

Implementing molecular marker technology in forage improvement

M. FAVILLE¹, B. BARRETT¹, A. GRIFFITHS¹, M. SCHREIBER², C. MERCER¹, I. BAIRD³,
N. ELLISON¹, G. BRYAN¹, D. WOODFIELD¹, J. FORSTER⁴, B. ONG⁴, T. SAWBRIDGE⁴,
G. SPANGENBERG⁴ and H.S. EASTON¹

¹AgResearch, Grasslands Research Centre, PB 11008, Palmerston North

²AgResearch, Joint Bioinformatics Institute, School of Biological Sciences, University of Auckland
PB 92019, Auckland

³AgResearch, Lincoln Science Centre, P.O. Box 60, Lincoln

⁴Plant Biotechnology Centre, Agriculture Victoria, Department of Primary Industries, LaTrobe
University, Bundoora, Victoria 3086, Australia
marty.faville@agresearch.co.nz

Abstract

Accelerated improvement of two cornerstones of New Zealand's pastoral industries, perennial ryegrass (*Lolium perenne* L.) and white clover (*Trifolium repens* L.), may be realised through the application of marker-assisted selection (MAS) strategies to enhance traditional plant breeding programmes. Genome maps constructed using molecular markers represent the enabling technology for such strategies and we have assembled maps for each species using EST-SSR markers – simple sequence repeat (SSR) markers developed from expressed sequence tags (ESTs) representing genes. A comprehensive map of the white clover genome has been completed, with 464 EST-SSR and genomic SSR marker loci spanning 1125 cM in total, distributed across 16 linkage groups. These have been further classified into eight pairs of linkage groups, representing contributions from the diploid progenitors of this tetraploid species. In perennial ryegrass a genome map based exclusively on EST-SSR loci was constructed, with 130 loci currently mapped to seven linkage groups and covering a distance of 391 cM. This map continues to be expanded with the addition of EST-SSR loci, and markers are being concurrently transferred to other populations segregating for economically significant traits. We have initiated gene discovery through quantitative trait locus (QTL) analysis in both species, and the efficacy of the white clover map for this purpose was demonstrated with the initial identification of multiple QTL controlling seed yield and seedling vigour. One QTL on linkage group D2 accounts for 25.9% of the genetic variation for seed yield, and a putative QTL accounting for 12.7% of the genetic variation for seedling vigour was detected on linkage group E1. The application of MAS to forage breeding based on recurrent selection is discussed.

Keywords: genome map, marker-assisted selection, perennial ryegrass, QTL, quantitative trait locus, SSR, simple sequence repeat, white clover

Introduction

Plant breeders have delivered significant genetic improvement in the performance of perennial ryegrass (*Lolium perenne* L.) and white clover (*Trifolium repens* L.) to New Zealand agriculture over the past 60 years (Woodfield 1999). However, selection to improve many economically-significant traits using classical plant breeding has been hindered by inefficiencies due to the complex (quantitative) genetic control of these traits.

Accelerated genetic improvement in white clover and perennial ryegrass may be realised via the enhancement of classical breeding practices by marker-assisted selection (MAS) strategies. MAS uses molecular markers linked with both superior and inferior genes as 'tags' for the location of those genes, enabling efficient, indirect selection for the superior genes (and against the inferior genes) by the plant breeder during germplasm and variety development (Lee 1995). A MAS-enhanced breeding strategy offers most potential for quantitative traits with low heritability and/or traits which are difficult or expensive to evaluate phenotypically (Dekkers & Hospital 2002; Sharma *et al.* 2002).

Two essential prerequisites to developing MAS technology are (a) the construction of a genome map using molecular markers; and (b) correlation of plant trait data to locations on the genome map, allowing identification on the map of genomic regions (quantitative trait loci, QTL), or in some cases single genes, contributing to the control of the trait. The inherent difficulties of genome mapping in outbreeding, heterogenous and heterozygous species, and a lack of high throughput molecular marker resources have combined to slow progress in forage genomes mapping, compared with more economically significant food crops such as rice. High throughput molecular marker systems, including amplified fragment polymorphisms (AFLP), and particularly simple sequence repeats (SSR), have become the marker systems of choice over low

throughput restriction fragment length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD) marker systems. High throughput markers are quickly and easily assayed, relatively inexpensive to use and amenable to automated analysis. First generation reference genetic linkage maps for perennial ryegrass have been constructed only recently, using a combination of high (SSR and AFLP) and low throughput (RFLP and RAPD) marker technologies (Hayward *et al.* 1998; Bert *et al.* 1999; Jones *et al.* 2002a, 2002b). For white clover, a low resolution framework map based on SSRs and AFLPs has been developed very recently (Jones *et al.* 2003).

