J. Dairy Sci. 99:4504–4511
http://dx.doi.org/10.3168/jds.2015-10456

Increasing the number of single nucleotide polymorphisms used in genomic evaluation of dairy cattle[1]

G. R. Wiggans,[2] T. A. Cooper, P. M. VanRaden, C. P. Van Tassell, D. M. Bickhart, and T. S. Sonstegard[3]
Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

GeneSeek (Neogen Corp., Lexington, KY) designed a new version of the GeneSeek Genomic Profiler HD BeadChip for Dairy Cattle, which originally had >77,000 single nucleotide polymorphisms (SNP). A set of >140,000 SNP was selected that included all SNP on the existing GeneSeek chip, all SNP used in US national genomic evaluations, SNP that were possible functional mutations, and other informative SNP. Because SNP with a lower minor allele frequency might track causative variants better, 30,000 more SNP were selected from the Illumina BovineHD Genotyping BeadChip (Illumina Inc., San Diego, CA) by choosing SNP to maximize differences in minor allele frequency between a SNP being considered for the new chip and the 2 SNP that flanked it. Single-gene tests were included if their location was known and bioinformatics indicated relevance for dairy cattle. To determine which SNP from the new chip should be included in genomic evaluations, genotypes available from chips already in use were used to impute and evaluate the SNP set. Effects for 134,511 usable SNP were estimated for all breed–trait combinations; SNP with the largest absolute values for effects were selected (5,000 for Holsteins, 1,000 for Jerseys, and 500 each for Brown Swiss and Ayrshires for each trait). To increase overlap with the 60,671 SNP currently used for genomic evaluation, 12,094 more SNP with the largest effects were added. After removing SNP with many parent–progeny conflicts, 84,937 SNP remained. Three cutoff studies were conducted with 3 SNP sets to determine reliability gain over that for parent average when evaluations based on August 2011 data were used to predict December 2014 performance. Across all traits, mean Holstein reliability gains were 32.5, 33.4, and 32.0 percentage points for 60,671, 84,937, and 134,511 SNP, respectively. After genotypes from the new chip became available, the proposed set was reduced from 84,937 to 77,321 SNP to remove SNP that were not included during manufacture, reduce computing time, and improve imputation performance. The set of 77,321 SNP was evaluated using August 2011 data to predict April 2015 performance. Reliability gain over 60,671 SNP was 1.4 percentage points across traits for Holsteins. Improvement over 84,937 SNP was partially the result of 4 mo of additional data and genotypes from the new chip. Revision of the SNP set used for genomic evaluation is expected to be an ongoing process to increase evaluation accuracy.

Key words: single nucleotide polymorphism, genomic evaluation, dairy cattle, beadchip, reliability

INTRODUCTION

Since genomic evaluation of dairy cattle began in the United States in April 2008 (Wiggans et al., 2011), a progression of genotyping chips has been employed, and genotypes from 18 chips had been submitted to the Council on Dairy Cattle Breeding (Bowie, MD) for use in national genomic evaluations by the end of August 2015. One factor that drives the creation of new chips is that the pool of bead types used to manufacture a chip becomes exhausted, which provides an opportunity to update the SNP set when creating a new pool. Even if the goal is to continue the same chip, a new bead pool may yield a slightly different set of SNP that pass performance standards.

Starting with the Illumina (San Diego, CA) Bovine3K BeadChip (Illumina Inc., 2011) in 2010, imputation has been necessary so that all SNP used for evaluation are represented for each animal (aside from a few SNP for individual animals for which imputation is not successful). Imputation has made it possible to include genotypes from a variety of chips with little change in the genomic evaluation process. Accuracy of imputation is critical to the accuracy of genomic evaluations. Validation of SNP sets for new versions of low-density chips

Received September 28, 2015.
Accepted February 14, 2016.
[1]The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by USDA or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.
[2]Corresponding author: george.wiggans@ars.usda.gov
[3]Current address: Acceligen of Recombinetics Inc., St. Paul, MN 55104.

