Animal (2014), **8:4**, pp 511–519
doi:10.1017/S1751731114000111

animal

Application of the *a posteriori* granddaughter design to the Holstein genome

J. I. Weller[1†], J. B. Cole[2], P. M. VanRaden[2] and G. R. Wiggans[2]

[1]*The Volcani Center, Institute of Animal Sciences, A.R.O., Bet Dagan 50250, Israel;* [2]*Animal Improvement Programs Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705-2350, USA*

(Received 8 September 2013; Accepted 20 December 2013)

An a posteriori *granddaughter design was applied to estimate quantitative trait loci genotypes of sires with many sons in the US Holstein population. The results of this analysis can be used to determine concordance between specific polymorphisms and segregating quantitative trait loci. Determination of the actual polymorphisms responsible for observed genetic variation should increase the accuracy of genomic evaluations and rates of genetic gain. A total of 52 grandsire families, each with* $\geqslant$*100 genotyped sons with genetic evaluations based on progeny tests, were analyzed for 33 traits (milk, fat and protein yields; fat and protein percentages; somatic cell score (SCS); 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 net merit). Of 617 haplotype segments spanning the entire bovine genome and each including ~*5×10^6 *bp, 5 cM and 50 genes, 608 autosomal segments were analyzed. A total of 19 335 unique haplotypes were found among the 52 grandsires. There were a total of 133 chromosomal segment-by-trait combinations, for which the nominal probability of significance for the haplotype effect was* $<10^{-8}$*, which corresponds to genome-wide significance of* $<10^{-4}$*. The number of chromosomal regions that met this criterion by trait ranged from one for rear legs (rear view) to seven for net merit. For each of the putative quantitative trait loci, at least one grandsire family had a within-family contrast with a* t*-value of >3. Confidence intervals (CIs) were estimated by the nonparametric bootstrap for the largest effect for each of nine traits. The bootstrap distribution generated by 100 samples was bimodal only for net merit, which had the widest 90% CI (eight haplotype segments). This may be due to the fact that net merit is a composite trait. For all other chromosomes, the CI spanned less than a third of the chromosome. The narrowest CI (a single haplotype segment) was found for SCS. It is likely that analysis by more advanced methods could further reduce CIs at least by half. These results can be used as a first step to determine the actual polymorphisms responsible for observed quantitative variation in dairy cattle.*

Keywords: granddaughter design, genomic evaluation, haplotype, quantitative trait locus, Holstein cow

Implications

An *a posteriori* granddaughter design was applied to the entire genome to determine haplotype effects for 33 traits of Holsteins. There were a total of 133 chromosomal segment-by-trait combinations, for which the nominal probability of significance for the haplotype effect was $<10^{-8}$, which corresponds to genome-wide significance of $<10^{-4}$. By trait, from one (rear legs, rear view) to seven (net merit) chromosomal segments met this criterion. Confidence intervals ranged from one to eight haplotype segments of ~75 markers each. These results can be used to determine the actual polymorphisms responsible for variation in quantitative traits.

† E-mail: weller@agri.huji.ac.il

Introduction

Since 2008, genomic evaluation has become a reality chiefly because of the development of high-density single nucleotide polymorphism (SNP) chips that allow for relatively inexpensive genotyping of individuals for tens of thousands of genetic markers. The methods developed for genomic evaluations are based on population-wide linkage disequilibrium between closely linked markers and the actual quantitative trait locus (QTL) that determine phenotypes for the traits of interest (e g. VanRaden *et al*., 2009). Since linkage disequilibrium is generally incomplete, even if a SNP has a major estimated effect on an economic trait, the SNP genotypes of individual animals do not necessary correspond to their QTL genotypes. With the exception of the *DGAT1* and *ABCG2* genes

(Grisart *et al.*, 2002; Winter *et al.*, 2002; Cohen-Zinder *et al.*, 2005), the actual polymorphisms responsible for detected QTL, the quantitative trait nucleotides (QTN) remain unknown. Determination of the QTN should result in increased rates of genetic gain (Weller and Ron, 2011). If the QTN are known, then these factors can be included directly in the genomic analysis model, resulting in more accurate genetic evaluations.

Ron and Weller (2007) presented a schematic strategy for determination if a genetic variant is a QTN in farm animals. The most convincing proof that the QTN has been determined is 'concordance,' that is determination for a group of animals that their genotypes for the putative QTN correspond to their inferred genotypes for the QTL. They proposed application of the '*a posteriori* granddaughter design' (APGD) to determine QTL genotypes for bulls from large populations of individuals genotyped using 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 '*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. For both *DGAT1* and *ABCG2*, the QTN was first determined by concordance.