An SSR molecular marker resource for genome mapping in white clover and perennial ryegrass was developed at AgResearch from proprietary EST (expressed sequence tag) gene databases generated at Agriculture Victoria within the joint Agriculture Victoria-AgResearch pasture plant genomics programme (Spangenberg *et al.* 2000; Barrett *et al.* 2001; Bryan 2001). These EST-SSR molecular markers are highly polymorphic, transportable between different populations (allowing for efficient tagging of genes in different genetic backgrounds), and amenable to high throughput analysis. Definition of location within the genome, as well as utility in QTL mapping, are qualities shared by all marker systems. In addition to this, and by virtue of their derivation from gene databases, EST-SSRs facilitate the location of specific genes, some of known function, on the genome maps of these forage species.

Barrett *et al.* (2001) provided an overview of genome mapping and MAS as it relates to forage improvement, and described the initiation of marker development and genome mapping at AgResearch. Here we report on the development and application of EST-SSR technology to the construction of genome maps in perennial ryegrass and white clover, and the initiation of QTL discovery in white clover.

Materials and methods

Plant material

Plants from diverse genetic backgrounds, exhibiting variation for important agronomic and forage traits were selected for use in the perennial ryegrass (PRG) and white clover (WC) genome mapping projects.

The PRG population (n=156) is derived from a cross between genotypes North African 6 (NA6, ecotype of Moroccan origin) and Aurora 6 (AU6, UK cultivar developed from a Swiss ecotype). Genomic DNA from PRG was supplied by Agriculture Victoria. The WC population (n=92) consists of progeny from the cross between a parental genotype (6525-5) from 'Sustain' cv. and a nematode resistant genotype (364/

7). These genotypes are genetically distinct at the *R* locus, which controls anthocyanin accumulation in the leaves – 6525-5 is heterozygous (genotype *R/r*) and 364/7 homozygous (genotype *r/r*), with *R* dominant to *r*. Genomic DNA was extracted from WC using the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA). Development and validation of EST-SSRs was conducted using DNA from the PRG and WC parental plants used to create our genome mapping populations.

Discovery and development of forage EST-SSRs

Identification of EST-SSRs in the forage EST database was conducted using a data mining bioinformatics programme. All marker assays were performed using a polymerase chain reaction (PCR) methodology based on three primers (Schuelke 2000) and data were collected using an ABI 3100 capillary electrophoresis array (Applied Biosystems, Foster City, California, USA) to resolve marker polymorphisms into discrete heritable characters. A subset of PRG and WC EST-SSRs were screened using DNA from the mapping population parents, to identify those EST-SSRs which produce discrete, polymorphic PCR product(s) of a length close to that predicted. Some were further screened in a subset (n=12) of the mapping population for clarification of polymorphism type. The EST-SSRs which passed one or both screens were used for construction of genome maps. A subset of genomic SSRs from a WC DNA sequence database was also included in the mapping project.

Construction of genetic linkage maps

Genome maps were developed for PRG and WC using genetic linkage analysis. All suitable screened EST-SSRs were mapped to unique locations in the PRG and WC genomes by tracking their inheritance from the parental to the F₁ generation using parent and progeny DNA samples. EST-SSR marker assays were carried out as described above. Linkage analysis was conducted with JoinMap[®] 3.0 software (www.kyazma.nl) using practices standard for mapping in this population architecture (Maliapaard *et al.* 1997).

QTL discovery pilot project

Clones of plants in the WC mapping population were evaluated for their performance. A replicated field trial was established at Lincoln in the 2001-2002 growing season using the clonally propagated plants from the mapping population. Seed production characters were measured using a spaced plant technique developed by our clover agronomy team for variety development pre-screening. Seedling characteristics

were measured in containerised plants sown outdoors in the autumn; seedling plant vigour was scored visually using a rating scale of 1 to 5 (1 = weak, 5 = vigorous) 8 weeks after planting, for each member of the mapping population. Analysis of variance (GenStat 6.0, VSN International Ltd., Oxford, United Kingdom), was used to determine if there was significant ($P < 0.05$) genotypic variation for measured characters, and MapQTL[®] 4.0 (www.kyazma.nl) was used to scan the genome for QTL. A preliminary logarithmic odds ratio (LOD) threshold of 2 was adopted since only a subset ($n = 92$) of the entire mapping population ($n = 187$) was evaluated in this pilot project.

The LOD threshold is applied to both QTL analysis and genetic linkage mapping. In QTL analysis, LOD score is a statistical measure of the probability that the region identified by the QTL contains a gene(s) associated with the measured trait. In genetic linkage mapping, LOD score represents a statistical estimate of whether two marker loci are likely to lie near each other on the same chromosome, and are therefore likely to be inherited together. In both applications, a higher LOD score represents a higher level of stringency.

