BEYOND STRUCTURAL GENOMICS FOR PLANT SCIENCE

Richard A. Dixon,[1] Joseph H. Bouton,[2] Brindha Narasimhamoorthy,[2] Malay Saha,[2] Zeng-Yu Wang[2] and Gregory D. May[3]

[1]Plant Biology Division, Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, Oklahoma 73401
[2]Forage Improvement Division, Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, Oklahoma 73401
[3]National Center for Genome Resources, 2935 Rodeo Park Drive East, Santa Fe, New Mexico 87505

Advances in Agronomy, Volume 95

0065-2113/07 $35.00
DOI: 10.1016/S0065-2113(07)95002-6

The past decade has seen unparalleled advances in our understanding of plant genomes, and genomic (primarily DNA sequence) information now underpins many aspects of plant trait improvement, through gene discovery to transgenesis and use of molecular markers in breeding. This chapter provides an overview of the genomic and postgenomic technologies that are likely to have the greatest impacts on agronomy over the next 10–20 years and describes a number of case studies of their application. Although the impacts of these technologies are already apparent, the amazing and still accelerating pace of technology development promises much, maybe more than can easily be assimilated into traditional plant improvement programs at present. A new breed of plant scientist with skills in understanding and integrating multiple disciplines, and making use of increasingly sophisticated computational approaches, is needed to take full advantage of even the present knowledge.

I. INTRODUCTION

The past several years have seen major advances in our ability to gather whole-genome-scale information from plants. Central to these developments, several projects have assembled working models of the complete or near complete genome sequences of the model crucifer *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000; Bevan *et al.*, 2001), rice (Goff *et al.*, 2002; Yu *et al.*, 2002), poplar (Tuskan *et al.*, 2006), and two model legumes (VandenBosch and Stacey, 2003). Several other projects have targeted a range of species for the sequencing of expressed sequence tags (ESTs) representing genes expressed in particular tissues or under particular developmental or environmental conditions.

For species with either sequenced genomes or extensive EST resources, commercial DNA microarrays are now available for global transcript profiling (Rensink and Buell, 2005). Technologies such as serial analysis of gene expression (SAGE; Matsumura *et al.*, 1999; Velculescu, 1999), massively parallel signature sequencing (MPSS; Brenner *et al.*, 2000), and cDNA-AFLP (Goossens *et al.*, 2003) provide tools for analysis of genome-wide changes in

transcriptional activity for plant species that lack sequenced genomes or even extensive EST resources. However, it has been increasingly realized that profiling changes in gene transcripts only provides part of the picture of the impacts of differential gene expression in plants (Hall *et al.*, 2002), and considerable efforts have therefore been put into developing robust and comprehensive methods for profiling the metabolome (the metabolite complement of a tissue, organ, or whole organism; Hall *et al.*, 2002) and the proteome (the complete set of proteins; Watson *et al.*, 2003). These latter technologies rely heavily on mass spectrometry (Roessner *et al.*, 2001; Yates, 1998).

The integration of data spanning the genome, proteome, and metabolome is a major goal of "postgenomics" biology. A whole new field of "Systems Biology" has been defined, encompassing the collection and interpretation of holistic data for biological systems. The goal is to understand the organizational principles that operate at the cellular and organismal levels and that relate individual components to the whole system. This new way of thinking about biological systems poses some major challenges, none more so than in the area of gene annotation. As discussed by Huang (2000), most gene products function as part of one or more complex regulatory systems, and exactly how they do this is often not apparent from the types of *in vitro* (enzymatic, interaction mapping) or simple genetic analysis currently utilized.

Although postgenomics biology faces major challenges in taking the next step to a full understanding of gene function at the organismal level, the techniques associated with this branch of biology are already finding application in agronomy. For example, holistic analysis of genome content and gene expression provides novel approaches to the design of markers for breeding and speeds the identification and isolation of genes associated with important traits. Metabolome analysis also holds promise as a tool for trait mapping (Dixon *et al.*, 2006).

In contrast to the exponential increase in genomic information, improvements in the efficiency of plant genetic transformation have occurred linearly over the past 10 years. Nevertheless, public perception and regulatory issues aside, transgenesis provides the most rapid means of introducing truly novel traits to crop plants, and is also a major technology for functional genomics in plants.

