Breeding and Genetics: Genomic methods and application—Dairy

W79 Common copy number variation regions affecting dairy traits in Gyr cattle. Gerson A. Oliveira Junior*[1], Adriana S. Carmo[2], Adam T. H. Utsunomiya[3], Tatiane C. S. Chud[3], Fernando S. B. Rey[3], Jose Bento S. Ferraz[1], and Marcos Vinicicus G. B. da Silva[2], *[1]University of São Paulo, Pirassununga, São Paulo, Brazil, [2]Embrapa Dairy Cattle, Juiz de Fora, Minas Gerais, Brazil, [3]São Paulo State University, Jaboticabal, São Paulo, Brazil.*

The Gyr cattle (*Bos indicus*) is a very important dairy breed in tropical countries like Brazil, mainly due to its tolerance to heat stress and parasites infestations as well as their use in crossbreeding schemes with other specialized dairy breeds, such as Holstein. In addition to the single nucleotide polymorphism (SNP), genomic structural variants such as copy number variants (CNV) have been revealed to be a substantial source of genetic and phenotypic variation in cattle, being an alternative to explain the missing heritability of complex traits. The aim of this work was to investigate common CNV regions associated with dairy traits in Gyr cattle. The Log R ratio profiles of 481 Gyr animals were determined using a high density SNP chip (Illumina BovineHD BeadChip assay), and the phenotypes evaluated were: age at first calving (AFC), milk (MY), fat (FY), protein (PY) and total solids yields (TSY), and protein (PP), fat (FP) and total solids percentage (TSP). The Log R ratio, predicted from Genome Studio software, was used in Golden Helix SNP & Variation Suite (SVS) 8.1, and the copy number analysis module (CNAM – multivariate algorithm) was used to identify common CNVs among animals. Linear regression was employed to identify CNVs associated with each production trait with significance level of FDR >0.05. A total of 47 CNV regions were detected that affected at least one trait, with 22 regions affecting 3 or more traits where one of them, located at chromosome 7, affected all traits except MY. Under these 22 regions, 38 structural variation, 18 genes and one pseudo-gene annotated in bovine genome (Biomart tool of Ensembl) were observed. The Panther Classification System divided these genes into 10 biological process categories highlighting localization, immune system and metabolic processes. The results suggest that common CNV regions can be biologically involved with more than one dairy trait.

Key Words: structural variant, genomics, dairy cattle

W80 Genome-wide association study on conception rate, milk production, and SCS in different stages of lactation for first three parities in US Holsteins. Shogo Tsuruta*[1], Daniela A. L. Lourenco[1], Ignacio Aguilar[2], and Ignacy Misztal[1], *[1]University of Georgia, Athens, GA, [2]INIA, Las Brujas, Canelones, Uruguay.*

The objectives of this study were to conduct genome-wide association studies (GWAS) on conception rate (CR), production traits, and SCS for Holstein cows and determine if the genetic architecture of these traits is different in 3 stages of lactation for the first 3 parities in US Holsteins. Genome-wide association studies were conducted for conception rate (CR), test-day milk, fat and protein yields, and test-day SCS. The data were split into 3 sets: early (<14wk), middle (14wk ≤ DIM ≤ 29wk), and end (29wk <). Heritability estimates for CR were lowest (0.03) in the middle and highest (0.05 to 0.08) in the end of lactation in all parities. Genetic correlations of CR with other production traits were low and all negative (−0.1 to −0.5). The SNP marker effects were divided into equal segments of 30 SNP. A segment on chromosome 14 was associated with CR only in early and middle stages of lactation in the third parity; the proportion of the total genetic variance explained by this segment for CR were 2.7% and 2.5%, respectively. The proportions for test-day milk and fat yields were highest in the middle of lactation (7% for milk and 8% for fat) in all 3 parities, whereas the proportions for test-day protein and SCS were low (<2%). The results suggest that gene expression for CR and milk production traits is stronger in early and middle lactation stages and similar over the lactations.

