Progress towards marker-assisted selection in forages

M.J. FAVILLE1, A.G. GRIFFITHS1, 2, M.Z.Z. JAHUFER1, 2 and B.A. BARRETT1
1AgResearch Ltd., Grasslands Research Centre, Private Bag 11008, Palmerston North 4474
2Pastoral Genomics, ℅ AgResearch Ltd., Grasslands Research Centre
marty.faville@agresearch.co.nz

Abstract
In developing marker-assisted selection (MAS) for herbage yield and persistence traits in forages, trait-linked markers were tested in elite breeding populations of complex parentage, to identify population-specific marker:trait associations. The effectiveness of these were subsequently evaluated in progeny generated by MAS. In white clover (*Trifolium repens*), markers detected significant (*P*<0.005) within-population differences in node number, a trait contributing to plant persistence. Progeny carrying beneficial marker variants exhibited a 19% increase in node number compared to those lacking these markers. In perennial ryegrass (*Lolium perenne*), a single marker detected a herbage yield differential of up to 23% within the populations tested, and progeny from divergent MAS exhibited a similar performance difference. These findings show that trait-linked markers can be used to manipulate quantitative traits in elite breeding populations of forage species and may subsequently contribute to development of improved cultivars. A next generation of marker technology will support expansion from single locus selection approaches to genomic selection, applied directly in breeding populations, offering enhanced contribution to genetic gain. The identification of markers associated with a trait is commonly achieved by QTL analysis in a biparental mapping population. A key step in the development of MAS in obligate outcrossing forages is the effective application of these marker:trait associations into genetically complex, multi-parent breeding populations. Our paper focuses on testing MAS in white clover and perennial ryegrass breeding populations, using simple sequence repeat (SSR) markers linked to QTL identified in mapping populations (Sartie et al. 2011; Faville et al. 2012). The approach taken is based on a prior study (Barrett et al. 2008) which showed that markers associated with seed yield QTL in white clover may be used to screen breeding populations, and identify marker alleles that have a significant effect on the trait in most populations.

Keywords: herbage yield, persistence, quantitative trait locus, simple sequence repeat, plant breeding

Introduction
Profitable production from New Zealand pastoral agriculture is dependent on reliable, high performance pastures. Perennial ryegrass (*Lolium perenne*) and white clover (*Trifolium repens*) based pastures are the primary source of nutrition for ruminant livestock on most New Zealand farms. Selective breeding in these species has delivered steady but modest genetic gain: 0.2–0.6% per annum (p.a.) for herbage yield in perennial ryegrass (Van Wijk & Reheul 1991; Easton et al. 2002; Sampoux et al. 2010) and 0.6–1.3% pa for white clover (Woodfield & Caradus 1994; Woodfield 1999). Expression of genetic gain at the pasture level may be further complicated by species interactions in the mixed sward, including the influence of the ryegrass endophytic fungus, *Neotyphodium lolii*. Acceleration of genetic gain in these species is needed to ensure pasture performance in the face of challenges including intensification, increasing ruminant genetic potential and abiotic and biotic stresses.

Accelerated genetic gain is achieved by increasing selection precision and efficiency in breeding programmes using appropriate genetic resources. Marker-assisted selection (MAS) is a breeding strategy designed to increase precision and efficiency (Collard et al. 2005) and uses molecular markers to tag a particular genome location influencing a trait (quantitative trait locus, QTL); plants that carry desirable gene variants for the trait can then be selected. Application of these markers enables indirect estimation of trait expression, in advance of the trait actually being expressed, and independent of confounding environmental factors that may compromise the precision of direct phenotypic measurement.

The identification of markers associated with a trait is commonly achieved by QTL analysis in a biparental mapping population. A key step in the development of MAS in obligate outcrossing forages is the effective application of these marker:trait associations into genetically complex, multi-parent breeding populations. Our paper focuses on testing MAS in white clover and perennial ryegrass breeding populations, using simple sequence repeat (SSR) markers linked to QTL identified in mapping populations (Sartie et al. 2011; Faville et al. 2012). The approach taken is based on a prior study (Barrett et al. 2008) which showed that markers associated with seed yield QTL in white clover may be used to screen breeding populations, and identify marker alleles that have a significant effect on the trait in most populations.