(G. R. Wiggans, unpublished data) has been done by creating a test set from a higher density chip, imputing, and comparing the results with the original. Imputation accuracy has been very high (>97%) if the set included most of the SNP in the Illumina BovineLD BeadChip (Illumina Inc., 2015b) and one or both parents have been genotyped.

In 2014, Neogen (Lexington, KY) decided to replace the GeneSeek Genomic Profiler HD for Dairy Cattle (**GHD**) beadchip, which had >77,000 SNP (Neogen Corporation, 2013b), with a second version of the chip (**GH2**) that would have approximately 140,000 SNP. When the GHD chip was designed, the goal was to replace less informative SNP from the Illumina BovineSNP50 Genotyping BeadChip (Illumina Inc., 2012) with SNP that were more informative and to increase the number of SNP (Wiggans et al., 2014). Subsequent testing of the selected 77,000 SNP revealed that higher reliability could be achieved by continuing to use some of the excluded BovineSNP50 SNP (Wiggans et al., 2014), possibly because SNP with a low minor allele frequency (**MAF**) may track causative variants better and because the added SNP had higher imputation errors as a result of relatively few animals with those genotypes at the time the SNP set was evaluated. VanRaden and O'Connell (2015) realized a gain in reliability of only 0.4 percentage points when increasing the number of SNP from 45,187 to 311,725. That finding suggests that some SNP can be added to increase reliability, but most added SNP will have little effect or be detrimental because of imputation errors.

With the opportunity to add thousands of SNP when creating the GH2 chip, any SNP currently used in genomic evaluations that were not on the GHD chip could be added, along with single-gene tests and causative variants that had become available. Other informative SNP from the Illumina BovineHD Genotyping BeadChip, which has >777,000 SNP (Illumina Inc., 2015a), could also be added (particularly those with lower MAF) as well as SNP for beef cattle to improve usefulness for those customers.

The first objective of this study was to document the selection of SNP for the GH2 chip. The goal was to include nearly all the SNP from the HD chip that are informative for dairy cattle and any newly discovered causative variants (or structural variants that might be causative variants). Other objectives were to determine which of the selected SNP to use in national genomic evaluations and to examine the benefit of increasing the number of SNP included in genomic evaluation by testing a set of SNP that was moderately larger than the 60,671 set used since December 2013.

MATERIALS AND METHODS

Chip Design

In addition to GHD SNP and any SNP currently used in genomic evaluation, 30,000 SNP from the BovineHD chip with MAF of >0.1 in Holsteins were selected for the GH2 chip so that MAF differences were maximized between the SNP being considered for GH2 inclusion (candidate SNP) and the 2 SNP that would flank the candidate if selected. This process added SNP with lower MAF than those already included from the GHD and other sources.

To select variants with potentially large effects, Illumina HiSeq 2× 100-bp paired-end reads of sequence data with 3 to 14× coverage for 25 Holstein bulls and 8 Jersey bulls (Supplemental Table S1; http://dx.doi.org/10.3168/jds.2015-10456) were aligned to the UMD3.1 reference genome (Zimin et al., 2009) using Burrows–Wheeler Alignment software (version 0.7.10-r789; Li and Durbin, 2009) to generate alignments in the Sequence Alignment/Map format. SAMtools software (version 0.1.19–44428cd; Li et al., 2009) was used to call SNP and indels. Putative biological effect was assessed by SnpEff (Cingolani et al., 2012) annotation using the UMD3.1.71 Ensembl annotation database (Ensembl, 2013). Variants with a SnpEff putative impact estimate of "high," ≥10 reads of coverage, and present in at least 3 individuals were selected to include on the GH2 chip, which resulted in the addition of 419 SNP. To identify SNP reference genome location via flanking sequence alignment, the 100-bp flanking sequence from each ambiguously mapped SNP was aligned to the UMD3.1 reference genome using the Burrows–Wheeler Alignment MEM algorithm (Li and Durbin, 2009). Consistency of flanking sequence alignment and site locations of assayed variants were assessed by using custom Perl scripts (https://github.com/njdbickhart/perl_toolchain/blob/master/snp_utilities/getSNPPositionFromSAM.pl).