Key Words: GWAS, conception rate, US Holsteins

W81 Single nucleotide polymorphisms associated with thermoregulation in lactating dairy cows exposed to heat stress. Serdal Dikmen*[1,3], Xian-zhong Wang[2,3], and Peter J. Hansen[3], *[1]Department of Animal Science, Faculty of Veterinary Medicine, University of Uludag, Bursa, Turkey, [2]College of Animal Science and Technology, Southwest University, Chongqing, China, [3]Department of Animal Sciences, University of Florida, Gainesville, FL.*

Dairy cows with increased rectal temperature during heat stress experience lower milk yield and fertility. Given that rectal temperature during heat stress is heritable in dairy cattle, genetic selection for regulation of body temperature should reduce effects of heat stress on production. One goal of the study was to validate the relationship between genotype and heat tolerance for SNPs previously related to resistance to heat stress. A second goal was to identify new candidate gene SNPs related to resistance to heat stress. Thermotolerance was assessed in 625 lactating Holstein cows during the hottest part of the day in summer by measuring rectal temperature (a direct measurement of body temperature regulation), respiration rate (an indirect measurement body temperature regulation) and sweating rate (the major evaporative cooling mechanism in cattle). Specific genetic markers and candidate genes responsible for genetic variation in these variables were identified. For SNPs previously related to heat tolerance in genome-wide analysis, a region of BTA6 was related to rectal temperature and 3 closely-located genetic markers on BTA24 and another on BTA29 were associated with sweating rate. New candidate gene SNPs were identified for rectal temperature (n = 7), respiration rate (n = 9), and sweating rate (n = 6). The largest effect on rectal temperature was for *PGR,* which explained 2.1% of the phenotypic variation after adjustment for dry bulb temperature. This SNP could affect heat loss via cutaneous cooling because progesterone regulates vasodilation in the skin during local heating. *ACAT2* (involved in lipid metabolism) explained the largest variation in respiration rate (3.3%) and *SERPINE2*, which regulates the enzyme thrombin that regulates epithelial cells in sweat glands, explained the largest variation in sweating rate (2.1%). These genetic markers could prove useful in genetic selection for heat tolerance in Holstein cattle.

Key Words: heat stress, thermoregulation, SNP

W82 Multi-generational imputation of SNP genotypes and accuracy of genomic selection. Sajjad Toghiani* and Romdhane Rekaya, *The University of Georgia, Athens, GA.*

Superiority of genomic selection (GS) is possible only when high density single nucleotide polymorphism (SNP) panels are used to track QTLs affecting traits. Even with the continuous decrease in genotyping costs, only a small fraction of the population has been genotyped with these high-density panels. To reduce the cost of GS, it is often the case that a larger portion of the population is genotyped with low-density SNP panels and then imputed to a higher density. Accuracy of SNP genotype

imputation tends to be high when minimum requirements are met. Nevertheless, a certain rate of errors is unavoidable. Such rate of errors tends to increase with the increase of the generational interval between reference and testing generations. Thus, it is reasonable to assume that accuracy of GEBVs will be affected by imputation errors. To evaluate the impact of multi-generational selection on the accuracy of SNP genotypes imputation on the reliability of resulting GEBVs, a simulation was carried out under varying updating of the reference population, distance between training and validation sets, and the approach used for the estimation of GEBVs. Using fixed reference populations, imputation accuracy decayed by around 0.5% per generations. In fact, after 25 generations, the accuracy was only 7% lower than the first generation. When the reference population was updated by either 1% or 5% of the top animals in the previous generations, decay of imputation accuracy was substantially reduced. These results indicate that low-density panels are useful, especially when the generational interval between reference and testing population is small. As the generational interval increases, the imputation accuracies decay, although not at an alarming rate. In absence of updating of the reference population, accuracy of GEBVs decays substantially in 1 or 2 generations with a decrease rate of around 20–25% per generation. When the reference population is updated by 1 or 5% every generation, the decay in accuracy was only 8 to 11% after 7 generations. These results indicate that imputed genotypes provide a viable alternative, as long the reference and training populations are appropriately updated.

Key Words: genotype imputation, genomic selection, accuracy

W83 Genome-wide association study for milk production traits in Russian dairy cattle. Alexander A. Sermyagin*[1], Elena A. Gladyr[1], Sergei N. Kharitonov[1], Alexander N. Ermilov[1,2], Ivan N. Yanchukov[2], Nikolai I. Strekozov[1], and Natalia A. Zinovieva[1], *[1]L.K.Ernst Institute of Animal Husbandry, Dubrovitsy, Moscow, Russia, [2]Regional Information Selection Center, Noginsk, Moscow, Russia.*