The objective of this research was to provide a proof of concept for MAS for herbage yield in perennial ryegrass and stolon node number in white clover. Herbage yield is a key trait for on-farm profitability and is therefore of high importance to farmers and breeders (Williams et al. 2007). Stolon node number is a determinant of persistence in white clover. Successful persistence in grazed swards is brought about by continuous stolon...
production (Brougham et al. 1978), with stolon nodes having the capacity to produce lateral stolon branches (Burdon 1983) and nodal roots (Thomas 1987) that support clonal spread of clover plants in the sward.

**Material and methods**

**Phenotyping trials**

Individual plants from perennial ryegrass breeding populations GA207 and GA208 (n=120 each), were established in a spaced plant field trial in row-column design with three replicates, at Palmerston North in May 2009. Soil type at the trial site was a Kairanga silt loam (Cowie 1978). The trial space was concurrently sown with a turf grass (Chewing’s fescue, Festuca rubra ssp. commutata) to provide a simulated sward environment within which individual ryegrass plants would be exposed to competition, but remain visually distinguishable. Herbage yield of individual plants was estimated by visual growth score (VGS) and direct measurement of dry matter (DM), with data collected for all plants on a seasonal schedule during 2009/2010 (Table 1). The sward within the experimental area was grazed following measurements and as required (DM above 1500 kg/ha), by a mob of sheep for a period of 2 to 3 hours per grazing. After grazing the sheep were immediately removed from the trial area. Soon after grazing, the sward was topped to a height of 3 cm using a lawn mower to even out any poorly grazed or un-grazed areas. The cut herbage was subsequently taken off site.

In a second trial, 448 randomly selected white clover seedlings from an early generation breeding population of cultivar ‘Kopu II’ were established outdoors in pots (15 cm diameter × 15 cm deep) at Palmerston North in November 2008. The trial was a non-replicated row-column spatial experimental design, which included one clonal check genotype repeated 40 times. In March 2009, two stolons of minimum 15 cm length with apical ends intact were removed from each plant for phenotype analysis. Node number (NN m⁻¹) was recorded by counting all nodes from the apex of each sampled stolon.

**Statistical analysis**

Analysis of phenotype data was conducted using the variance component analysis procedure, Residual Maximum Likelihood (REML) option, in GenStat 7.1 (GenStat 2003). A completely random linear model was used in the analysis using the REML algorithm. The final adjusted phenotypic means, best linear unbiased predictors (BLUPs), were based on adjustment for random error across columns, rows and (clover experiment only) the repeated clonal checks.

**Genotyping and marker:trait associations**

In ryegrass, DNA extracted (FastDNA kit, QBiogene, following manufacturer’s instructions) from leaf material was assayed by up to 24 SSR markers (Faville et al. 2004) associated with QTL controlling herbage production, leaf morphogenetic and structural traits (Sartie et al. 2011; Faville et al. 2012). DNA extracted (Red Extract-N-Amp Plant Kit, Sigma) from leaf material of ‘Kopu II’ plants was screened with six SSR markers (Griffiths et al. unpublished data) subtending

