Future Directions in the Molecular Breeding of Forage and Turf

G. C. Spangenberg
La Trobe University, Australia

J. W. Forster
La Trobe University, Australia

D. Edwards
Molecular Plant Breeding Cooperative Research Centre, Australia

U. John
La Trobe University, Australia

A. Mouradov
La Trobe University, Australia

See next page for additional authors

The XX International Grassland Congress took place in Ireland and the UK in June-July 2005. The main congress took place in Dublin from 26 June to 1 July and was followed by post congress satellite workshops in Aberystwyth, Belfast, Cork, Glasgow and Oxford. The meeting was hosted by the Irish Grassland Association and the British Grassland Society.

Proceedings Editor: D. A. McGilloway

Publisher: Wageningen Academic Publishers, The Netherlands

Presenter Information

G. C. Spangenberg, J. W. Forster, D. Edwards, U. John, A. Mouradov, M. Emmerling, J. Batley, S. Felitti, N. O. I. Cogan, K. F. Smith, and M. P. Dobrowolski

Future directions in the molecular breeding of forage and turf

G.C. Spangenberg[1,3], J.W. Forster[1,3], D. Edwards[1,3], U. John[1,3], A. Mouradov[1,3], M. Emmerling[1,3], J. Batley[1,3], S. Felitti[1,3], N.O.I. Cogan[1,3], K.F. Smith[2,3] and M.P. Dobrowolski[2,3]

[1]*Primary Industries Research Victoria, Plant Biotechnology Centre, La Trobe University, Bundoora, Victoria 3086, Australia, Email: german.spangenberg@dpi.vic.gov.au*

[2]*Primary Industries Research Victoria, Hamilton Centre, Mount Napier Road, Hamilton, Victoria 3300, Australia*

[3]*Molecular Plant Breeding Cooperative Research Centre, Australia*

Key points

1. Molecular breeding of forage and turf plants and their endosymbionts has entered the post-genomic era with a large amount of structural genomics information and genomic resources available for key forage and turf species and relevant model systems.
2. A primary future challenge is the conversion of this information into useful functional knowledge for the development of molecular breeding technologies and products that address a range of high impact outcome scenarios in forage and turf.
3. High-throughput approaches for spatial and temporal analysis, from genome to phenome, and the respective data integration in a systems biology context will be critical for the establishment of stringent gene-function correlations.
4. Translational genomics will permit results obtained using model systems to have major impact on the understanding of the molecular basis of plant processes and direct application to the molecular breeding of forage and turf plants.
5. These developments will be enhanced through applications of transgenesis and functionally-associated genetic markers in forage and turf molecular breeding building on genomic and post-genomic discoveries in these target species.

Keywords: systems biology, functional and translational genomics, bioinformatics, transgenesis, candidate gene based marker systems

Introduction

Major advances in genome technologies have revolutionised biology. High-throughput gene discovery through expressed sequence tag (EST) sequencing and genome shotgun sequencing provides a powerful tool for the analysis of biological systems at different hierarchical levels, from organism to ecosystem (or, from genomes to biomes). As well as providing access to sequence information from the transcribed component of genomes, EST-derived resources also provide the basis for the analysis of gene expression, genetic diversity and molecular evolution. Large-scale EST discovery has been performed for key forage and turf grasses and legumes allowing the generation of unigene cDNA microarrays for transcriptome analysis and the development of functionally-associated molecular marker systems. The analysis of genomic and post-genomic data and the integration of information from the related fields of transcriptomics, proteomics, metabolomics and phenomics has been facilitated by developments in bioinformatics. This has enabled the identification of genes and gene products, and the elucidation of functional relationships between genotype and phenotype, thereby allowing a system-wide analysis from genome to phenome. The effective application of systems biology approaches to forage and turf plants will close gaps in our understanding of the underlying genetics, physiology and biochemistry of many complex plant pathways and developmental processes. This is likely to advance progress in many applications of transgenesis in forage and turf plant improvement, building on gene technology as a powerful tool for the generation of the required functional genomics knowledge. It will further require

improved transformation methodologies for the development of 'market-ready' transformation events of likely 'pyramided' gene technologies following a sensible choice of high impact targets. Representative examples of these developments for perennial ryegrass (*Lolium perenne* L.), white clover (*Trifolium repens* L.), and grass fungal endophytes of the genera *Epichloë* and *Neotyphodium* are outlined in this chapter. Analogous developments in the forage model system, barrel medic (*Medicago truncatula* Gaertn.), are described elsewhere in this volume.

Systems biology and translational genomics for forage and turf molecular breeding

Genomics

High-throughput gene discovery by EST sequencing, initiated in 1991 (Adams *et al*., 1991), led to the development of significant genomic resources for a range of species, including the chief temperate forage and turf grasses such as perennial ryegrass (Sawbridge *et al*., 2003a), forage legumes such as white clover (Sawbridge *et al*., 2003b), and forage and turf plant endosymbionts such as *Epichloë* and *Neotyphodium* grass fungal endophytes (Felitti *et al*., 2004; Spangenberg *et al*., 2005). It has also established bioinformatics requirements for large and searchable sequence databases including tools for data analyses, sequence annotation and mining of complex interacting datasets. Although EST sequencing is still the standard procedure for gene discovery in many crops, a reduction in the cost of DNA sequencing has led to a move towards whole-genome sequencing and sequencing of gene-rich genomic regions.

Plant genomics was revolutionised by the release of the complete *Arabidopsis thaliana* genome sequence by the *Arabidopsis* Genome Initiative in 2000, four years ahead of schedule (The Arabidopsis Genome Initiative, 2000). Two years later, the completion of the rice (*Oryza sativa* L ssp. *japonica* Nipponbare) genome sequence by public consortia was announced. This research was complemented by rice sequencing work undertaken by the agribusinesses Monsanto (Barry, 2001) and Syngenta (Goff *et al*., 2002), and a separate research project at the Beijing Genomics Institute that sequenced the rice subspecies *indica* (Yu *et al*., 2002). Particularly relevant for translational genomics in forage legumes have been the ongoing efforts to completely sequence of gene-rich genomic regions of the model legume *Medicago truncatula* (Roe & Kupfer, 2004) and the development of other integrated *Medicago* functional genomics resources for improvement of alfalfa (May, 2004; this volume). Due to the anticipated similarities at the genomic level between the model plant species arabidopsis, barrel medic and rice and important forage crops such as white clover, alfalfa and perennial ryegrass, completion of these genome sequences is expected to have significant future impact on forage and turf molecular breeding. In this context, bioinformatics developments will play a fundamental role in enabling translational genomics for forage and turf plants, by translating and allowing exploitation of genomic information from the model to the agriculturally relevant species. Examples of such developments in gene discovery and associated establishment of bioinformatics tools and databases for representative forage and turf species are outlined below.

