J. Dairy Sci. 96:3336–3339
http://dx.doi.org/10.3168/jds.2012-6208

Short communication: Relationship of call rate and accuracy of single nucleotide polymorphism genotypes in dairy cattle[1]

T. A. Cooper,[2] G. R. Wiggans, and P. M. VanRaden
Animal Improvement Programs Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705-2350

ABSTRACT

Call rates on both a single nucleotide polymorphism (SNP) basis and an animal basis are used as measures of data quality and as screening tools for genomic studies and evaluations of dairy cattle. To investigate the relationship of SNP call rate and genotype accuracy for individual SNP, the correlation between percentages of missing genotypes and parent-progeny conflicts for each SNP was calculated for 103,313 Holsteins. Correlations ranged from 0.14 to 0.38 for the BovineSNP50 and BovineLD (Illumina Inc., San Diego, CA) and GeneSeek Genomic Profiler (Neogen Corp., Lincoln, NE) chips , with lower correlations for newer chips. For US genomic evaluations, genotypes are excluded for animals with a call rate of <90% across autosomal SNP or <80% across X-specific SNP. Mean call rate for 220,175 Holstein, Jersey, and Brown Swiss genotypes was 99.6%. Animal genotypes with a call rate of ≤99% were examined from the US Department of Agriculture genotype database to determine how genotype call rate is related to accuracy of calls on an animal basis. Animal call rate was determined from SNP used in genomic evaluation and is the number of called autosomal and X-specific SNP genotypes divided by the number of SNP from that type of chip. To investigate the relationship of animal call rate and parentage validation, conflicts between a genotyped animal and its sire or dam were determined through a duo test (opposite homozygous SNP genotypes between sire and progeny; 1,374 animal genotypes) and a trio test (also including conflicts with dam and heterozygous SNP genotype for the animal when both parents are the same homozygote; 482 animal genotypes). When animal call rate was ≤80%, parentage validation was no longer reliable with the duo test. With the trio test, parentage validation was no longer reliable when animal call rate was ≤90%. To investigate how animal call rate was related to genotyping accuracy for animals with multiple genotypes, concordance between genotypes for 1,216 animals that had a genotype with a call rate of ≤99% (low call rate) as well as a genotype with a call rate of >99% (high call rate) were calculated by dividing the number of identical SNP genotype calls by the number of SNP that were called for both genotypes. Mean concordance between low- and high-call genotypes was >99% for a low call rate of >90% but decreased to 97% for a call rate of 86 to 90% and to 58% for a call rate of <60%. Edits on call rate reduce the use of incorrect SNP genotypes to calculate genomic evaluations.

Key words: call rate, genotype accuracy, genomic evaluation

Short Communication

Call rate has been used as a measure of quality on both a SNP and animal basis since SNP genotypes were first used in genomic evaluation of dairy cattle (Wiggans et al., 2011). For an individual SNP to be used in US genomic evaluation, the SNP call rate must be above a threshold that increases from 90 to 100% as minor allele frequency declines from 50 to 0% (Wiggans et al., 2010). Animal genotypes with call rates below a cut-off threshold are excluded from genomic evaluation. Worldwide, that threshold ranges from 80 to 95% (Interbull, 2011); the US threshold is 90% for autosomal SNP and 80% for X-specific SNP (Wiggans et al., 2010). If an animal is genotyped more than once, the genotype with higher animal call rate is used. The animal genotype will refer to an observed set of called SNP from 1 chip applied to 1 animal.

Animal genotypes that fail initial quality control by genotyping laboratories [GeneSeek Inc., Lincoln, NE; Genetic Visions Inc., Middleton, WI; DNA LandMarks Inc., Saint-Jean-sur-Richelieu, QC, Canada; and Zoetis Genetics (formerly Pfizer Animal Genetics), Kalamazoo, MI] are not submitted to the US Department of Agriculture for use in national genomic evaluations of dairy cattle. However, some laboratories have submitted genotypes with call rates of <90% for potential

Received September 27, 2012.
Accepted January 17, 2013.
[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 the US Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.
[2]Corresponding author: Tabatha.Cooper@ars.usda.gov

research use or to be provided to other countries that have lower call rate thresholds for genotypes that are used in genomic evaluation. Changing the edit limit to accept or reject 1% more animal genotypes than currently allowed would have a financial impact of 0.01 × 300,000 × $50 = $150,000 (with >300,000 genotyped animals in the database) when applied to the current database, assuming a $50 average value across chips purchased. Additional information on the tradeoff between call rate and accuracy is needed.