### Table 1.

Means (best linear unbiased predictors), ranges and genotypic variance component (σ²) with associated standard errors (±SE), based on perennial ryegrass visual growth score (VGS) and dry matter (DM) data for GA207 and GA208 individual spaced plants at Palmerston North in 2009 - 2010.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Date</th>
<th>Season</th>
<th>GA207 mean</th>
<th>range</th>
<th>GA208 mean</th>
<th>range</th>
<th>LSD (P&lt; 0.05)</th>
<th>σ² ± SE</th>
<th>Correlation with DM (r²)</th>
<th>χ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGS</td>
<td>Sept 2009</td>
<td>Winter/spring</td>
<td>3.45</td>
<td>2.46 - 4.56</td>
<td>3.00</td>
<td>1.59 - 4.22</td>
<td>0.8</td>
<td>0.36 ± 0.04</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oct 2009</td>
<td>Spring</td>
<td>3.09</td>
<td>2.23 - 3.67</td>
<td>2.87</td>
<td>2.16 - 3.66</td>
<td>1.0</td>
<td>0.18 ± 0.04</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jan 2010</td>
<td>Summer</td>
<td>2.66</td>
<td>1.78 - 3.50</td>
<td>2.63</td>
<td>1.46 - 3.59</td>
<td>1.0</td>
<td>0.22 ± 0.04</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feb 2010</td>
<td>Summer</td>
<td>2.56</td>
<td>1.43 - 3.66</td>
<td>2.46</td>
<td>1.47 - 3.54</td>
<td>1.2</td>
<td>0.27 ± 0.05</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mar 2010</td>
<td>Autumn</td>
<td>2.87</td>
<td>1.05 - 4.07</td>
<td>2.81</td>
<td>1.30 - 4.42</td>
<td>1.4</td>
<td>0.53 ± 0.07</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>May 2010</td>
<td>Autumn</td>
<td>3.18</td>
<td>1.59 - 4.11</td>
<td>3.04</td>
<td>1.54 - 4.60</td>
<td>1.4</td>
<td>0.05 ± 0.02</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>June 2010</td>
<td>Winter</td>
<td>3.02</td>
<td>1.54 - 4.10</td>
<td>2.98</td>
<td>1.66 - 4.24</td>
<td>1.0</td>
<td>0.28 ± 0.03</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>DM (g/plant)</td>
<td>Oct 2009</td>
<td>Spring</td>
<td>8.48</td>
<td>6.50 - 10.39</td>
<td>8.21</td>
<td>6.63 - 10.45</td>
<td>3.0</td>
<td>1.90 ± 0.50</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feb 2010</td>
<td>Summer</td>
<td>10.77</td>
<td>4.61 - 19.49</td>
<td>9.82</td>
<td>5.20 - 18.00</td>
<td>7.0</td>
<td>12.00 ± 2.00</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>May 2010</td>
<td>Autumn</td>
<td>10.50</td>
<td>4.68 - 16.38</td>
<td>10.05</td>
<td>5.21 - 18.98</td>
<td>7.2</td>
<td>10.00 ± 2.00</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>June 2010</td>
<td>Winter</td>
<td>8.04</td>
<td>3.77 - 13.13</td>
<td>7.65</td>
<td>4.25 - 13.81</td>
<td>5.8</td>
<td>0.30 ± 0.07</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

§ Pearson correlation coefficient
previously identified QTL regions contributing to node number. These QTL had been identified in a different population in multi-year, multi-environment mixed sward field trials and corroborated in a subsequent pot trial (Griffiths et al. unpublished data). SSR genotyping was performed as described in Barrett et al. (2008) for white clover and Sartie et al. (2011) for ryegrass.

Marker:trait associations were determined in datasets comprising SSR genotype and BLUP phenotype data, using linear regression (GenStat 2010) of phenotypic trait values on binary marker data (allele or genotype) from each individual within a population.

Marker-assisted crosses
In clover a series of selections (n=22 individuals each) was developed based on phenotype or marker indices: 1) high phenotype value (P+); 2) markers associated with high phenotype (M+); and 3) markers associated with low phenotype (M-). High phenotypic refers to the top 5% of individuals in the random sample for NN expression. M+ and M- selections were made randomly within the population, without recourse to phenotypic data. Plants within each of the selections were polycrossed using pollen-free bumble bees during summer 2009/2010, and seed from the half-sibling (HS) families was harvested within each of the crossing categories.

In ryegrass, marker-based selections (marker present, M+, and marker absent, M-) were made from amongst the subset of GA208 individuals in the top 30% for 2010 annual DM production. Selection was based on a single marker on linkage group (LG) 6 and selected plants were phenotypically balanced, with mean trait values in the M+ and M- selections differing by less than 5%, and not statistically different. Selections were polycrossed in summer 2010/11 and seed harvested by parent within selections to enable HS progeny testing.

Half-sibling family evaluation
In the white clover experiment, 10 progeny selected randomly from each of 10 HS families within each crossing category were established in a pot trial in November 2010. A random sample of 50 individuals from the ‘Kopu II’ early generation breeding population was also included. All seedling, trial establishment and phenotyping procedures were similar to those described above, and NN data were collected in March 2011. The experimental design was a row-column design with repeated checks and data analysis performed as described above.

In the perennial ryegrass experiment, balanced bulk samples of M+ and M- HS progenies were generated by mixing equal quantities of seed from each HS family within M+ and M- selection categories. In May 2011 a row trial (n=3 replicates), including M+ and M- bulks and the GA208 base population, was established at Palmerston North. A row-column design with repeated checks was used and data analysis performed as described above. Growth score and DM measurements were made for each row on a seasonal basis, between November 2011 and May 2012. Plants were defoliated between measurements by sheep grazing.