Results and discussion

Discovery and development of EST-SSR markers

Amongst the 3261 EST-SSRs discovered in the EST database, screening of 480 PRG markers identified 259 (54%) functional EST-SSRs. Of these, 50% (130) were useful for genome mapping in the NA6 x AU6 population, while the remainder will be evaluated in subsequent populations. Amongst the WC EST-SSRs 554 (70%) were functional, and 64% of these were informative in the Sustain x 364/7 mapping population. The 221 PRG and 237 WC non-functioning EST-SSRs are undergoing assay optimisation or redevelopment prior to deployment. The different levels of polymorphism observed in the PRG and WC marker sets may be attributed to the polyploid white clover genome, which may allow for an inherently higher level of polymorphism to be maintained within genes.

Perennial ryegrass genome map

A genetic linkage map consists of marker loci arranged in a specific order. It is further divided into distinct linkage groups, each linkage group corresponding to a specific chromosome within the genome. Genome mapping in plants most commonly utilises populations of plants derived from crossing inbred genotypes. Populations suitable for genome mapping in outbreeding plant species, which are resistant to self-pollination, are generated via a cross

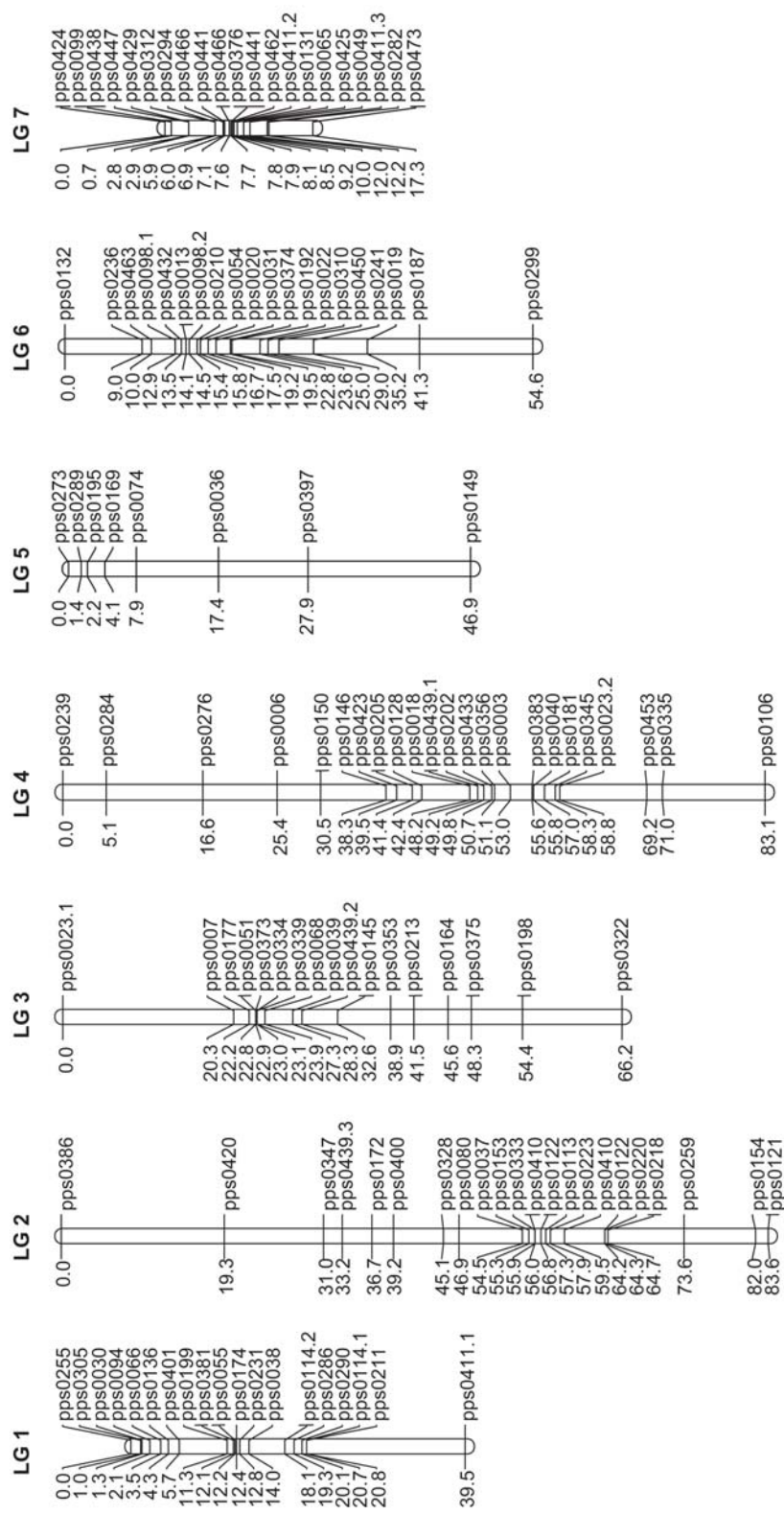
between two heterozygous genotypes (Grattapaglia & Sederoff 1994). One advantage of this architecture over the more commonly reported inbred lines or population structures is that it minimises the risk of both inbreeding depression and segregation distortion in the mapping population, both of which can significantly reduce the power and utility of a genome map. When mapping using a cross between heterozygous parents, the initial construction and examination of discrete maternal and paternal maps is essential. This enables identification of the matching, or homologous, linkage groups from each parent, and it is necessary to ensure that there is consistency of loci order between groups, and recombination rates between parents, before proceeding to the generation of a consensus map of the genome (Maliepaard *et al.* 1997).