Genotypes already available from other chips were used to impute and evaluate SNP selected for the GH2 chip for their usefulness in genomic evaluation. Effects for 134,511 usable SNP were estimated for all breed–trait combinations using iteration for marker effects based on the linear model and approximate Bayes A algorithm of VanRaden (2008), and the SNP with the largest absolute values for effects were selected (5,000 for Holsteins, 1,000 for Jerseys, and 500 each for Brown Swiss and Ayrshires for each trait). To increase overlap with the set of SNP currently used for genomic evaluation, 12,094 more SNP with the largest effects were

added. After removing SNP with many parent–progeny conflicts after imputation, 84,937 SNP remained. Based on previous experience with the GHD chip (Wiggans et al., 2014), using a subset of the proposed GH2 SNP was expected to give higher reliability than using all SNP. In addition, limiting the number of SNP used would reduce computation time.

Validation Studies

Proposed GH2 SNP. Three cutoff studies (VanRaden et al., 2009) with Holstein data were conducted with 60,671, 84,937, and 134,511 SNP to determine gain in reliability over that for parent average when evaluations based on data from August 2011 were used to predict performance in December 2014. Table 1 shows counts of animals in training and validation sets by trait. Training sets ranged from 17,024 to 49,229 bulls and cows, and validation sets ranged from 1,306 to 3,936 bulls.

To assess the effect of using estimates of SNP effects from current data to determine which SNP to select as most informative, 2 cutoff studies (VanRaden et al., 2009) were conducted. To minimize computational effort, an existing file of 132,587 imputed Holstein genotypes was used. The SNP effects were estimated using October 2011 and April 2015 data. Five yield traits (milk, fat, and protein yields and fat and protein percentages) were evaluated, and the absolute values of their SNP solutions were ranked. For each cutoff date, the highest rank for each SNP across traits was determined. The 77,000 SNP with the highest ranks for each cutoff date were selected and used in cutoff studies in which October 2011 data were used to predict April 2015 performance.

SNP for National Genomic Evaluations. The GH2 chip (manifest available at https://support.illumina.com/downloads/geneseek-ggp-bovine-150k-product-files.html) became commercially available in early 2015. The first GH2 genotypes with 138,942 re-

Table 1. Counts of animals in training and validation sets for cutoff studies for 134,511 SNP proposed for version 2 of the GeneSeek Genomic Profiler HD BeadChip (Neogen Corp., Lexington, KY) and for 77,321 SNP proposed for use in national genomic evaluation by trait and breed

	134,511 SNP		77,321 SNP			
Trait	Holstein training	Holstein validation	Holstein training	Holstein validation	Jersey training	Jersey validation
Milk yield	49,229	3,936	49,674	4,383	10,573	685
Fat yield	49,229	3,936	49,674	4,383	10,573	685
Protein yield	49,218	3,936	49,663	4,383	10,568	685
Fat percentage	49,229	3,936	49,674	4,383	10,573	685
Protein percentage	49,218	3,936	49,663	4,383	10,568	685
Net merit	34,982	3,936	35,238	4,383	6,676	685
Productive life	30,499	3,936	30,738	4,383	6,556	682
Somatic cell score	44,918	3,933	45,339	4,383	10,251	685
Daughter pregnancy rate	30,604	3,649	30,850	4,101	6,457	643
Service-sire calving ease	23,243	1,311	22,236	1,718	—	—
Daughter calving ease	19,074	1,311	19,071	1,718	—	—
Service-sire stillbirth rate	18,873	1,311	18,866	1,718	—	—
Daughter stillbirth rate	17,024	1,306	17,021	1,712	—	—
Final score	39,070	2,873	39,414	3,132	9,389	661
Stature	39,188	2,873	39,533	3,132	9,423	661
Strength	38,913	2,873	39,256	3,132	9,369	661
Body depth	39,131	2,873	39,475	3,132	—	—
Dairy form	39,017	2,871	39,360	3,132	9,418	661
Rump angle	39,104	2,873	39,449	3,132	9,417	661
Rump width	36,319	2,873	36,658	3,132	9,195	661
Rear legs (side view)	38,981	2,873	39,324	3,132	9,339	657
Rear legs (rear view)	36,382	2,873	36,721	3,132	—	—
Foot angle	37,153	2,873	37,493	3,132	9,190	661
Feet and legs	38,564	2,873	38,905	3,132	—	—
Fore udder attachment	39,039	2,873	39,383	3,132	9,636	667
Rear udder height	38,968	2,873	39,311	3,132	9,097	662
Udder cleft	38,499	2,873	38,841	3,132	9,381	661
Udder depth	39,157	2,873	39,502	3,132	9,781	672
Front teat placement	39,147	2,873	39,491	3,132	9,405	661
Rear teat placement	38,833	2,871	39,177	3,132	—	—
Teat length	39,115	2,873	39,459	3,132	9,413	661

