Revisiting the "A Posteriori" Granddaughter Design

G.R. Wiggans[1] and J.I. Weller[2]

[1] Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350, USA

[2] Institute of Animal Science, ARO, The Volcani Center, Bet-Dagan 50250, Israel

Abstract

An updated search for quantitative trait loci (QTLs) in the Holstein genome was conducted using the a posteriori granddaughter design. The number of Holstein sires with ≥100 genotyped and progeny-tested sons has increased from the previous 52 to 71 for a total of 14 246 sons. The bovine genome was divided into 621 segments of ~100 markers each. The sons of each bull were divided into two groups based on which paternal haplotype was transmitted to each son for each chromosomal segment. Significance was tested for each economic trait for each chromosomal segment by a linear model that included the effect of paternal haplotype nested within father. Thirty-three traits were analyzed: yield (milk, fat and protein and component percentages), milk somatic cell score, productive life, daughter pregnancy rate, heifer and cow conception rates, service-sire and daughter calving ease, service-sire and daughter stillbirth rates, 18 conformation traits and the net merit genetic-economic index. Fifty-five chromosomal regions met a significance criterion of probability (*P*) of $<10^{-14}$ compared with 30 regions in the previous analysis based on 52 grandsire families with 9 178 sons. All traits had at least one significant effect, except for protein yield, daughter stillbirth rate and four conformation traits. Confidence intervals (CIs) of 90% were determined for all effects by application of a non-parametric bootstrap. Length of CIs ranged from 2 to 15 chromosomal segments. In all cases, the CI included only part of the chromosome. No significant relationship between log *P* of the effect and CI length was found, even though *P*s ranged from 10^{-14} to 10^{-41} on chromosome 3 for protein percentage. At least six of the regions displayed a bimodal effect distribution in the bootstrap analysis, which indicates more than a single QTL segregating on the chromosome. Results for yield traits were compared with those recently reported for Australian Holsteins, which found effects with a nominal *P* of $<10^{-20}$ on five chromosomes (excluding effects on chromosome 14, which clearly result from the *DGAT1* gene) when each single-nucleotide polymorphism (SNP) effect was estimated as a fixed effect. For U.S. Holsteins, a nominal *P* of $<10^{-6}$ was found in this study for the same traits in nearly the same chromosomal locations, except for effect of fat percentage on chromosome 27. The identified CIs provide promising locations for study of sequence data to identify causative polymorphisms.

Key words: granddaughter design, quantitative trait locus, Holstein, confidence interval, genome

Introduction

With the exception of *DGAT1* and *ABCG2* (Grisart *et al.*, 2002; Winter *et al.*, 2002; Cohen-Zinder *et al.*, 2005), the quantitative trait nucleotides (QTNs), the actual polymorphisms that are responsible for detected QTLs, remain unknown. Determination of QTNs should result in increased rates of genetic gain (Weller and Ron, 2011). If the QTNs are known, then their effects can be included directly in the genomic analysis model, which would increase accuracy of genetic evaluations.

Ron and Weller (2007) presented a schematic strategy for farm animals to determine if a genetic variant is a QTN. The most convincing proof that the QTN has been determined is "concordance"; i.e., determination for a group of animals that their genotypes for the putative QTN correspond to their inferred genotypes for the QTL. Ron and Weller proposed application of the a posteriori granddaughter design (APGD) to determine QTL genotypes for bulls from large populations of cattle genotyped using mid- or high-density SNP chips. Similar to the original granddaughter design, sires with many

progeny-tested sons are analyzed. However, rather than genotype the sons specifically for application of a granddaughter design, the data generated by genotyping many bulls for high-density SNP chips are utilized. Thus, the design is considered to be a posteriori. The sons of each bull are divided into two groups based on which paternal haplotype was passed to each son for the chromosomal region with the putative QTL.

With APGD, each haplotype is based on the genotypes of tens of tightly linked SNPs, and the paternal haplotype of nearly all sons can be determined (Weller *et al.*, 2013). Compared with the application of granddaughter designs based on microsatellites, APGD is more powerful for detection of segregating QTLs. Furthermore, APDG is potentially much more extensive than previous granddaughter design analyses, both in the number of animals included in the analysis and the number of traits analyzed.