A resource of 44,524 perennial ryegrass ESTs was generated from single pass DNA sequencing of randomly selected clones from 29 cDNA libraries that represent a range of plant organs and developmental stages (Sawbridge *et al*., 2003a). EST redundancy was resolved through assembly with CAP3, identifying a total of 12,170 ryegrass unigenes within this dataset. Similarly, a resource of 42,017 white clover ESTs from 16 cDNA libraries and

corresponding to 15,989 unigenes was established (Sawbridge *et al*., 2003b). Each of these sequences has been annotated by comparison to GenBank and SwissProt public sequence databases and automated intermediate Gene Ontology (GO) annotation has also been determined. All sequences and annotation are maintained within ASTRA format MySQL databases, with web-based access for text searching, BLAST sequence comparison and GO hierarchical tree browsing. Each ryegrass sequence was mapped onto an EnsEMBL genome viewer for comparison with the complete genome sequence of rice and expressed sequences from related species. Similarly, each white clover sequence was mapped onto an EnsEMBL genome viewer for comparison with the complete *A. thaliana* genome sequence and expressed sequences from related legumes. Analogous developments in forage and turf symbio-genomics, e.g. large-scale gene discovery in grass endophytes (Spangenberg *et al*., 2001; 2005) have led to the production of 13,964 ESTs collectively from the grass endophytes *Neotyphodium coenophialum, N. lolii* and *Epichloë festucae* thus identifying at total of 7,585 unigenes (Felitti *et al*., 2004; Spangenberg *et al*., 2005).

Complementing these genomic resources, large insert genomic libraries of white clover (50,302 BAC clones with 101 kb average insert size, corresponding to 6.3 genome equivalents and 99% genome coverage), perennial ryegrass (50,304 BAC clones with 113 kb average insert size, corresponding to 3.4 genome equivalents and 97% genome coverage) and the ryegrass fungal endophyte *N. lolii* (6,000 BAC clones with 120 kb average insert size and 15-fold genome coverage) have been established for physical mapping, map-based cloning, novel gene and promoter discovery.

The availability of large sequence datasets permits mining for biological features, such as single nucleotide polymorphism (SNP) (Barker *et al*., 2003; Batley *et al*., 2003; Somers *et al*., 2003) and simple sequence repeat (SSR) (Robinson *et al*., 2004) molecular markers, that can be then applied to molecular breeding research such as genetic trait mapping and the development of functionally-associated genetic markers (Morgante *et al*., 2003; Andersen & Lübberstedt, 2003). The representative examples of EST resources for forage and turf molecular breeding described above, led to the identification of 3,214 ryegrass, 5,407 white clover, and 1,047 endophyte SSR molecular markers using SSRPrimer and to the design of associated specific PCR amplification primers. In addition, AutoSNP permitted the identification of: 1,817 candidate SNPs and 1,706 insertion-deletion (InDel) molecular markers across 1,409 white clover loci; 2,716 candidate SNPs and 345 InDels in 493 perennial ryegrass loci; and 1,636 candidate SNPs and 326 InDels across 300 grass endophyte loci. The integration of different analysis tools and the resulting data within central resources permits researchers to identify novel candidate genes and associated molecular genetic markers that can be applied to functional genomics and germplasm enhancement in these forage and turf plants and their endosymbionts.

The availability of complete-genome sequences enables further mining for novel promoter sequences (Qui, 2003; Ettwiller *et al*., 2003) and other regulatory features such as micro-RNAs (Rhoades *et al*., 2003; Nelson *et al*., 2003). This tertiary level annotation provides links to both the phenotype and the complex regulatory mechanisms that govern development and response to the environment (Edwards & Batley, 2004).

Transcriptomics

Knowledge of the expression pattern of genes provides a valuable insight into gene function and role in determining the observed heritable phenotype. The application of microarrays and

sequence-based methods to transcriptome analysis and expression profiling has added an extra dimension to current genomic data and has founded several statistics-based disciplines within bioinformatics. Owing to their extended linear dynamics, sequence-based methods have the potential to determine more accurately quantitative levels of gene expression. Furthermore, they do not require prior sequence information and so have the advantage of being able to identify novel genes or to assess gene expression in uncharacterized plant species. With the increasing scale of EST sequencing efforts, it is becoming possible to mine these datasets to estimate expression information (Rafalski *et al*., 1998), although this remains more a by-product of EST discovery than a true transcriptome analysis tool.

While the predominant methods for sequence-based expression analysis are serial analysis of gene expression (SAGE) (Velculescu *et al*., 1995) and massively parallel signature sequencing (MPSS) (Brenner *et al*., 2000a, 2000b), hybridisation-based microarrays have become the transcriptomic tool of choice. High–density cDNA and oligonucleotide microarrays represent powerful tools for transcriptome analysis to gain an understanding of gene expression patterns for thousands of genes. The rapid implementation of microarrays has been followed by a growth in the bioinformatics of microarray data analysis (Moreau *et al*., 2003; Goodman, 2002). Microarray technology continues to expand: cDNA arrays are being produced for gene expression analysis in many plant species, and complete oligonucleotide-based unigene arrays have been developed for major crop plants such as rice, wheat and barley.

Internationally coordinated efforts in transcriptome analyses and sharing of microarray resources will benefit the advancement of our understanding of gene function in forage and turf species. In this context, high-density cDNA microarrays representing approximately 15,000 unique genes for each of perennial ryegrass and white clover (Sawbridge *et al*., 2003a; b) and unigene microarrays allowing the interrogation of over 5,000 *Neotyphodium* and *Epichloe* genes (Nchip™ microarray and EndoChip™ microarray) (Felitti *et al*., 2004; Spangenberg *et al*., 2005) have been developed. These microarrays have been applied in hybridisations with labelled total RNA isolated from a variety of genotypes, plant organs, developmental stages, and growth conditions. The collated results enabled validation of functions predicted through comparative sequence annotation and suggested roles for novel genes lacking comparative sequence annotation. This data allowed expression analysis of genes associated with selected metabolic pathways and developmental processes, to dissect these at the transcriptome level, and to identify novel genes co-regulated with template genes known to be involved in these processes. Furthermore, these microarrays have enabled applications for gene and promoter discovery when used in concert with BAC libraries established for each of these target species. An *International Transcriptome Initiative for Forage and Turf* (ITIFT) to facilitate international efforts in microarray-based transcriptome analyses for key forage and turf plants and their endosymbionts is proposed here. Support of ITIFT will be provided through contributed access to and transcriptional profiling with unigene microarrays for key forage and turf species, namely perennial ryegrass, white clover and grass endophytes (*Neotyphodium*/*Epichloe*), as well as access to a range of platforms for microarray spotting, hybridisation and scanning operationally integrated through a Scierra laboratory workflow system. An ITIFT database with a web-based front-end portal for secure access by the research community to appropriate data and information will be developed and maintained.

Proteomics

The term ‘proteomics’ was coined in the mid 1990s (Wilkins *et al*., 1996, 1997) on the back of the success of ‘genomics’ and has since come to incorporate many aspects of protein

biochemistry. Knowledge of the structure and function of every protein would revolutionise the field of proteomics. Proteomics has significant prospects for advancing our understanding in plant biotechnology owing to its direct relationship with gene and transcript data. Proteomes also have a strong influence on the measured phenotype of the plant, either directly through protein content or function or indirectly through the relationship of a protein with the metabolome. The potential for bioinformatics to structure and integrate –omic data, therefore, relies on an ability to model both the proteome and its interactions (Edwards & Batley, 2004). With the first reports on comprehensive proteome analysis in model plant systems now becoming available (Canovas *et al*., 2004), there is still substantial scope for analogous developments in forage and turf plants and significant challenges ahead in the translation of complete genome DNA sequence data into protein structures and predicted functions.

