534 Use of genomic recursions in single-step genomic BLUP with a large number of genotypes. Breno D. Fragomeni*[1], Daniela A. L. Lourenco[1], Shogo Tsuruta[1], Yutaka Masuda[1], Ignacio Aguilar[2], Andres Legarra[3], Thomas J. Lawlor[4], and Ignacy Misztal[1], *[1]Department of Animal and Dairy Science, University of Georgia, Athens, GA, [2]Instituto Nacional de Investigacion Agropecuaria, Las Brujas, Canelones, Uruguay, [3]INRA, UMR1388 GenePhySE, Castanet Tolosan, France, [4]Holstein Association USA Inc., Brattleboro, VT.*

The purpose of this study was to evaluate accuracy of genomic selection in single-step genomic BLUP (ssGBLUP) when the inverse of the genomic relationship matrix (G) is derived by the APY (algorithm for proven and young animals). This algorithm implements genomic recursions on a subset of "proven" animals. Only a matrix due to the subset needs to be inverted and extra costs of adding "young" animals are linear. Analyses involved 10,102,702 Holsteins with final scores on 6,930,618 cows. A total of 100k animals with genotypes included 23k sires (16k with more than 5 progenies), 27k cows, and 50k young animals. Genomic EBV (GEBV) were calculated with a regular inverse of G, and with the G inverse approximated by APY. Initially, animals in the "proven" subset included only sires or cows. Later, animals in the "proven" subset were randomly sampled from all genotyped animals in sets of 5k, 10k, and 20k; each sample was replicated 4 times. Genomic EBV with APY were accurate when the number of animals in the "proven" subset was ≥10k, with little difference between the ways of creating the subset. Numerical properties as shown by the number of rounds to convergence were best with random subsets. The ssGBLUP with APY can accommodate a large number of genotypes at low cost and with high accuracy.

Table 1 (Abstr. 534). Correlations (or range of correlations) between genomic EBV with regular and APY ssGBLUP for young genotyped animals and rounds to convergence for different subset of animals used in recursions

Definition of subset	Animals in subset	Correlation	Rounds to convergence
All	100,000	1.00	567
Sires	23,174	0.99	432
Cows	27,215	0.99	797
2k random	2,000	0.94	356
5k random	5,000	0.97	360
10k random	10,000	0.99	396
20k random	20,000	0.99	420

Key Words: single-step method, genomic selection, genomic recursion

535 Genomic predictions with approximated G-inverse for a large number of genotyped animals. Yutaka Masuda*[1], Ignacy Misztal[1], Shogo Tsuruta[1], Daniela A. L. Lourenco[1], Breno Fragomeni[1], Andres Legarra[2], Ignacio Aguilar[3], and Tom J. Lawlor[4], *[1]University of Georgia, Athens, GA, [2]INRA, Castanet-Tolosan Cedex, France, [3]Instituto Nacional de Investigación Agropecuaria, Canelones, Uruguay, [4]Holstein Association USA Inc., Brattleboro, VT.*

The objective of this study was to compare the accuracy of genomic predictions in final score for young Holstein bulls calculated from single-step GBLUP models with the regular G^{-1} and approximated G^{-1} (G^{-1}_{ap}) matrices. The G^{-1}_{ap} was calculated with recursions on a small subset of animals. The regular G^{-1} has a quadratic memory cost and cubic computational cost as the number of genotyped animals increases, whereas G^{-1}_{ap} has a linear cost for animals outside the subset. The predictor data set consisted of 77,066 genotyped animals, 9,009,998 pedigree animals, and 6,384,859 classified cows born in 2009 or earlier. For calculation of G^{-1}_{ap}, 9,406 high accuracy bulls or 16,828 high accuracy bulls and cows were used as the small subset. Genomic predictions (GEBV2009) were calculated for predicted bulls that had no classified daughters in 2009 but did in 2014. The validation data set contained phenotypes and pedigree recorded up to March 2014. Daughter yield deviations (DYD2014) were calculated for the predicted bulls with at least 30 daughters in 2014 (n = 2,948). Coefficient of determination (R^2), calculated from a linear regression of DYD2014 on GEBV2009, was 0.44 with the regular G^{-1} and 0.45 with G^{-1}_{ap} for either subset. Genomic predictions using all available 569,404 (570K) genotypes were also calculated with G^{-1}_{ap} on 9,406 bulls. The computation was performed using 16 CPU cores and 61 G bytes of working memory. Setting up G^{-1}_{ap} took 1.8 h, setting up matrices associated with A_{22}^{-1} took 7 min, and iterations took 4.6 h, resulting in 6.4 total hours. In contrast, BLUP computations without genotypes took 1.6 h in total. The R^2 value from 570K genotypes was similar to the result from GEBV2009. Genetic predictions can be obtained with substantially less computational cost but without loss of reliability using G^{-1}_{ap}. The single-step GBLUP with G^{-1}_{ap} is applicable to very large genotyped populations.