Genome-wide association study has been proven as a powerful tool for identifying genomic variants associated with economically important traits in domestic animal breeds. Our study is the first step toward the creating the reference population of Russian Holsteins to utilize the genomic information in the dairy cattle breeding programs in Russia. The objective of our study was to evaluate the association between SNPs and estimated breeding values (EBVs) for milk production traits. The genomic data were obtained by genotyping 195 progeny-tested and 61 young bulls using the Illumina Bovine SNP50 v2 BeadChip. The SNP's quality control was performed by using Plink (1.07) software. BLUP AM approach has been used to estimate the marker effects which were applied then to calculate the genomic EBVs for young bulls to increase the prediction reliability of the associations. Direct genomic and genomic EBVs (by GBLUP) were estimated for 305-d milk yield (MY), milk fat yield (FY), milk protein yield (PY), fat percent (FP) and protein percentage (PP). After the quality control, 41370 SNPs were selected for the association analysis. The average number of daughters per sire was about 240 and the reliability of EBVs amounted 87%. The linkage disequilibrium was $r^2 = 0.41$. The Bonferroni correction for detection significant associations was applied as $P < 1.2 \times 10^{-6}$. Two SNPs which had the most significant effect for MY were identified: ARS-BFGL-NGS-50172 on BTA17 ($P = 7.6 \times 10^{-8}$) and Hapmap54246-rs29017970 on BTA13 ($P = 1.7 \times 10^{-7}$). The association analysis for milk components revealed 3 SNPs significantly associated with FP: BTA-104917-no-rs on BTA9 ($P = 4.1 \times 10^{-8}$), ARS-BFGL-NGS-107379 on BTA14 ($P = 6.0 \times 10^{-7}$) and BTB-01604502 on BTA9 ($P = 1.1 \times 10^{-6}$). The effect on PP was shown for SNP Hapmap43278-BTA-50082 on BTA20 ($P = 8.0 \times 10^{-7}$). Few SNPs were found to have the effects on FY and PY traits. The significant effects of SNPs explained from 9.0 to 11.3% of additive genetic variances. Supported by the Russian Ministry of Education and Science (RFMEFI60414X0062).

Key Words: genome-wide association, breeding value, milk production

W84 Identification of copy number variable gene families in Holstein and Jersey cattle. Derek M. Bickhart*[1], Lingyang Xu[2,1], Jana L. Hutchison[1], Harris A. Lewin[3], and George E. Liu[1], *[1]United States Department of Agriculture, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD, [2]University of Maryland, Department of Animal and Avian Sciences, College Park, MD, [3]University of California, Department of Evolution and Ecology, Davis, CA.*

Copy number variants (CNV) represent a large proportion of genetic variation within the cattle genome that has yet to be accurately characterized by SNP genotyping arrays. While significant progress has been made in the identification of CNVs within individual animals using next generation sequence data, CNV frequencies within larger populations have not yet been estimated in cattle. In this study, we sequenced 28 individual bulls from 2 dairy breeds of cattle (22 Holstein bulls; 6 Jersey bulls) to identify dairy breed-specific copy number variation. Using a read depth method of CNV detection, we identified 1359 non-redundant CNV regions within all 28 animals. The number of variable bases contained within these CNV regions accounts for ~2% of the cattle genome, and the average CNV region frequency was 37.67%. This high average frequency suggests that a large proportion of CNVs were present in the ancestral population of both breeds of cattle rather than as a result of a large number of de novo events arising in subsequent generations after breed formation. We also assigned copy number values to each gene within each individual sequenced using the normalized sequencing read depth of non-overlapping genomic windows. Using a Vst approach on these gene copy number values, we identified 27 gene families with breed specific copy number expansions/contractions. We identified a Jersey-exclusive expansion of the CLEC5A gene, which is a regulator of osteoclastogenesis. Additionally, we identified a Holstein-exclusive duplication of the ASAP1 gene, which may be involved in cell membrane trafficking and the differentiation of fibroblasts into adipocytes. CNVs identified by this survey intersected gene families that may play a role in productive traits in dairy cattle and are therefore good candidates for novel genetic marker design.

Key Words: copy number variant, genomics, sequencing

W85 Single nucleotide polymorphisms in specific candidate genes are associated with phenotypic differences in days open for first lactation in Holstein cows. M. Sofia Ortega*[1], Anna C. Denicol[1], Daniel J. Null[2], John B. Cole[2], and Peter J. Hansen[1], *[1]Department of Animal Sciences, University of Florida, Gainesville, FL, [2]Animal Genomics and Improvement Laboratory, Agriculture Research Service, United States Department of Agriculture, Beltsville, MD.*