Weller *et al.* (2014) applied APGD to the U.S. Holstein population using August 2012 U.S. evaluations. A total of 9 180 bulls, sons of 52 sires with ≥100 sons per sire, were analyzed for 33 economic traits. Since then, the number of bulls genotyped with a mid-density SNP chip has increased dramatically. Furthermore, additional studies based on SNP chip analyses have also located segregating QTLs based on stringent criteria (Daetwyler *et al.*, 2014; Kemper *et al.*, 2015). The objectives of this study were to reapply APGD to the U.S. Holstein population using the more extensive data currently available and to compare the results to other recent studies that have identified segregating QTLs in the U.S. and Australian dairy cattle populations.

Materials and Methods

Data

The current APGD application was based on April 2015 U.S. evaluations and included analysis of 71 grandsires with a total of 14 246 sons. The number of genotyped sons per grandsire ranged from 791 to 100. Numbers of sons and granddaughters for the 71 sires analyzed are shown in Figure 1.

The entire bovine genome, which included the 60 671 SNPs used in U.S. genomic evaluation, was divided into 621 segments of ~100 markers each. The specific number of markers was adjusted to achieve near equality within chromosome. Haplotypes were determined using findhap (VanRaden, 2015). The SNPs located on the sex chromosomes were not analyzed, because all sons receive the Y (not the X) chromosome of their sire.

Thirty-three economic traits were analyzed, including all traits for which genomic

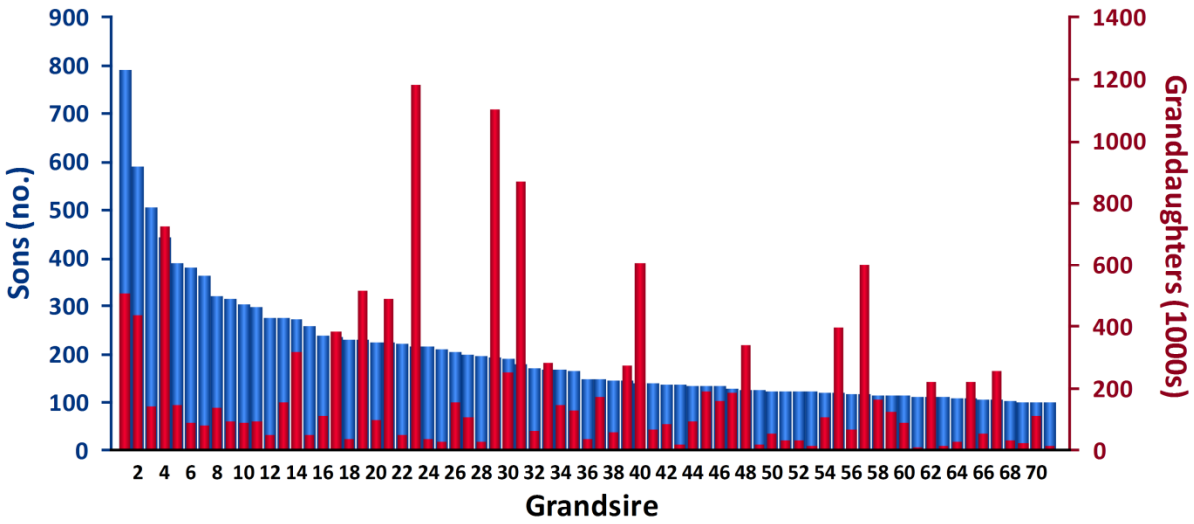

Figure 1. Numbers of sons and granddaughters for APGD analysis of 71 U.S. Holstein bulls.

evaluations are computed for U.S. Holsteins. The traits analyzed included five milk production traits (milk, fat and protein yields as well as fat and protein percentages), somatic cell score, productive life, three fertility traits (daughter pregnancy rate as well as cow and heifer conception rates), four calving traits (service-sire and daughter calving ease as well as service-sire and daughter stillbirth rates), 18 conformation traits and the net merit genetic-economic index. Genomic estimated breeding values were analyzed.

Statistical Analysis

The model for APGD analysis was

$$y_{ijk} = s_i + h_{ij} + e_{ijk},$$

where y_{ijk} is the genetic evaluation of bull k, son of sire i that received sire haplotype j, s_i is the effect of sire i, h_{ij} is the effect of haplotype j of sire i and e_{ijk} is the random residual associated with each record. Analysis of this model was by the GLM procedure of SAS. Overall significance for the haplotype effect indicates that a QTL is segregating within the haplotype segment or is in close proximity. Significance of a specific within-sire haplotype effect indicates that the specific sire is segregating for the QTL.

