Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Genomic selection as a tool to decrease greenhouse gas emission from dual purpose New Zealand Sheep

S. Rowe*, J.C. McEwan*, S.M. Hickey†, R. A. Anderson*, D. Hyndman*, E. Young*, H. Baird*, K.G. Dodds*, C.S. Pinares-Patiño‡, and N.K. Pickering*.

*AgResearch Limited, PB 50034, Mosgiel, New Zealand, †AgResearch Limited, Hamilton, New Zealand, ‡AgResearch Limited, Palmerston North, New Zealand

ABSTRACT: In New Zealand, enteric emissions of methane (CH_4) from livestock account for ~33% of greenhouse gas emissions. Breeding animals for lower CH_4 emissions could provide a sustainable means of mitigation. Genomic prediction and genome wide association analyses were performed using estimated breeding values for CH_4 emissions adjusted for feed intake (gCH_4/kgDMI) in 1,872 dual purpose sheep. Analyses were performed using genotypes from the Illumina Ovine 50K BeadChip. SNP effects were estimated using a genomic best linear unbiased prediction model. Results show molecular breeding values with moderate accuracies of 0.37 for gross CH_4 and 0.43 for gCH_4/kgDMI respectively. Two markers located on chromosome 9 (P<9.42E-05) and 25 (*P*<1.03E-05) were associated with CH_4 and gCH_4/kgDMI respectively. These markers are within or close to plausible physiological candidates, but did not meet the genome-wide significance threshold. With the availability of high-density SNP chips there is further potential to identify loci associated with CH_4 emissions.

Keywords: Methane; Emissions; Greenhouse gas; Sheep; Genomic

Introduction

There is a growing body of research determining methods of mitigating methane emissions from ruminants. (Hristov (2013)), but the underpinning physiology is complex (Oddy et al. (2014)). Manipulation of digestive processes has the potential to affect many downstream traits involved in health and production. Ruminant digestion is based on a symbiotic relationship with rumen microbes and CH_4 is largely produced when excess hydrogen from the fermentation process is utilized by methanogenic archaea. This CH_4 is released from the animal by eructation. The volume of CH_4 produced follows a strong diurnal pattern and is closely related to feed intake. Recent evidence in cattle and sheep, however, has shown that there is variation amongst individuals beyond that which can be explained by nutrition. This suggests that there is potential to select animals with lower rates of emission. Furthermore, variation among ruminant species, indicates that rumen morphology and physiology can vary widely in nature, thus, there is scope to alter rumen parameters through breeding (Hoffman (1989)). The initial and perhaps the greatest challenge is to define a phenotype that captures individual variation in methane emission independent of the physiological and nutritional factors such as live-weight, nutritional quality and feed intake. Current estimates for the heritability of gross CH_4 (g) emissions in sheep are ~0.29 (± 0.05) and for yield of g CH_4/kg dry matter intake (DMI) ~0.13 (±0.03) (Pinares-Patiño et al (2013)). Crucially, these estimates have been shown to be repeatable over time, across ages, and nutritional status (Pinares-Patiño et al (2011)). Although this is promising for genetic selection, thus far, parameters have been estimated by the use of respiratory chambers under controlled conditions. These methods make obtaining phenotypic measures of the trait, on a commercial basis, prohibitively expensive. For selection to take place in a commercial setting there is a need for direct or indirect phenotypic predictors. Development of new technologies to measure direct CH_4 emission proxies such as “sniffers” and portable accumulation chambers is currently underway (Garnsworthy et al. (2012); Lassen et al. (2012); Goopy et al. (2013)), but the greatest potential to date is the use of genomic prediction. Genetic variation amongst individuals is likely to be underpinned by many variants, however, genomic breeding values should lower the number of phenotypic measures required, shorten generation intervals and increase the accuracy of breeding low methane emitting animals. In addition this methodology has already been implemented in many livestock sectors, therefore, the marginal cost of calculating genomic breeding values is relatively low, only incurring the ongoing phenotyping costs of the training population.

