Genomic selection and scan for major genes for a new lamb survival trait for the New Zealand sheep industry

Benoit Auvray[1]
Sylvie Vanderick[2]
Sheryl-Anne Newman[1]
Ken Dodds[1]
Julie Everett-Hincks[1]

[1]AgResearch Ltd Invermay Agricultural Centre, Mosgiel, New Zealand
[2]Gembloux Agro-Bio Tech University of Liège, Gembloux, Belgium

jeh@agresearch.co.nz

agresearch   Université de Liège

Background

- Lambing percentage is one of the most significant factors affecting profitability on New Zealand sheep farms. Since the early 1990s, lambing percentage has increased at about 1% per year from a relatively stable level of approximately 100%. Top performing sheep farms are now consistently achieving 150% or more.

- As lambing percentage increases, the proportion of ewes having twin and triplet litters increases accordingly. Lamb mortality rate in multiples is higher than in singles, with triplets being particularly susceptible.

- Sheep Improvement Ltd (SIL, New Zealand's national sheep genetic evaluation system) records lamb survival to weaning but genetic improvement has been limited due to the low heritability of the trait and the current method of farmer recording.

- To replace the current SIL survival trait (SIL SUR), a new lamb survival trait (NEWSUR) has been defined based on using a combination of lamb birth fate codes and presence/absence of live weight measurements to assess whether a lamb has survived until weaning.

Objectives

- **Derive molecular breeding value (mBV) prediction equations and accuracies for NEWSUR direct (NEWSURd) and maternal (NEWSURm).**

- **Perform a genome wide association study (GWAS) scan to identify any region of the genome with a major effect on NEWSUR.**

Data sources

- The phenotype data was exported from the SIL database in October 2011 and consisted of 3,427,496 lamb records, born between 1990 and 2010 from 596 performance recorded flocks.

- The genotype data consisted of 8,644 animals genotyped with the OvineSNP50 BeadChip, QC checked and edited using the standard AgResearch process. The chip included 53,903 SNPs for which genotypes could be reported. The QC process reduced this number to 47,116 SNPs. SNPs were coded as the number of copies (0, 1 or 2) of the 'A' allele (using AB format).

Current SIL survival trait

- SIL lamb survival to weaning is a binary DEAD (0) / ALIVE (1) trait and assumes lambs are alive at weaning unless told otherwise by the breeder or a lamb birth fate code (BFATE) describing the time and cause of lamb death.

BFATE	Description
F	Fostered
H	Hand-reared
J	Born dead
J3	Died within 3 days of birth (autopsy)
K	Died between birth and rearing
M	Died of misadventure
P	Born dead - Premature
R	Born dead - Rotten
1	Died between rearing and weaning
4	Culled at birth (alive but not tagged)

SIL's rules to define SIL SUR
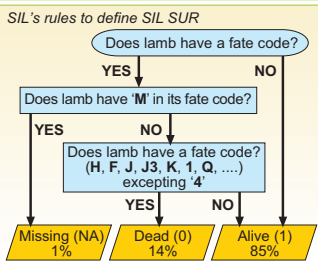
New survival trait

- The definition of NEWSUR is based on using a combination of lamb BFATE codes and presence/absence of live weight measurements to assess whether a lamb has survived until weaning. The weight measurements used are weaning weight, live weight at 6, 8, 10, 12 and 18 months, and fleece weight at 12 months of age.