A total of 19 932 combinations (604 chromosomal segments × 33 traits) were analyzed. In this case, nominal significance levels of 0.05 or 0.01 are meaningless. To correct for multiple combinations, only chromosomal segments with a nominal P of $<10^{-14}$ were considered to be significant.

A non-parametric bootstrap analysis (Visscher *et al.*, 1996) was applied to each chromosome that included significant haplotype segments. A total of 100 samples were generated for each trait × chromosome combination by sampling the 14 246 sons with repeats. For each bootstrap sample, all haplotype segments along the chromosome were analyzed by APGD, and the segment with the lowest P was selected. A 90% CI then was determined by the distribution of the segments with the lowest P-value. The regression of CI on –log P was computed.

Results & Discussion

Excluding *DGAT1* and *ABCG2*, for which causative polymorphisms have been identified, 55 trait × chromosome combinations were significant ($P < 10^{-14}$). Weller *et al.* (2014) found only 30 effects that met this criterion. Ordinal numbers of the first SNP among the 60 671 SNPs sorted by chromosome and location are in Table 1 for the chromosomal segment with lowest P along with P for that segment by trait and chromosome number. All traits had at least one significant effect, except protein yield, daughter stillbirth rate and four conformation traits. Lowest P (2.4×10^{-42}) was for protein percentage on chromosome 3.

The Manhattan plot for net merit is in Figure 2. Highest peaks were found on chromosomes 14 and 18. The peak on chromosome 14 corresponds to the position of *DGAT1*. The large effect on chromosome 18 was found previously by Cole *et al.* (2011) and Weller *et al.* (2013).

Kemper *et al.* (2015) discovered QTLs for milk production traits of Australian dairy cattle. Their analysis of Holsteins included 8 478 cows and 3 049 bulls. They only considered effects that were significant by two criteria for further analysis:

1. Single SNP regression for each trait using EMMAX software (Kang *et al.*, 2010).

2. Average local genetic evaluation variance for chromosomal segments including most significant SNP compared with distribution of variances among all segments. This differs from criterion 1 in that all SNPs within the segment are fitted simultaneously to estimate SNP marker effects.

They found effects with a nominal P of $<10^{-20}$ on six chromosomes (including chromosome 14, which clearly is a result of *DGAT1*) when each SNP effect was estimated as a fixed effect. For U.S. Holsteins, a nominal P of $<10^{-6}$ was found using APGD for the same trait in nearly the same chromosomal location, except for the effect of fat percentage on chromosome 27. Results of the two studies are compared in Table 2.

Table 1. APGD significant effects ($P < 10^{-14}$) by trait, chromosome (Chr.) and ordinal number of first SNP in chromosomal segment with lowest *P* among 60 671 SNPs sorted by chromosome and location.

Trait	Chr.	SNP	*P*	Trait	Chr.	SNP	*P*
Milk yield	5	13 937	1.56×10^{-16}	Final score	5	13 838	4.36×10^{-27}
	15	36 887	7.63×10^{-15}		19	44 321	1.95×10^{-18}
Fat yield	5	13 937	1.12×10^{-37}	Dairy form	5	13 838	4.09×10^{-42}
	15	37 182	3.27×10^{-21}		6	16 674	1.80×10^{-24}
Fat percentage	5	13 937	9.85×10^{-40}		7	18 735	3.16×10^{-17}
	15	37 182	1.11×10^{-16}	Feet and legs score	18	42 846	7.13×10^{-16}
Protein percentage	3	6 849	2.36×10^{-42}	Stature	5	14 036	3.50×10^{-34}
	20	45 897	2.44×10^{-33}		7	18 735	1.32×10^{-22}
Somatic cell score	5	13 244	1.79×10^{-19}		11	29 074	7.62×10^{-34}
	6	16 674	1.39×10^{-39}		14	34 088	4.88×10^{-25}
	14	34 278	2.43×10^{-18}		19	44 321	7.24×10^{-23}
	16	38 739	1.09×10^{-16}	Strength	5	14 036	1.57×10^{-25}
Productive life	5	13 244	5.05×10^{-19}		10	26 024	4.82×10^{-27}
	6	16 674	2.47×10^{-26}		14	34 088	1.49×10^{-26}
Daughter pregnancy rate	1	1 863	7.69×10^{-17}	Udder depth	5	13 739	2.69×10^{-19}
	5	13 838	1.59×10^{-26}	Body depth	5	14 036	1.43×10^{-38}
	18	42 650	2.87×10^{-27}		10	26 024	4.26×10^{-16}
Cow conception rate	5	13 937	1.75×10^{-29}		14	34 088	7.46×10^{-27}
	18	42 748	4.14×10^{-29}	Foot angle	5	14 236	6.06×10^{-18}
Heifer conception rate	5	13 343	5.88×10^{-23}	Fore udder	5	13 739	8.85×10^{-17}
	6	16 965	4.87×10^{-22}		28	57 199	1.37×10^{-16}
Daughter calving ease	18	42 944	2.70×10^{-16}	Rear udder height	13	32 747	9.26×10^{-17}
Service-sire calving ease	5	13 244	1.41×10^{-19}	Rump angle	2	770	5.90×10^{-16}
	18	42 944	2.27×10^{-18}		7	18 735	1.42×10^{-22}
Service-sire stillbirth rate	5	13 046	5.14×10^{-16}		8	2 072	4.08×10^{-16}
Net merit	18	42 258	1.28×10^{-17}	Thurl (rump) width	5	14 036	4.03×10^{-21}
					10	26 624	9.09×10^{-16}
					19	44 321	3.01×10^{-21}
				Teat length	5	12 353	8.65×10^{-38}