Previously, a candidate gene approach identified 51 single nucleotide polymorphisms (SNP) associated with genetic merit for reproductive traits and 26 associated with genetic merit for production in dairy bulls. We evaluated association of these 77 SNP with days open (DO) for first lactation in a population of Holstein cows grouped based on predicted

transmitting ability for daughter pregnancy rate (DPR): ≤ −1 (n = 1220) and ≥1.5 (n = 1053), and located on 11 farms in Florida and California. Cows were genotyped using a Sequenom MassARRAY assay. To evaluate phenotypes, farm records were retrieved from on-farm computers and combined with records from the national genetic evaluation system. The association of the genetic variants with DO was evaluated using the MIXED procedure of SAS V9.4 (SAS Institute, Inc., Cary, NC). The model included farm, number of copies of the minor allele, and the numerator relationship matrix to account for (co)variances among animals. For each SNP, the genotype was treated as a categorical variable to estimate additive and heterosis effects. Days open was lower ($P < 0.0001$) for cows in the high DPR group as compared with the low DPR group (97.8 ± 2.6 d vs 163.0 ± 2.9 d). There were 6 SNP with significant additive effects ($P < 0.05$) on DO (*COQ9, FCER1G, FST, GPLD1, MRGPRF* and *OCLN*) and an additional 6 SNPS with a tendency ($P < 0.10$) for an association (*ACAT2, CD14, PCCB, PMM2, RABEP2* and *SREBF1*). For example, DO for cows with 0, 1, or 2 copies of the minor allele for *COQ9* averaged 139.4 ± 3.5, 134.3 ± 2.8, and 123.6 ± 3.5 d, respectively. The DO for cows with 0, 1, or 2 copies of the minor allele for *FST* averaged 124.9 ± 3.3, 134.8 ± 2.6 and 135.8 ± 4.4 d, respectively. For 9 of 12 genes, the favorable allele for DO was also the favorable allele in the earlier report based on bulls. The SNP related to genetic and phenotypic estimates of fertility are likely to be informative markers for genetic selection. Moreover, the study of the role of these genes could provide new insights into the physiological regulation of fertility in dairy cattle (USDA AFRI 2013–68004–20365).

Key Words: single nucleotide polymorphism, days open, dairy cattle

W86 Animal selection for whole-genome sequencing by quantifying the unique contribution of homozygous haplotypes sequenced. Jana L. Hutchison*, John B. Cole, and Derek M. Bickhart, *United States Department of Agriculture, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD.*

Major whole-genome sequencing projects promise to identify rare and causal variants within livestock species; however, the efficient selection of animals for sequencing remains a major problem within these surveys. The goal of this project was to develop a library of high accuracy genetic variants found within diverse haplotypes that were in a homozygous state identified from animal genotypes in the national database. An inverted weight function that calculated the value of sequencing an animal based on the sum of the rarity of the haplotypes it had in its SNP-based genotype was used to calculate the estimate, as more common haplotypes would likely be represented within animals already sequenced in subsequent iterations. A weight value was assigned to each 75-SNP haplotype based on the inverse of its frequency within genotyped animals in the national database. Each individual's haplotype weights were summed, and the highest scoring animal was selected for sequencing. Haplotypes from selected animals were removed from future consideration, and the cumulative scores of all remaining animals were recalculated in the absence of those selected haplotypes. This iteration continued until all haplotypes above a frequency threshold of 4% had been selected for sequencing. There were a total of 3,680 75-SNP haplotypes above a frequency of 4% in the national database and 484,522 genotyped Holstein animals. We compared this method against the selection of animals for sequencing based on 3 additional algorithms: (1) an ascending relatedness weight function, (2) an unbiased predictor of imputation accuracy, and (3) a random selection of animals from the population. By calculating an iterative summed score based on the inverse value of an animal's unsequenced haplotypes, one can quickly determine the value of sequencing a new individual and avoid data redundancy that plagues projects that focus on sequencing highly related individuals in a population.

Key Words: sequencing, haplotype

W87 A GWAS on heat tolerance phenotypes for Italian Holstein bulls. Stefano Biffani[1], Umberto Bernabucci[2], Nicola Lacetera[2], Andrea Vitali[2], Paolo Ajmone Marsan[3], Nicolo PP Macciotta*[4], and Alessandro Nardone[2], *[1]IBBA-CNR, Lodi, Italy, [2]Dipartimento di Scienze e Tecnologie per l'Agricoltura, le Foreste, la Natura e l'Energia Università degli Studi della Tuscia, Viterbo, Italy, [3]Istituto di Zootecnica, Università CAttolica del Sacro Cuore, Piacenza, Italy, [4]Dipartimento di Agraria, Università di Sassari, Sassari, Italy.*