The granddaughter design has been applied to nearly all major commercial dairy cattle populations (Weller, 2007). So far, 7091 QTL for 441 economical traits have been reported based on marker–QTL linkage analyses (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index), the overwhelming majority by granddaughter designs. However, relatively little practical use has been made of the results. The reasons were twofold. First, because of the very large number of tests performed, the usual *P*-values of 0.05 or 0.01 will be obtained many times purely by chance. Lander and Kruglyak (1995) developed a formula that accounts for linkage among sites along a chromosome. For a complete analysis of the bovine genome with 30 chromosomes and a map length of 30 M, a pointwise *P*-value of 5×10^{-5} is required to obtain a genome-wide *P*-value of 0.05; that is, the probability is 0.05 that a nominal *P*-value of 5×10^{-5} ($\log_{10}(5 \times 10^{-5}) = -4.3$) for a specific genome location will be obtained at least once in the genome. However, very few of the effects actually detected by daughter or granddaughter designs met that stringent criterion (e.g. Georges *et al.*, 1995).

The second difficulty in application of daughter or granddaughter design results to commercial breeding programs was that confidence intervals (CIs) for QTL location, except for the largest effects such as *ABCG2* and *DGAT1*, were very wide (if computed at all) and generally included most of the chromosome (Ron *et al.*, 2004). Again, because sites along the chromosome are linked, CI cannot be computed by an analytical formula. Visscher *et al.* (1996) proposed that CI could be estimated by application of the nonparametric bootstrap. Multiple samples are drawn from the data with repeats; those samples are analyzed, and CI are derived from the distributions of the samples.

Compared with application of granddaughter designs based on microsatellites (e.g. Georges *et al.*, 1995), the APGD, in which each haplotype is based on the genotypes of tens of tightly linked SNP and the paternal haplotype of nearly all sons can be determined (Weller *et al.*, 2013), is more powerful for detection of segregating QTL. Furthermore, the 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. Georges *et al.* (1995) analyzed 1518 sons of 14 sires for five milk production traits. Ashwell *et al.* (2004) analyzed 1415 sons of 10 sires for the same milk production traits, pregnancy rate, somatic cell score (SCS) and productive life. Approximately 70 000 US Holstein bulls have already been genotyped, but most of these bulls do not yet have genetic evaluations based on progeny test.

For all 33 traits evaluated for US Holsteins, Weller *et al.* (2013) applied APGD to the specific autosomal region that included the SNP with the greatest effect for each trait as determined by genome-wide analysis including all valid SNP. The effects of nearly all of the chromosomal segments analyzed were nominally significant at $P < 0.001$. Because only a single haplotype segment was analyzed per trait, multiple comparisons were not a serious problem. The objective of this study was to apply APGD to the entire genome for all 33 traits and to estimate CI for large effects of interest.

Material and methods

Animals, genotypes and traits analyzed

Similar to Weller *et al.* (2013), of 19 365 US Holstein bulls with genotypes for the BovineSNP50 BeadChip (Illumina Inc., 2011), and $\geqslant$10 valid daughter records; 9178 bulls that were sons of 52 bulls with $\geqslant$100 sons per sire were retained for APGD application. These sons had a total of 7 334 778 daughters with valid records for the milk production traits. As in Weller *et al.* (2013), 45 188 SNP from the BovineSNP50 BeadChip were retained for analysis, and haplotypes with a maximum of 75 markers per haplotype segment were determined using the findhap.f90 Fortran program (VanRaden, 2011). Over the entire genome of $\sim 3 \times 10^9$ bp and a length of 3000 cM, there were 617 nonoverlapping haplotype segments; each haplotype segment included $\sim 5 \times 10^6$ bp and 5 cM. Only 608 autosomal segments were included in the analysis, because the other nine segments were located on the X chromosome in the region that is nonhomologous to the Y chromosome, and sons did not receive the paternal X chromosome. A total of 19 335 unique haplotypes were found among the 52 grandsires analyzed, for a mean of 32 unique haplotypes per segment.

The 33 traits analyzed by Weller *et al.* (2013) were included: milk, fat and protein yields; component percentages; SCS; 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. Genomic

Table 1 *Heritabilities and mean reliabilities of the sons included in the analysis by trait*

		Reliabilities of the sons (%)	
Trait	Heritability	Mean	Minimum
Milk yield	0.30	87.5	47.0
Fat yield	0.30	87.5	47.0
Protein yield	0.30	87.5	47.0
Fat percentage	0.50	87.5	47.0
Protein percentage	0.50	87.5	47.0
Net merit	0.20	77.7	38.0
Productive life	0.08	69.5	29.0
SCS	0.12	80.4	36.0
Daughter pregnancy rate	0.04	65.2	28.0
Heifer conception rate	0.01	46.5	24.0
Cow conception rate	0.016	53.1	26.0
Sire calving ease	0.086	70.6	25.0
Sire stillbirth rate	0.03	51.2	25.0
Daughter calving ease	0.064	62.5	25.0
Daughter stillbirth rate	0.065	53.0	25.0
Final score	0.29	75.4	28.0
Stature	0.42	83.9	30.0
Strength	0.31	78.0	28.0
Dairy form	0.29	78.6	25.0
Foot angle	0.15	73.7	25.0
Rear legs (side view)	0.21	78.2	29.0
Body depth	0.37	76.9	28.0
Rump angle	0.33	82.7	30.0
Thurl width	0.26	78.8	29.0
Fore udder attachment	0.29	80.0	29.0
Rear udder height	0.28	77.6	29.0
Udder depth	0.28	83.4	25.0
Udder cleft	0.24	76.3	29.0
Front teat placement	0.26	79.2	30.0
Teat length	0.26	80.0	25.0
Rear legs (rear view)	0.11	68.8	25.0
Feet and legs score	0.17	70.6	25.0
Rear teat placement	0.18	75.3	29.0