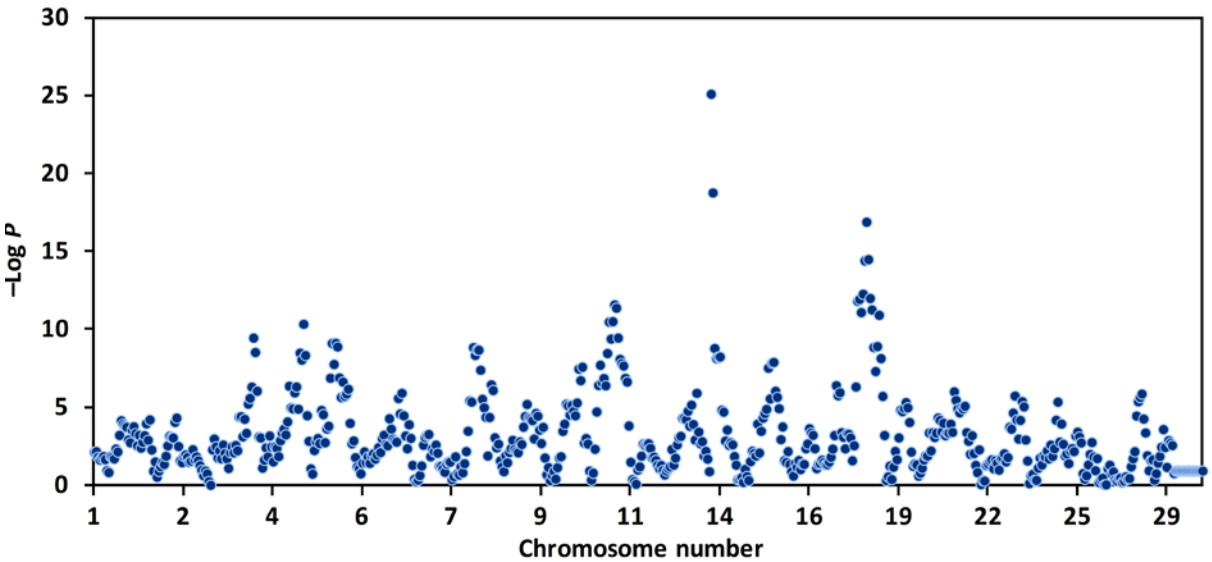

Figure 2. Manhattan plot of –log *P* for net merit.

Table 2. Comparison of U.S. and Australian Holstein QTLs that affect milk production traits and have significant[a] effects in the Australian population by trait and chromosome.

		Location (base pairs)[b]		*P*	
Trait	Chromosome	Australia	United States	Australia	United States
Protein percentage	3	15 632 410	16 097 418	3.2×10^{-30}	2.4×10^{-42}
	20	31 228 912	31 393 193	1.3×10^{-34}	2.4×10^{-33}
	29	41 989 397	42 770 336	7.9×10^{-41}	5.6×10^{-07}
Fat percentage	5	93 945 655	92 115 327	2.0×10^{-38}	9.8×10^{-40}

[a]$P < 10^{-20}$; *DGAT1* excluded.

[b]Refers to the SNP with the greatest effect for Australia and to the first SNP in the segment with the greatest effect for the United States.