ported SNP were submitted to the Council on Dairy Cattle Breeding for use in genomic evaluation in April 2015. A new set of SNP for possible use in national genomic evaluation was created based on the 138,942 GH2 SNP as well as those on version 3 of the GeneSeek Genomic Profiler for Dairy Cattle (**GP3**; Neogen Corporation, 2013a) and version 2 of the Zoetis (Kalamazoo, MI) low-density chip. The SNP included were restricted to these chips that were widely used in April 2015 to ensure that selected SNP would continue to be supplied. The number of GH2 SNP between 2 consecutive GP3 SNP was determined by sorting SNP proposed for genomic evaluations by chromosome and base pair position. If more than 10 of the GH2 SNP were between 2 consecutive GP3 SNP, only the 10 GH2 SNP with the largest magnitudes of effect were retained to improve imputation performance. A total of 77,321 SNP remained for possible use in genomic evaluations.

Cutoff studies with those SNP were conducted for Holsteins and Jerseys. Data from August 2011 were used to predict performance in April 2015. Table 1 shows counts of animals in training and validation sets by trait. Training sets ranged from 17,021 to 49,674 bulls and cows for Holsteins and from 6,457 to 10,573 bulls and cows for Jerseys; validation sets ranged from 1,712 to 4,383 bulls for Holsteins and from 643 to 685 bulls for Jerseys. These analyses included GH2 genotypes that had been received by the Council on Dairy Cattle Breeding for April 2015 evaluations.

Composite SNP

When creating a new chip, some important SNP are included several times to minimize the likelihood that the SNP will be lost during manufacture. New SNP are proposed from many research projects, and sometimes a SNP is proposed that duplicates an existing one but has a different name. Duplication was determined by comparison of SNP locations or flanking sequences. To extract information for a SNP across chips, a composite SNP was defined that creates a single name for the preferred SNP across chips regardless of its reported

Table 2. Gains in reliability of US Holstein genetic evaluations from including genomic information in the evaluation compared with traditional parent average by trait and number of SNP included

	Reliability gain (percentage points)		
Trait	60,671 SNP	84,937 SNP	134,511 SNP
Milk yield	34.0	35.1	33.9
Fat yield	33.8	34.2	33.2
Protein yield	24.9	26.3	24.3
Fat percentage	58.5	58.7	58.5
Protein percentage	49.0	49.7	49.7
Net merit	30.7	31.3	31.4
Productive life	33.8	32.0	34.8
Somatic cell score	36.2	36.5	36.4
Daughter pregnancy rate	25.3	24.5	24.9
Service-sire calving ease	24.4	25.2	24.6
Daughter calving ease	34.1	37.0	34.1
Service-sire stillbirth rate	7.8	8.4	7.5
Daughter stillbirth rate	57.0	60.9	56.8
Final score	25.2	25.6	24.2
Stature	35.5	36.1	35.0
Strength	34.2	35.2	33.4
Body depth	35.5	36.1	34.4
Dairy form	37.7	37.9	36.8
Rump angle	36.7	37.6	37.0
Rump width	32.3	32.9	31.6
Rear legs (side view)	24.0	24.7	23.6
Rear legs (rear view)	24.2	25.5	24.1
Foot angle	20.3	22.2	20.5
Feet and legs	19.4	20.6	19.1
Fore udder attachment	37.8	38.6	37.3
Rear udder height	24.7	25.0	23.5
Udder cleft	24.1	25.2	24.0
Udder depth	45.5	46.7	45.8
Front teat placement	34.5	35.2	34.6
Rear teat placement	34.3	35.2	34.0
Teat length	33.2	33.8	32.6
All traits	32.5	33.4	32.0