Heat stress is a key factor that negatively affects livestock productive and reproductive performance. A genome-wide scan was performed on a sample of 1,592 Italian Holstein bulls using 2 different measures of heat tolerance for milk yield and protein percentage. The first was a temperature-humidity index breeding value (THI EBV) recently proposed for the Italian Holstein breed. The latter was obtained by a principal component analysis carried out on milk test-day records corrected for environmental effects except for THI. Only the second principal component (PC2), which describes the individual patterns of corrected production data across different THI levels, was considered. Animals were genotyped with the Illumina BovineSNP 50 BeadChip. Monomorphic SNPs (7,140) and SNPs with a call-rate <95% (1,045) were discarded. In total, 45,546 SNPs were retained for the analysis. All bulls had a THI EBV, whereas the PC2 was available only for a sub-sample of bulls (423). Genome-wide scan was performed fitting the GRAMMAR approach through the GenABEL R package. Then, a Gene discovery analysis was carried out considering windows of 0.5 Mb surrounding the significant marker (0.25Mb up and down stream respectively). No significant associations were detected for milk THI EBV, apart from a weak signal on BTA 2 at about 32 Mb. In this region is located the solute carrier family 38, member 11 (SLC38A11) gene, reported to be involved in folliculogenesis in cattle. For PC2, 3 SNPs were detected on BTA 6, 16 and 26, respectively. The SNP on BTA26 is located in a region that hosts genes involved in the ovarian activity (FGF8). An interesting candidate for the SNP located on BTA16 at approximately 42.1 Mb is the dehydrogenase/reductase member 3 (DHRS3), involved in the embryonic development in humans. No significant associations were found for protein percentage THI EBV. However, 3 significant markers were detected for PC2 on BTAs 20, 14 and 8. Interestingly the BTA14 region hosts the junctophilin 1 (JPH1) gene, whose expression has been found to be upregulated in the hypothalamus of chickens subjected to heat stress. These preliminary findings suggest potential genomic regions linked to heat stress resistance in dairy cattle.

Key Words: heat stress, GWAS, principal component analysis

W88 A genome-wide association study of mastitis in US Holstein and the relationship to mammary microbiome profile identifies novel QTL. Heather Huson*[1] and Rodrigo Bicalho[2], *[1]College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, [2]College of Veterinary Medicine, Cornell University, Ithaca, NY.*

One of the most prevalent and costly obstacles facing dairy producers is the occurrence of mastitis. Mastitis is a worldwide endemic disease causing both short and long-term cow health and economic repercussions with production losses in terms of reduced milk yield, clinical treatment, culling of animals, and discarded milk. The objective of this study was

to conduct a genome-wide association study (GWAS) for mastitis and explore the use of milk microbiome profiles to identify novel mastitis QTL using the Illumina Bovine High-Density (777K SNPs) Beadchip on a cohort of US Holstein dairy cows. To this end, 2 GWAS were run comparing Holstein cows; the first compared cows phenotypically characterized as having clinical mastitis (CM) or healthy based on their somatic cell score (SCS) and the second GWAS used a linear score based on their mammary microbiome profile. The mammary microbiome profile was generated sequencing 16S rRNA of the milk microbiota, directly correlated with mastitis incidence, to generate a cow specific profile. Four genomic regions demonstrating significant association to CM incidence were identified on BTA 5, 16, 26, and 29 using SCS and 5 regions on BTA 5, 9, 16, 26, and 29 were significantly associated with microbiome profile GWAS (corrected *P*-value < 0.05). Of particular interest to this study was to identify regions of similar association and those different between the 2 trait GWAS, CM as opposed to microbiome profile. Regions on BTA 5, 9, and 26 validated previous mastitis QTL findings. Both GWAS approaches in this study identified novel QTL on BTA 16 and 29 potentially being the product of a more generalized immune response to infection. In contrast, marker associations on BTA 1, 2, and 11 show more variation between the 2 GWAS and may reflect a genetic predisposition to specific microbiota. In all, the study validated previously discovered QTL and identified novel QTL with both similarities and variation in the comparison of the 2 trait GWA studies. Further studies are required to validate specific candidate genes and their influence on mastitis and the potential identification of diagnostic markers for the selection of mastitis resistant dairy cows.