Metabolomics

Like proteomics, metabolomics was derived from the field of biochemistry and involves the analysis (usually high-throughput or broad scale) of small-molecule metabolites. The foundations of metabolomics are descriptions of biological pathways, and current metabolomic databases such as Kyoto Encyclopedia of Genes and Genomes (Kanehis *et al*., 2000, 2002) are frequently based on well-characterized pathways. Metabolomics might be considered to be the key to integrated systems biology because it frequently provides direct measure of the desired phenotype (Fiehn, 2002), with relevance to both quantitative and qualitative traits. Moreover, metabolomes can be correlated with genetics through proteomes, transcriptomes and genomes and complement the more traditional quantitative trait locus (QTL) approach to molecular breeding. One of the challenges for bioinformatics will be the structuring and integration of these diverse types of data for the emerging field of systems biology (Fernie, 2003; Sumner *et al*., 2003; Weckworth, 2003). Emerging applications of metabolome analysis in forage and turf are currently focused primarily on *M. truncatula* ecotypes and elicited cell cultures (Sumner *et al*., 2003; May, 2004), as well as on metabolic fingerprinting and metabolic profiling of the mutualistic interaction between perennial ryegrass and its fungal endophyte within the context of a spatial and temporal systems biology approach (Spangenberg *et al*., 2005).

Bioinformatics, phenomics and data integration from genome to phenome

Bioinformatics arose from the need to structure and to interrogate the ever-increasing quantity and types of biological data that is generated by the developing '–omic' technologies. Genomics has spawned a plethora of related –omics terms that frequently relate to established fields of research. Of these terms, 'phenomics', the high-throughput analysis of phenotypes, has probably the biggest application in molecular breeding. Phenomics platforms using large scale detailed phenotypic data collection strategies based on series of distinct growth stages that describe the entire plant life cycle have been developed in recent years by agribusinesses for *A. thaliana* (e.g. Paradigm Genetics; K. R. Davies, pers. comm.) and rice (e.g. CropDesign; W. Broekaert, pers. comm.). As these technologies continue to grow, so does the need for interrogation across the various types of data and scientific disciplines. Precise data integration requires the formal annotation of data with relational terms, and this is an essential driver behind applications of bioinformatics in the development of systems biology (Edwards & Batley, 2004). Recent bioinformatics developments for systems biology in forage and turf plants have primarily focused on tools for data integration and target selection in *M. truncatula* (Wang & Zhang, 2004), perennial ryegrass and white clover (Love *et al*., 2005).

Transgenesis for forage and turf molecular breeding

Transgenesis offers the opportunity to generate unique genetic variation for molecular breeding of forage and turf, specifically when pre-existing variability is either absent or shows very low levels of heritability (Spangenberg *et al*., 2001). Transgenic forage and turf plants with simple 'engineered' traits have been generated and characterised over the last decade (Hartman *et al*., 1994; Austin-Phillips and Ziegelhoffer, 2001; Kalla *et al*., 2001; Wang *et al*., 2001; Xu *et al*., 2001; Chen *et al*., 2003; Chen *et al*., 2004; Luo *et al*., 2004; Petrovska *et al*., 2004), with few selected transformation events having already reached the stage of field-evaluation and transgenic germplasm development (Harriman *et al*., 2003; Emmerling *et al*., 2004; Smith *et al*., 2005).

In the near future, applications of transgenesis to forage and turf plant improvement are likely to remain primarily focused on the development of transformation events with unique genetic variation, on studies on the molecular genetic dissection of key developmental processes (e.g. leaf senescence, transition to flowering, pollen development), metabolic pathways (e.g. lignin biosynthesis, fructan metabolism, fatty acid biosynthesis, proanthocyanidin biosynthesis), and on the assessment of candidate genes for enhancing tolerance to biotic and abiotic stresses.

An emphasis on key target traits for the application of transgenesis to forage and turf improvement will be retained for forage quality, disease and pest resistance, nutrient acquisition efficiency, tolerance to abiotic stresses and the manipulation of growth and development. Corresponding experimental approaches and representative examples in temperate forage and turf grasses and forage legumes have been reviewed in detail over recent years (Spangenberg *et al*., 1998; 2001; 2005; Wang *et al*., 2001; Gruber *et al*., 2001; Humphreys & Abberton, 2004; Dixon, 2004) and are consequently not considered in any further detail here.

Increasingly, transgenesis applications in forage and turf improvement will address outcome scenarios associated with animal health and welfare, such as transgenic white clover and alfalfa with 'bloat safety' through foliar accumulation of proanthocyanidins, transgenic alfalfa with foliar expression of foot and mouth disease antigens, as well as environmental and human health, such as transgenic ryegrasses with down-regulation of main pollen allergens, transgenic forage grasses and legumes with modified fatty acid biosynthesis for animal products with enhanced levels of 'healthier fats', transgenic turf grasses for bioremediation applications, transgenic alfalfa with salt tolerance, and transgenic white clover with enhanced phosphorus acquisition efficiency.

As described above, significant advances in gene discovery for key forage and turf plants and their endosymbionts (Sawbridge *et al*., 2003a; 2003b; Felitti *et al*., 2004), and opportunities arising from translational genomics derived from EST discovery and sequencing of gene-rich regions of relevant model systems such as *M. truncatula* (May, 2004; Roe & Kupfer, 2004) and whole genome sequencing of rice (Goff *et al*., 2002; Yu *et al.,* 2002) have established an overwhelming amount of information concerning the physical structure of thousands of specific genes. This has led to the requirement of efficient experimental approaches for high-throughput functional analysis of the large number of genes for which limited or no functional information is available.

While robust methodologies for genetic transformation of key forage and turf grasses and forage legumes have been established over the last decade (for reviews refer to McKersie &

Brown, 1997; Spangenberg *et al*., 1997; 1998; 2001; Forster & Spangenberg, 1999; Wang *et al*., 2001), these methods are generally low-throughput in nature and often lead to multiple gene insertions. They are thus largely inadequate for use in forward and reverse genetic approaches to determine plant gene function. Consequently, efficient methods and tools for high-throughput gene silencing (e.g. high-throughput construction hairpin RNA vectors, virus-induced gene silencing, biolistic delivery of siRNAs), for large-scale, fast and low-cost transient gene function assays as well as the development of T-DNA insertional mutant collections and T-DNA activation tagging lines are required for key forage and turf plants and their endosymbionts, such as white clover, perennial ryegrass and *Neotyphodium* grass endophytes.

The effective application of systems biology approaches to forage and turf plants will close gaps in our understanding of the underlying genetics, physiology and biochemistry of many complex plant pathways and developmental processes. This is likely to advance progress in many applications of transgenesis in forage and turf plant improvement, building on gene technology as a powerful tool for the generation of the required functional genomics knowledge. Furthermore, requisite improved transformation methodologies for the development of 'market-ready' (i.e. single transgene inserts; free of selectable markers; deploying primarily host gene sequences) transformation events of prospective 'pyramided' gene technologies (that is, deployment of multiple transgenes in a single transformation event) following a sensible choice of high impact targets, and permitting the stable transfer of large inserts (e.g. BAC clones for QTL transfer) will be developed and applied in molecular breeding of forage and turf plants in the not too distant future.

Accompanying these developments, biosafety research including modelling of transgenic pastoral production systems and the development of transgene tracking and tracing assays will be required to meet needs of regulatory frameworks currently based on extreme interpretations of the precautionary principle.