Table 3. Mean reliabilities for October 2011 traditional parent averages and reliability gains[1] for genomic evaluations of Holstein bulls without daughter information in October 2011 but with a traditional evaluation by April 2015 using SNP effects selected from evaluations based on October 2011 or April 2015 data by yield trait

Yield trait	Parent average reliability (%)	Reliability gain (percentage points)	
		SNP from October 2011 data	SNP from April 2015 data
Milk yield	36.8	33.3	38.3
Fat yield	36.8	33.5	38.1
Protein yield	36.8	27.3	30.7
Fat percentage	36.8	52.5	59.6
Protein percentage	36.8	45.5	54.4

[1]Genomic reliability − parent average reliability.

name. This composite SNP was used for parent–progeny comparisons as well as extracting data for genomic evaluations. Composites for 65 SNP were defined.

Single-Gene Tests and Structural Variants

As GH2 genotypes accumulated, sufficient data became available for Holsteins to allow consideration of some GH2 single-gene tests and structural variants for inclusion in genomic evaluation. Some single-gene tests had been included on previous GeneSeek chips; therefore, substantial numbers of genotypes were already available for those tests. The tests were considered for inclusion in genomic evaluation if their location was known and bioinformatics information indicated relevance for dairy cattle. All structural variants that were not monomorphic also were considered. Because SNP usability in genomic evaluations may change as additional animals are genotyped, all proposed SNP except those specifically retained were subjected to standard edits for MAF, Mendelian consistency, and call rate.

RESULTS AND DISCUSSION

Validation Studies

Proposed GH2 SNP. Mean Holstein reliability gains over traditional parent average for the 60,671, 84,937, and 134,511 SNP cutoff studies (Table 2) were 32.5, 33.4, and 32.0 percentage points, respectively, across all traits. The lowest gain for the largest SNP set was likely the result of imputation errors because some of the new SNP were present in only a few genotypes. The gain may be overestimated because data used to select the most informative SNP were also the data used to determine gain. The findhap program (version 3; VanRaden, 2015) used for imputation does not impute a genotype for a SNP with insufficient information for an accurate allele call. Because the percentage of missing alleles is considered in the calculation of reliability, differences in imputation accuracy by chip affect reliability.

Table 3 compares reliability gains when SNP were selected from an evaluation using October 2011 data or

Table 4. Gains in reliability of US genomic evaluations based on 77,321 SNP compared with evaluations based on the 60,671 SNP currently used in national genomic evaluation by trait and breed

Trait	Reliability gain (percentage points)	
	Holstein	Jersey
Milk yield	1.7	0.1
Fat yield	1.3	0.5
Protein yield	1.4	0.4
Fat percentage	1.4	−0.1
Protein percentage	2.2	0.1
Net merit	2.0	1.6
Productive life	1.7	0.3
Somatic cell score	0.8	0.5
Daughter pregnancy rate	1.6	0.0
Service-sire calving ease	0.7	—
Daughter calving ease	2.7	—
Service-sire stillbirth rate	1.7	—
Daughter stillbirth rate	2.7	—
Final score	1.1	0.6
Stature	1.3	1.0
Strength	1.4	−0.3
Body depth	0.7	—
Dairy form	1.0	2.3
Rump angle	0.7	1.1
Rump width	0.5	−0.1
Rear legs (side view)	1.4	−0.3
Rear legs (rear view)	1.7	0.4
Foot angle	1.6	—
Feet and legs	1.2	1.3
Fore udder attachment	0.7	−0.7
Rear udder height	0.2	1.1
Udder cleft	1.6	4.6
Udder depth	0.8	—
Front teat placement	2.0	−1.2
Rear teat placement	1.8	—
Teat length	1.5	1.7
All traits	1.4	0.6

an evaluation using April 2015 data. Selecting the SNP based on the more recent data resulted in a reliability advantage of 3.4 to 8.9 percentage points. Using current data may give better predictions because the SNP represent the current population better, but it also may overstate the gain due to genomics. The relative benefits of the SNP sets evaluated in this study should not be affected by the data used to select the SNP.