Key Words: mastitis, GWAS, microbiome

W89 The accuracy of genomic predictions for Japanese Holsteins using by GBLUP and ssGBLUP methods. Yusaku Gotoh*[1], Toshimi Baba[1], Satoshi Yamaguchi[2], and Takayoshi Kawahara[1], *[1]Holstein Cattle Association of Japan, Hokkaido Branch, Sapporo, Japan, [2]Hokkaido Dairy Milk Recording and Testing Association, Sapporo, Japan.*

Several strategies including GBLUP and ssGBLUP methods have been proposed to predict genomic breeding values (GEBV) for dairy cattle. Many studies have reported that the accuracy of GEBV predicted from ssGBLUP was higher than from GBLUP. There is no report applying same analysis to Japanese Holsteins. The objective of this study was to compare the accuracies of GPI predicted by GBLUP and ssGBLUP. The traits used in this analysis were 305-d milk, fat and protein yields, and feet and legs, udder and final score. In total, 3,787 bulls genotyped by the Illumina BovineSNP50 BeadChip were used. Validation bulls were prediction bulls with no daughters in 2009 data but more than 20 daughters in 2013 data. In multistep evaluation, training bulls were required to have 10 daughters in 2009 data. Number of training bulls were 517 and 554 for the analysis of milk and type traits, respectively. For multistep evaluation, GPI were estimated by selection index blending after calculating DGV by linear GBLUP. Coefficients of determination (R^2) by regression analysis of GPI for 2009 data on deregressed EBV for 2013 data were used as indicator of accuracy. The R^2 ranged from 0.19 (protein) to 0.27 (milk) and from 0.17 (feet and legs) to 0.28 (milk) for DGV and GPI predicted by GBLUP and from 0.21 (protein) to 0.30 (milk) for GEBV predicted by ssGBLUP, respectively. Hence, it is suggested that the GEBV predicted by ssGBLUP is higher accuracy than DGV and GPI predicted by GBLUP.

Key Words: GBLUP, ssGBLUP, accuracy

W90 Identification of loci associated with fertility in Holstein heifers. Joao G. N. Moraes*[1], Joseph Dalton[2], Thomas E. Spencer[1], Jennifer N. Kiser[1], Gregory W. Burns[1], Andrzej Wojtowicz[1], Mahesh Neupane[1], and Holly L. Neibergs[1], *[1]Department of Animal Science, Washington State University, Pullman, WA, [2]Department of Animal and Veterinary Sciences, University of Idaho, Caldwell, ID.*

Selection for higher milk production in United States dairy cattle has been very successful during the past 50 years, however modern lactating dairy cows exhibit a high incidence of subfertility and infertility with a national pregnancy rate of only 15 to 20%. The objective of this study was to identify genomic loci associated with fertility in nulliparous Holstein heifers. Breeding and health records of Holstein heifers (n = 2,333) were analyzed from a commercial heifer raising facility in Southern Idaho. Of these, 1,114 heifers were classified as highly fertile (conceived on first AI service) and 209 were identified as subfertile (did not conceive until after the fourth AI service or culled due to failure to conceive). Blood samples were obtained from the fertility-classified heifers, and DNA was extracted from 497 high fertile and 209 subfertile heifers. The DNA was genotyped with the Illumina Bovine HD Genotyping BeadChip. Quality control consisted of removing SNPs with < 90% call rate, and a MAF < 1% and removing heifers with a genotyping rate < 90%, leaving 575,959 SNPs and 470 fertile and 189 subfertile heifers for analysis. A genome wide association analysis (GWAA) and heritability estimate was conducted with the Efficient Mixed-Model Association expedited (EMMAX) software. This mixed model program empirically estimated a genomic relationship matrix and used it to model the correlation between the fertility phenotypes. Correction for population stratification was done by variance components and resulted in $\lambda_{GC} = 1.0$. The GWAA identified a QTL on BTA4 with a strong association with fertility ($P = 2.9 \times 10^{-9}$), while loci on BTA1, BTA2, BTA5, BTA6, BTA10, BTA11, BTA18, BTA23, BTA26, BTA27 and BTA28 were identified with a moderate association with fertility ($P < 5.0 \times 10^{-5}$). The heritability estimate for fertility in Holstein heifers was 0.52. These results indicate that there is ample opportunity to make significant gains in fertility in nulliparous Holstein heifers with genomic selection. This project was supported by Agriculture and Food Research Initiative Competitive Grant no. 2013–68004–20365 from the USDA National Institute of Food and Agriculture.