Functionally-associated genetic markers for forage and turf molecular breeding

As previously described (Forster *et al*., 2004), the majority of research to date on molecular marker development and validation in outcrossing forage species has been based on anonymous genetic markers, such as genomic DNA-derived SSRs and amplified fragment length polymorphisms (AFLPs) (Jones *et al*., 2002a,b; Jones *et al*., 2003). The paradigm for marker-assisted selection (MAS) that was established in autogamous plant species such as tomato, rice and wheat involves the use of such markers to construct linkage maps, genetic trait dissection through QTL analysis, and selection of linked markers in selection schemes such as donor-recipient recurrent selection. The obligate outbreeding nature of many pasture grasses and legumes clearly presents major limitations to the ready implementation of the inbreeding paradigm.

The use of markers that are in linkage of varying strength, rather than directly associated with the gene of interest, is a problem for both inbreeding and outbreeding species. For this reason, closely linked markers, ideally flanking the target region, are preferred. However, given fixation of the target region in an inbred background, to generate a homogeneous variety, the problem of potential reversal of linkage between favourable gene variant and selected marker allele is eliminated. In the context of a genetically heterogeneous synthetic population, complete fixation of a target genomic region will be difficult and slow to achieve, and consequently, the probability of recombination to decouple the favourable marker-trait allele

combination will be high, especially in the absence of closely linked flanking markers. This logic implies that diagnostic genetic markers are of even higher potential value for outbreeding than inbreeding crops.

A further problem for the implementation of inbreeding paradigm MAS for pasture species is the large number of parental genotypes that are generally used in the polycross design for synthetic development. The number of foundation individuals varies between restricted base varieties (4-6 parents) and non-restricted base varieties (6-100+ parents) (Guthridge *et al*., 2001; Forster *et al*., 2001). Even for restricted base varieties, the process of tagging each gene variant in the parental genotypes with linked markers would imply multiple cycles of genetic trait-dissection in pair cross-derived mapping families. This is in contrast to the situation for inbreeders, in which the conduct of the trait-dissection process in a sib-ship derived from crossing of the future donor and recipient lines provides all relevant information for subsequent recurrent selection. One way to overcome this multiplicity of marker-trait allele associations would be to pre-introgress the desired combination into each of the selected parents. However, this implies a prior round of MAS, and does not adequately address the logistical complexity problem for molecular breeding of outcrossing forages.

The most obvious solution to these problems is to develop candidate gene-based markers that show a functional association with the target trait region (Andersen & Lübberstedt, 2003). Based on the population biology of perennial ryegrass (outbreeding with relatively large effective population sizes, at least for ecotypic populations), linkage disequilibrium (LD) is expected to extend over relatively short molecular distances, although variations of recombination frequency in different genotypes and different genomic regions may complicate this scenario (Forster *et al*., 2004). In this instance, it should be possible to identify diagnostic variants for the selection of individual parental genotypes on the basis of superior allele content, given selection of the correct candidate gene. This will allow more efficient use of germplasm collections for parental selection. In addition, such 'perfect' markers will allow highly effective progeny selection.

The large-scale gene sequence collections generated by both incremental and EST discovery in perennial ryegrass and white clover provide the resource for functionally-associated marker development, with c. 15,000 unigenes currently defined for each species (Sawbridge *et al*., 2003a; b). Selected genes have already been mapped as gene-associated RFLP and SSR loci (Barrett *et al*., 2004; Faville *et al*., 2005). RFLP markers are not readily implemented in molecular breeding, and SSRs are only present in a sub-set (generally less than 10%) of target genes. However, genic SNP markers can in principle be developed for any gene, and show the benefits of locus-specificity, high data fidelity and high-throughput analysis.

Our gene-associated SNP discovery process is based on a four-part strategy. The 'ultra fast-track' approach is based on validation of SNPs detected *in silico* in EST contigs. The perennial ryegrass and white clover EST collections are derived from multiple heterogeneous individuals from cultivars Grasslands Nui (Sawbridge *et al*., 2003a) and Grasslands Huia (Sawbridge *et al*., 2003b), respectively. SNPs may be distinguished from sequencing errors based on co-segregation data generated by applications such as AutoSNP (Barker *et al*., 2003), and are validated by transfer and segregation analysis in pair cross-derived mapping families such as $F_1(NA_6 \times AU_6)$ (Faville *et al*., 2005). Current data from perennial ryegrass suggests that c. 15% of *in silico* SNPs detect polymorphic loci in the mapping family. However, the majority of targeted genes may not been accessible to this approach. The other three approaches are based on *in vitro* discovery and are differentiated by the number of SNPs

obtained: 'fast-track' involves short ESTs, providing single SNP loci for structured map enhancement; 'medium-track' involves full-length cDNAs, providing several SNP loci and partial haplotypic data; 'slow-track' is based on full-length genes with intron-exon structure, providing multiple SNP loci and determination of complete haplotype structures.

The experimental method for SNP discovery is based on cloning and sequencing of gene-specific amplicons from the heterozygous parents of two-way pseudo-testcross mapping families. Although this method is costly and time-consuming compared to direct sequencing of PCR products (Zhu *et al*., 2003; Rickert *et al*., 2003), it has several important advantages. Ambiguities in sequence traces due to heterozygous InDels may be readily distinguished, and haplotype structures within amplicons are directly determined (Zhang & Hewitt, 2003). In addition, paralogous sequence amplified by conserved primers are discriminated, and may be used to distinguish between members of multigene families. The putative SNPs are then validated in the progeny set, and cross-validated in other sib-ships and diverse germplasm.

'Proof-of-concept' for this process was obtained with the perennial ryegrass *Lp*ASRa2 gene, which shows 89% nucleotide identity with the rice *OsAsr1* gene. The *Asr* gene family encodes a group of proteins that are transcriptionally-induced by ABA treatment and water stress, and during fruit ripening. Osmotic and saline stress leads to up-regulation of the rice gene (Vaidyanathan *et al*., 1999), and the maize *Zm-Asr1* gene co-locates with QTLs for traits responsive to mild water stress (Jeanneau *et al*., 2002). The *Lp*ASRa2 gene consequently provides an excellent candidate for the assessment of correlation between genic sequence polymorphism and phenotypic variation. The full-length cDNA (890 bp) was tiled with 4 amplicons, covering 716 bp of the 5'-UTR, CDS and 3'UTR, and including a single 100 bp intron. A total of 9 SNPs were detected within and between the parents of the $F_1(NA_6 \times AU_6)$ mapping family. Of these, 7 SNP loci showing segregation in the progeny were assigned to coincident locations on linkage group (LG) 4 of the NA_6 parental genetic map, directly adjacent to the corresponding RFLP locus (Figure 1A). Partial haplotypic data for *Lp*ASRa2 reveals the maximum variant number of four, three of which are closely related, while the fourth is more divergent, defining two putative haplogroups (Olsen *et al*., 2004). Although the majority of the exon-located changes define synonymous amino acid changes (Figure 1B), two SNPs defined amino acid substitutions, one of which (glutamate to glutamine at coordinate 136) produces a radical change, and may be functionally significant. Alternatively, the characterised SNPs may be in LD with functionally-significant changes in the transcriptional control regions (Paran & Zamir, 2003), given haplotype stability over gene-length distances.