Key Words: computing, Holstein, single-step genomic BLUP (ssGBLUP)

536 Theoretical aspects of the APY algorithm for inverting a large genomic relationship matrix. Ignacy Misztal*, *University of Georgia, Athens, GA.*

The algorithm for proven and young animals (APY) implements the inverse of the genomic relationship matrix by recursion on a subset of animals. If the subset is small, storage and computations are approximately linear with the number of genotyped individuals, allowing for processing of practically an unlimited number of animals. The APY algorithm was tested with many subsets, including proven bulls, bulls and cows, cows only and random subsets. GEBVs calculated with APY were accurate when the number of animals in the subset was ≥10k, with little difference between different subsets. Best convergence rates when solving equations with APY were obtained with subsets composed of randomly chosen animals. The properties of the APY algorithm can be explained using the concept of a finite number of independent chromosome segments. Assume that each segment has a fixed value, that a fraction of each segment has a value proportional to the length of that segment (infinitesimal model), and that a genome of an individual is composed of a fraction of each segment. Subsequently, in the absence of confounding, n animals allow for identification of a population with n segments. Assuming some errors, lowest estimation errors are with heterogeneous animals. For traits where genes are distributed unequally across the genome, a conceptual division of segments into smaller segments with quasi-equal distribution would result in a larger subset required for the same accuracy of GEBV. If genome sequencing allows for identification of all m QTN, each QTN may be treated as one segment and, assuming a purely additive model, a recursion on m sufficiently heterogeneous animals will capture all variability in the genome. Computation in APY assume ability to compute parts of such a genomic relationship matrix that reflects the genetic architecture for

each trait. The APY algorithm may allow routine genomic evaluation for any number of genotyped animals with any model at a cost not much above BLUP.

Key Words: genome selection, genomic recursion, genomic relationship matrix

537 Effect of increasing the number of single nucleotide polymorphisms from 60,000 to 85,000 in genomic evaluation of Holsteins. George R. Wiggans*, Tabatha A. Cooper, Paul M. VanRaden, Curt P. Van Tassell, Derek M. Bickhart, and Tad S. Sonstegard, *Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD.*

The periodic need to restock reagent pools for genotyping chips provides an opportunity to increase the number of single nucleotide polymorphisms (SNP) on a chip. As an improved replacement for the GeneSeek Genomic Profiler HD for dairy cattle, a set of >140,000 SNP was selected that included all SNP on the current chip, all SNP used in genomic evaluations, SNP that are possible functional mutations, and other informative SNP. Most added SNP were selected from the Illumina Bovine HD Genotyping BeadChip based on the magnitude of effects on evaluated traits. Some SNP with lower minor allele frequency were considered because of their potential for better tracking of causative variants. Genotypes already available from other chips 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), which resulted in 78,032 SNP after removing duplicates. An additional 9,130 SNP with many parent-progeny conflicts after imputation were removed, which resulted in 72,843 SNP. Of those, 38,515 were among the 60,671 SNP currently used in genomic evaluation. To minimize possible accuracy loss,12,094 of the SNP currently used but not already selected and with the largest effects were added for a total of 84,937 SNP. Three cutoff studies were conducted with 60,671, 84,937, and 134,511 SNP to determine gain in reliability over parent average when evaluations based on data from August 2011 were used to predict genetic merit from December 2014. Across all traits, mean gains were 32.5, 33.4, and 32.0 percentage points, respectively. Previous experience indicates that gains from the highest number of SNP will increase as the number of genotypes from the new SNP set increases. The gain of 0.9 percentage points from adding nearly 25,000 SNP justifies the extra computation time needed. However, the gain may be overestimated because data used to select the most informative SNP were also the data used to determine gain.