SCS = somatic cell score.

evaluations were analyzed, but they were expected to be very similar to traditional genetic evaluations computed with a standard animal model (VanRaden and Wiggans, 1991), because each son was required to have ⩾10 daughters with a record usable for genetic evaluation. Heritabilities, and mean and minimum reliabilities of the sons included in the analysis by trait are given in Table 1. Heritabilities ranged from 0.01 for heifer conception rate to 0.5 for fat and protein percentage. Mean reliabilities ranged from 46.5% for heifer conception rate to 87.5% for milk production traits. Minimum reliabilities ranged from 24% for heifer conception rate to 47% for milk production traits.

Statistical analyses

The statistical model of Weller *et al.* (2013) was used for each haplotype segment:

$$Y_{ijk} = S_i + H_{ij} + e_{ijk}$$

where Y_{ijk} is the genetic evaluation of bull k, which is a 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 evaluation. All sons were given equal weight in the analysis, even though there were differences in their reliabilities. Model analysis was by the GLM procedure of SAS (SAS Institute Inc., 2012). An overall haplotype effect indicates that a QTL is segregating within or near the haplotype segment. A specific within-family haplotype effect ($P < 0.001$) indicates that the specific bull is segregating for the QTL.

A nonparametric bootstrap analysis (Visscher *et al.*, 1996) was applied to the chromosome that included the haplotype segment with the greatest effect (lowest *P*-value) for nine traits: milk and protein yields; SCS; net merit; productive life; daughter pregnancy rate; heifer and cow conception rates; and final score. A total of 100 bootstrap samples were generated for each trait-by-chromosome combination by sampling the 9178 sons with repeats. For each bootstrap sample, all haplotype segments along the chromosome were analyzed by APGD, and the segment with the lowest *P*-value was selected. A 90% CI then was determined by the distribution of the segments with the lowest *P*-value. Because distributions generally were nonsymmetric, especially if QTL were located near the end of a chromosome, the shortest chromosomal segment that included 90% of the samples was selected as the CI.

Results and discussion

Bos taurus chromosome (BTA) for the SNP with the greatest effect by all-SNP analysis (Weller *et al.*, 2013) and by the APGD, nominal *P*-value ($-\log_{10}$) for APGD haplotype effects, and the number of chromosomes with a nominal $P < 10^{-8}$ for the APGD haplotype effect are shown in Table 2. For six traits, the SNP with the greatest effect was on the sex chromosome. This could either be because of a major QTL located on the sex chromosome with effects on several traits or a possible bias in the effect estimates due to the fact that sires always pass their single X chromosome to all daughters. For those traits, the autosome with the greatest SNP effect was shown in Table 2. For 23 of the 33 traits analyzed the greatest APGD haplotype effect was not on the chromosome with the greatest SNP effect. This is not surprising, considering that the effect associated with a specific SNP on average underestimates the actual QTL by a factor of five (Weller *et al.*, 2013).

A nominal $P < 10^{-8}$ for the APGD haplotype effect corresponds to genome-wide $P < 10^{-4}$. There were a total of 133 chromosomal segment-by-trait combinations that met this criterion. Thus, many more QTL were detected with genome-wide significance, as compared with the previous granddaughter designs for the traits analyzed (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index). Rear legs (rear view) had only a single autosomal region with a nominal $P < 10^{-8}$ for the APGD haplotype effect; all other traits had more than a single region. Genetic correlations from Interbull

Table 2 *Chromosomes with the SNP with the greatest effect by all-SNP analysis (Weller* et al.*, 2013) and by the APGD, $-log_{10}$ P-values for this effect and the number of chromosomes with APGD probabilities $<10^{-8}$*

Trait	Chromosome with		P ($-\log_{10}$) for APGD haplotype effect	Chromosomes with APGD probabilities of $<10^{-8}$ (number)
	Greatest SNP effect	Greatest APGD haplotype effect		
Milk yield	14	14	37.5	4
Fat yield	14	14	93.6	3
Protein yield	5	10	10.8	5
Fat percentage	14	14	226.0	3
Protein percentage	6	14	40.0	6
Net merit	18	18	17.4	7
Productive life	18	6	22.4	5
SCS	6	6	25.0	6
Daughter pregnancy rate	18	18	16.1	4
Heifer conception rate	6	6	13.1	5
Cow conception rate	18	6	31.6	4
Sire calving ease	18	5	14.1	2
Sire stillbirth rate	18	5	17.6	4
Daughter calving ease	18	18	10.0	3
Daughter stillbirth rate	18	10	8.6	2
Final score[1]	11	5	15.6	5
Stature	18	5	25.6	6
Strength	18	5	23.9	5
Dairy form[1]	6	5	22.0	3
Foot angle[1]	5	5	14.7	2
Rear legs (side view)	19	3	12.7	3
Body depth	18	5	28.3	5
Rump angle	14	2	10.9	5
Thurl width	18	5	17.8	4
Fore udder attachment	5	1	12.6	6
Rear udder height[1]	7	28	10.6	4
Udder depth	5	1	12.5	4
Udder cleft	18	14	9.9	4
Front teat placement	5	1	8.7	2
Teat length	5	5	26.5	3
Rear legs (rear view)[1]	13	1	9.9	1
Feet and legs score[1]	17	18	10.0	4
Rear teat placement	26	1	10.3	4