Preliminary analysis of haplotype variation has been performed using a diversity panel including representatives of all *Lolium* taxa, including the three outbreeding cultivated species and several wild inbreeding species, with 8 validated *Lp*ASRa2 SNPs. Because of the uncertainty over *cis-trans* relationships between polymorphic SNP loci in heterozygous genotypes, a minimum-maximum range of haplotype numbers must be defined. However, inferences based on the prevalence of certain haplotype combinations in the homozygous state, as proposed by Clark (1990), suggest that the ryegrasses, like maize, may possess limited numbers of diverse haplotypes (Rafalski & Morgante, 2004).

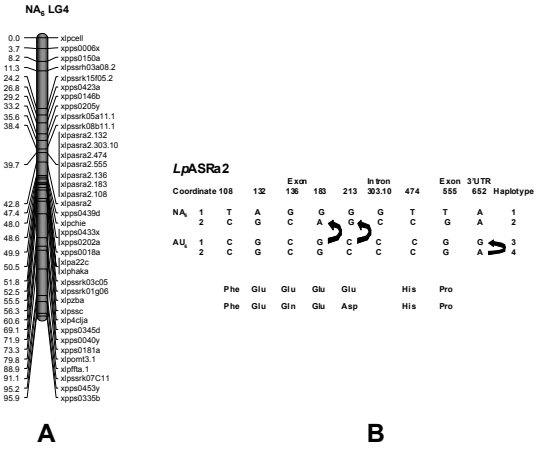

Figure 1 A. Genetic linkage map of LG4 of the NA_6 parental map, showing the SNP loci (indicated as xlpasra2.coordinate number) in close linkage with the corresponding RFLP locus (xlpasra2). B. *Lp*ASRa2 haplotype structures within and between the NA_6 and AU_6 parental genotypes. Arrows show putative mutational changes between members of the second haplogroup (haplotypes 2, 3 and 4), and predicted translation products of exon-located SNP loci are indicated.

By the start of 2005, over 150 genes had been introduced into the *in vitro* SNP discovery, of which 98 were sequenced and aligned, with a total of 1542 putative SNPs across 78 genes. SnuPe-validated SNPs were detected for 64 genes. Over a total of 83 kb of re-sequenced DNA, a relatively high SNP frequency (for four haplotypes) of 1/54 bp was observed, with an anticipated higher incidence in intron compared to exon sequences. For full-length herbage quality genes (the lignin biosynthetic genes *Lp*CCR1 and *Lp*CAD2 and the fructan metabolism genes *Lp*FT1 and *Lp*1-SST), a large number of SNPs have been identified (up to 265). Polymorphic SNPs have been validated at positions up to 6 kb distant in these genes, providing the basis for LD studies over gene-length distances.

Success in the detection of functionally-associated SNP variation must be followed by demonstration of causal correlation between genotypic diversity and phenotypic variation. An important pre-requisite for such analysis is the observation of candidate gene-QTL co-location. For perennial ryegrass, such coincidences have been observed for the orthologue of wheat *VRN1* with a QTL for vernalisation response on LG4 (Jensen *et al*., 2005), the orthologue of the rice semi-dwarf gene *SD1* with a QTL for plant height on LG3 (T. Yamada, pers. comm.), and the lignin biosynthetic genes *Lp*OMT1, *Lp*CCR1 and *Lp*CAD2 with QTLs for herbage digestibility on LG7 (Cogan *et al*., 2005). Experiments are currently being performed to evaluate the association between haplotype variation in genes such as *Lp*ASRa2 and *Lp*CCR1 and various aspects of phenotypic variation, such as gene transcription levels, metabolite synthesis and field phenotypes such as drought tolerance and herbage digestibility, respectively.

Apart from practical applications, candidate gene-based marker development provides an auxiliary source of information for functional genomics analysis of gene function. Given success in the validation of functionally-associated markers, an optimised strategy for forage biotechnology may be devised, in which the transgene-based technologies described in previous sections are deployed in highly adapted germplasm derived by marker-assisted breeding.

Positional cloning of genes from forage and turf species

The development of high-resolution genetic maps and functionally-associated markers, along with BAC libraries with high genome coverage, provides the basis for highly targeted physical mapping and eventual positional cloning of forage plant genes. The criteria for adoption of such an approach will be identification of genes controlling specific (initially qualitative) variation for important agronomic traits, for which minimal biochemical data to infer gene function is available. The *Festuca pratensis* 'stay-green' gene *sid* belongs to this category, and has been the subject of physical mapping studies (Donnison *et al*., 2003). The highest priority targets in our studies are the *S* and *Z* genes that control gametophytic self-incompatibility (SI) in grasses (Cornish *et al*., 1979). Isolation and construction of allele-specific markers for *S* and *Z* will be crucial for the effective implementation of molecular breeding, particularly in the context of restricted-base varietal development. The SI genes have been mapped on LGs 1 and 2 of perennial ryegrass in locations predicted through conserved synteny with rye and *Phalaris coerulescens* (Thorogood *et al*., 2002). Fine mapping data from *P. coerulescens* (Bian *et al*., 2004), including the location of a thioredoxin gene in close linkage to *S* (Baumann *et al*., 2000), suggests a strategy for BAC contig construction in the equivalent *L . perenne* region. Similar data is now available for the Z gene-region of cereal rye (*Secale cereale* L.) (Hackauf & Wehling, 2005). Other genomic regions will be targeted as more data on gene-trait variation associations becomes available.

Conclusion

Forage and turf plants molecular breeding has now clearly entered the post-genomic era, reliant on high-throughput approaches for spatial and temporal analysis from genome to phenome, the respective data integration in a systems biology context for the establishment of stringent gene-function correlations. A primary challenge ahead is now the conversion of the vast amount of structural genomics information available into useful functional information. Model plant systems such as *Arabidopsis*, rice and *M. truncatula* will play a major role in the elucidation of gene function in higher plants. Translational genomics will allow for results obtained using these model systems to have major impact on understanding the molecular basis of plant processes and have direct application to the molecular breeding of forage and turf plants. These developments will be enhanced through applications of transgenesis and functionally-associated genetic markers in forage and turf molecular breeding building on genomic and post-genomic discoveries in these target species.

Acknowledgements

The support by the Victorian Department of Primary Industries, the Molecular Plant Breeding Cooperative Research Centre, Dairy Australia Ltd., Meat & Livestock Australia Ltd., the Gardiner Dairy Foundation, the Victorian Microarray Technology Consortium, the Victorian Bioinformatics Consortium, the Victorian Centre for Plant Functional Genomics, the Australian Centre for Plant Functional Genomics and the Argentine Research Council is gratefully acknowledged.

References

Adams, M.D., J.M. Kelley, J.D. Gocayne, M. Dubnick, M.H. Polymeropoulos, H. Xiao, C.R. Merril, A. Wu, B. Olde, R.F. Moreno *et al*. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252, 1651-1656.

Andersen, J.R. & T. Lübberstedt (2003) Functional markers in plants. *Trends in Plant Science*, 8, 554-560.