For all 55 significant effects, a 90% CI that spanned only part of the chromosome was determined. The CI included only two segments for fat yield on chromosome 5 and protein percentage on chromosome 3. Each chromosomal segment includes ~100 markers and 5 million base pairs. At least 6 regions had bimodal effect distributions in the bootstrap analyses, including net merit on chromosome 18. A bimodal distribution is expected if more than a single QTL affecting the analyzed trait is segregating on the chromosome. The result for net merit on chromosome 18 is consistent with that of Cole *et al*. (2011).

The 90% CIs as a function of −log *P* are shown in Figure 3. Bimodal CIs were not included. Although the CIs narrowed as −log *P* increased, the regression of CI on −log *P* was not significant, even though the bimodal CIs were excluded.

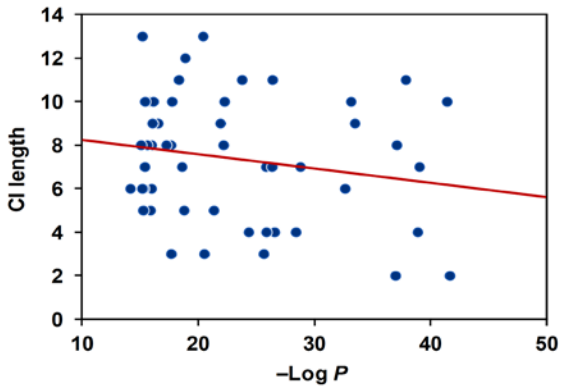

Figure 3. The 90% CIs as a function of –log *P*; bimodal CIs were not included.

VanRaden *et al.* (2011) found three haplotypes with major negative effects on fertility in Holsteins: HH1, HH2 and HH3 on chromosomes 5, 1 and 8, respectively. All three effects are caused by recessive lethals, which results in observed reduced fertility for heterozygotes. Thus, the effects associated with those haplotypes are from genes with major effects, not QTLs. Causative mutations have been identified for HH1 and HH3 but not for HH2. Results of VanRaden *et al.* (2011) are compared with those from this study in Table 3. Significant APGD effects were found for cow conception rate and daughter pregnancy rate on chromosomes 5 (HH1) and 1 (HH2) but not on chromosome 8. For HH1, CIs were the same for cow conception rate and daughter pregnancy rate but did not include the position of the causative mutation (Adams *et al.*, 2012). The CI for daughter pregnancy rate included the location of the HH2 haplotype. A CI was not computed for cow conception rate because minimum *P* was $>10^{-14}$.

As proposed by Ron and Weller (2007), the next step to find the QTNs for these effects will be determination of concordance between effects in individual families and specific polymorphisms within the CI. This will require genomic sequencing of the grandsires. Of the 71 grandsires analyzed, 42 have been sequenced, and their sequence data are available through the 1000 Bull Genomes Project (Daetwyler *et al.*, 2014). The remaining 29 bulls will be sequenced as part of a Binational Agricultural Research and Development project between Israel and the United States. Initially they will be sequenced

Table 3. Comparison of HH1 and HH2 results from VanRaden *et al.* (2011) with those from this study

Study	HH1 (chromosome 5)	HH2 (chromosome 1)
VanRaden *et al.* (2011)		
Location (base pairs)	63 150 400	94 860 836 – 96 553 339
Effect, conception rate (%)	−3.0 ± 0.8	−3.2 ± 0.4
Haplotype frequency (%)	1.92	1.66
APGD, cow conception rate (%)		
Segment with greatest effect (base pairs)	92 115 327 – 96 166 308	64 592 861 – 68 997 018
APGD *P*	1.7×10^{-29}	6.9×10^{-14}
CI (base pairs)	65 922 088 – 96 166 308	—
APGD, daughter pregnancy rate (%)		
Segment with greatest effect (base pairs)	88 359 142 – 92 115 327	88 167 139 – 92 958 471
APGD *P*	1.6×10^{-26}	7.7×10^{-17}
CI (base pairs)	65 922 088 – 96 166 308	64 592 861 – 111 573 593

to depth of 10–15×. Haplotype determination will enable a nearly complete, accurate sequence for most bulls. Additional sequencing will be performed as required to determine the complete genomic sequence.