SNP = single nucleotide polymorphism; APGD = *a posteriori* granddaughter design; SCS = somatic cell score.
[1]The SNP with the greatest effect was on the sex chromosome.

are lower for this trait than for other conformation traits, averaging 0.76 for United States with other countries (http://www-interbull.slu.se/conform/c-appen4-061.html) compared with 0.85 to 0.95 for most conformation traits, indicating that the trait is defined less consistently or evaluated less accurately.

Net merit had seven chromosomes with a nominal $P < 10^{-8}$, which was more than any other trait. For daughter stillbirth rate, the greatest APGD haplotype effect, as measured by the negative log of the nominal probability, was 8.6. For all other traits, the greatest APDG haplotype effect was larger. Daughter stillbirth rate had the sixth lowest heritability among the traits analyzed (Table 1).

The effects for milk and fat yields and fat and protein percentages on BTA14 had by far the greatest *F*-values and lowest *P*-values. Those effects correspond to the effect associated with the *DGAT1* gene located near the beginning of that chromosome (Grisart *et al.*, 2002; Winter *et al.*, 2002). The effect associated with the *ABCG2* gene on BTA6 that chiefly affects protein percentage (Cohen-Zinder *et al.*, 2005) was not the effect associated with the lowest probability for that trait. This may be due to the fact that the frequency of the favorable allele that decreases milk yield but increases protein concentration is now at a very high frequency in the North American Holstein population. In addition, the effect of the QTL may be spread over several closely linked markers, or alternatively no single marker may be in strong linkage disequilibrium with the QTL.

Manhattan plots of the APGD haplotype effects are depicted in Figure 1 for milk, fat and protein yields; SCS; net merit; and productive life; Manhattan plots for the remaining traits are in the Supplementary Figures S1–S4.

