0.12	0.00	−0.06	0.05
Feet and legs	0.34	0.28	0.13	**0.83**	0.11	0.07	0.00	−0.02
Rear teat placement	**0.89**	0.11	−0.05	0.07	0.08	−0.01	−0.10	0.02
Variance explained (%)	0.18	0.12	0.10	0.09	0.09	0.08	0.02	0.01

Values in bold are significantly higher than 0.60 ($p \leq 0.01$).

factor pattern extracted from G_CHROM of BTA5 (Table S2). The yield factor showed large correlations only for milk and protein, while fat yield had a large loading in the same factor as fat percentage. The US Holstein population has large SNP effects on BTA5 for milk, fat, and protein yields and fat percentage (Cole & VanRaden 2010). QTLs affecting milk fat content located on BTA5 were reported for German (Wang *et al.* 2012) and Australian (Hayes *et al.* 2010; Raven *et al.* 2014) Holsteins. Epidermal growth factor receptor pathway substrate 8 (EPS8), a gene involved in the fat metabolism of mammals, has been suggested as a candidate gene for that QTL region. Moreover, a QTL affecting milk, protein and fat yield was reported on BTA5 for the Fleckvieh breed (Awad *et al.* 2011).

On BTA11 (Table S3), protein percentage exhibited large loadings both in factor 4, mainly associated with measures of longevity, and in factor 7, with fat content. The US Holstein population has large SNP effects on BTA11 for protein and fat content (Cole & VanRaden 2010). A QTL affecting milk protein content on BTA11 has been detected in Holstein Friesians by Schopen *et al.* (2009) in a position close to the beta-lactoglobulin (BLG) gene.

A different behaviour of fat percentage, in comparison with the results obtained for the G_GEN, was observed on BTA27. The fourth factor (Table S4) showed large correlation values with milk and protein yield, and fat content, but not with fat yield. In the G_GEN (Table 3), yield and composition traits were associated with distinct factors. BTA27 has a large signal for fat percentage in the US Holstein (Cole & VanRaden 2010). Wang *et al.* (2012) reported a major QTL for fat content on this chromosome. These authors suggested the glycerol-3-phosphate acyltransferase 4 (GPAT4) as neighbouring gene for this QTL. Raven *et al.* (2014), in a multibreed study reported a SNP associated with fat content on BTA27, hypothesizing the GINS complex subunit 4 as a candidate gene.

Table 9 Regression analysis of communalities extracted from the genomic correlation matrix on those extracted from the different chromosome matrices

BTA	Intercept	p^1	Slope	p^2
6	0.32 ± 0.09	0.01	0.66 ± 0.10	0.02
14	0.30 ± 0.10	0.02	0.68 ± 0.12	0.03
18	0.51 ± 0.03	<0.001	0.48 ± 0.03	<0.001
20	0.41 ± 0.01	<0.001	0.58 ± 0.02	<0.001
3	−0.12 ± 0.13	0.390	1.12 ± 0.16	0.453

p^1 = Statistical significance of the test H0: intercept = 0; Ha: intercept $\neq 0$.
p^2 = Statistical significance of the test H0: slope = 1; Ha: slope $\neq 1$.
Test are declared statistically significant if $p < 0.05$.

Finally, on BTA28 (Table S5), daughter pregnancy rate had a large correlation in the same factor of yield traits (Factor 2). The US Holstein population exhibits large SNP effects on BTA28 for daughter pregnancy rate and heifer conception rate (Cole & VanRaden 2010). A SNP significantly associated with calving ease has been detected on BTA28 in Italian Holsteins (Minozzi *et al.* 2013). The bone morphogenetic protein receptor type 1A (BMPRA1) and the growth differentiation factor 2 (GDF2) genes are plausible candidates that could underlie the QTL effect (Pennington & Ealy 2012).

Discussion

Large correlation matrices (31 traits) of genomic breeding values were dissected using MFA. This technique was able to analyse their deep structure, extracting factors with biologically interpretable meanings. These new variables can be considered as indicators of aggregate traits as conformation, longevity, feet and legs, yield, body size, milk composition, respectively. Such a feature is of particular interest for matrix comparisons because most proposed methodologies are unable to give biological explanations of results. The basic assumption of the factorial model, that is, that the (co)variance of a multivariate system is generated by causes that may affect either one or many variables, seemed to be adequate to fit the structure of the genomic correlation matrices. This model has previously been used to generate covariance matrices that are both simple and biologically reasonable (Houle *et al.* 2002) and has been used for finding the dimension of variance–covariance matrices (Hine & Blows 2006).