For PRG, 122 EST-SSR markers were used to construct discrete genetic linkage maps for the genomes of the parental genotypes NA6 and AU6. The PRG genome contains DNA packaged into seven chromosome pairs, therefore we expect to detect seven linkage groups. There were approximately equal numbers of loci detected in both parents; the AU6 map was composed of 75 loci and detected seven linkage groups, while 84 loci were located on the NA6 map within eight linkage groups (although two of these formed a single group under a less stringent analysis). Matching maternal and paternal linkage groups were identified and aligned using 37 'bridging' markers, which detected loci on both parental maps. All bridging loci showed conservation of order between parents and observed variation in recombination rate between parents was minimal (data not shown).

With those criteria fulfilled, a consensus map of the PRG genome was estimated using all 130 informative EST-SSRs. A total of 130 loci were located on a map spanning 391 cM, with clear resolution of seven distinct linkage groups (Figure 1). A further three loci were located to a small satellite NA6 linkage group, and nine markers were unlinked to any other marker. Our map currently shows an irregular distribution of marker loci both within and between linkage groups, and is shorter than previously published maps which utilised larger numbers of markers (930 cM, Bert *et al.* 1999; 811-814 cM, Jones *et al.* 2002a,b). The on-going addition of marker loci, including those which remain unlinked, will address these issues.

The use of a gene-based marker system, such as EST-SSRs, imparts a further dimension to the genome map, with 73 (56%) of the marker loci located on our PRG map representing genes of known function. The

Figure 1 A consensus map of the perennial ryegrass genome based on genetic linkage analysis of 130 EST-SSR markers in the NA6 x AU6 mapping population. On each linkage group (LG 1–7) map distance (cM, Kosambi function) is shown at left, with EST-SSR loci on the right. Assignment of linkage group numbers is based on the reference International *Lolium* Genome Initiative (ILGI) perennial ryegrass genome map (Jones *et al.* 2002a), using genomic SSR linkage data provided by Agriculture Victoria. Thresholds for grouping and ordering were $LOD > 8.0$ and $LOD > 2.0/r < 0.4$, respectively.



potential of this feature will be realised where a mapped gene associates with a QTL underlying a trait of interest. This information can be used to further our understanding of the genetic control of that trait, as well as representing an extremely efficient 'perfect marker' to tag the trait for use in a MAS programme.

White clover genome mapping

A comprehensive map of the 16 linkage groups within the white clover genome was constructed using SSR markers to identify discrete locations within the genome. The white clover genome is tetraploid, meaning that it contains two ancient diploid genomes, one each from the two ancestors of white clover. This knowledge creates the expectation that the 16 white clover linkage groups may be further organised into 2 sets each with 8 pairs of linkage groups, reflecting the evolutionary relationship between the two diploid progenitors of modern white clover. This pairwise organisation can be identified using SSR markers that identify locations in both groups in the pair, and is very useful for understanding the role of pairwise duplicated genes in the pattern of genetic control of economically significant traits.

The 356 EST-SSRs that were informative in this population, as well as 31 genomic SSRs, were used to create the first comprehensive map of the white clover genome (Figure 2). A total of 464 SSR loci were distributed across the genome, with a total estimated map length of 1 125 cM. This is substantially longer than the map of lucerne (754 cM, Kaló *et al.* 2000) which is also a tetraploid forage, but shorter than the map reported for the diploid model forage legume *Medicago truncatula* (1 225 cM, Thoquet *et al.* 2002). Our analysis has clearly identified the 16 linkage groups that comprise the white clover genome, and we have used the map to describe the location of the *R* locus (Figure 2), a genetic factor which provides temperature dependent control of anthocyanin accumulation in white clover leaves (Corkill 1971). Furthermore, unambiguous classification of the 16 groups into 8 pairs has been achieved, based on information from 107 SSR markers which were mapped to discrete locations in both genomes (Figure 2). The remainder of the markers were shared nearly equally between the two parents, indicating that balanced parental marker datasets are available for undertaking discovery in the clover genome of QTL controlling economically significant traits in this cornerstone of the New Zealand pastoral industry.