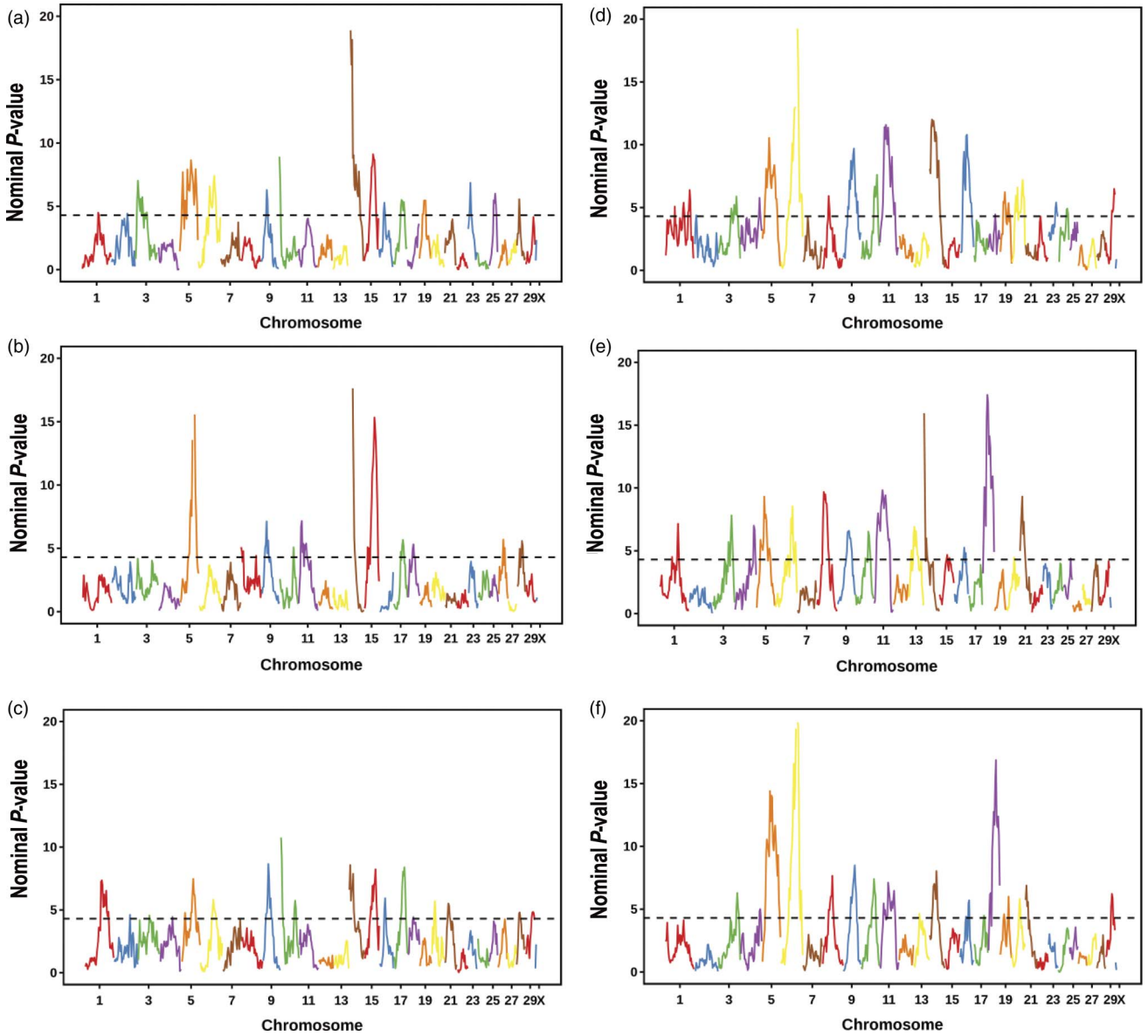

Figure 1 Manhattan plots of the *a posteriori* granddaughter design for milk yield (a), fat yield (b), protein yield (c), SCS (d), net merit (e), and productive life (f), where nominal P ($-\log_{10}$) is plotted as a function of haplotype segment. The dotted line at $P = 4.3$ corresponds to genome-wide $P = 0.05$. Each chromosome is plotted in a different color. SCS = somatic cell score.

In the figures, $-\log_{10}$ nominal P for each chromosome segment is plotted as a function of genome location. In addition to the major effect on BTA14 associated with *DGAT1*, large effects were found for milk yield on BTA5, 6, 10 and 15 and for protein yield on BTA9 and 10. Of the 29 autosomes, 17 had a nominal $P < 5 \times 10^{-5}$ for protein yield. Thus, numerous segregating QTL can be detected, even though milk and protein yields have been under intense selection for 50 years. (Although protein has been measured directly only since 1977, protein yield is highly correlated to milk yield.) For net merit, the largest effect was on BTA18, which also was the chromosome that had the SNP with the greatest effect; however, a very large effect also was found on BTA14. For both productive life and SCS, the largest effects were on BTA6, even though the SNP with the greatest effect on productive life was on BTA18, which had the second largest effect for that trait. Cole *et al.* (2009) reported the discovery of a large QTL associated with net merit on BTA18. The QTL on BTA14 is in the region that includes *DGAT1*, which has an effect on net merit because of the economic value associated with milk components (Cole *et al.*, 2010). The SCS findings are consistent with a QTL on BTA6 as previously reported by Bennewitz *et al.* (2004) and Lund *et al.* (2008).

Chromosomes with nominal APDG $P < 10^{-8}$ and numbers of grandsire families with t-values >3 for within-family contrasts are shown in Table 3 by chromosome. For each of these putative QTL, at least one grandsire family had a within-family contrast with t-value >3. The effect for fat percentage on BTA14 had the greatest number (26) of segregating families followed by fat yield (21), protein

Table 3 *Chromosomes with nominal* $P < 10^{-8}$ *for APGD haplotype effects*

Trait	Chromosomes with $P < 10^{-8}$ for APGD haplotype effect (number of grandsire families with *t*-values > 3)
Milk yield	5 (4), 10 (4), 14 (14), 15 (1)
Fat yield	5 (6), 14 (21), 15 (5)
Protein yield	9 (2), 10 (5), 14 (2), 15 (4), 17 (4)
Fat percentage	5 (10), 14 (26), 15 (5)
Protein percentage	3 (6), 6 (8), 11 (4), 14 (15), 15 (2), 20 (9)
Net merit	5 (4), 6 (4), 8 (4), 11 (3), 14 (7), 18 (8), 21 (3)
Productive life	5 (5), 6 (7), 9 (3), 14 (2), 18 (5)
SCS	5 (5), 6 (8), 9 (3), 11 (4), 14 (3), 16 (4)
Daughter pregnancy rate	1 (8), 5 (7), 6 (2), 18 (7)
Heifer conception rate	4 (2), 5 (5), 6 (5), 9 (5), 29 (4)
Cow conception rate	5 (3), 6 (8), 18 (4), 19 (1)
Sire calving ease	5 (4), 18 (4)
Sire stillbirth rate	5 (6), 10 (3), 13 (4), 18 (2)
Daughter calving ease	5 (3), 6 (5), 18 (4)
Daughter stillbirth rate	10 (3), 18 (2)
Final score[1]	1 (3), 5 (5), 18 (4), 19 (3), 28 (3)
Stature	5 (7), 7 (7), 10 (4), 11 (7), 14 (2), 19 (4)
Strength	4 (4), 5 (7), 10 (4), 14 (3), 18 (4)
Dairy form[1]	1 (5), 5 (8), 6 (9)
Foot angle[1]	5 (6), 25 (2)
Rear legs (side view)	3 (4), 5 (4), 8 (5)
Body depth	5 (6), 10 (4), 11 (6), 14 (4), 18 (3)
Rump angle	1 (3), 2 (5), 7 (3), 8 (7), 14 (3)
Thurl width	5 (7), 9 (3), 10 (6), 15 (5)
Fore udder attachment	1 (6), 5 (6), 10 (4), 11 (3), 20 (2), 28 (4)
Rear udder height[1]	1 (6), 13 (3), 14 (4), 28 (3)
Udder depth	1 (7), 5 (3), 10 (4), 17 (5)
Udder cleft	1 (3), 8 (4), 11 (2), 14 (4)
Front teat placement	1 (3), 14 (2)
Teat length	5 (7), 11 (3), 15 (4)
Rear legs (rear view)[1]	1 (4)
Feet and legs score[1]	1 (4), 8 (5), 18 (5), 28 (2)
Rear teat placement	1 (4), 8 (4), 11 (3), 14 (4)