Here we describe the use of a resource population measured in respiratory chambers to predict genomic breeding values for g CH_4 and g CH_4/kg DMI in progeny of these animals. We explore how genetic variation in methane emissions can be captured and harnessed by estimating the effects of ~50k SNP genotypes using GBLUP. We subsequently describe genome-wide association analyses using these de-regressed breeding values to identify regions of the genome significantly associated with the trait. The aim of this study was to determine the accuracy with which methane emissions in the New Zealand sheep flock can

be predicted with genomic breeding values and to determine whether there are major loci that could enable marker assisted selection and provide insights into the underlying biological mechanisms involved. It is anticipated that similar mechanisms will affect emissions in all ruminants thus results will inform future studies in other breeds of sheep, and in deer, buffalo and cattle.

Materials and Methods

In total 1225 animals were measured in respiration chambers between 5 - 10 months of age (30 - 40kg live weight). The animals were progeny of 99 maternal, dual-purpose sires generated by 5 flocks from the New Zealand industry progeny test program from 2007-2012. Pedigree data was available from these 5 flocks for all animals born 1990 to 2012 (n = 71,745) as previously described in Pinares-Patiño et al., (2013).

Animals were placed into "lots" of 48 or 96 and acclimatized over a 21 day period to a lucerne based pellet feed and to a metabolic crate environment. After acclimatization they were placed in respiration chambers (n = 24, "group") for 2 days ("round" 1). The animals were fed equal amounts at 9am and 3pm each day. The quantity of feed for each individual was estimated at 2.2 times the maintenance requirement; based on live weight (LW). Measures from any animal that did not eat >95% of the ration were discarded from the analysis. Daily CH_4 outputs were calculated based on measures at 6 minute intervals over each 24 hour period for CH_4 (g/day) and for g CH_4/kg DMI). Measurements were repeated 14 days later for a further 2 days ("round" 2). Extreme animals were selected for repeat measures for a further 2 consecutive years using the same protocol.

The Illumina OvineSNP50 BeadChip was used to genotype 1872 animals. These samples included the phenotyped progeny (1225), their sires and samples from additional relatives available from the 71,745 pedigree recorded animals.

Quality control was performed as described by Dodds et al. (2009). Briefly, SNPs were discarded if they had a call rate <97%, appeared non-autosomal (including pseudoautosomal), were fixed for one allele, i.e. had a minor allele frequency (MAF) =0, or had a weighted Gencall10 score <0.422. Any SNPs that were not retained as part of the Ovine HapMap study (Kijas (2012)) were also discarded.

Estimating breeding values and calculating the dependent variables. Breeding values (EBVs) were estimated with a univariate linear mixed model using ASReml3 software (Gilmour (2009)). Random polygenic or "animal" effects (σ^2_u) were fitted together with permanent environmental effects accounting for day, within year across rounds, and across year effects. Fixed effects included contemporary group 1: birth year + flock + sex and contemporary group 2: recording year + lot + group + round. $\sigma^2_u \sim N(0, A\ \sigma^2_u)$ where A is the standard numerator relationship matrix estimated using pedigree relationships. Reliabilities were calculated as $\sqrt{1-(SEP^2/\sigma^2_u)}$ where SEP is the standard error of the BVs and σ^2_u is the additive genetic variance (Falconer (1996)).

Dependent variables (y) and associated reliabilities for subsequent analyses were obtained by removing parental contributions (removing from the EBV the contribution of the parent average information, but not of the individual or its descendants) and then de-regressing (dividing by the reliability of the EBV with parent-average removed). The method used is described by Garrick et al. (2009).

Breed designation and reliability threshold. Data was filtered on breed and reliability prior to analysis. Breed was designated by the following conditions: Romney, Coopworth and Perendale were reported if their breed composition was greater or equal to 75%. The remainder of the animals were reported as composite (CompALL) if they were greater than 30% combined Romney, Coopworth and Perendale. Only animals with a reliability of the dependent variable greater or equal to 80% of the heritability of the trait were used for further analysis. The resulting 1284 Parent averaged de-regressed estimated breeding values (DREBV) estimated for CH_4 g/d and the 1350 DREBV for g CH_4/kg DMI were used as phenotypes to estimate individual SNP effects for genomic prediction and for genome wide association analyses. For genomic selection, individuals with DREBVs were further assigned into training and validation sets. (Table 1).

Table 1. Numbers of animals assigned to training and validation sets.