Genotypes in the US Department of Agriculture genotype database were examined to determine how call rate relates to SNP genotype accuracy. Because genotypes from the Illumina Bovine3K BeadChip (Illumina Inc., 2011b) are known to have lower call rates and more parent-progeny conflicts (Boichard et al., 2012; Wiggans et al., 2012), only genotypes from the Illumina BovineSNP50 (Illumina Inc., 2011a), Illumina BovineHD (Illumina Inc., 2010), Illumina BovineLD (Boichard et al., 2012), and GeneSeek Genomic Profiler (**GGP**; Neogen Corporation, 2012) BeadChips were included.

To investigate the relationship of SNP call rate and genotype accuracy for individual SNP, the correlation between percentages of missing SNP genotypes and parent-progeny conflicts for each SNP (Wiggans et al., 2010) was calculated for 103,313 genotyped Holsteins (Table 1) by genotyping chip. Correlations ranged from 0.14 to 0.38 for the BovineSNP50, BovineLD, and GGP chips. Positive correlations indicate that SNP with a higher rate of missing SNP genotypes also have a higher rate of parent-progeny conflicts. The correlation was highest for version 1 of the BovineSNP50 chip. As new versions and genotyping chips became available, only the best-performing SNP were chosen to be included, which accounts for the correlation decrease because SNP more subject to genotyping errors were eliminated.

For each animal genotype, call rate was determined by dividing the number of called autosomal and X-specific SNP genotypes by the total number of SNP from that chip that were usable for genomic evaluation (43,593, BovineSNP50, version 1; 43,289, BovineSNP50, version 2; 40,241, BovineHD; 6,836, Bovine LD; and 8,031, GGP). The mean call rate for 220,175 Holstein, Jersey, and Brown Swiss genotypes was 99.6%. BovineSNP50, BovineLD, and GGP genotypes had similar call rates; for those chips, 80 to 90% of animals had a call rate that rounded to 100%, and 93 to 96% had a call rate of 99 or 100% (Table 2). For BovineHD genotypes, 67% had a 100% call rate, and 89% had a call rate of 99 or 100%. Animals genotyped by different SNP chips were grouped in call rate groups from 100 to 90% with a decrease of 1%. The number of animals within these groups decreased rapidly with a decrease in call rate. Animals with a call rate below the 90% exclusion limit exist, although most of the data submitted for evaluation is truncated at that value.

Table 1. Correlations between the rate of missing genotypes and the number of parent-progeny conflicts on a SNP basis by Illumina (San Diego, CA) or GeneSeek (Neogen Corp., Lincoln, NE) genotyping chip for 103,313 Holsteins

Genotyping chip	SNP[1] (no.)	Correlation
Illumina BovineSNP50K, version 1	42,906	0.38
Illumina BovineSNP50K, version 2	42,608	0.14
Illumina BovineLD	6,632	0.24
GeneSeek Genomic Profiler	7,779	0.14

[1]Autosomal SNP usable for US genomic evaluation.

To investigate the relationship of animal call rate and parentage validation, Mendelian conflicts between a genotyped animal and its sire or dam were determined using the method of Wiggans et al. (2012). Parent-progeny conflicts were counted in 2 ways: only including opposite homozygous SNP genotypes between sire and progeny (duo test), or also including opposite homozygous SNP between dam and progeny and heterozygous SNP for the animal when both parents are the same homozygote (trio test). Many genotyping errors are not detectable even with the trio test. If one parent is homozygous and the other parent heterozygous, both of the parental genotypes are valid for a progeny, and switches between these 2 will not be detected. If both parents are heterozygous, any genotype is valid for a progeny, and no conflict will be detected.

Conflict rate was the number of conflicts detected divided by the number of parent genotypes that allowed detection, which was number of SNP where sire was homozygous (duo) or either parent was homozygous (trio) for the SNP tested. The SNP included were those called for both the sire and progeny (duo) or called for both parents and the progeny (trio).