As expected, differences between the genome-wide and the chromosome-wide correlation matrices of direct genomic predictions were detected. Under a geometrical perspective, basic elements of a genetic correlation matrix are (i) its orientation, which can be represented by the structure of its eigenvectors, and (ii) its length, which is related to the magnitude of its eigenvalues. Multivariate factor analysis was able to describe these two aspects of the matrices examined in the present study. In particular, the orientation was described by the factor pattern, while the length was summarized by the amount of variance explained by each factor. Differences between G_GEN and G_CHROM were found in both aspects, but most interesting were those detected in factor patterns. Biologically, latent factors may be regarded as a sort of mirror of genes or pools of genes that affect sets of traits. The clustering of traits across different latent variables followed a biologically and technically coherent pattern when genome-wide covariances were examined. Differences detected at the chromosome level involved those traits for which chromosomes were known to harbour significant genes as, for example, the behaviour of morphology and calving ease traits for BTA18. Mezey & Houle (2003) pointed out that two genetic correlation matrices are similar when they present the same modular organization, that is, when pleiotropic effects of genes are associated with the same set of traits in both matrices. If this concept is reversed, different factor patterns yielded by MFA may indicate variation in modular organization, that is, in the genetic architecture of groups of traits, of the compared matrices.

Some differences were detected among groups of traits. Milk yield and composition were associated with distinct factors at the genome-wide level, and they tended to join in chromosomes where genes affecting milk yield are located, such as BTA14. On the other hand, many morphological traits clustered in the same latent variables both at genome and chromosome level. They were also frequently associated with the first or second extracted factor, whereas milk traits had relevant loadings on the later factors in terms of explained variance. Such behaviour could be related to the genetic regulation of the two groups of traits: mainly attributable to a relatively small number of genes with a moderate effect for milk composition or due to a polygenic background for conformation traits, respectively (Hayes *et al.* 2010).

The MFA also provides an estimate of the amount of variance each variable shares with the others. The lowest communalities were obtained for rump angle, calving traits and some indicators of teat placement, while the highest values were associated with milk production traits. The uniqueness of each variable (that can be calculated as 1 – communality) expresses its specific variability, and it seems to be related to the

nature of the trait (either measured directly, or evaluated by an expert). However, variation within the same trait has been observed. The largest communalities were usually found for chromosomes where QTL or genes affecting the trait were located, such as sire calving ease and stillbirth for BTA18. Thus, the communality also yields useful information for the detection of chromosomal regions that affect a specific set of traits. Moreover, also, the pattern of variation in this parameter across chromosomes (large variability for functional and conformation traits, low for yield traits) could provide additional information about the genetic background of traits.

Finally, the proposed approach allows for a preliminary scan across the whole genome to identify regions of potential interest associated with genetic control of a group of traits using only the information that are currently produced by genomic selection programs. An example is represented by results for pregnancy rate on BTA28. Although it is quite easy to perform, being based upon routine calculations that are normally implemented in most commercial and free statistical software packages, MFA also is able to flag groups of traits that are characterized by different genetic architectures, such as milk yield, composition or conformation traits (Hayes *et al.* 2010). In the present study, the method was tested on chromosomes known to harbour some important candidate genes to check its reliability. It could be further tested on less investigated chromosomes within the same population, applied to new phenotypes or used to compare the same chromosome in different breeds.

Acknowledgements

The cooperation of the Council on Dairy Cattle Breeding (Reynoldsburg, OH) in supplying pedigree, performance and genotypic data for US dairy cattle is acknowledged. J.B. Cole and D.J. Null were supported by appropriated project 1265-31000-096-00, 'Improving Genetic Predictions in Dairy Animals Using Phenotypic and Genomic Information'. The research was also funded by the INNOVAGEN project of the Italian Ministry of Agriculture.