	CH_4 Training (Validation)	g CH_4/kg DMI Training (Validation)
Romney	48	50
Coopworth	360	406
Perendale	15	16
CompALL	524 (337)	541 (337)
Total N	**1284**	**1350**

Estimation of SNP effects Analyses were carried out using R software (R core team (2013)). To estimate SNP effects, fixed heritabilities of 0.32 (±0.05) for CH_4 and 0.15 (±0.03) for CH_4/kg DMI were

used with the assumption that σ^2_u was equal to the value used to calculate the EBVs. All animals, Romney, Coopworth, Perendale and CompALL (training and validation) were used in this analysis.

The SNP effects (b_i) were calculated using the genomic BLUP (GBLUP) model described by VanRaden (2008) where the genomic relationship matrix (G1) is estimated using $G1 = \frac{zz'}{2\sum p_i(1-p_i)}$ Z is the SNP matrix $-2p_i$, $1-2p_i$, and $2-2p_i$ for BB, AB, and AA respectively, and p_i is the frequency of A allele of the i^{th} SNP. A linear mixed model was fitted to y, including the first 6 principal components from G1 to account for population structure. The individual genetic variation (animal) was fitted as a random term with effects distributed as $N(0, G1\sigma^2_u)$, where σ^2_u is the additive genetic variance, and residual effects distributed as N(0, R) R is a diagonal matrix with diagonal elements $(1-r^2)/r^2$ and r^2 is the reliability of the DREBV's.

Fitting the first six principal components (PCs), calculated using animals in the training set, accounted for the majority of the population structure (0.67 of the variation contained in the genomic relationship). The first 4 PCs, calculated using all animals to show the full population in the analysis are shown in Figure 1.

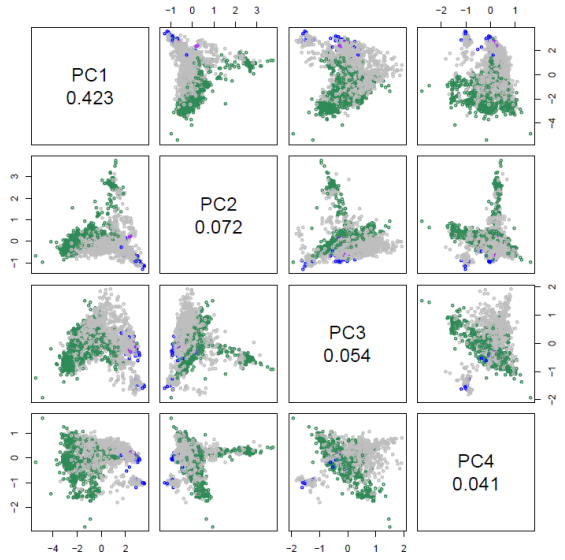

Figure 1. The first 4 principal components (PC) from the genomic relationship matrix calculated from 46, 912 SNP genotypes from 1872 animals. Romney (blue), Coopworth (green), Perendale (purple), CompALL (grey).

Significance values for SNP effects were calculated assuming the b_i follow a normal distribution with mean zero and variance $var = \frac{2pi(1-pi)m*var(bi)}{\sum(2pi(1-pi))}$ where m is the number of SNPs. The -log10(*P*) values are given in Figure 2 using positions from Ovine genome v3 (ISGC et al. (2010)). To account for multiple testing, nominal significance thresholds for type 1 error of less than 5% were approximated using the Bonferroni correction (Rice (1989)). This is stringent as it assumes that each SNP genotype is independent.

To estimate the accuracy of genomic prediction, molecular breeding values (MBV) were obtained by multiplying b_i by SNP genotypes (BB=0, AB=1, AA=2) and summing over all SNPs within an individual. Assuming that the effective heritability (h_g^2) is equal to the average reliability (r^2), prediction accuracy was estimated as $cor(y, MBV)/h_g$.