Austin-Phillips, S. & T. Ziegelhoffer (2001) The production of value-added proteins in transgenic alfalfa. In: G. Spangenberg, (ed.). Molecular Breeding of Forage Crops. Kluwer Academic Press, 285-301.
Barker, G., J. Batley, H. O'Sullivan, K.J. Edwards & D. Edwards (2003) Redundancy-based detection of sequence polymorphisms in expressed sequence tag data using AutoSNP. *Bioinformatics*, 19, 412-422.
Barrett, B., A. Griffiths, M. Schreiber, N. Ellison, C. Mercer, J. Bouton, B. Ong, J. Forster, T. Sawbridge, G. Spangenberg, G. Bryan & D. Woodfield, D. (2004) A microsatellite map of white clover (*Trifolium repens* L.). *Theoretical and Applied Genetics*, 109, 596-608.
Barry, G.F. (2001) The use of the Monsanto draft rice genome sequence in research. *Plant Physiol*., 125, 1164-1165.
Batley, J., G. Barker, H. O'Sullivan, K.J. Edwards & D. Edwards (2003) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiol*., 132, 84-91.
Baumann, U., J. Juttner, B. Xueyu & P. Langridge (2000) Self incompatbility in the grasses. *Annals of Botany*, 85 (Supplement A), 203-209.
Bian, X.Y., A. Friedrich, J.R. Bai, U. Baumann, D.L. Hayman, S.J. Barker & P. Langridge (2004) High-resolution mapping of the *S* and *Z* loci in *Phalaris coerulescens*. *Genome*, 47, 918-930.
Brenner, S., M. Johnson, J. Bridgham, G. Golda, D.H. Lloyd, D. Johnson, S. Luo, S. McCurdy, M. Foy, M. Ewan, R. Roth, D. George, S. Eletr, G. Albrecht, E. Vermaas, S.R. Williams, K. Moon, T. Burcham, M. Pallas, R.B. DuBridge, J. Kirchner, K. Fearon, J. Mao & K. Corcoran (2000a) Gene expression analysis by massively parallel signal sequencing (MPSS) on microbead arrays. *Nature Biotechnology*, 18, 630-634.
Brenner, S., S.R. Williams, E.H. Vermaas, T. Storck, K. Moon, C. McCollum, J.I. Mao, S. Luo, J.J. Kirchner, S. Eletr, R.B. DuBridge, T. Burcham & G. Albrecht (2000b) *In vitro* cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci*., 97, 1665-1670.
Canovas, F.M., E. Dumas-Gaudot, G. Recorbet, J. Jorrin, H.P. Mock & M. Rossignol (2004) Plant proteome analysis. *Proteomics*, 4, 285-298.
Chen, L., C.K. Auh, P. Dowling, J. Bell, F. Chen, A. Hopkins, R.A. Dixon & Z.Y. Wang (2003) Improved forage digestibility of tall fescue (*Festuca arundinacea*) by transgenic down-regulation of cinnamyl alcohol dehydrogenase. *Plant Biotechnology Journal*, 1, 437-449.
Chen, L., C.K. Auh, P. Dowling, J. Bell, D. Lehman & Z.Y. Wang (2004) Transgenic down-regulation of caffeic acid *O*-methyltransferase (COMT) led to improved digestibility in tall fescue (*Festuca arundinacea*). *Functional Plant Biology*, 31, 235-245.
Clark, A.G. (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. *Molecular Biology and Evolution*, 7, 111-122.
Cogan, N.O.I., K.F Smith, T. Yamada, M.G. Francki, A.C. Vecchies, E.S. Jones, G.C Spangenberg & J.W Forster (2005) QTL analysis and comparative genomics of herbage quality traits in perennial ryegrass (*Lolium perenne* L.). *Theoretical and Applied Genetics*, 110, 364-380.
Cornish, M.A., M.D. Hayward & M.J. Lawrence (1979) Self-incompatibility in ryegrass. I. Genetic control in diploid *Lolium perenne* L. *Heredity*, 43, 95-106.
Dixon, R.A.(2004) Molecular improvement of forages – from genomics to GMOs. In: A. Hopkins, Z.-Y. Wang, M. Sledge, R.E. Barker (eds.). Molecular Breeding of Forage and Turf. Kluwer Academic Press, 1-19.
Donnison, I., D. O'Sullivan, A. Thomas, H. Thomas, K. Edwards, H.-M. Thomas, I. King (2003) Plant and Animal Genome XI, San Diego, California, W218.
Edwards, D. & J. Batley (2004) Plant bioinformatics: from genome to phenome. *Trends in Biotechnology*, 22, 232-237.
Emmerling, M., P. Chu, K. Smith, R. Kalla & G. Spangenberg (2004) Field evaluation of transgenic white clover with AMV Immunity and development of elite transgenic germplasm. In: A. Hopkins, Z.-Y. Wang, M. Sledge, R.E. Barker (eds.). Molecular Breeding of Forage and Turf. Kluwer *Academic Press, 359-366.*
Ettwiller, L.M., J. Rung & E. Birney (2003) Discovering novel cis-regulatory motifs using functional networks. *Genome Res*., 13, 883-895.
Faville, M., A.C. Vecchies, M. Schreiber, M.C. Drayton, L.J. Hughes, E.S. Jones, K.M. Guthridge, K.F. Smith, T. Sawbridge, G.C. Spangenberg, G.T. Bryan & J.W. Forster (2005) Candidate gene-based molecular marker map construction in perennial ryegrass (*Lolium perenne* L.). *Theoretical and Applied Genetics*, 110, 12-32.
Felitti, S., K. Shields, M. Ramsperger, P. Tian, T. Webster, B. Ong, T. Sawbridge & G. Spangenberg (2004) Gene discovery and microarray-based transcriptome analysis in grass endophytes. In: A. Hopkins, Z.-Y. Wang, M. Sledge, R.E. Barker (eds.). Molecular Breeding of Forage and Turf. Kluwer Academic Press, 145-153.
Fernie, A.R. (2003) Metabolome characterisation in plant systems analysis. *Funct. Plant Biol*., 30, 111-120.
Fiehn, O. (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Biol*., 48, 155-171.
Forster, J.W. & G. Spangenberg (1999) Forage and turf grass biotechnology: principles, methods and prospects. In: J.K. Setlow (ed.) Genetic Engineering: Principles and Methods, Vol 21, Kluwer Academic Publishers, 191-237.

Forster, J.W., E.S. Jones, R. Kölliker, M.C. Drayton, J. Dumsday, M.P. Dupal, K.M. Guthridge, N.L. Mahoney, E. van Zijll de Jong & K.F. Smith, (2001) Development and implementation of molecular markers for forage crop improvement. In: G. Spangenberg, (ed.). Molecular Breeding of Forage Crops. Kluwer Academic Press, 101-133.

Forster, J.W., E.S. Jones, J. Batley & K.F. Smith (2004) Molecular marker-based genetic analysis of pasture and turf grasses. In: A. Hopkins, Z.-Y. Wang, M. Sledge, R.E. Barker (eds.). Molecular Breeding of Forage and Turf. Kluwer Academic Press, 197-239.

Goodman, N. (2002) Biological data becomes computer literate: new advances in bioinformatics. *Curr. Opin. Biotechn*., 13, 68-71.

Gruber, M.Y., H. Ray & L. Blahut-Beatty (2001) Genetic manipulation of condensed tannin synthesis in forage crops. In: G. Spangenberg, (ed.). Molecular Breeding of Forage Crops. Kluwer Academic Press, 189-201.

Guthridge, K.M., M.D. Dupal, R. Kölliker, E.S. Jones, K.F. Smith & J.W. Forster (2001) AFLP analysis of genetic diversity within and between populations of perennial ryegrass (*Lolium perenne* L.). *Euphytica*, 122, 191-201.