Conclusions

Fifty-five chromosomal regions met a significance criterion of $P < 10^{-14}$ compared with 30 regions in the previous analysis of 52 grandsire families. At least one significant effect was found for all but six traits. Results for yield traits corresponded to those for Australian Holsteins, and results for fertility traits generally corresponded to previous results for U.S. Holsteins. A CI that included only part of the chromosome could be determined for all significant effects, but distribution of the bootstrap sample was bimodal for at least six effects. Results will be used to identify promising regions of sequence data for discovery of causative mutations. Determination of QTNs should increase rates of genetic gain and aid in understanding the biological pathways that determine these traits.

Acknowledgements

This research was supported the Binational Agricultural Research and Development Fund (BARD). We thank the Ministero delle Politiche Agricole Alimentari e Forestali (MIPAAF, Rome, Italy) for funding the HD genotypes contributed by the Innovagen project (DM 10750-7303-2011), Defra (London, UK) for funding the HD genotypes contributed by the United Kingdom as part of the Ruminant Genetic Improvement Network and the Council on Dairy Cattle Breeding (Reynoldsburg, OH, USA) for providing the genetic evaluation data. The figures were created using Daniel's XL Toolbox add-in for Excel (version 6.53) by Daniel Kraus, Würzburg, Germany.

References

Adams, H.A., Sonstegard, T., VanRaden, P.M., Null, D.J., Van Tassell, C. & Lewin, H. 2012. Identification of a nonsense mutation in APAF1 that is causal for a decrease in reproductive efficiency in dairy cattle. *Proceedings of Plant and Animal Genome XX Conference*, abstract P0555.

Cohen-Zinder, M., Seroussi, E., Larkin, D.M., Loor, J.J., Everts-van der Wind, A., Lee, J.-H., Drackley, J.K., Band, M.R., Hernandez, A.G., Shani, M., Lewin, H.A., Weller, J.I. & Ron, M. 2005. Identification of a missense mutation in the bovine *ABCG2* gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research 15*, 936–944.

Cole, J.B., Wiggans, G.R., Ma, L., Sonstegard, T.S., Lawlor Jr., T.J., Crooker, B.A., Van Tassel, 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.

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R.F., Liao, X., Djari, A., Rodriguez, S.C., Grohs, C., Esquerré, D., Bouchez, O., Rossignol, M.-N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P.J., Coote, D., Chamberlain, A.M., Anderson, C., Van Tassell, C.P., Hulsegge, I., Goddard, M.E., Guldbrandtsen, B., Lund, M.S., Veerkamp, R.F., Boichard, D.A., Fries, R. & Hayes, B.J. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics 46*, 858–865.

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 Research 12*, 222–231.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-Y., Freimer, N.B., Sabatti, C. & Eskin, E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics 42*, 348–354.

Kemper, K.E., Hayes, B.J., Daetwyler, H.D. & Goddard, M.E. 2015. How old are quantitative trait loci and how widely do they segregate? *Journal of Animal Breeding and Genetics 132*, 121–134.

Ron, M. & Weller, J.I. 2007. From QTL to QTN identification in livestock – winning by points rather than knock-out: A review. *Animal Genetics 38*, 429–439.

VanRaden, P.M. 2015. findhap.f90: Find haplotypes and impute genotypes using multiple chip sets and sequence data. http://aipl.arsusda.gov/software/findhap/. Accessed July 20, 2015.

VanRaden, P.M. Olson, K.M., Null, D.J. & Hutchison, J.L. 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of Dairy Science 94*, 6153–6161.

Visscher, P.M., Thompson, R. & Haley, C.S. 1996. Confidence intervals in QTL mapping by bootstrapping. *Genetics 143*, 1013–1020.

Weller, J.I. & Ron, M. 2011. *Invited review:* Quantitative trait nucleotide determination in the era of genomic selection. *Journal of Dairy Science 94*, 1082–1090.

Weller, J.I., Cole, J.B., VanRaden, P.M. & Wiggans, G.R. 2014. Application of the *a posteriori* granddaughter design to the Holstein genome. *Animal 8*, 511–519.

Weller, J.I., VanRaden, P.M. & Wiggans, G.R. 2013. Application of a posteriori granddaughter and modified granddaughter designs to determine Holstein haplotype effects. *Journal of Dairy Science 96*, 5376–5387.

Winter, A., Kramer, W., Werner, F.A.O., Kollers, S., Kata, S., Durstewitz, G., Buitkamp, J., Womack, J.E., Thaller, G. & Fries, R. 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (*DGAT1*) with variation at a quantitative trait locus for milk fat content. *Proceedings of the National Academy of Sciences of the United States of America 99*, 9300–9305.