SNP for National Genomic Evaluations. The reliability gain from using 77,321 rather than 60,671 SNP (Table 4) was 1.4 percentage points across all traits for Holsteins. This improvement compared with that for 84,937 SNP (0.9 percentage points, Table 2) was the result of including several additional months of genotypes from the GH2 chip as well as data from 4 additional months for the training set and validation bulls. The reliability gain across all traits was smaller (0.6 percentage points) for Jerseys, and no gain was realized for 7 traits. The benefit of additional markers is less with a smaller reference population.

Single-Gene Tests and Structural Variants

Table 5 lists the single-gene tests that were selected for use in routine genomic evaluations because they were identified as causative variants or closely linked to

Table 5. Description of single-gene tests on GeneSeek beadchips (Neogen Corp., Lexington, KY) that were identified as causative variants or closely linked to them for dairy cattle by breed

			Genotypes (no.)[2]			Minor allele frequency		
Gene-test name[1]	Chromosome	Location (bp)	Holstein	Jersey	Brown Swiss	Holstein	Jersey	Brown Swiss
ABCG2_1	6	38,027,010	3,580	371	33	0.012	0.003	0.000
Arachnomelia-BS	5	57,641,340	195,779	47,127	2,286	0.000	0.000	0.001
BCN_8219	6	87,181,501	2,872	211	33	0.000	0.000	0.000
BCN-A2	6	87,181,620	2,872	210	33	0.385	0.195	0.227
BCNAB	6	87,181,503	175,471	45,608	1,980	0.026	0.145	0.158
BetaLact	11	103,304,758	194,206	46,800	2,263	0.443	0.405	0.261
BGHR	1	72,129,955	102,326	21,009	1,051	0.156	0.265	0.100
BLAD	1	145,119,004	213,570	47,077	2,287	0.002	0.000	0.000
CAPN1_1	29	44,069,063	226,703	19,524	13,989	0.449	0.451	0.203
CAPN1_2	29	45,685,980	445,753	35,027	14,270	0.238	0.240	0.443
DGAT_2	14	1,802,265	3,087	354	30	0.238	0.428	0.033
Dominant_Red	3	9,479,761	50,856	8,214	576	0.001	0.000	0.000
DUMPS	1	69,757,801	202,227	47,207	2,287	0.000	0.000	0.000
EXON2FB	4	93,262,107	195,545	47,135	2,286	0.417	0.483	0.258
GNSC319	6	87,390,663	195,454	47,138	2,285	0.324	0.086	0.215
GNSC355	6	87,390,632	198,130	47,349	2,467	0.135	0.006	0.000
HH1	5	63,150,400	249,997	61,234	3,867	0.011	0.000	0.000
HH3	8	95,410,507	3,567	371	33	0.019	0.000	0.000
HH4	1	1,277,227	3,580	371	33	0.002	0.000	0.000
JH1	15	15,707,169	199,150	48,109	3,288	0.000	0.106	0.000
KappaCasein12951	6	87,390,459	2,873	211	33	0.000	0.000	0.000
LeptinA59V	4	93,264,030	194,152	47,138	2,260	0.281	0.037	0.140
Leptin_C963T	4	93,248,896	193,962	47,079	2,260	0.416	0.483	0.137
Leptin_T945M	3	80,072,356	194,379	47,136	2,262	0.296	0.270	0.351
MC1R358	18	14,757,910	195,065	46,951	2,276	0.070	0.003	0.001
MC1R373	18	14,757,925	113,623	27,286	1,535	0.047	0.010	0.011
MC1R_EBR	18	14,757,740	2,827	209	32	0.005	0.117	0.063
SDM	11	14,742,058	196,107	47,206	2,287	0.000	0.000	0.022
SMA	24	62,138,835	195,406	47,060	2,240	0.000	0.000	0.020
YellowFat-FB	15	22,877,552	2,874	240	33	0.001	0.063	0.000