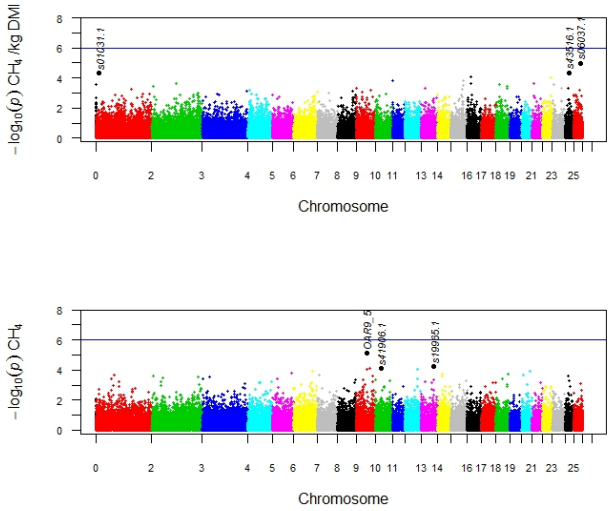

Figure 2. Manhattan plots for significance of individual SNPs for genome wide association analyses (-log10 P-value) for methane yield (CH_4 g/kg DMI) top, and gross methane emission (CH_4) bottom. Blue line is genome-wide significance threshold based on bonferroni correction.

Results and Discussion

1872 animals were genotyped using the Illumina Ovine 50K SNP chip. After quality control 46,912 autosomal loci were used for further analysis.

The analysis did not detect any genes of major effect (Figure 2); this was not unexpected given the complex nature of the trait. This implies that genomic prediction and selection methods that take advantage of combining many small effects across the entire genome are likely to be the most suitable breeding tool to reduce methane emissions in the NZ sheep flock. The

accuracy for animals in the validation set was 0.37 for CH_4 g/d and 0.43 for g CH_4/kg DMI.

Genome-wide association analyses. Quantile - quantile (QQ) plots were used to visualize the distribution of observed versus expected genome-wide -log10(*P*) values under the null hypothesis of no genetic association and no LD between SNPs (Figure 3). Although the QQ plots do not show any departure from the expected distribution, and therefore no indication of genes of major effect, the lambda value of 1.001 does indicate that fitting the 6 principal components accounted for the considerable population substructure. It is possible that the use of de-regressed breeding values to estimate SNP effects is not optimal. Ekine et al. (2013) show that this may lead to a reduction in power due to the masking of the Mendelian sampling term by greater weighting on information from relatives. Although here we have taken steps to account for this by using DREBV, comparisons should be made with an approach based on raw phenotypes as more data becomes available. Although no single SNP reached genome-wide significance due to stringent multiple testing thresholds, Figure 2 shows that the markers of greatest significance for gross methane and methane yield were on chromosomes 9 and 25 respectively. SNP positions on the Ovine genome v3 (ISGC et al. (2010)) were used to identify flanking genes. The *s06037.1* marker ranked first in the GWAS for g CH4/kg DMI (P<2.13E-05) and 49th for g CH_4 (P<0.0008). The marker is located on chromosome 25 and lies within the *Tetraspanin14* gene. This gene is highly expressed in the human colon and has been significantly associated with inflammatory bowel disease (Jostins et al. (2013)). The *OAR9_55642688.1* marker on chromosome 9, similarly, ranks highly for both traits and is 1.5 MB away from the *PEX2* gene which is hypothesized to be important in lipid metabolism and fatty acid oxidation in cattle (Mach (2013)). As rumen fermentation involves the metabolism of fatty acids and the human colon has a complex microbial community, both of these genes are good physiological candidates and warrant further investigation. High density genotypes for sheep are now available with the Illumina Ovine 700k SNP chip. It is possible that the greater marker density achieved by these genotypes coupled with a more comprehensive distribution of minor allele frequencies will identify markers in higher LD with causative variants thereby localizing variants affecting the trait and leading to more information on the underlying biological mechanisms.

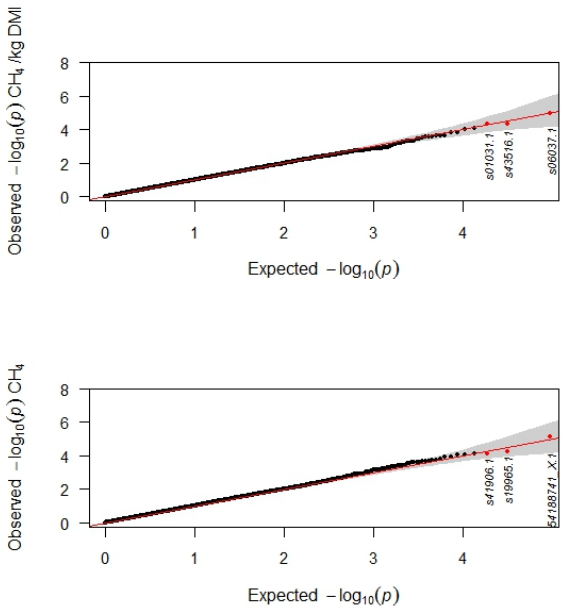

Figure 3. Quantile Quantile plots for observed versus expected significance values –log10 P, under the null hypothesis of no association between phenotype and genotype. Grey shaded area is confidence interval based on test statistic distribution ~χ^2_1. Lambda value was 1.0.