The percentage of parent-progeny conflicts based on the duo test were calculated for 1,374 animal genotypes with a call rate of <99.0% and a genomically validated sire. A subset of those data (482 animal genotypes) that also had a genomically verified dam was used to calculate the percentage of parent-progeny conflicts based on both duo and trio tests. The percentage of parent-progeny conflicts increased as animal call rate decreased (Table 3). Within call rate group, the percentage of parent-progeny conflicts was higher with the trio test than the duo test for animal call rates of >70%, whereas the reverse was true for call rates of <70%. That result may indicate that called SNP do not change from homozygous for one allele to homozygous for the other allele (e.g., AA to BB); instead, they are

Table 2. Percentages of genotypes[1] by animal call rate for BovineSNP50 and Bovine HD BeadChips (Illumina Inc., San Diego) and the GeneSeek Genome Profiler (GGP; Neogen Corp., Lincoln, NE)

	Genotyping chip				
	BovineSNP50				
Call rate (%)	Version 1 (n = 64,352)	Version 2 (n = 56,891)	BovineLD (n = 62,493)	GGP (n = 33,255)	BovineHD (n = 3,184)
100	79.9	86.8	81.0	89.6	67.3
99	13.2	6.5	11.5	6.4	22.0
98	2.9	1.3	2.8	1.6	6.0
97	1.3	1.9	1.4	0.7	1.8
96	0.7	0.4	0.8	0.5	0.7
95	1.0	0.3	0.5	0.4	0.3
94	0.3	2.1	0.4	0.3	0.2
93	0.2	0.3	0.3	0.2	0.3
92	0.2	0.1	0.2	0.2	<0.1
91	0.1	0.1	0.2	0.1	<0.1
90	<0.1	0.1	0.1	<0.1	0.1
≤89	0.2	0.1	0.9	<0.1	1.3

[1]Includes animal genotypes not used for national genomic evaluations of dairy cattle.

called as heterozygous (i.e., AB), except for the few cases of very low-quality genotypes. A parent-progeny conflict rate of ≤0.46% for animal genotypes is considered to confirm a parent-progeny relationship (Wiggans et al., 2010). On average, when the animal call rate was ≤80% (Table 3), parentage validation was no longer reliable using the traditional homozygous SNP test. If heterozygous SNP also were used to determine parent-progeny conflicts, then parentage validation was no longer reliable when the genotype call rate was ≤90%.

To investigate how animal call rate was related to genotyping accuracy for individual animals, 1,216 animals that had a genotype with a call rate of ≤99% (low call rate) as well as a genotype with a call rate of >99% (high call rate) were used to determine concordance between those genotypes. Concordance was calculated by dividing the number of identical SNP genotype calls (with "no calls" excluded) by the number of SNP that were called for both genotypes. Mean concordance (Table 4) between low- and high-call genotypes was 99.9% when the low call rate ranged from 95 to 99%; concordance decreased slightly to 99.2% when the low call rate was 91 to 95%. When the low call rate was ≤90%, however, concordance began to decrease: 97% for a low call rate of 86 to 90%, 94% for 81 to 85%, 87% for 71 to 80%, 77% for 61 to 70%, and 58% for <60%. To control data quality for US genomic evaluations, 2 animal genotypes were considered to be from different animals if their SNP differed by ≥2.3% for called SNP (Wiggans et al., 2010). Results in Table 4 indicate that genotypes for the same animal could appear to be from different animals if one of the genotypes has a call rate of ≤90% and the other has a call rate of ≥99%.

Genotype call rate is useful as a screening tool for data quality for genomic studies and genomic evaluations and is related to genotype accuracy on a SNP and animal basis. For SNP and for animals, as the percentage of missing genotypes increased, the percentage of parent-progeny conflicts also increased. The accuracy of SNP genotypes that is related to genotype call rate

Table 3. Means, SD, minimums, and maximums for parent-progeny conflicts of genotypes for animals with genomically validated parents by animal call rate group

Genomically validated parent	Call rate (%)	Genotypes (no.)	Parent-progeny conflict (%)			
			Mean	SD	Minimum	Maximum
Sire	≤60	73	20.56	12.54	0.62	55.84
	61 to 70	35	7.89	10.64	0.07	44.72
	71 to 80	71	1.19	1.63	0.00	6.93
	81 to 90	172	0.13	0.25	0.00	1.86
	91 to 99	1,023	0.04	0.06	0.00	0.49
Sire and dam	61 to 70	6	2.87	1.99	1.14	6.14
	71 to 80	14	1.72	1.42	0.01	5.80
	81 to 90	51	0.35	0.50	0.00	2.55
	91 to 99	411	0.09	0.15	0.00	1.81

Table 4. Means, SD, minimums, and maximums for concordance between low- and high-call genotypes[1] for an animal by low call rate group