QTL mapping

Most economically-significant traits in forage species

exhibit continuously distributed trait values and are controlled by several or many genes, termed QTL. An immediate utility of the forage genome maps described here will be in the discovery of genes controlling agronomic or non-traditional traits, by identification and mapping of QTL to specific locations. In order to explore the efficacy of the WC genome map for this purpose, the map was used in a pilot project to discover putative QTL associated with seed production and seedling vigour. The seed production trial at Lincoln in 2001 indicated that the yield potential of individual genotypes of the mapping population in that environment ranged from 0.63 to 6.0 t/ha. An initial analysis indicates the presence of multiple regions of the genome that influenced seed yield; one QTL (LOD 3.3) on linkage group D2 accounted for 25.9% of the genetic variation for seed yield (Figure 3a). A putative QTL (LOD 2.2) for seedling vigour was identified on linkage group E1; this QTL explained just 12.7% of the genetic variation for seedling vigour. This pilot project for QTL discovery in the white clover genome will be expanded by genotyping additional progeny to increase the power of the statistical analysis and by extending the analysis to additional traits measured in both seed production and mixed-sward environments.

The genome maps presented here represent a resource of molecular markers, of known position in the genome, which can be easily transferred to other populations designed expressly for the investigation of specific traits. In PRG we are currently using this framework marker resource to construct genome maps for populations from which trait data including flowering time, tillering characteristics, forage quality parameters, seed characteristics and disease resistance have been collected for QTL analysis.

Moving toward MAS

Recurrent selection is a commonly used breeding strategy for genetic improvement in forage species, whereby repeated cycles of phenotypic evaluation, selection of superior individuals, and poly-crossing amongst selected individuals are performed in order to increase the frequency of favourable alleles in an elite germplasm background (Figure 4). Each cycle may take several years using phenotypic selection alone. However, recurrent selection may be inadequate or uneconomical for complex traits with low heritability, traits that are strongly influenced by environment, recessively inherited traits, and it may be prohibitively expensive for traits requiring sophisticated phenotypic evaluation. The implementation of MAS will enhance and

Figure 2 A genetic linkage map of the white clover genome. The 8 homeologous groups are designated A - H and the homeologs within each group designated 1 and 2, respectively. The map is based on 464 SSR loci covering 1,125 cM of the clover genome. Markers informative in both homeologs are identified by **underlining** of the marker names and lines indicating the position of the marker in each homeolog. Markers informative in both parents are indicated by an "x", "y", or "z" suffix.