Conclusions

The current work indicates that as a proof of principle, genomic selection has potential to assist selection for reduced methane emissions from ruminants. Results indicate polygenic inheritance with the majority of the genetic variation explained by many loci each with small effects on the traits. Future work will involve two areas. The first is to increase the number of animals in the training population to improve the resolution and accuracy of genomic prediction, the second will involve genotyping new animals and key sires with the 600K ovine HD SNP chip combined with imputation of the current 50k genotypes to the HD density in an attempt to better refine individual loci affecting the trait.

Acknowledgements

This work was funded by the Pastoral Greenhouse Gas Research Consortium, Sustainable Land Management and Climate Change and the New Zealand Agricultural Greenhouse Gas Research Centre. The animals were part of the Ovita partnership and the related Central Progeny Test: both funded in part or whole by Beef + Lamb New Zealand. The New

Zealand Government in support of the Livestock Research Group of the Global Research Alliance (GRA) on Agricultural Greenhouse Gases funded Natalie Pickering's postdoctoral fellowship.

Literature cited

Dodds K.G., Auvray B., Pickering N., et al. (2009). Proceedings of the Association for the Advancement of Animal Breeding and Genetics: 296-299.

Ekine, C., Rowe, S.J., Bishop, S. et al. (2013). G3. 2:341-347

Falconer and Mackay. (1996). Introduction to Quantitative Genetics. Longmann, Essex, UK

Garnsworthy, P. C., Craigon, J., Hernandez-Medrano, et al. (2012). J. Dairy Sci. 95:3166-3180.

Garrick, D.J., Taylor, J.F., Fernando, R.L. (2009). Genet. Sel. Evol. 41:55.

Gilmour A.R., Gogel B.J., Cullis B.R., et al. (2009). ASReml User Guide Release 3.0. VSN International Ltd

Goopy, J. P., Donaldson, A., Hegarty, R. et al. (2013). Brit. J. Nutr. 8:1-8.

Hofmann R.R. (1989). Oecologia 78: 443-457.

Hristov, A. N., Oh, J., Firkins, J. L., Dijkstra, et al. (2013). J. Anim. Sci. 91: 5045-5069.

Jostins et al. (2012). Nature . 491 (7422): 119–124.

Kijas J.W., Lenstra J.A., Hayes B., et al. (2012). PLoS Biology 10(2):e1001258.

Lassen, J., Løvendahl, P., Madsen, J. (2012). J. Dairy Sci. 95: 890-898.

Mach, N.,Goselink, R.M.A.,van Baal, J. et al (2013). Livestock Science. v151 pp 92-96

MfE (2012). Ministry for the Environment. New Zealand Government. INFO 644.

Oddy, V. H., de Haas, Y., Basarab, J. et al. (2014). WCGALP. (submitted)

Pinares-Patiño, C. S., Ulyatt, M, J., Lassey, K. R. et al. (2003). J. Agr. Sci, Cambridge. 140: 205-214.

Pinares-Patiño, C. S., Waghorn, G, C., Hegarty, R. S. et al. (2009). NZ Vet. J. 57: 252-261.

Pinares-Patiño C.S, McEwan JC, Dodds KG, et al. (2011). Animal Feed Science and Technology 166: 210-218

Pinares-Patiño, C.S., Hickey, S. M., Young, E. A. et al. (2013). Animal. 7:s2: 316-321.

Raadsma H.W. (2010). Plant & Animal Genomes XVIII: San Diego, CA W134.

Rice W.R. (1989) Evolution 43(1):223-225.

R Core Team (2013). R Fondation for Statistical Computing. Vienna, Austria

The International Sheep Genomics Consortium et al. (2010). Animal Genetics 41(5):449-453.

VanRaden P.M. (2008). Journal of Dairy Science 91:4414-4423.