Results
Perennial ryegrass
Significant genotypic variation was detected amongst the GA207 and GA208 plants for all yield measurements (Table 1). A total of 130 alleles were detected at 24 assayed SSR loci in GA208 and 60 alleles at nine SSR loci tested in GA207. Nine and four significant (P<0.001) marker:trait associations were detected for GA208 and GA207, respectively. In GA208, significantly different DM (g/plant) of up to 23% (Figure 1) was observed between the two subpopulations of plants with alternative genotypes at a single marker locus from LG6 (Sartie et al. 2011). The same marker was significant in GA207, with a maximum differential of 11% between M+ and M- genotype groups in May 2010 (data not shown). Selections (M+ and M-) based on genotype at the LG6 locus were made for phenotypically similar groups of GA208 plants (M+ mean = 12.8 g annual DM/plant ±1.10 g ; M- mean = 12.2 g annual DM/plant ±0.61 g).

Following polycrosses within M+ and M- groups, limited quantities of seed per parent plant were available for evaluation and therefore balanced bulks of HS seed for M+ and M- selections were used to evaluate
effectiveness of marker-based selection. Differences in plant growth were observed between M+ and M- bulked HS progenies in November 2011, January 2012 and March 2012 (Table 2), similar in magnitude to the differentials observed amongst the parental generation. This indicated that the marker effect observed in the parent plants was heritable and expressed in a row trial environment, but the low statistical power of the format applied meant that the differences were not significant.

White clover
There was significant (P<0.05) genotypic variation among the 448 ‘Kopu II’ plants for NN (\(\sigma^2_g = 189 \pm 31\)). A total of 158 marker alleles from six SSRs were detected in the population and marker:trait associations were identified. Five SSR alleles with highly significant (P<0.0001) trait associations were used to develop divergent marker selection indices for increased or reduced NN. The plants carrying the beneficial marker alleles exhibited a 19% increase in mean NN compared with those without the alleles (data not presented).

Following polycrosses within P+, M+ and M- selection groups, significant (P<0.05) differences were observed for NN among the three HS progeny groups (Table 3). Mean NN for HS progeny from polycrosses based on genotype or markers associated with increased NN (P+ and M+) were significantly (P<0.05) higher than that based on markers associated with reduced NN (M-). Progeny from P+ had mean NN significantly (P<0.05) higher than M+ progeny, highlighting that selections made with phenotype indices had greater genetic gain. The M+ and M- selections differed significantly, exhibiting a 19% difference in mean trait value indicating that marker-based selection was effective. The selection group M+ had mean NN significantly (P<0.05) higher than the genotypes representing the early generation ‘Kopu II’ breeding population. This indicates potential genetic improvement for increased expression of NN relative to the original base population.

Discussion
The scientific literature hosts a large catalogue of QTL that have been identified, influencing a range of agronomically important traits in forage legumes and grasses. There are few reported examples where markers linked to these QTL have been implemented in a MAS application (Humphreys & Turner 2001; Dolstra et al. 2007; Stendal et al. 2006), particularly in complex, multi-parent breeding populations. Barrett et al. (2008) described an approach wherein seed yield QTL-associated markers were re-tested to identify marker:trait associations in breeding populations. Marker:trait associations discovered by this process may be used to select parent plants, facilitating reduction in the frequency of deleterious alleles or genotypes. This will increase frequency of favourable combinations in the subsequent generation of the population, improving its genetic merit.

The two proof-of-concept experiments reported here support the feasibility of this strategy for complex traits, in both white clover and perennial ryegrass. An index based on QTL-linked markers for NN showed a 19% differential in NN for ‘Kopu II’ clover plants with and without the favourable marker index. A single marker in GA208 ryegrass was responsible for up to 23% difference in herbage yield performance between plants with alternative marker genotypes, in a population already proven as elite breeding material (Dr Alan Stewart, pers. comm.). Ryegrass data was collected only from a first year of evaluation but a second year of data has since been obtained and will inform further identification of marker:trait associations, and subsequent marker-based selections.

Results from progeny assessment indicate that marker effects on trait values are heritable, evidence that QTL-informed MAS in forage breeding populations is a viable option.

Herbage yield of spaced grass plants is not always well-correlated with sward yield (Hayward & Vivero, 2012).