SNP = single nucleotide polymorphism; APGD = *a posteriori* granddaughter design; SCS = somatic cell score.
Numbers of grandsire families with *t*-values >3 for within-family contrasts in parentheses.
[1]The SNP with the greatest effect was on the sex chromosome.

percentage (15) and milk yield (14) on this chromosome. As noted, all of those appear to be related to the *DGAT1* gene. BTA5 had the next highest number of segregating families for fat percentage (10). That QTL apparently also affects milk and fat yields. Both protein percentage and dairy form had effects segregating in nine families (BTA20 and six, respectively); however, for both traits, those chromosomes were not the one with the overall lowest *P*-value.

Weller *et al.* (1990) considered 20 grandsire families each with 200 sons. In that case, power to detect a QTL with a substitution effect of 0.1 phenotypic standard deviations is 0.61 with 50 daughters per son and equal frequency for the two QTL alleles. Increasing the number of grandsires to 50 results in nearly complete power of detection. Because a QTL of that magnitude and allelic frequencies accounts for only 0.5% of phenotypic variance, the detection of several QTL for each trait is not surprising.

Table 4 shows 90% CI for selected effects of nine traits using the nonparametric bootstrap method as well as nominal *P*. Generally, CI narrowed as $-\log_{10} P$ increased. A bimodal distribution was observed only for net merit. That is two peaks were observed for the distribution of the bootstrap samples. Furthermore, this trait had the widest CI (eight haplotype segments). Thus, two QTL that are segregating on BTA18 likely are affecting net merit, which is consistent with the findings of Cole *et al.* (2011). For all other chromosomes, the CI spanned less than a third of the chromosome. The narrowest CI was for SCS; the maximum effect was found in the same haplotype segment for more than 90% of bootstrap samples. In previous attempts to estimate CI using a nonparametric bootstrap method for either daughter or granddaughter designs, CI generally spanned almost half of the chromosome (Weller *et al.*, 2002).

Lander and Botstein (1989) proposed the use of support intervals to estimate CI for QTL location; that is the points at

Table 4 *Confidence intervals (CI) of 90% and nominal* P*(–log$_{10}$) for selected* a posteriori *granddaughter design effects using a nonparametric bootstrap*

Trait	Chromosome	Haplotype segment[1]	P (–log$_{10}$)	Haplotype segments on chromosome	CI
Milk	15	389	9.1	377–396	387–392
Protein	10	264	10.8	264–288	264–265
Somatic cell score	6	177	25.0	156–185	177
Net merit[2]	18	441	17.4	438–450	440–447
Productive life	6	177	22.4	156–185	176–179
Daughter pregnancy rate	18	445	16.1	438–450	444–448
Heifer conception rate	6	176	13.1	156–185	177–183
Cow conception rate	6	177	31.6	156–185	177–179
Final score	5	149	15.6	131–155	147–151

[1]Genome was divided into 617 segments of ~75 markers each. Haplotype segments were numbered consecutively beginning with BTA1 and concluding with the sex chromosome.
[2]Bimodal distribution for bootstrap results.

which the score for the logarithm of the odds falls below a certain value relative to the maximum. However, Darvasi *et al.* (1993) demonstrated that this method does not reliably reflect actual CI, and Bennewitz *et al.* (2002) found that the bootstrap tends to overestimate actual CI. Furthermore, this method is not effective if more than a single QTL is segregating on the chromosome for the trait of interest.

A CI of three haplotype segments included ~1.5×10^7 bp, 15 cM and 150 genes. For a dense genome scan, Darvasi and Soller (1997) and Weller and Soller (2004) developed formula to estimate CI for QTL location as a function of experimental design, QTL substitution effect and the number of individuals analyzed. However, none of their experimental designs correspond to the granddaughter design, in which multiple families are analyzed but only a fraction of those families are segregating for the QTL. The backcross design presented in those studies is roughly parallel to the granddaughter design for a single family. A total of 4917 backcross progeny are required to obtain a CI of 95% for a substitution effect of 0.25 phenotypic standard deviations. The width of the CI is inversely proportional to the square of the substitution effect relative to the residual standard deviation (Weller and Soller, 2004). For the granddaughter design, in which genetic evaluations are analyzed, the variance among sons will be approximately equal to three-fourths of the genetic variance, but only half of the substitution effect of the QTL will be passed from sons to their daughters. Thus, for a given substitution effect, fewer sons are required to obtain the same CI as compared with a backcross design, but the CI will also be a function of heritability.