References

Awad A., Russ I., Forster M., Medugorac I. (2011) Mapping of a milk production quantitative trait locus to a 1.056 Mb region on bovine chromosome 5 in the Fleckvieh dual purpose cattle breed. *Genet. Sel. Evol.*, **43**, 8.

Blott S., Kim J.J., Moisio S., Schmidt-Kuntzel A., Cornet A., Berzi P., Cambisano N., Ford C., Grisart B., Johnson D., Karim L., Simon P., Snell R., Spelman R., Wong J., Vilkki J., Georges M., Farnir F., Coppieters W. (2003) Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the Bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics*, **163**, 253–266.

Bongiorni S., Mancini G., Chillemi G., Pariset L., Valentini A. (2012) Identification of a short region on chromosome 6 affecting direct calving ease in Piedmontese cattle breed. *PLoS ONE*, **7**, e50137.

Brand B., Baes C., Mayer M., Reinsch N., Seidenspinner T., Thaller G., Kühn Ch (2010) Quantitative trait loci mapping of calving and conformation traits on Bos taurus autosome 18 in the German Holstein population. *J. Dairy Sci.*, **93**, 1205–1215.

Cerny B.A., Kaiser H.F. (1977) A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behav. Res.*, **12**, 43–47.

Chamberlain A., Hayes B.J., Savin K., Bolormaa S., McPartlan H.C., Bowman P.J., Van Der Jagt C., MacEacherm S., Goddard M.E. (2012) Validation of a single nucleotide polymorphisms associated with milk production traits in dairy cattle. *J. Dairy Sci.*, **95**, 864–875.

Cole J.B., Null D.J. (2013) Visualization of the transmission of direct genomic values for paternal and maternal chromosomes for 15 traits in US Brown Swiss, Holstein, and Jersey cattle. *J. Dairy Sci.*, **96**, 2713–2726.

Cole J.B., VanRaden P.M. (2010) Visualization of results from genomic evaluations. *J. Dairy Sci.*, **93**, 2727–2740.

Cole J.B., VanRaden P.M., O'Connell J.R., Van Tassell C.P., Sonstegard T.S., Schnabel R.D., Taylor J.F., Wiggans G.R. (2009) Distribution and location of genetic effects for dairy traits. *J. Dairy Sci.*, **92**, 2931–2946.

Cole J.B., Wiggans G.R., Ma L., Sonstegard T.S., Lawlor T.J. Jr, Crooker B.A., Van Tassell C.P., Yang J., Wang S., Matukumalli L.K., Da Y. (2011) Genome-wide association analysis of thirty-one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics*, **12**, 408.

Cole J.B., Waurich B., Wensch-Dorendorf M., Bickhart D.M., Swalve H.H. (2014) A genome-wide association study of calf birth weight in Holstein cattle using single nucleotide polymorphisms and phenotypes predicted from auxiliary traits. *J. Dairy Sci.*, **97**, 3156–3172.

Flisikowski K., Venhoranta H., Nowacka-Woszuk J., McKay S.D., Flyckt A., Taponen J., Schnabel R., Schwarzenbacher H., Szczerbal I., Lohi H., Fries R., Taylor J.F., Switonski M., Andersson M. (2010) A novel mutation in the maternally imprinted PEG3 domain results in a loss of MIMT1 Expression and Causes Abortions and Stillbirths in Cattle (Bos taurus). *PLoS ONE*, **5**, 11.

Flori L., Fritz S., Jaffrézic F., Boussaha M., Gut I., Heath S., Foulley J.L., Gautier M. (2009) The genome response to