Key Words: dairy cattle, genomic evaluation, single nucleotide polymorphisms

538 Genome-wide association study of fertility traits in dairy cattle using high-density single nucleotide polymorphism marker panels. Kristen L. Parker Gaddis[1] and John B. Cole*[2], *[1]Department of Animal Sciences, University of Florida, Gainesville, FL, [2]Animal Genomics and Improvement Laboratory, ARS, USDA, Beltsville, MD.*

Unfavorable genetic correlations between production and fertility traits are well documented. Genetic selection for fertility traits is slow, however, due to low heritabilities. Identification of single nucleotide polymorphisms (SNP) involved in reproduction could improve reliability of genomic estimates for these low heritability traits. Additionally, high-density marker panels can increase the power of resultant GWAS by providing increased coverage and stronger linkage disequilibrium between markers and causal variants. The objective of this study was to identify SNP associated with 3 fertility traits in dairy cattle, daughter pregnancy rate (DPR), heifer conception rate (HCR), and cow conception rate (CCR), using high-density marker panels. Deregressed predicted transmitting abilities were available for 10,000 bulls sampled from the National Dairy Database that had high-density genotypes. Of those, 725 had been genotyped with the Illumina BovineHD Genotyping BeadChip. The remaining bulls had genotypes from various chip densities that were imputed up to the same level. After editing, 312,614 markers were included in the analyses. Univariate analyses were performed for DPR, HCR, and CCR using REMLF90 (version 1.79) with genomic options. postGSf90 (version 1.170) was used to calculate SNP effects and 10-SNP window variances. The largest proportion of variance explained for DPR (0.126%) was located on chromosome 6. Peaks were also identified on chromosomes 5, 18, and 28 associated with DPR. For HCR, the region explaining the largest proportion of variance (0.155%) was located on chromosome 1. Large peaks were also identified for HCR on chromosomes 6, 8, 14, and 17. The largest proportion of variance explained for CCR (0.181%) was located on chromosome 18. Large peaks associated with CCR were also identified on chromosomes 6, 15, and 19. Numerous markers and regions aligned with those previously identified. Significant SNPs could be used in genomic selection programs as well as in identification of genes and networks involved with fertility.

Key Words: fertility, genomic evaluation, high-density genotype

539 Segment-based methods to calculate weights for weighted single-step GBLUP. Xinyue Zhang*, Daniela A. L. Lourenco, and Ignacy Misztal, *University of Georgia, Athens, GA.*

The purpose of this study was to explore additional options for calculating weights in weighted single-step GBLUP (WssGBLUP). In GWAS by ssGBLUP, GEBV are converted to marker (SNP) effects. Unequal variances for markers are then derived from SNP solutions and subsequently incorporated into a weighted genomic relationship matrix. Improvements on the SNP weights were obtained iteratively by recomputing both the SNP effects and the GEBV. Six options were used to calculate the weights: (1) proportional to u_i^2 where u_i is the effect of the i-th SNP; (2) proportional to u_i^2 + constant; (3) weights as $\nu^{|s-2|}$, where ν is a scale standing for the departure from normality, and *s* is number of standard deviation from mean for each u_i^2 where p_i is frequency of the second allele; (4) as the largest effect (u_i^2) among every 20 SNP; (5) as the mean effect of every 20 SNP; (6) as the summation of effects of every 20 SNP. A simulated data set was used that included 15,600 animals in 5 generations, of which 1,540 were genotyped for 50k SNP. The simulation involved phenotypes for a trait with heritability of 0.5 and affected by 5, 100, and 500 QTL. Accuracy between TBV and GEBV for genotyped animals in the last generation was used for evaluation. Comparisons also involved BayesB and BayesC with deregressed proofs or EBV from BLUP, and $\pi = 0.99$, 0.9 or 0.5. In single-step, SNP effects were tracked along 10 iterations and weights were equal to 1.0 in the first iteration. Option 5 was the best in identifying simulated QTL without background noise and with precision in most of the regions. Option 2 kept accuracy of GEBV at the plateau after 2 iterations and was 0.81 as opposed to 0.70 for BayesC and 0.48 for BayesB under 500 QTL scenario. All methods reached better accuracies than BayesB and BayesC when number of QTL approached or exceeded 100 (0.2% of all SNP) due to automatically including PA in GEBV. Weights based on a sum of SNPs may be superior to those based on individual SNPs.