Key Words: genomics, fertility, heifer

W91 Accuracy of genomic imputation in a Thai multibreed dairy cattle population. Danai Jattawa*[1,2], Skorn Koonawootrittriron[1], Mauricio A. Elzo[2], and Thanathip Suwanasopee[1], *[1]Kasetsart University, Chatuchak, Bangkok, Thailand, [2]University of Florida, Gainesville, FL.*

The objective of this study was to investigate the accuracy of imputation from low (LDC) to moderate density SNP chips (MDC) in a Thai Holstein-Other multibreed dairy cattle population. Dairy cows with complete pedigree information (n = 1,110) from 129 dairy farms were genotyped with GeneSeek GGP20K (n = 570) and GGP26K (n = 540) BeadChips. After checking for genotypic quality, 16,387 SNP in common between the GGP20K and GGP26K were used to represent MDC in this study. Cows were divided into 2 groups, a reference group (n = 778) and a test group (n = 332). The SNP genotypes chosen for the test group were those SNP located in positions corresponding to GeneSeek GGP9K (n = 7,356). The LDC to MDC genomic imputation was carried out using 3 different methods, namely a population-based algorithm in the Beagle software (PBG), a population-based algorithm in the FImpute software (PFI), and a combined family and population-based algorithm in FImpute (CFI). Imputation accuracies within and

across chromosomes were calculated as ratios of correctly imputed genotypes to overall imputed genotypes. Imputation accuracy for the 3 methods ranged from 76.31% to 93.91%. The CFI had slightly higher imputation accuracy (93.91%) than PFI (93.56%) and both methods were substantially more accurate than PBG (76.31%). Noticeably most chromosomes that showed either high or low imputation accuracies were the same chromosomes that had high and low average linkage disequilibrium (defined here as the correlation between pairs of adjacent SNP within chromosomes less than 5 MB apart). This suggested that choosing sets of SNP with high levels of average linkage disequilibrium would improve imputation accuracy. Results clearly indicated that FImpute software (population or combined family-population) were more suitable than Beagle for genotype imputation in this Thai multibreed population. Perhaps additional increments in imputation accuracy could be achieved by discarding SNP with low levels of average linkage disequilibrium, and by increasing the completeness of pedigree information.

Key Words: imputation accuracy, linkage disequilibrium, multibreed dairy cattle

W92 Identification of copy number variation in Brazilian synthetic dairy cattle breed. T. C. S. Chud[1], M. V. G. B. da Silva[2], A. S. Carmo[2], T. B. R. Silva*[1], G. A. Oliveira Junior[3], F. S. Baldi Rey[1], and D. P. Munari[1], *[1]Univ Estadual Paulista "Júlio de Mesquita Filho," Jaboticabal, SP, Brazil, [2]Embrapa - Brazilian Corporation of Agricultural Research, Juiz de Fora, MG, Brazil, [3]Universidade de São Paulo, Pirassununga, SP, Brazil.*

Copy number variation (CNV) refers to genomic segments that present a type of structural variation, such as duplications or deletions. CNVs have been observed as an important source of genetic and phenotypic variation for production traits and animal health. The aim of this work was to identify CNVs in a synthetic breed (Gyr × Holstein) dairy cattle population (Girolando cattle). The data set contained 417 animals genotyped with the Illumina 50K SNP panel (~54.609 SNPs). An algorithm based on the Hidden Markov Model was implemented using PennCNV software (Wang et al., 2007) for CNV identification. PennCNV perl script was used to eliminate calls from low quality samples, based on the standard deviation of LRR (<0.30), the BAF drift (<0.01) and waviness factor (less than 0.05). The final data set was composed of 384 animals. Gene content of cattle CNV was assessed using Ensembl genes. We used the PANTHER classification system for testing the hypothesis ($P < 0.05$) that the GO terms of the molecular function, biological process, and pathway terms were under or overrepresented in the CNV. An account of 1,986 CNVs were found along the genome, of which 84% were duplications and 16% were deletions. The chromosomes BTA3, BTA17 and BTA23 presented higher frequencies (10.52%, 11.53%, 8.30%, respectively) of CNV. Chromosomes that showed lower frequency of CNV (<1%) were BTA27, BTA14 and BTA29. A total of 861 genes were found within these regions and they are involved in biological processes, such as development (105 genes), growth (2 genes), immune system (83 genes), metabolism (343 genes) and reproduction (12 genes). This study showed evidences of structural variations in the genome of Girolando cattle and the genes found in CNV may be involved in the expression of production and animal health traits.