Rules used to score NEWSUR.
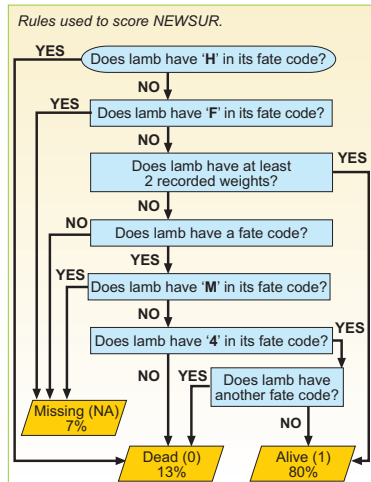
Breeding values for NEWSUR

- Direct and maternal breeding values (eBVs) were estimated for NEWSUR using the following BLUP model:

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{b} + u_i + m_{di} + l_{di} + e_i$$

where, for animal i and its dam d_i, y_i is the NEWSUR phenotype, $\mathbf{x}_i$ and $\mathbf{z}_i$ are incidence vectors linking the phenotype with the corresponding effects, $\boldsymbol{\beta}$ is a vector of fixed effects (flock, lamb birth year, lamb birth rank, age of dam, sex, flock * birth year, flock * birth rank, birth year * birth rank, birth year * age of dam and flock * birth year * birth rank), b is a vector of random birth day class * flock * birth year effects, u_i is the random direct genetic effect, m_{di} is the maternal genetic effect, l_{di} is the litter (dam * birth year) random effect and e_i is the residual.

- The variance components used were:

component	NEWSUR
σ^2 bdayclass	0.007 ± 0.003
σ^2 litter	0.014 ± 0.002
σ^2 genetic direct	0.005 ± 0.003
Cov (Gd, Gm)	-0.002 ± 0.002
σ^2 genetic maternal	0.002 ± 0.002
σ^2 total genetic	0.004 ± 0.002
σ_e^2	0.098 ± 0.002
σ^2 total phenotypic	0.123 ± 0.004
h^2 direct	0.040 ± 0.000
Corr (Gd, Gm)	-0.438 ± 0.003
h^2 maternal	0.019 ± 0.000
h^2 total	0.035 ± 0.000

Genomic selection for NEWSUR

The genomic selection pipeline used to obtain prediction equations and accuracies for NEWSUR is as follows:

- The direct and maternal eBVs were de-regressed (divided by their reliability) after removing the contribution from the parent average.

- Animals with a reported breed composition over 50% of Romney (R), Coopworth (C), Perendale (P), a composite of these (Comp), or Texel (T) were considered for the analysis. Only animals with a reliability of the de-regressed eBVs (DR-eBVs) at least equal to 80% of the heritability of the trait were kept.

- The Romney, Coopworth and Perendale animals were split into training and validation animals based on birth year. A cut-off year was chosen so that at least 200 animals per breed were used for validation. The Composite group was added to the training set and the estimation of the SNP coefficients was done using all the training animals together (mixed breed training set).

Cut off years and number of animals in training and validation sets.

	1st validation year			Training sets				Validation sets			
	R	C	P	R	C	P	Comp	R	C	P	T
NEWSURd	2008	2009	2005	1,021	1,252	120	121	343	412	210	353
NEWSURm	2008	2009	2005	960	1,130	122	54	219	244	207	341

- A relationship matrix ($\mathbf{G}$) that accounts for breed composition and reduces sampling error for the small relationships was estimated from the SNP markers for all animals used.

- mBVs were obtained directly from the following gBLUP model:

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + g_i + e_i$$

where, for animal i, $y_i = \frac{\hat{u}_{i-PA}}{r_{\hat{u}_{i-PA}}^2}$, the DR-eBV with parent average contribution removed, $\mathbf{x}_i$ is a vector of the first 6 principal components of the $\mathbf{G}$ matrix relating to animal i, $\boldsymbol{\beta}$ is a vector of fixed effects of the principal components, g_i is the genetic effect and e_i is the residual. In addition, $\text{var}(\mathbf{g}) = \sigma_u^2\mathbf{G}$ and $\text{var}(\mathbf{e}) = \mathbf{WR}$, where $\mathbf{R} = \sigma_e^2\mathbf{I}$ and $\mathbf{W}$ is another diagonal matrix of weights $W_{ii} = \frac{1 - r_{\hat{u}_{i-PA}}^2}{r_{\hat{u}_{i-PA}}^2}$.