Key Words: weighted SNP, single-step genomic BLUP (ssGBLUP), BayesB

540 Multi-allelic haplotype model based on genetic partition for genomic prediction and variance component estimation. Yang Da*, *Department of Animal Science, University of Minnesota, Saint Paul, MN.*

Functional genomic information has been growing rapidly but remains largely unused in genomic selection. Genomic prediction and estimation using haplotypes in genome regions with functional elements such as all genes of the genome can be an approach to integrate functional genomic information with genomic selection. Toward this goal, a multi-allelic haplotype model treating each haplotype as an 'allele' was developed for genomic prediction and estimation based on the partition of a multi-allelic genotypic value into additive and dominance values. Each additive value is expressed as a function of h-1 additive effects, where h = number of alleles or haplotypes, and each dominance value is expressed as a function of h(h − 1)/2 dominance effects. For a sample of q individuals, the limit number of effects is 2q − 1 for additive effects and is the number of heterozygous genotypes for dominance effects. Additive values are factorized as a product between the additive model matrix and the h − 1 additive effects, and dominance values are factorized as a product between the dominance model matrix and the h(h-1)/2 dominance effects. Genomic additive relationship matrix is defined as a function of the haplotype model matrix for additive effects, and genomic dominance relationship matrix is defined as a function of the haplotype model matrix for dominance effects. Based on these results, a mixed model implementation for genomic prediction and variance component estimation that jointly use haplotypes and single SNPs is established, including 2 computing strategies for genomic BLUP (GBLUP) and genomic REML (GREML) with identical results. The multi-allelic genetic partition fills a theoretical gap in genetic partition by providing general formulations for partitioning multi-allelic genotypic values and provides a haplotype method based on the quantitative genetics model toward the utilization of functional genomic information for genomic selection.

Key Words: haplotype, GBLUP, GREML

541 Revisiting allelic frequencies estimation: A decision theory approach to derive Bayes, minimax, and admissible estimators. Carlos A. Martinez*[1,2], Kshitij Khare[2], and Mauricio A. Elzo[1], *[1]Department of Animal Sciences, University of Florida, Gainesville, FL, [2]Department of Statistics, University of Florida, Gainesville, FL.*

Decision theory was used to derive point estimators of allelic frequencies with optimal statistical properties. Uniparameter (2 alleles) and multiparameter (multiple alleles) estimation problems were addressed for an arbitrary locus, then results were extended to multiple loci. First, estimators satisfying the Bayes principle of average risk optimality were obtained using a multinomial sampling model, a Beta (biallelic loci) and a Dirichlet (multiallelic loci) prior and 3 different loss functions: Squared Error Loss (SEL), Kullback-Leibler Loss (KLL) and a Quadratic Error Loss (QEL). Second, these Bayes estimators were used to obtain minimax estimators by finding values of the hyperparameters such that the frequentist risk functions were constant, a condition that implies minimaxity. Finally, the admissibility of the estimators was checked using standard theorems from decision theory. The frequentist risk function of the Bayes estimator derived from KLL involved a finite sum without a closed form, hence this risk function could not be written as a simple algebraic expression. However, this does not prevent its computation. Under SEL and QEL it was possible to find Bayes-minimax-admissible estimators (BMAE). Sufficient conditions for the usual maximum likelihood estimator to be BMAE and for the risk functions tending to infinite or converging to zero were also found. In addition to optimal statistical properties, these estimators have the appealing feature of taking into account random variation in allelic frequencies. The impact of using these estimators in different areas of quantitative and population genetics needs be assessed either empirically or theoretically and this poses a problem for further research.