[1]ABCG2_1 = ATP binding cassette, subfamily G, member 2/*ABCG2*; BCN_8219 = β casein A3/*CSN2* (casein β); BCN-A2 = β casein A2/*CSN2* (casein β); BCNAB = β casein A/B/*CSN2* (casein β); BetaLact = β lactoglobulin, aberrant low expression/*LGB* (progestagen-associated endometrial protein); BGHR = bovine growth-hormone receptor; BLAD = bovine leukocyte adhesion deficiency/*ITGB2* (integrin, β 2); CAPN1_1, CAPN1_2 = calpain 1, large subunit/*CAPN1*; DGAT_2 = diacylglycerol O-acyltransferase 1/*DGAT1*; Dominant_Red = Holstein HDR haplotype for dominant red coat color/*COPA* (coatomer protein complex, subunit α); DUMPS = deficiency of uridine monophosphate synthase/*UMPS* (uridine monophosphate synthetase); EXON2FB, LeptinA59V, Leptin_C963T = leptin/*OB*; GNSC319, GNSC 355 = kappa casein GNSC 319/355/*CSN3* (casein kappa); JH1 = Jersey fertility haplotype 1/*CWC15* (CWC15 spliceosome-associated protein homolog); HH1 = Holstein fertility haplotype 1/*APAF1* (apoptotic peptidase activating factor 1); HH3 = Holstein fertility haplotype 3/*SMC2* (structural maintenance of chromosomes 2); HH4 = Holstein fertility haplotype 4/*GART* (glycinamide ribonucleotide formyltransferase); KappaCasein12951 = κ casein 12951/*CSN3* (casein κ); Leptin_T945M = leptin receptor/*OB-R*; MC1R358, MC1R373, MC1R_EBR = red/black coat color (Holstein haplotype HBR)/*MC1R* (melanocortin 1 receptor); SDM = spinal dysmyelination (Brown Swiss haplotype BHD)/*SPAST* (spastin); SMA = spinal muscular atrophy (Brown Swiss haplotype BHM)/*KDSR* (3-ketodihydrosphingosine reductase); YellowFat-FB = yellow fat/*BCO2* (β-carotene oxygenase 2).

[2]Counts as of September 16, 2015.

one. The well-known QTL *DGAT1* (diacylglycerol O-acyltransferase 1) and *ABCG2* (ATP binding cassette, subfamily G, member 2) were included. For SNP that were on several chips, the number of genotypes was much higher. Some SNP had MAF of 0 (e.g., DUMPS, deficiency of uridine monophosphate synthase) but were still included because the recessive condition is routinely reported. Some SNP were relevant to specific breeds, such as the haplotypes that affect fertility (HH1, HH3, HH4, and JH1). The current number of GH2 genotypes that have been submitted for Holsteins is sufficient for imputation; for other breeds, imputation accuracy is lower but will increase because the number of GH2 genotypes is increasing. Building on the 77,321 SNP set tested in the cutoff studies, the addition of single-gene tests and structural variant SNP increased the total to 77,531 SNP to be included in genomic evaluations. With genotypes for causative variants (particularly *DGAT1* and *ABCG2*) now available, SNP effect estimation should be adjusted to reflect their greater accuracy. A procedure to give their genotypes greater weight is under consideration.

The addition of SNP that are causative variants is expected to continue as they are revealed by investigation of sequence data. Many causative variants for a trait may exist, even within the same gene. Better predictive tools based on biological knowledge would speed genotype-to-phenotype association. To avoid the delay from including such SNP on a chip and then receiving enough genotypes from the new chip to allow accurate imputation, the sequence data may be used directly in evaluations. A 2-step process might be implemented where sequence data are used to impute the new SNP for the bull predictor population using a large number of SNP to maximize imputation accuracy. Those imputed values would then be included in the normal imputation process to provide imputed genotypes for all animals. Brøndum et al. (2015) reported an increase in reliability of 4 percentage points by including 1,623 QTL markers when using genomic BLUP to evaluate production traits of Nordic Holsteins. VanRaden and O'Connell (2015) reported a substantial increase in accuracy in a simulation as causative variants were included and found that the total number of SNP could be reduced to increase accuracy further.

