and dominance effects for both a traditional quantitative trait and fitness along with genomic covariance among traits. The second stage generates new individuals across generations based on a variety of selection scenarios. The selection stage can be performed using a wide variety of relationship matrices including pedigree, independent markers, haplotypes, or run of homozygosity based haplotypes. Relationship matrices and their associated inverse are generated using computationally efficient algorithms based on updating matrices from previous generations. Complex population structures can be generated that allow for a differential contribution of gametes to the next generation as well as mating constraints. To demonstrate the program, we present a small application that mimics a dairy cattle and swine population to describe some of the metrics that are generated. Scenarios were generated based on a 12,000 SNP marker panel spread across 3 chromosomes and a population size of 650 animals (sires = 50; dams = 600) per generation. A scenario with selection on a quantitative trait occurring for 5 generations and breeding values estimated from pedigree or independent SNP had a running time for the dairy cattle scenario of 4.85 and 5.82 min, respectively. GenoDriver allows for a wide range of selection strategies to be evaluated in the presence of a fitness trait and is available at https://github.com/jeremyhoward/GenoDriver.

Key Words: genetic simulation, quantitative traits, genomic selection

---

0302 Identifying and calling insertions, deletions, and single-base mutations efficiently from sequence data. P. M. VanRaden[*1], D. M. Bickhart[2], and J. R. O'Connell[3], *[1]Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, [2]Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, [3]University of Maryland School of Medicine, Baltimore.*

Whole-genome sequencing studies can identify causative mutations for subsequent use in genomic evaluations, but sequence alignment and variant identification are lengthy and sometimes inaccurate processes. Speed and accuracy of identifying small insertions and deletions (indels) of sequence can be improved by calling variants while aligning sequence reads. Previous algorithms separated alignment and calling steps, whereas program findmap stores previously known variants in memory, calls alleles for those variants, and identifies other potential new variants during alignment. The algorithm uses a string-pattern hash to store the reference genome in a rapidly accessed table. If both ends of a paired-end read do not align fully, the length of a potential indel within the read is calculated from the map location difference for two partial matches. The algorithm then finds the indel location and checks if the full read matches after accounting for the indel. Potential variants detected by findmap are checked and edited by program findvar for consistency across reads. New variants from findvar were compared with those from the Genome Analysis Toolkit (GATK) UnifiedGenotyper and from SamTools after Burrows-Wheeler Aligner (BWA) alignment. Detection accuracy was examined using reads simulated for 10 animals at 10X coverage from cattle reference map UMD3.1 with variants derived from run 5 (July 2015) of the 1000 bull genomes project that included 38,062,190 SNP, 532,179 insertions, and 1127,620 deletions. Half of variants were simulated as heterozygous, one-fourth as homozygous alternate, and one-fourth as homozygous reference. For homozygous alternate variants, findvar found 99.8% of SNP, 79% of insertions, and 67% of deletions; GATK found 99.4, 90, and 89%; and SamTools found 99.8, 12, and 18%, respectively. For heterozygotes, findvar found 99.1, 75, and 62%; GATK found 99.0, 90, and 88%; and SamTools found 98.2, 9, and 11%, respectively. False positives as percentages of true variants were 14, 0.4, and 0.3% from findvar; 12, 8.4, and 2.9% from GATK; and 37, 1.3, and 0.4% from SamTools, respectively. Read depth was 85.9 from findmap/findvar, 96.1 from BWA/GATK, and 84.4 from BWA/SamTools. With 10 processors, clock times were 106 h for BWA, 25 h for GATK, 11 h for SamTools, 3 h for findmap, and 1 h for findvar. The new software is freely available, with algorithms 10 to 30 times faster than current strategies for calling known and identifying new variants. Accuracy is improved by accounting for DNA variants while aligning sequence data.

Key Words: sequence alignment, variant calling, indel

0303 Issues in commercial application of single-step genomic BLUP for genetic evaluation in American Angus. D. A. L. Lourenco[*1], S. Tsuruta[1], B. D. Fragomeni[1], Y. Masuda[1], I. Pocrnic[1], I. Aguilar[2], J. K. Bertrand[1], D. W. Moser[3], and I. Misztal[1], *[1]University of Georgia, Athens, [2]INIA, Las Brujas, Uruguay, [3]Angus Genetics Inc., St. Joseph, MO.*

American Angus Association (AAA) has been using genomic information for genetic evaluations in a multistep approach since 2009. To improve accuracy while simplifying procedures, AAA is transitioning to single-step genomic BLUP (ssGBLUP) in the middle of 2016. Initial tests with ssGBLUP showed an increase in prediction accuracy of 25% for growth traits compared with traditional evaluations. Besides evaluation for growth traits, the goal of this study was to update the full pipeline for genetic evaluation with ssGBLUP methodology. The pipeline includes multi-trait models with linear and categorical traits, maternal effects, multibreed evaluations with external information, and a large number of genotyped animals but most of them with low EBV accuracy. Data included 9.7 M animals in the pedigree, 184,354 genotyped animals, and at most 8.2 M phenotypes for growth traits, calving ease (categorical), and carcass traits. The first issue during the implementation was the increasing number of genotyped animals. Single-step GBLUP requires the inverse of the genomic