Key Words: admissibility, allele frequency, minimaxity

542 Strategies for estimating hyperparameters based on single-step Bayesian models. Lei Zhou* and Robert J. Tempelman, *Michigan State University, East Lansing, MI.*

Single-step BLUP (SS-BLUP) genomic prediction has the advantage of combining phenotypic information on both genotyped and nongenotyped animals. Recently developed single-step Bayesian regression models (SSBR) provide potentially even greater flexibility (e.g., heavy-tailed or variable selection alternatives) for modeling the prior distribution of SNP markers in tandem with polygenic effects. We discuss and present strategies for inferring upon these hyperparameters, particularly when data from either category (i.e., genotyped or nongenotyped) of animals is limiting. For example, when most animals in a genetic evaluation are not genotyped, inferences on key hyperparameters (i.e., marker variance, scale parameters) are compromised if these inferences are primarily based on marker data from genotyped animals only (strategy 1) whereas information on these hyperparameters might be readily borrowed from polygenic inferences involving nongenotyped animals as well (strategy 2). To compare these 2 strategies, a simulation study with 10 replicates was conducted. Five generations and a total of 2000 animals were simulated in each replicate. The heritability of the trait was 0.5 based on 1500 SNPs. The proportion of animals genotyped ranged from 10 to 90%. Results showed that strategy 2 estimated hyperparameters with greater accuracy than strategy 1 (Table 1), particularly when the proportion of animals that were genotyped was low (i.e., 10%). Nevertheless both strategies lead to similar accuracies of estimated breeding values for both genotyped and non-genotyped animals under the various genotyping rate scenarios. We also present methodology for both strategies using REML to infer upon these hyperparameters based on Gaussian specifications as well as how to infer upon hyperparameters in heavy-tailed (BayesA) and/or variable selection (BayesB) specifications. In conclusion, our proposed strategy had some advantage upon inferring hyperparameters, and further research is necessary for SSBR models.

Table 1 (Abstr. 542). Genetic variance by scenario and strategy

Scenario	Strategy	Total genetic variance (mean ± SD)
		24.60 ± 4.48 (True)
10% genotyped	1	24.92 ± 5.64
	2	25.70 ± 4.33
50% genotyped	1	22.35 ± 4.70
	2	24.05 ± 2.87
90% genotyped	1	20.73 ± 7.81
	2	24.26 ± 3.22

Key Words: single-step, Bayesian, hyperparameter

543 Reassessing hierarchical Bayesian genome-wide association analyses. C. Chen*, J. P. Steibel, and R. J. Tempelman, *Michigan State University, East Lansing, MI.*

Genomic best linear unbiased prediction (GBLUP) analyses have been increasingly adapted for genome-wide association (GWA) analyses. A

currently popular modification of GBLUP for GWA, which we label as classic GBLUP, is to treat all genetic markers as random, except for the marker being tested; conversely, shrinkage GBLUP treats all markers as random in a whole genome prediction (WGP) analysis. The classic GBLUP modification has been demonstrated to preserve Type I error rates whereas shrinkage GBLUP leads to a very conservative GWA test. Nevertheless, shrinkage estimation has recently been shown to have GWA properties under alternative prior specifications. Some popular WGP model specifications are heavy-tailed (i.e., BayesA) or involve variable selection (i.e., BayesSSVS) that do not shrink large marker effects as much as shrinkage GBLUP. Given that MCMC implementations of these models are computationally onerous, we propose inferences under these alternative priors based on the EM algorithm (i.e., EMBayesA and EMBayesSSVS). In a simulation involving 10 replicated data sets, each involving about 2000 individuals and 5000 SNP markers with average pairwise LD $r^2 = 0.30$ across 5 chromosomes, we discovered that EMBayesA and EMBayesSSVS shrink the majority of the posterior *z*-score based *P*-values to be larger relative to classic GBLUP, whereas markers in QTL regions tend to have substantially smaller *P*-values in EMBayesA and EMBayesSSVS compared with classic GBLUP. In an application involving backfat data from a Duroc-Pietrain F2 cross, we determined that EMBayesSSVS inferred SNP effects in albeit fewer putative QTL regions compared with classical GBLUP, although GWA using EMBayesA did not detect any such association. We suggest that EMBayesSSVS or other hierarchical variable selection models represent promising alternatives for GWA analyses of complex traits for which null marker effects might not appropriately represent a global null hypothesis; however, we also demonstrate that recently developed regularization techniques are vitally important in helping avoid posterior multimodality concerns in large dimensional EM-based inferences as well.