Ron and Weller (2007) showed that if 50 grandsire families are analyzed and $\geqslant$5 grandsires are segregating for a QTL, the probability of obtaining concordance by chance for even a single nucleotide would still be very low. Thus, assuming that a QTL is the result of a single point mutation, determining the causative polymorphism for the 52 grandsires included in this analysis should be possible by DNA sequencing of their genomes. A possible pitfall is that significance of the within-grandsire effect may not be definitive for all families. Grandsires with dubious QTL genotype status should be deleted from the concordance analysis as suggested by Weller *et al.* (2013). The causative polymorphism may also be the result of a more complicated mechanism such as copy number variation (e.g. Seroussi *et al.*, 2010). In that case, ascertaining the causative DNA sequence probably would be more difficult.

Considering that CI generally spanned individual chromosomal segments, it is likely that CI could be reduced by analysis of shorter segments. In addition, CI could be further reduced by various methods that have been proposed in previous studies. Korol *et al.* (1995) demonstrated that CI could be decreased by multitrait analysis of correlated traits. Of the two genes for which the QTN have been determined in dairy cattle, *ABCG2* and *DGAT1*, both affected multiple traits. Thus, if a common peak location is observed in the Manhattan plots for more than a single trait, it would be reasonable to concentrate the search on the trait with the lowest probability, based on the hypothesis that the same polymorphism is likely to be affecting other correlated traits.

Meuwissen *et al.* (2002) presented an algorithm for joint linkage disequilibrium and linkage analysis (LDLA) mapping. They used a Gibbs sampling algorithm but included only those haplotypes that could be determined with near certainty. Using this method, they were able to map a QTL affecting twinning rate to a chromosomal region of <1 cM in the middle part of BTA5. Olsen *et al.* (2005) used this method to map the QTL affecting protein concentration on BTA6 (later identified to be *ABCG2*) to a region of 420 kb, ~0.5 cM. With linkage mapping via a granddaughter design, the CI for QTL location was 7.5 cM. LDLA mapping was extended by Meuwissen and Goddard (2004) to multitrait and multiple QTL analysis and by Druet and Georges (2010) to include reconstruction of haplotypes by a hidden Markov model. Druet and Georges (2010) applied their method to analyze the effect of *DGAT1* in a Dutch Holstein–Friesian pedigree of 199 sires and their 1502 sons. To date, LDLA methods have not been applied to a complete genome analysis.

Finally, Weller *et al.* (2013) demonstrated that residual variances for the APGD can be reduced to approximately one-quarter if the genetic evaluations of sons are corrected

for SNP effects associated with all chromosomes other than the one under analysis. Because, as noted previously, the width of the CI is inversely proportional to the square of the substitution effect relative to the residual standard deviation, this procedure may reduce CI by half. It is also possible that shorter CI may be obtained by using smaller haplotypes.

Conclusions

For 33 traits analyzed with the APGD, each trait had at least one significant ($P<10^{-8}$) within-family haplotype, which corresponds to genome-wide significance of $P<10^{-4}$. Net merit had seven chromosomes with nominal $P<10^{-8}$ for APGD haplotype effect. For each of those putative QTL, at least one grandsire family had a within-family contrast with a *t*-value of >3. The bootstrap distribution generated by 100 samples was bimodal only for net merit, which was the trait with the widest 90% CI (eight haplotype segments). For all other chromosomes, the CI spanned less than a third of the chromosome. The narrowest 90% CI was for SCS, with the maximum effect found in the same haplotype segment. A CI of three haplotype segments will include ~1.5×10^{7} bp, 15 cM and 150 genes. If 50 grandsire families were analyzed and ⩾5 grandsires were segregating for a QTL, the probability of obtaining concordance by chance would be very low for even a single nucleotide. Thus, these results can be used as a first step to determine the actual polymorphisms responsible for observed quantitative variation in dairy cattle.

Acknowledgments

This research was supported by a grant from the Binational Agricultural Research and Development Fund Research Project IS-4394-11R. The authors thank S. M. Hubbard, Animal Improvement Programs Laboratory, Agricultural Research Service, US Department of Agriculture (Beltsville, MD, USA), for technical manuscript review.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1751731114000111.

References

Ashwell MS, Heyen DW, Sonstegard TS, Van Tassell CP, Da Y, VanRaden PM, Ron M, Weller JI and Lewin HA 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. Journal of Dairy Science 87, 468–475.

Bennewitz J, Reinsch N and Kalm E 2002. Improved confidence intervals in quantitative trait loci mapping by permutation bootstrapping. Genetics 160, 1673–1686.

Bennewitz J, Reinsch N, Guiard V, Fritz S, Thomsen H, Looft C, Kühn C, Schwerin M, Weimann C, Erhardt G, Reinhardt F, Reents R, Boichard D and Kalm E 2004. Multiple quantitative trait loci mapping with cofactors and application of alternative variants of the false discovery rate in an enlarged granddaughter design. Genetics 168, 1019–1027.