CONCLUSIONS

A new set of 77,531 SNP for use in national genomic evaluation was created based on 138,942 SNP from the GH2 chip as well as SNP on the GP3 chip and the Zoetis low-density chip. Evaluation of 77,321 of those SNP using data from August 2011 to predict April 2015 performance resulted in a reliability gain over using 60,671 SNP of 1.4 percentage points across traits for Holsteins. Revision of the set of SNP included in genomic evaluation is expected to be an ongoing process to increase evaluation accuracy by substituting more informative SNP for less informative ones, adding causative variants as they are discovered, and eliminating SNP found to contribute mostly noise.

ACKNOWLEDGMENTS

The authors thank the Council on Dairy Cattle Breeding (Bowie, MD) for supplying pedigree, performance, and genotypic data; D. Pomp (University of North Carolina, Chapel Hill) for insight, advice, and leadership on developing version 2 of the GeneSeek Genomic Profiler HD for Dairy Cattle BeadChip; and S. M. Hubbard (Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD) for technical manuscript review. The project was supported by USDA Agricultural Research Service appropriated project 1245-31000-101-00, "Improving Genetic Predictions in Dairy Animals Using Phenotypic and Genomic Information."

REFERENCES

Brøndum, R. F., G. Su, L. Janss, G. Sahana, B. Guldbrandtsen, D. Boichard, and M. S. Lund. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. J. Dairy Sci. 98:4107–4116.

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) 6:80–92.

Ensembl. 2013. Cow assembly and gene annotation. Accessed Sep. 24, 2015. http://apr2013.archive.ensembl.org/Bos_taurus/Info/Annotation.

Illumina Inc. 2011. GoldenGate Bovine3K Genotyping BeadChip. Accessed Nov. 18, 2015. http://www.illumina.com/documents/products/datasheets/datasheet_bovine3k.pdf.

Illumina Inc. 2012. BovineSNP50 Genotyping BeadChip. Accessed Sep. 24, 2015. http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp5O.pdf.

Illumina Inc. 2015a. BovineHD Genotyping BeadChip. Accessed Sep. 24, 2015. http://www.illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf.

Illumina Inc. 2015b. BovineLD v2.0 Genotyping BeadChip. Accessed Nov. 18, 2015. http://www.illumina.com/documents/products/datasheets/datasheet_bovineLD.pdf.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Neogen Corporation. 2013a. GeneSeek® Genomic Profiler™ for Dairy Cattle. Accessed Sep. 24, 2015. http://www.neogen.com/Genomics/pdf/Slicks/GGP-LD_Dairy.pdf.

Neogen Corporation. 2013b. GeneSeek® Genomic Profiler™ HD for Dairy Cattle. Accessed Sep. 24, 2015. http://www.neogen.com/Genomics/pdf/Slicks/GGP_HD_Dairy.pdf.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.
VanRaden, P. M. 2015. findhap.f90, Find haplotypes and impute genotypes using multiple chip sets and sequence data. Accessed Nov. 18, 2015. http://aipl.arsusda.gov/software/findhap/.
VanRaden, P. M., and J. R. O'Connell. 2015. Strategies to choose from millions of imputed sequence variants. Interbull Bull. 49:10–13.
VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24.
Wiggans, G. R., T. A. Cooper, D. J. Null, and P. M. VanRaden. 2014. Increasing the number of single nucleotide polymorphisms used in genomic evaluations of dairy cattle. Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada, Comm. 301. Am. Soc. Anim. Sci., Champaign, IL. Accessed Sep. 24, 2015. https://asas.org/docs/default-source/wcgalp-proceedings-oral/301_paper_9522_manuscript_742_0.pdf.
Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. J. Dairy Sci. 94:3202–3211.
Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 10:R42.