Goff, S.A., D. Ricke, T.H. Lan, G. Presting, R. Wang, M. Dunn, J. Glazebrook, A. Sessions, P. Oeller, H. Varma, D. Hadley, D. Hutchinson, C. Martin, F. Katagiri, B.M. Lange, T. Moughamer, Y. Xia, P. Budworth, J. Zhong, T. Miguel, U. Paszkowski, S. Zhang, M. Colbert, W.L. Sun, L. Chen, B. Cooper, S. Park, T.C. Wood, L. Mao, P. Quail, R. Wing, R. Dean, Y. Yu, A. Zharkikh, R. Shen, S. Sahasrabudhe, A.Thomas, R. Cannings, A. Gutin, D. Pruss, J. Reid, S. Tavtigian, J. Mitchell, G. Eldredge, T. Scholl, R.M. Miller, S. Bhatnagar, N. Adey, T. Rubano, N. Tusneem, R. Robinson, J. Feldhaus, T. Macalma, A. Oliphant & S. Briggs (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). *Science* 296, 92-100.

Hackauf, B. & P. Wehling (2005) Approaching the self-incompatibility locus Z in rye (*Secale cereale* L.) via comparative genetics. *Theoretical and Applied Genetics*, available on-line.

Harriman, R.W., E. Nelson & L. Lee (2003) Enhance turfgrass performance with biotechnology. In: I.K. Vasil (ed.) Plant Biotechnology 2002 and Beyond, Proceedings of the 10th IAPTC7B Congress, Kluwer Academic Publishers, 503-506.

Hartman, C.L., L. Lee, P.R. Day & E.T. Nilgun (1994) Herbicide resistant turfgrass (*Agrostis palustris* Huds.) by biolistic transformation. *Bio/Technology*, 12, 919-923.

Humphreys, M.O. & M.T. Abberton (2004) Molecular breeding for animal, human and environmental welfare. In: A. Hopkins, Z.-Y. Wang, M. Sledge, R.E. Barker (eds.). Molecular Breeding of Forage and Turf. Kluwer Academic Press, 165-180.

Jeanneau, M., D. Gerentes, X. Foueillassar, M. Zivy, J. Vidal, A. Toppan, P. Perez (2002) Improvement of drought tolerance in maize: towards the functional validation of the *Zm-Asr1* gene and increase of water use efficiency by over-expressing C4-PEPC. *Biochimie*, 84, 1127-1135.

Jensen, L.B., J.R. Andersen, U. Frei, Y. Xing, C. Taylor, P.B. Holm, T. Lübberstedt (2005) QTL mapping of vernalization response in perennial ryegrass (*Lolium perenne* L.) reveals co-location with an orthologue of wheat *VRN1*. *Theoretical and Applied Genetics*, 110, 527-536.

Jones, E.S., N.L. Mahoney, M.D. Hayward, I.P. Armstead, J.G. Jones, M.O. Humphreys, I.P. King, T. Kishida, T. Yamada, F. Balfourier, C. Charmet & J.W. Forster (2002) An enhanced molecular marker-based map of perennial ryegrass (*Lolium perenne* L.) reveals comparative relationships with other Poaceae species. *Genome*, 45, 282-295.

Jones, E.S., M.D. Dupal, J.L. Dumsday, L.J. Hughes & J.W. Forster (2002) An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). *Theoretical and Applied Genetics*, 105, 577-584.

Jones, E.S., L.J. Hughes, M.C. Drayton, M.T. Abberton, T.P.T. Michaelson-Yeates & J.W. Forster (2003) An SSR and AFLP molecular marker-based genetic map of white clover (*Trifolium repens* L.). *Plant Science,* 165, 531-539.

Kalla, R., P. Chu & G. Spangenberg (2001) Molecular breeding of forage legumes for virus resistance. In: G. Spangenberg (ed.) Molecular Breeding of Forage Crops, Kluwer Academic Publishers, 219-237.

Kanehisa, M. & S. Goto (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*., 28, 27-30.

Kanehisa, M., S. Goto, S. Kawashima and A. Nakaya (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res*., 30, 42-46.

May, G.D. (2004) From model to crops: integrated *Medicago* genomics for alfalfa improvement. In: A. Hopkins, Z.-Y. Wang, M. Sledge, R.E. Barker (eds.). Molecular Breeding of Forage and Turf. Kluwer Academic Press, 325-332.

Moreau, Y., S. Aerts, B. De Moore, B. De Strooper & M. Dabrowski (2003) Comparison and meta-analysis of microarray data: from the bench to the computer desk. *Trends Genet*., 19, 570-577.

Morgante, M. & F. Salamini (2003) From plant genomics to breeding practice. *Curr. Opin. Biotechnol*., 14, 214-219.

Love, C.L., T.A. Erwin, E.G. Logan, G. Barker, N. James, S. May, G. Spangenberg & D. Edwards (2005) ASTRA: integrated bioinformatics tools for ryegrass and white clover genomics. Plant and Animal Genome XIII, San Diego, California, W098.

Luo, H., Q. Hu, K. Nelson, C. Longo & A.P. Kausch (2004) Controlling transgene escape in genetically modified grasses. In: A. Hopkins, Z.-Y. Wang, M. Sledge, R.E. Barker (eds.). Molecular Breeding of Forage and Turf. Kluwer Academic Press, 245-254.

McKersie B.D. & D.C.W. Brown (1997) Biotechnology and the improvement of forage legumes. In: Biotechnology in Agricultural Series No. 17. CAB International, Wallingford, 444.

Nelson, P., M. Kiriakidou, A. Sharma, E. Maniataki & Z. Mourelatos (2003) The microRNA world: small is mighty. *Trends Biochem. Sci*., 28, 534-540.

Olsen, K.M., S.S. Halldorsdottir, J.R. Stinchcombe, C. Weinig, J. Schmitt & M.D. Purugganan, (2004) Linkage disequilibrium mapping of *Arabidopsis CRY2* flowering time alleles. *Genetics*, 167, 1361-1369.

Paran, I. & D. Zamir (2003) Quantitative traits in plants: beyond the QTL. *Trends in Genetics*, 19, 303-306.

Petrovska, P., X. Wu, R. Donato, Z.Y. Wang, E.K. Ong, E. Jones, J. Forster, M. Emmerling, A. Sidoli, R. O'Hehir & G. Spangenberg (2004) Transgenic ryegrasses (*Lolium* spp.) with down-regulation of main pollen allergens. *Molecular Breeding*, 14, 489-501.

Qui, P. (2003) Recent advances in computational promoter analysis in understanding the transcriptional regulatory network. *Biochem. Biophys. Res. Commun*., 309, 495-501.

Rafalski, J.A., M. Hanafey, G.H. Miao, A. Ching, J.M. Lee, M. Dolan & S. Tingey (1998) New experimental and computational approaches to the analysis of gene expression. *Acta Biochim. Pol*., 45, 929-934.

Rafalski, A. & M. Morgante, (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics*, 20, 103-111.

Rhoades, M.W., B.J. Reinhart, L.P. Lim, C.B. Burge CB, B. Bartel & D.P. Bartel (2003) Prediction of plant microRNA targets. *Cell*, 110, 513-520.

Rickert, A.M., J.H. Kim, S. Meyer, A. Nagel, A. Ballvora, P.J. Oefner & C. Gebhardt, (2003) First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome. *Plant Biotech. J.*, 1, 399-410.

Robinson, A.J., Love, C.G., Batley, J., Barker, G., Edwards, D. (2004) Simple sequence repeat marker loci discovery using SSR primer. *Bioinformatics* 10.1093/bioinformatics/bth104 (www.bioinformatics.oupjournals.org).