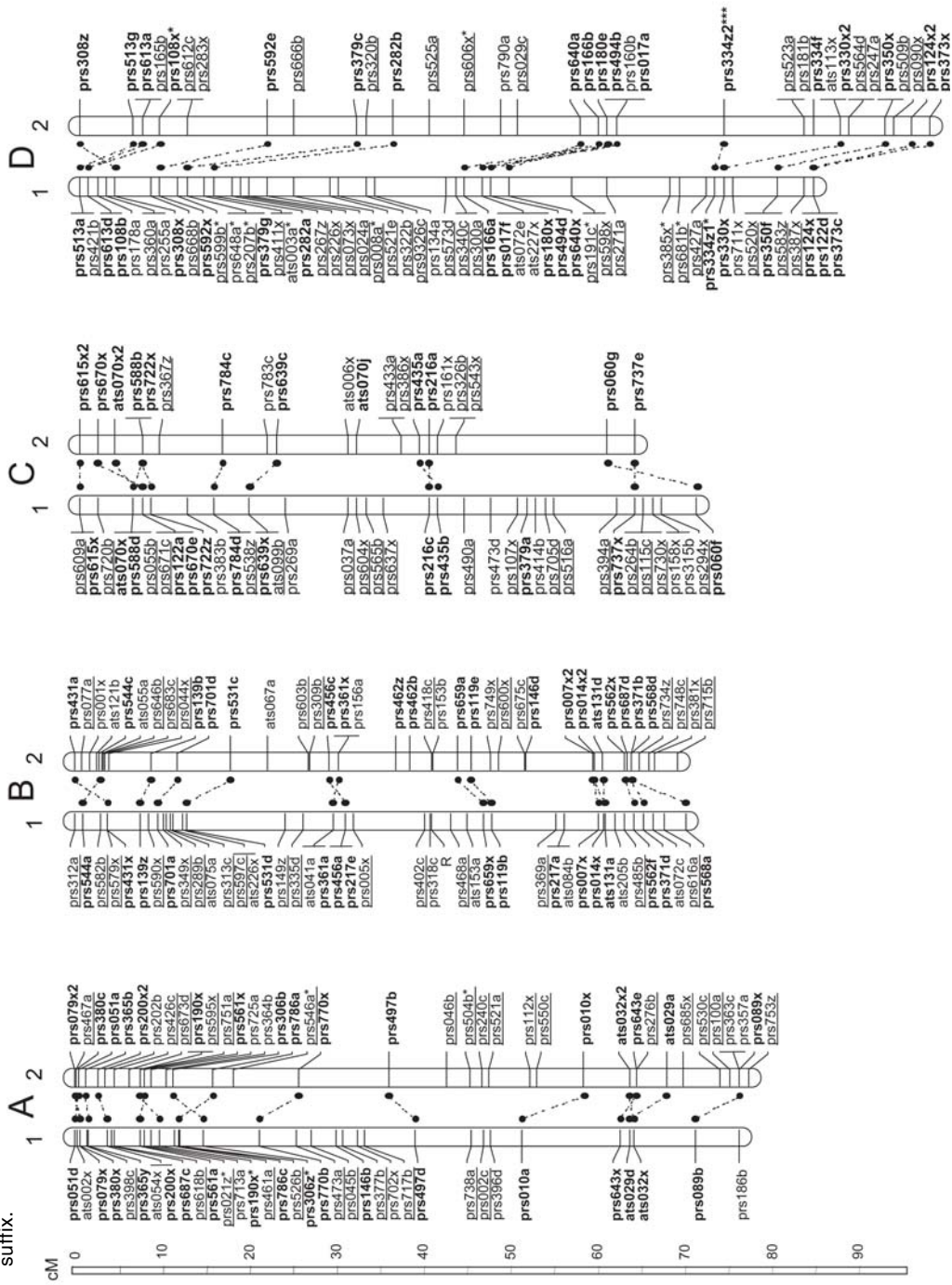


Figure 3 Putative QTL discovered in the white clover genome. a) A QTL on group D2 accounting for 26% of the genetic variation for seed yield in a replicated field trial. b) A QTL (LOD 2.2) on group E1 controlling 12% of the genetic variation for seedling vigour in a containerised plant experiment. The X axis indicates SSR marker positions along one of the 16 white clover linkage groups. The Y axis indicates LOD ratio, the measure of probability that a marker is correlated with genetic performance for the trait of interest. Each data point is the LOD score at a marker locus.