Cohen-Zinder M, Seroussi E, Larkin DM, Loor JJ, Everts-van der Wind A, Lee JH, Drackley JK, Band MR, Hernandez AG, Shani M, Lewin HA, Weller JI and 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 JB, VanRaden PM and Multi-State Project S-1040 2010. Net merit as a measure of lifetime profit: 2010 revision AIPL Res Rep NM$4(12-09). Retrieved April 4, 2013 from http://aipl.arsusda.gov/reference/nmcalc-2010.htm

Cole JB, VanRaden PM, O'Connell JR, Van Tassell CP, Sonstegard TS, Schnabel RD, Taylor JF and Wiggans GR 2009. Distribution and location of genetic effects for dairy traits. Journal of Dairy Science 92, 2931–2946.

Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ Jr, Crooker BA, Van Tassel CP, Yang J, Wang S, Matukumalli LK and Da Y 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. BMC Genomics 12, 408.

Darvasi A and Soller M 1997. A simple method to calculate resolving power and confidence interval of QTL map location behavior. Genetics 27, 125–132.

Darvasi A, Weinreb A, Minke V, Weller JI and Soller M 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134, 943–951.

Druet T and Georges M 2010. A hidden Markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. Genetics 184, 789–798.

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X, Womack JE and Hoeschele I 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139, 907–920.

Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, Spelman R, Georges M and 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.

Illumina Inc 2011. BovineSNP50 Genotyping BeadChip. Retrieved April 4, 2013 from http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp50.pdf

Korol AB, Ronin YI and Kirzhner VM 1995. Interval mapping of quantitative trait loci employing correlated trait complexes. Genetics 140, 1137–1147.

Lander E and Kruglyak L 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics 11, 241–247.

Lander ES and Botstein D 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185–199.

Lund MS, Guldbrandtsen B, Buitenhuis AJ, Thomsen B and Bendixen C 2008. Detection of quantitative trait loci in Danish Holstein cattle affecting clinical mastitis, somatic cell score, udder conformation traits, and assessment of associated effects on milk yield. Journal of Dairy Science 91, 4028–4036.

Meuwissen THE and Goddard ME 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. Genetics Selection Evolution 36, 261–279.

Meuwissen THE, Karlsen A, Lien S, Olsaker I and Goddard ME 2002. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics 161, 373–379.

Olsen HG, Lien S, Gautier M, Nilsen H, Roseth A, Berg PR, Sundsaasen M, Svendsen KK and Meuwissen TH 2005. Mapping of a milk production QTL to a 420 kb region on bovine chromosome 6. Genetics 169, 275–283.

Ron M and Weller JI 2007. From QTL to QTN identification in livestock – winning by points rather than knock-out: a review. Animal Genetics 38, 429–439.

Ron M, Feldmesser E, Golik M, Tager-Cohen I, Kliger D, Reiss V, Domochovsky R, Alus O, Seroussi E, Ezra E and Weller JI 2004. A complete genome scan of the Israeli Holstein population for quantitative trait loci by a daughter design. Journal of Dairy Science 87, 476–490.

SAS Institute Inc 2012. Base SAS® 9.3 Procedures Guide, 2nd edition, SAS Institute Inc, Cary, NC.

Seroussi E, Glick G, Shirak A, Yakobson E, Weller JI, Ezra E and Zeron Y 2010. Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. BMC Genomics 11, 673.

VanRaden P 2011. findhap.f90. Retrieved April 4, 2013 from http://aipl.arsusda.gov/software/findhap/

VanRaden PM and Wiggans GR 1991. Derivation, calculation, and use of national animal model information. Journal of Dairy Science 74, 2737–2746.

VanRaden PM, Van Tassel CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF and Schenkel FS 2009. Invited review: reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science 92, 16–24.

Visscher PM, Thompson R and Haley CS 1996. Confidence intervals in QTL mapping by bootstrapping. Genetics 143, 1013–1020.

Weller JI 2007. Marker-assisted selection in dairy cattle. Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish (ed. EP Guimarães, J Ruane, BD Scherf, A Sonnino and JD Dargie), 199–228. Food and Agriculture Organization of the United Nations, Rome, Italy.

Weller JI and Ron M 2011. Invited review: quantitative trait nucleotide determination in the era of genomic selection. Journal of Dairy Science 94, 1082–1090.

Weller JI and Soller M 2004. An analytical formula to estimate confidence interval of QTL location with a saturated genetic map as a function of experimental design. Theoretical and Applied Genetics 109, 1224–1229.

Weller JI, Kashi Y and Soller M 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. Journal of Dairy Science 73, 2525–2537.

Weller JI, VanRaden PM and Wiggans GR 2013. Application of a posteriori granddaughter and modified granddaughter designs to determine Holstein haplotype effects. Journal of Dairy Science 96, 5376–5387.

Weller JI, Weller H, Kliger D and Ron M 2002. Estimation of quantitative trait locus allele frequency via a modified granddaughter design. Genetics 162, 841–849.

Winter A, Krämer W, Werner FAO, Kollers S, Kata S, Durstewitz G, Buitkamp J, Womack JE, Thaller G and 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. Proceeding of the National Academy of Sciences USA 99, 9300–9305.