Key Words: genomics, single nucleotide polymorphism, structural variation

W93 Linkage disequilibrium in a Thai dairy cattle population with different Holstein fractions. Thawee Laodim[1], Skorn Koonawootrittriron*[1], Mauricio A. Elzo[2], and Thanathip Suwanasopee[1], *[1]Kasetsart University, Bangkok, Thailand, [2]University of Florida, Gainesville, FL.*

Linkage disequilibrium (LD) is important for gene mapping, accuracy of genomic prediction, and understanding of recombination biology in dairy cattle populations. The level of LD can vary among populations depending on their genetic structure, selection and recombination rates. The objective of this study was to estimate and compare levels of LD in dairy cattle with different Holstein fractions under tropical conditions. Blood samples of 2,643 dairy cattle (89 bulls and 2,554 cows) from 304 farms located in Central, Northern, Western and Southern Thailand were extracted for DNA. The DNA samples were genotyped with one of 4 GeneSeek Genomic Profiler BeadChips (9K, 20K, 26K, or 80K). Only SNPs from autosomes in common among the 4 chips were considered. In addition, SNPs with a minor allele frequency (MAF) lower than 0.01 and a call rate lower than 90% were excluded. This resulted in a set of 7,123 SNPs used in this study. Animals were classified into 7 groups based on their Holstein fraction (HF): HF <75%, 75% ≤ HF <80%, 80% ≤ HF <85%, 85% ≤ HF <90%, 90% ≤ HF <95%, 95% ≤ HF <100%, and purebred HF. Distribution of MAF and estimation of LD were done using Haploview. All HF groups had similar patterns of MAF across autosomes (fraction of SNPs increased with an increase in MAF). However, means of MAF across autosomes differed among HF groups and it tended to decrease with an increase in H fraction (from 0.376 for HF <75% to 0.362 for purebred HF). Conversely, the mean r^2 across autosomes tended to increase as HF increased from 0.081 for HF <75% to 0.109 for purebred HF. Results from this study will be useful for genome wide association studies and for genomic prediction and selection of crossbred Holstein cattle in tropical regions.

Key Words: linkage disequilibrium, Holstein, tropics

W94 Improving the genotyping-by-sequencing (GBS) approach for the identification of SNPs associated with Johne's disease. Émilie Constant[1,2], Eveline M. Ibeagha-Awemu[1], Filippo Miglior[3,4], Gilles Robitaille[5], and Nathalie Bissonnette*[1,2], *[1]Dairy & Swine Research and Development Centre Agriculture and Agri-Food Canada, Sherbrooke, Quebec, Canada, [2]Department of Biology, Université de Sherbrooke, Sherbrooke, Quebec, Canada, [3]Canadian Dairy Network, Guelph, Ontario, Canada, [4]CGIL, University of Guelph, Guelph, Ontario, Canada, [5]Food Research and Development Centre, Agriculture and Agri-Food Canada, Saint-Hyacinthe, Québec, Canada.*

Bovine paratuberculosis is a disease caused by *Mycobacterium avium* ssp. *paratuberculosis* (MAP). Most infected cows are culled before they reach clinical stage, leading to the premature slaughter of many animals and significant economic losses. Several genetic variants have been reported associated with host susceptibility to MAP. The objective of this study was to validate a whole genome genotyping-by-sequencing (GBS) method to identify single nucleotide polymorphisms (SNPs) associated with bovine paratuberculosis. Animals were selected from 10 farms in the province of Quebec. Fecal and blood samples were collected to identify 24 MAP infectious status by fecal culture and serum ELISA and 24 healthy cows. Two GBS methods were compared: a conventional (restriction enzymes *Pst*I and *Msp*I at 5′/3′ used to construct DNA libraries) method (CM) and CM with more selective primers to reduce the complexity of the libraries (RM). DNA was extracted from isolated peripheral blood monocyte cells. Multiplexed libraries (48 libraries per lane) were subjected to 100-bp single-end sequencing on an Illumina HiSeq 2000 system. Reads that passed all filtering criteria were mapped to the bovine genome (Bta_4.6.1). The SNPs were called using the Universal Network-Enabled Analysis Kit and 30,266 and 82,593 passed

the quality control steps for the CM and the RM methods, respectively. Using the SVS GoldenHelix software, 3,653 (CM) and 17,413 (RM) SNPs were associated with regions that are not intergenic. Associated genes are linked to biological processes related to the immune system and to responses to stimuli, such as defense responses to bacteria. Analysis confirmed that 8 (CM) and 19 (RM) SNPs were associated with MAP infectious status ($P < 0.05$). Validation of their genetic association with bovine paratuberculosis are currently performed by SEQUENOM mass spectrometry using a larger group of animals (n = 800). The reduction of the complexity of the genome produced a larger number of qualified genotypes compared with the conventional method suggesting that this GBS strategy provides a greater reading depth, which increases the quality of each genotype.

Key Words: genotyping-by-sequencing, *Mycobacterium avium* ssp. *paratuberculosis*, genetic predisposition to disease