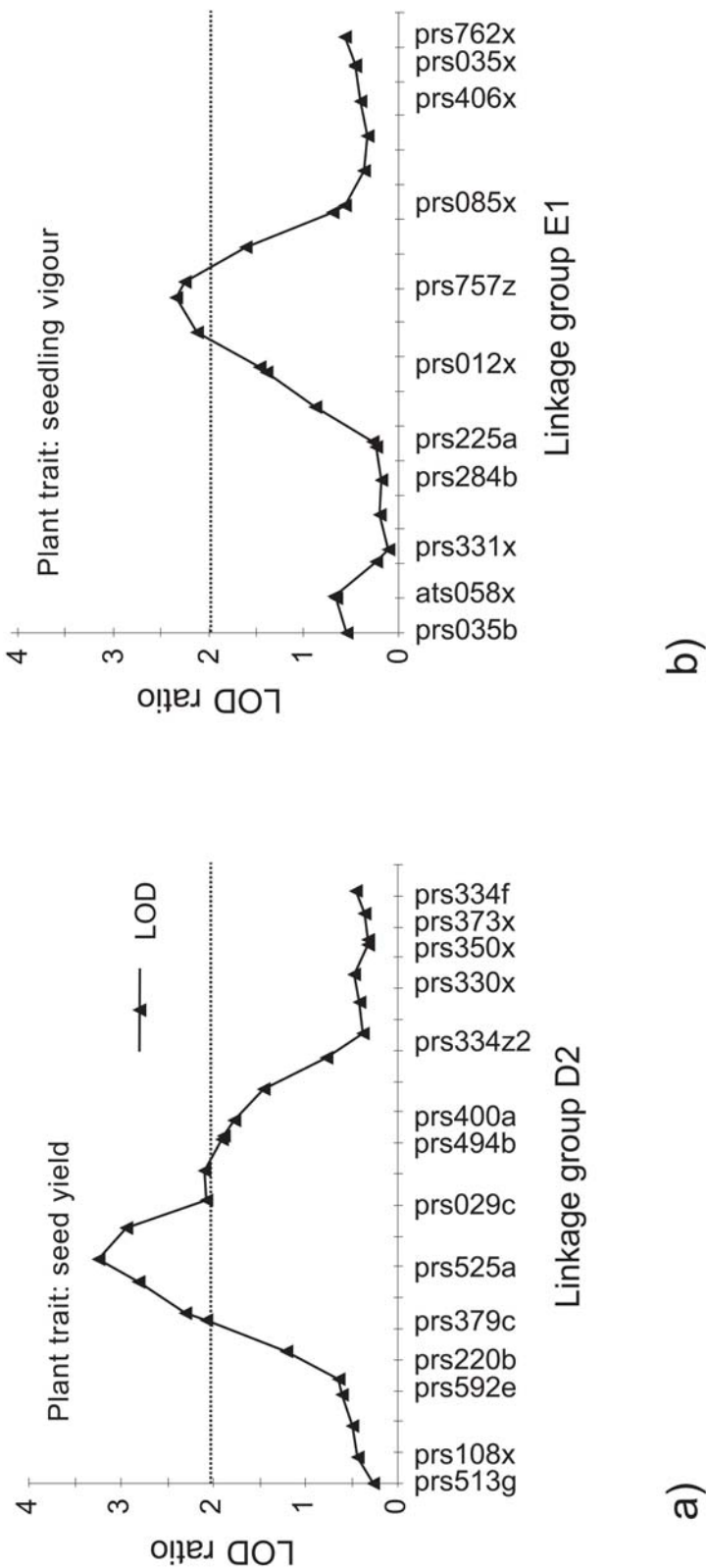
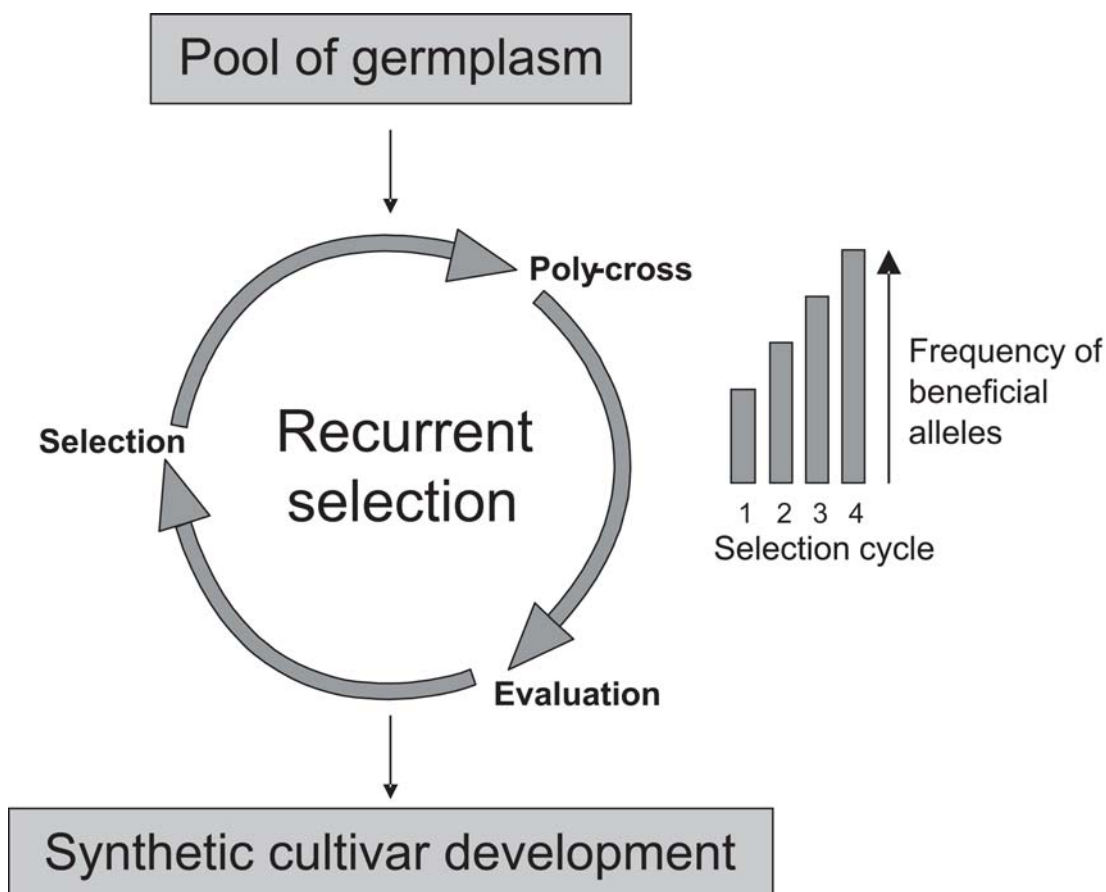


Figure 4 A simplified forage species breeding scheme based on recurrent selection. Superior genotypes selected from a pool of germplasm are subject to repeated cycles of poly-crossing, evaluation and selection, enabling the accumulation of beneficial alleles within that pool of germplasm. The application of MAS, alongside phenotypic selection, during each cycle will accelerate this accumulation.



complement, rather than replace, traditional selection programmes, overcoming many of the shortfalls outlined above and increasing efficiency by allowing for indirect selection for the trait of interest. The general strategy for MAS of quantitative traits involves the identification of QTL controlling the trait, determination of which QTL variants have a large and desirable effect on the trait, and use of the marker-QTL associations as indirect selection criteria in the recurrent selection process. Our long term objective is to identify several suitable QTL candidates for MAS in WC and PRG, and to implement MAS-based selections in elite germplasm.