Key Words: expectation maximization, genome-wide association (GWA)

544 Approximating realized additive relationships in absence of genomic information. Romdhane Rekaya*, Sajjad Toghiani, and L. Y. Chang, *The University of Georgia, Athens, GA.*

Superiority of genomic selection is due to the use of the realized relationships. Unfortunately, only a small portion of animals included in the evaluation are genotyped. Consequently, the majority of animals will still be evaluated based on their expected relationships. With the availability of genomic information, it is possible to assess the variation of expected co-ancestry coefficients for different orders (degrees) of relationships. Thus, this available information represents a good prior and could be used to better assess the Mendelian sampling in absence of genomic data. Furthermore, a sizeable portion of non-genotyped animals are already phenotyped for several traits with moderate to high heritabilities. Phenotypic records contain information about the true genetic relationship between animals. Using these 2 sources of information, we can improve the pedigree based relationship without using genomic data. A simulation was carried out to investigate the ability to infer realized additive relationships in absence of genomic information. A single trait with heritability of either 0.5 or 0.8 was simulated. A pedigree with large number of full (FS) and half (HS) sibs reflective of a chicken population was also simulated. All animals in the pedigree were assigned 60K SNP genotypes. The average (A) and the realized (G) matrices were computed using the pedigree and the genomic information, respectively. Using our procedure a realized relationship matrix (G*) was computed without use of the simulated genomic information. The results showed that the similarity between G and G*, measured by the absolute value of difference between corresponding elements of both matrices, was greater than between G and A. The average difference between G and A was 0.07 compared with 0.03 between G and G*. The realized relationships for FS and HS estimated using our proposed method overlap largely with those obtained using genomic information. Additionally, the similarity between G and G* increased, as expected, with the increase of heritability. The fact that the difference between G and G* is smaller than between G and A indicate that the matrix G* is a better approximation of G than A.

Key Words: realized, relationship, genomic

545 Imputation using whole-genome sequence data in Brown Swiss and Original Braunvieh. Christine F. Baes*[1,2], Beat Bapst[2], Franz R. Seefried[2], Heidi Signer-Hasler[1], Christine Flury[1], Dorian Garrick[3], Christian Stricker[4], and Birgit Gredler[2], *[1]Bern University of Applied Sciences, Zollikofen, Bern, Switzerland, [2]Qualitas AG, Zug, Zug, Switzerland, [3]Iowa State University, Ames, IA, [4]agn Genetics, Davos, Grison, Switzerland.*

The distinct half-sib population structure common in dairy cattle populations permits imputation of array-based genotypes up to sequence level using sequence information of key ancestors. This approach provides a cost-effective alternative to sequencing all animals. The process works well in single breeds (e.g., Holstein), however it is less accurate when validation and reference animals are not of the same breed or population (e.g., Brown Swiss or Original Braunvieh). The objective of this study was therefore to investigate imputation strategies for composing reference and validation sets using Brown Swiss (BS) and Original Braunvieh (OB) animals. Whole genome sequence information (WGS) of 70 BS, 17 BS × OB and 8 OB animals was available for analysis. WGS was masked to mimic medium (50K; 54,609 SNP), and high-density (HD; 777,962 SNP) SNP arrays and then imputed back up to sequence level. Imputation was conducted using fimpute, which employs an overlapping sliding window approach to exploit haplotype similarities between validation and reference animals. Principal component analysis was used to determine which animals constituted reference and validation sets; scenarios involving various breed compositions were compared. The accuracy of imputation from 50K and HD to WGS was evaluated for each scenario by calculating the number of correctly imputed genotypes and concordance between true and imputed genotypes. Furthermore, imputation accuracy when the reference population was solely comprised of the breed to be imputed was compared with results obtained when the reference population was comprised of multiple breeds. Validation of OB genotypes resulted in 64.6–69.6% (50K to WGS) and 79.5–83.8% (HD to WGS) correctly imputed genotypes. BS validation resulted in 78.0–78.9% (50K to WGS) and 85.5–86.1% (HD to WGS) correctly imputed genotypes. Validation of intermediary animals using both BS and OB as reference resulted in the highest percentage of correctly imputed genotypes (79.1–79.6% for 50K to WGS and 86.5–86.7% for HD to WGS). Results show that breed composition of reference and validation sets has a considerable effect on imputation accuracy, and that imputation quality is improved if reference animals from similar populations are included.

Key Words: imputation, whole-genome sequence, accuracy