- The accuracies of the mBVs were estimated in each of the 3 breeds separately using the validation sets and were also estimated in the Texel breed. Two methods were used:

a) The breed accuracies were calculated as: $r_A = \frac{\text{cor}(\mathbf{y}_{breed}, \hat{\mathbf{g}}_{breed})}{h_g^2}$, where $h_g^2 = \sigma_u^2 E(\frac{1}{1+W})$. $h_g^2(\text{NEWSURd}) = 0.271$ and $h_g^2(\text{NEWSURm}) = 0.167$.

b) Individual accuracies were calculated as: $r_i = \sqrt{1 - \frac{PEV_i}{\sigma_g^2}}$, by inverting the left hand (LHS) of the mixed model equations to obtain prediction error variances (PEV). Those accuracies were then averaged per breed to form $r_B = E(r_{breed})$.

Correlations between mBVs and DR-eBVs in the validation sets and their lower 90% confidence limit, and breed accuracies estimated using the 2 methods, r_A and r_B.

Breed	Trait	cor(mBV,DR-eBV)	90% CI	r_A	r_B
Romney	NEWSURd	0.203	0.116	0.389	0.357
	NEWSURm	0.113	0.002	0.277	0.295
Coopworth	NEWSURd	0.067[NS]	-0.014	0.128	0.375
	NEWSURm	0.218	0.115	0.533	0.295
Perendale	NEWSURd	0.148	0.035	0.284	0.324
	NEWSURm	0.122	0.007	0.298	0.286
Texel	NEWSURd	-0.008[NS]	-0.096	-0.016	0.177
	NEWSURm	0.099	0.009	0.241	0.134

NS = not significant at 10% confidence level.

Genome-wide association study for NEWSUR

- Using all animals (3,832 for NEWSURd and 3277 for NEWSURm), the residuals from the gBLUP were used to perform a GWAS for NEWSURd and NEWSURm. For each SNP s of the 47,116 SNPs, the following model was fitted:

$$y_i = x_{is}\beta_S + e_i$$

where, for animal i, y_i is the residual (e_i) from the gBLUP model, x_{is} is the number of copies of the 'A' allele for SNP s, β_S is the effect of SNP s and e_i is the residual. Moreover, $\text{var}(\mathbf{e}) = \hat{\sigma}^2\mathbf{W}$ and $\hat{\sigma}^2$ is the OLS estimate of the residual variance and $\mathbf{W}$ is the diagonal matrix of weights used in gBLUP.

- The statistic $t_s = \frac{\hat{\beta}_s}{\sqrt{\mathbf{x}'(\hat{\sigma}^2\mathbf{W})^{-1}\mathbf{x}}}$ was computed for each SNP s and converted into a p-value using a t-distribution with a number of degrees of freedom equal to the number of animals –1. The –$\log_{10}$(p-value) for NEWSURm and NEWSURd are reported below.
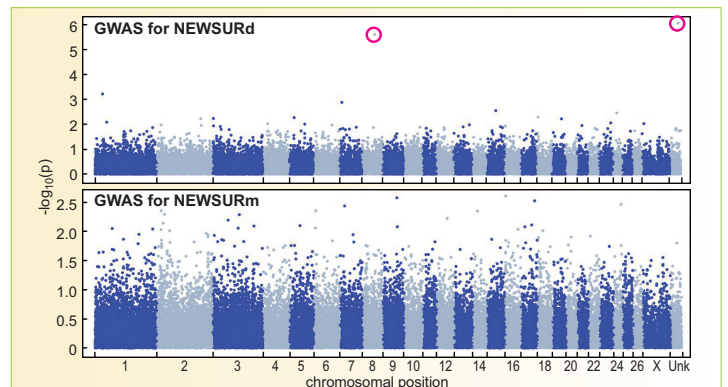
- The GWAS for NEWSURd uncovered a couple of potentially interesting regions that could be followed up, narrowed down and validated in a subsequent experiment.