Low call rate (%)	Animals (no.)	Concordance (%)			
		Mean	SD	Minimum	Maximum
≤60	137	57.5	0.15	31.8	86.3
61 to 70	65	77.0	0.16	28.5	96.9
71 to 80	62	87.3	0.08	64.3	99.4
81 to 85	41	93.6	0.08	72.2	99.9
86 to 90	106	96.7	0.06	77.2	100.0
91 to 95	178	99.2	0.02	81.1	100.0
96 to 99	627	99.9	0.00	99.0	100.0

[1]Low-call genotypes had a call rate of ≤99%; high-call genotypes had a call rate of >99%.

can be detected through parentage validation when the call rate is ≤80% for animals with a genotyped sire. The most common calling error was a single-allele change (i.e., a homozygous allele to a heterozygous allele or the reverse), which can only be detected if both parents are genotyped; parentage was most affected when animal call rate was ≤90%. For animals with 2 genotypes, a concordance of >99% was only found if one genotype had a call rate of >90%. To fill in missing SNP genotypes, accurate input genotypes are required for imputation (Boichard et al., 2012). Genotyping errors interfere with phasing and imputation because the progeny's genotype does not match the haplotypes actually transmitted by the parents. Genomic evaluations based on the regression of copy number of a particular allele at each locus are also affected by accuracy of SNP calls. Errors due to low call rate are most often a one-allele change, which will have an effect equal to the effect size of that particular allele.

Mattalia et al. (2012) also found that a low call rate was associated with lower genotype accuracy when using a stricter threshold of 98% for animal call rate. To balance the risk of lowered accuracy with the inclusion of the maximum number of genotypes, the threshold of 90% for animal call rate will continue to be used for US genomic evaluations. Edits on call rate reduce the use of incorrect SNP genotypes to calculate genomic evaluations. If the parent(s) are genotyped, Mendelian conflicts can be detected to determine if an animal genotype is unreliable (Wiggans et al., 2012) and eliminate conflicts on an individual SNP basis. In the absence of genotyped parents, call rate is even more important. These results allow for selecting an edit threshold to balance call rate, the cost of eliminating data, and accuracy.

REFERENCES

Boichard, D., H. Chung, R. Dassonneville, X. David, A. Eggen, S. Fritz, K. J. Gietzen, B. J. Hayes, C. T. Lawley, T. S. Sonstegard, C. P. Van Tassell, P. M. VanRaden, K. A. Viaud-Martinez, and G. R. Wiggans. 2012. Design of a bovine low-density SNP array optimized for imputation. PLoS ONE 7:e34130.

Illumina Inc. 2010. BovineHD Genotyping BeadChip. Accessed Aug. 31, 2012. http://www.illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf.

Illumina Inc. 2011a. BovineSNP50 Genotyping BeadChip. Accessed Aug. 31, 2012. http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp5O.pdf.

Illumina Inc. 2011b. GoldenGate Bovine3K Genotyping BeadChip. Accessed Aug. 31, 2012. http://www.illumina.com/Documents/products/datasheets/datasheet_bovine3K.pdf.

Interbull. 2011. GEBV test—November 2011. Accessed Sep. 17, 2012. http://www.interbull.org/index.php?option=com_content&view=article&id=93&Itemid=157.

Mattalia, S., S. Barbier, M. Y. Boscher, D. Boichard, and S. Fritz. 2012. French experience with recognition of laboratories providing genotypes to the official genomic evaluation of dairy cattle. 38th ICAR Session, Cork (Ireland). Accessed Sep. 27, 2012. http://www.icar.org/Cork_2012/Manuscripts/Published/Mattalia.pdf.

Neogen Corporation. 2012. GeneSeek Genomic Profiler for Dairy Cattle. Accessed Aug. 31, 2012. http://www.neogen.com/GeneSeek/pdf/Catalogs/DairyGenomicProfiler.pdf.

Wiggans, G. R., T. A. Cooper, P. M. VanRaden, K. M. Olson, and M. E. Tooker. 2012. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. J. Dairy Sci. 95:1552–1558.

Wiggans, G. R., P. M. VanRaden, L. R. Bacheller, M. E. Tooker, J. L. Hutchison, T. A. Cooper, and T. S. Sonstegard. 2010. Selection and management of DNA markers for use in genomic evaluation. J. Dairy Sci. 93:2287–2292.

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.