Roe, B.A. & D.M. Kupfer (2004) Sequencing gene rich regions of *Medicago truncatula*, a model legume. In: A. Hopkins, Z.-Y. Wang, M. Sledge, R.E. Barker (eds.). Molecular Breeding of Forage and Turf. Kluwer Academic Press, 333-344.

Sawbridge, T., E.-K. Ong, C. Binnion, M. Emmerling, R. McInnes, K. Meath, N. Nguyen, K. Nunan, M. O'Neill, F. O'Toole, C. Rhodes, J. Simmonds, P. Tian, K. Wearne, T. Webster, A. Winkworth & G. Spangenberg, (2003a) Generation and analysis of expressed sequence tags in perennial ryegrass (*Lolium perenne* L.). *Plant Science*, 165, 1089-1100.

Sawbridge, T., E.-K. Ong, , C. Binnion, M. Emmerling, K. Meath, K. Nunan, M. O'Neill, F. O'Toole, J. Simmonds, K. Wearne, A. Winkworth & G. Spangenberg, (2003b) Generation and analysis of expressed sequence tags in white clover (*Trifolium repens* L.). *Plant Science*, 165, 1077-1087.

Smith, K.F., J.W. Forster, M.P. Dobrowolski, N.O.I. Cogan, N.R. Bannan, E. van Zijll de Jong, M. Emmerling & G.C. Spangenberg (2005) Application of molecular technologies in forage plant breeding. In: Molecular Breeding of Forage and Turf, Aberystwyth.

Somers, D.J., R. Kirkpatrick, M. Moniwa & A. Walsh (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. *Genome*, 46, 431-437.

Spangenberg, G., Z.Y. Wang, R. Heath, V. Kaul & R. Garrett (1997) Biotechnology in pasture plant improvement: methods and prospects. In: Proceedings XVIII International Grassland Congress, Winnipeg and Saskatoon, Vol 3.

Spangenberg, G., Z.Y. Wang & I. Potrykus (1998) Biotechnology in forage and turf grass improvement. In: R. Frankel (ed.) Monographs on Theoretical and Applied Genetics, Vol 23, Springer Verlag, Heidelberg, pp192.

Spangenberg, G., R. Kalla, A. Lidgett, T. Sawbridge, E.K. Ong & U. John (2001) Breeding forage plants in the genome era. In: G. Spangenberg (ed.) Molecular Breeding of forage crops, Kluwer Academic Publishers, 1-39.

Spangenberg, G., M. Emmerling, U. John, R. Kalla, A. Lidgett, E.K. Ong, T. Sawbridge & T. Webster (2003) Transgenesis and genomics in molecular breeding of temperate pasture grasses and legumes. In: I.K. Vasil (ed.) Plant Biotechnology 2002 and Beyond, Proceedings of the 10th IAPTC7B Congress, Kluwer Academic Publishers, 497-502.

Spangenberg, G., S. Felitti, K. Shields, M. Ramsperger, P. Tian, E.K. Ong, D. Singh, E. Logan & D. Edwards (2005) Gene discovery and microarray-based transcriptome analysis of the grass-endophyte association. In: C.A. Roberts, C.P.West & D.E. Spiers (eds.). *Neotyphodium* in Cool-Season Grasses. Blackwell Publishing, 103-121.

Sumner, L.W., P. Mendes & R.A Dixon (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*, 62, 817-836.

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome of *Arabidopsis thaliana*. *Nature*, 408, 796-815.
Thorogood, D., W.J. Kaiser, J.G. Jones & I. Armstead (2002) Self-incompatibility in ryegrass. 12. Genotyping and mapping the S and Z loci of *Lolium perenne* L. *Heredity*, 88, 385-390.
Vaidyanathan, R., S. Kuruvilla & G. Thomas (1999) Characterisation and expression pattern of an abscisic acid and osmotic stress responsive gene from rice. *Plant Science*, 140, 25-36.
Velculescu, V.E., L. Zhang, B. Vogelstein & K.W. Kinzler (1995) Serial analysis of gene expression. *Science*, 270, 484-487.
Wang, L. & Y. Zhang (2004) Data integration and target selection for *Medicago* genomics. In: A. Hopkins, Z.-Y. Wang, M. Sledge, R.E. Barker (eds.). Molecular Breeding of Forage and Turf. Kluwer Academic Press, 275-288.
Wang, Z.Y., X.D. Ye, J. Nagel, I. Potrykus & G. Spangenberg (2001) Expression of a sulphur-rich sunflower albumin gene in transgenic tall fescue (*Festuca arundinacea* Schreb.) plants. *Plant Cell Rep*., 20, 213-219.
Weckworth, W. (2003) Metabolomics in systems biology. *Annu. Rev. Plant Physiol*., 54, 660-689.
Wilkins, M.R., C. Pasquali, R.D. Appel, K. Ou, O. Golaz, J.C. Sanchez, J.X. Yan, A.A. Gooley, G. Hughes, I. Humphrey-Smith, K.L. Williams & D.F. Hochstrasser (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nature Biotechnology*, 14, 61-65.
Wilkins, M.R., K.L. Williams, R.D. Appel and D.F. Hochstrasser (eds.) (1997) Proteome Research: New Frontiers in Functional Genomics, Springer Verlag
Xu, J., J. Schubert & F. Altpeter (2001) Dissection of RNA mediated virus resistance in fertile transgenic perennial ryegrass (*Lolium perenne* L.). *Plant Journal*, 26, 265-274.
Yu, J. S. Hu, J. Wang, G. KS Wong, S. Li, B. Liu, Y. Deng, L. Dai, Y. Zhou, X. Zhang, M. Cao, J. Liu, J. Sun, J. Tang, Y. Chen, X. Huang, W. Lin, C. Ye, W. Tong, L. Cong, J. Liu, Q. Qi, J. Liu, L. Li, T. Li, X. Wang, H. Lu, T. Wu, M. Zhu, P. Ni, H. Han, W. Dong, X. Ren, X. Feng, P. Cui, X. Li, H. Wang, X. Xu, W. Zhai, Z. Xu, J. Zhang, S. He, J. Zhang, J. Xu, K. Zhang, X. Zheng, J. Dong, W. Zeng, L. Tao, J. Ye, J. Tan, X. Ren, X. Chen, J. He, D. Liu, W. Tian, C. Tian, H. Xia, Q. Bao, G. Li, H. Gao, T. Cao, J. Wang, W. Zhao, P. Li, W. Chen, X. Wang, Y. Zhang, J. Hu, J. Wang, S. Liu, J. Yang, G. Zhang, Y. Xiong, Z. Li, L. Mao, C. Zhou, Z. Zhu, R. Chen, B. Hao, W. Zheng, S. Chen, W. Guo, G. Li, S. Liu, M. Tao, J. Wang, L. Zhu, L. Yuan & H. Yang (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). *Science*, 296, 79-92.
Zhang, D.-X. & G.M. Hewitt (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, 12, 563-584.
Zhu, Y.L., Q.J. Song, D.L. Hyten, C.P. Van Tassell, L.K. Matukumalli, D.R. Grimm, S.M. Hyatt, E.W. Fickus, N.D. Young & P.B. Cregan, (2003) Single-nucleotide polymorphisms in soybean. *Genetics*, 163, 1123-1134.