Conclusion

The development and evaluation of an EST-SSR molecular marker resource has advanced the construction of genome maps for white clover and

perennial ryegrass. The deployment of these tools into QTL and gene mapping programmes has been initiated, and the effectiveness of the white clover map for this purpose has been demonstrated in a pilot study of seed yield and seedling vigour. The advancement of this programme, and its equivalent in perennial ryegrass, will lead us closer to the implementation of MAS strategies to enhance classical selection programmes in these important forage species.

ACKNOWLEDGEMENTS

We are grateful to Benjamin Franzmayr, Deborah Knox and Alieu Sartie for technical assistance. This research is funded by the Foundation for Research, Science and Technology, contract number C10X0203, and was conducted within the Plant Genomics Programme, a joint venture between AgResearch Ltd. and Agriculture Victoria.

REFERENCES

- Barrett, B.; Griffiths, A.; Mercer, C.; Ellison, N.; Faville, M.; Easton, S.; Woodfield, D.R. 2001. Marker-assisted selection to accelerate forage improvement. *Proceedings of the New Zealand Grassland Association* 63:241-245.
- Bert, P.F.; Charmet, G.; Soudille, P.; Hayward, M.D.; Balfourier, F. 1999. A high-density molecular map for ryegrass (*Lolium perenne*) using AFLP markers. *Theoretical and Applied Genetics* 99: 445-452.
- Bryan, G.T. 2001. Biotechnology in forage crops – capturing our potential. *Proceedings of the New Zealand Grassland Association* 63: 235-239.
- Corkill, L. 1971. Leaf markings in white clover. *Journal of Heredity* 62 (5): 307-310.
- Dekkers, J.C.M.; Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* 3: 22-32.
- Grattapaglia, D.; Sederoff, R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137: 1121-1137.
- Hayward, M.D.; Forster, J.W.; Jones, J.G.; Dolstra, O.; Evans, C.; McAdam, N.J.; Hossain, K.G.; Stammers, M.; Will, J.; Humphreys, M.O.; Evans, G.M. 1998. Genetic analysis of *Lolium*. I. Identification of linkage groups and the establishment of a genetic map. *Plant Breeding* 17: 451-455.
- Jones, E.S.; Mahoney, N.L.; Hayward, M.D.; Armstead, I.P.; Gilbert Jones, J.; Humphreys, M.O.; King, I.P.; Kishida, T.; Yamada, T.; Balfourier, F.; Charmet, G.; Forster, J.W. 2002a. An enhanced molecular marker based genetic map of perennial ryegrass (*Lolium perenne*) reveals competitive relationships with other Poaceae genomes. *Genome* 45: 282-295.
- Jones, E.S.; Dupal, M.P.; Dumsday, J.L.; Hughes, L.J.; Forster, J.W. 2002b. An SSR-based genetic linkage map for perennial ryegrass. *Theoretical and Applied Genetics* 105: 577-584.
- Jones, E.S.; Hughes, L.J.; Drayton, M.C.; Abberton, M.T.; Michaelson-Yeates, T.P.T.; Bowen, C.; Forster, J.W. (2003). An SSR and AFLP molecular marker-based map of white clover (*Trifolium repens* L.). *Plant Science*: (in press).
- Kaló, P.; Endre, G.; Zimányi, L.; Csanádi, G.; Kiss, G.B. 2000. Construction of an improved linkage map of diploid alfalfa (*Medicago sativa*). *Theoretical and Applied Genetics* 100: 641-657.
- Lee, M. 1995. DNA markers and plant breeding programs. *Advances in Agronomy* 55: 265-344.
- Maliepaard, C.; Jansen, J.; Van Ooijen, J.W. 1997. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Genetical Research* 70: 237-250.
- Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233-234.
- Sharma, H.C.; Crouch, J.H.; Sharma, K.K.; Seetharama, N.; Hash, C.T. 2002. Applications of biotechnology for crop improvement: prospects and constraints. *Plant Science* 163: 381-395.
- Spangenberg, G.; Kalla, R.; Lidgett, A.; Sawbridge, T.; Ong, E.K.; John, U. 2000. Breeding forage plants in the genome era. pp 1-39. *In*: Proceedings of the 2nd International Symposium, Molecular Breeding of Forage Crops, Lorne and Hamilton, Victoria, Australia. Ed. Spangenberg, G. Kluwer Academic Publishers, Developments in Plant Breeding Volume 10.
- Thoquet, P.; Ghéardi, M.; Journet, E.-P.; Kereszt, A.; Ané, J.-M.; Prosperi, J.-M.; Huguet, T. 2002. The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. *BMC Plant Biology* 2: 1-13.
- Woodfield, D.R. 1999. Genetic improvements in New Zealand forage cultivars. *Proceedings of the New Zealand Grassland Association* 61: 3-7.