Table 2. Best linear unbiased predictors of dry matter yield (DM) from half-sibling progeny composite row plots, based on balanced bulks from M+ and M- parental selections in ryegrass population GA208. “GA208 base” is a random sample from the foundation population.

<table>
<thead>
<tr>
<th>Selection category</th>
<th>DM (g/row)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nov 2011</td>
</tr>
<tr>
<td>GA208 M+</td>
<td>85.2</td>
</tr>
<tr>
<td>GA208 M-</td>
<td>67.0</td>
</tr>
<tr>
<td>GA208 base</td>
<td>52.7</td>
</tr>
<tr>
<td>L.S.D.(\text{0.05})</td>
<td>26.0</td>
</tr>
</tbody>
</table>

Table 3. Genotypic variance (\(\sigma^2_g\)) among half-sibling (HS) family progeny derived from phenotype or marker-directed parental selection classes and HS family means (best linear unbiased predictors) for the trait node number (NN) (no. m-1).

<table>
<thead>
<tr>
<th>Selection category</th>
<th>NN</th>
<th>(\sigma^2_g) (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top Phenotypic (P+)</td>
<td>85.4(^a)</td>
<td>209±97</td>
</tr>
<tr>
<td>Marker + (M+)</td>
<td>72.7(^b)</td>
<td>△19%</td>
</tr>
<tr>
<td>Marker - (M-)</td>
<td>61.2(^c)</td>
<td></td>
</tr>
<tr>
<td>Unselected early generation Kopu II</td>
<td>46.4(^d)</td>
<td></td>
</tr>
<tr>
<td>L.S.D.(\text{0.05})</td>
<td>10.0</td>
<td></td>
</tr>
</tbody>
</table>

Means with identical superscript letters are not significantly (P<0.05) different.
1984; Waldron et al. 2008), although some studies have demonstrated stronger relationships between the two (Fujimoto & Suzuki 1975; Humphreys 1989). Here, a contrasting grass background was used to simulate a competitive sward environment for the ryegrass spaced plants during field evaluation. Progeny assessment conducted in rows indicated spaced plant marker effects were expressed in a sward format, but the critical test will be the level of expression achieved in sown pasture plots. Testing at this scale will be initiated following seed increases from the GA208 M- and M+ populations.

The 19% NN performance differential between clover progenies, based purely on marker selections (M+ and M-), indicates genetic gain enhancements are achievable through marker-only selection cycles that would circumvent costly and time-consuming rounds of progeny testing and reduce generation interval. As phenotyping for this trait in sward conditions is complex and laborious, progeny testing after each crossing cycle adds a significant time and cost burden to the breeding process. Marker-directed genetic gain can be utilised by alternating marker-only selection rounds, where candidates are identified at the seedling stage, with conventional selection cycles where the marker-trait association can be reconfirmed. This process allows for “off season” marker-directed selection cycles to increase genetic gain between conventional selections and reduce the generation interval. Additionally, while the marker effects have transferred successfully from mixed sward multi-environment multi-year grazed field trials to a pot environment, to further assess the gains indicated, the white clover populations have been transferred to a mixed sward field trial managed under grazing, to enable evaluation under more relevant agronomic conditions.

In the ryegrass GA208 progeny assessment, numerical differences between M+ and M- selections were not statistically significant, due to the diminished statistical power that could be achieved in the experiment. However, the clear data trend indicates the potential for significant trait improvement in already-elite breeding populations, based on only one or a few marker assays. Seed increases from M+ and M- subpopulations of GA208 are being generated and will support statistically robust, multi-site evaluation of these selections.

The successful application of MAS in forage plant breeding programmes indicates a pathway for cost-effective acceleration of genetic gain in forages. Next generation marker technologies, such as genotyping-by-sequencing (Elshire et al. 2011), will support expansion from this single locus selection approach to genomic selection (Hayes & Goddard 2010), applied directly in breeding populations for significantly enhanced efficiencies and genetic gain.

ACKNOWLEDGEMENTS
This work was funded by the Pastoral Genomics Consortium Ltd, Grasslands Innovation Ltd and the Ministry of Science and Innovation. We are grateful for the excellent technical support provided by Craig Anderson, Anthony Dunn, Casey Flay, Benjamin Franzmayr, Michael Hickey and Jana Schmidt.

REFERENCES