artificial selection: a case study in dairy cattle. *PLoS ONE*, **4**, e6595.
Flury B.N. (1984) Common principal components in K groups. *J. Am. Stat. Assoc.*, **79**, 892–898.
Grisart B., Coppieters W., Farnir F., Karim L., Ford C., Berzi P., Cambisano N., Mni M., Reid S., Simon P., Spelman R., Georges M., Snell R. (2002) Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res.*, **12**, 222–231.
Hayes B.J., Pryce J., Chamberlain A.J., Bowman P.J., Goddard M.E. (2010) Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milkfat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genet.*, **6**, 1–11.
Hine E., Blows M.W. (2006) Determining the effective dimensionality of the genetic variance-covariance matrix. *Genetics*, **173**, 1135–1144.
Houle D., Mezev J., Galpern P. (2002) Interpretation of the results of common principal component analyses. *Evolution*, **56**, 433–440.
Kaupe B., Brandt H., Prinzenberg E.-M., Erhardt G. (2007) Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. *J. Anim. Sci.*, **85**, 11–21.
Mezey J.G., Houle D. (2003) Comparing G matrices: are common principal components informative? *Genetics*, **165**, 411–425.
Minozzi G., Nicolazzi E.L., Stella A., Biffani S., Negrini R., Lazzari1 B., Ajmone-Marsan P., Williams J.L. (2013) Genome wide analysis of fertility and production traits in Italia Holstein cattle. *PLoS ONE*, **8**, e80219.
Morrison F. (1976) Multivariate Statistical Methods. McGraw-Hill, New York, NY.
Olsen H.G., Hayes B.J., Kent M.P., Nome T., Svendsen M., Lien S. (2009) A genome wide association study for QTL affecting direct and maternal effects of stillbirth and dystocia in cattle. *Anim. Genet.*, **41**, 273–280.
Pausch H., Flisikowski K., Jung S., Emmerling R., Edel C., Gotz K.U., Fries R. (2011) Genome-wide association study identifies two major loci affecting calving ease and growth-related traits in cattle. *Genetics*, **187**, 289–297.
Pennington K.A., Ealy A.D. (2012) The expression and potential function of bone morphogenetic proteins 2 and 4 in bovine trophectoderm. *Reprod. Biol. Endocrinol.*, **10**, 12.
Qanbari S., Gianola D., Hayes B., Schenkel F., Miller S., Moore S., Thaller G., Simianer H. (2011) Application of site haplotype frequency based approaches for detecting selection signatures in cattle. *BMC Genomics*, **12**, 318.
Raven L.A., Coks B., Hayes B.J. (2014) Multibreed genome-wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genomics*, **15**, 62.
SAS Institute (2008) SAS/STAT Version 9.2. SAS Inst. Inc., Cary, NC.
Schnabel R.D., Sonstegard T.S., Taylor J.F., Ashwell M.S. (2005) Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. *Anim. Genet.*, **36**, 408–416.
Schopen G.C.B., Koks P.D., van Arendonk J.A.M., Bovenhuis H., Visker M.H.P.W. (2009) Whole genome scan to detect quantitative trait for bovine milk protein composition. *Anim. Genet.*, **40**, 524–537.
Sodeland M., Kent M.P., Olsen H.G., Opsal M.A., Svendsen M., Sehested E., Hayes B.J., Lien S. (2011) Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. *Anim. Genet.*, **42**, 457–465.
Steppan S.J., Phillips P.C., Houle D. (2002) Comparative quantitative genetics: evolution of the G matrix. *Trends Ecol. Evol.*, **17**, 320–327.
Tetens J., Seidenspinner T., Buttchereit N., Thaller G. (2012) Whole-genome association study for energy balance and fat/protein ratio in German Holstein bull dams. *Anim. Genet.*, **44**, 1–8.
VanRaden P.M., O'Connell J.R., Wiggans G.R., Weigel K.A. (2011) Genomic evaluations with many more genotypes. *Genet. Sel. Evol.*, **43**, 10.
Wang X., Wurmser C., Pausch H., Jung S., Reinhardt F., Tetens J., Thaller G., Fries R. (2012) Identification and dissection of four major QTL affecting milk fat content in the German Holstein Friesian population. *PLoS ONE*, **7**, e40711.
Wiggans G.R., VanRaden P.M., Cooper T.A. (2011) The genomic evaluation system in the United States: past, present, future. *J. Dairy Sci.*, **94**, 3202–3211.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Correlation matrix between direct genomic values for 31 production, conformation and functional traits.

Table S2 Factor pattern of the correlation matrix between chromosomal values for 31 production, conformation and functional traits for BTA5.

Table S3 Factor pattern of the correlation matrix between chromosomal values for 31 production, conformation and functional traits for BTA11.

Table S4 Factor pattern of the correlation matrix between chromosomal values for 31 production, conformation and functional traits for BTA27.

Table S5 Factor pattern of the correlation matrix between chromosomal values for 31 production, conformation and functional traits for BTA28.