This chapter provides only a brief overview of the technical bases of the important new genomic and postgenomic technologies. Our major aim is to present the reader with a feeling for how the genomics revolution is set to impact plant science in general, and agronomy in particular, over the next 10–20 years. To provide case studies of the uses of several of the outlined technologies, we describe four projects in Sections VI and VII, three ongoing at the Noble Foundation, in which genomic approaches have been utilized for introducing important agronomic traits into legumes.

Much of the work applying postgenomic technologies to agriculture requires collaboration and understanding between scientists with quite

different academic backgrounds. Multidisciplinary approaches of this type will play an increasing role in basic and applied agronomy in the future.

II. SEQUENCED GENOMES, MODEL SYSTEMS, AND COMPARATIVE GENOMICS

A. Introduction

By definition, postgenomics technologies take as their starting point the availability of genome-wide information for any particular target species. The partially annotated sequence of the model crucifer *A. thaliana* first appeared in 2000 (Arabidopsis Genome Initiative, 2000) as the result of a major international effort that took nearly 10 years to bring to fruition. *Arabidopsis* was chosen primarily for its small genome size, self-fertility, short time to flowering and seed set, and the availability of many naturally occurring ecotypes (geographical variants of the species; Koornneff *et al.*, 2004). While the sequencing program was ongoing, dramatic improvements in the genetic transformation efficiency of *Arabidopsis* (Clough and Bent, 1998) accelerated the rate of development of genetic resources through T-DNA tagging based on random insertion of the *Agrobacterium tumefaciens* transfer DNA following transformation (Alonso *et al.*, 2003; Azpiroz-Leehan and Feldmann, 1997).

The combination of a sequenced genome and availability of a range of genetic resources such as defined ecotypes or mutant populations derived from chemical or DNA-based (deletion or insertion) mutagenesis greatly facilitates the use of the model system for gene discovery and annotation. However, this does not mean that postgenomics technologies can only be applied to species with sequenced genomes. A number of approaches can be used for global transcript profiling, mining mutations, and developing molecular markers in species that do not have extensive genome, or even EST, sequence information. Furthermore, comparative genomic approaches that link genomic data from less well-defined systems to the well-defined model species are becoming increasingly useful for gene discovery.

B. *A. thaliana*

Several articles have reviewed the strategies used for sequencing the *Arabidopsis* genome (Arabidopsis Genome Initiative, 2000; Bevan *et al.*, 1999, 2001). Essentially, the approach relied on a robust physical map of yeast artificial chromosome clones of genomic DNA fragments to "place" the emerging sequence in its genomic context, since this was primarily

obtained by sequencing bacterial artificial chromosomes harboring *Arabidopsis* genomic DNA that had been anchored to the physical map. Whole-genome "shotgun" sequencing was not attempted for *Arabidopsis*, since at the time this approach had several drawbacks, particularly as regards final sequence assembly, especially of the highly repetitive heterochromatic regions around the centromeres. Advances that greatly increase the power and throughput of both sequencing technology and computational analysis of genome sequence information have since made the shotgun approach more feasible, and this approach was therefore taken as the principal technique for obtaining the genome sequences of rice and poplar (Goff *et al.*, 2002; Tuskan *et al.*, 2006; Yu *et al.*, 2002). In the *Arabidopsis* project, as in all subsequent large-scale genome projects, the availability of a large set of EST sequences was invaluable for genome assembly and annotation. Table I summarizes the current status of plant genome sequencing projects.

Table I
A Summary of Plant Species Genome Projects

Species	Common name	Chromosome number (n)	Genome size (~Mbp)	Number of TIGR ESTs	Project status
Arabidopsis thaliana	Thale cress	5	120	616,064	Complete[a]
Coffee arabica	Coffee	11	640	1064	Initiated
Glycine max cv Williams 82	Soybean	20	1115	351,935	Initiated
Lotus japonicus	Lotus	6	470	148,617	Near Complete
Manihot esculenta	Cassava	18	765	17,910	Initiated
Medicago truncatula cv Jemalong A17	Barrel medic	8	500	217,148	Near Complete
Oryza sativa cv Nipponbare	Rice	12	430	1,169,591	Complete[b]
Populus trichocarpa cv Nisqually-1	Black cottonwood	19	550		Complete[c]
Lycoperiscon esculentum	Tomato	12	950	200,248	Initiated
Solanum tuberosum	Potato	12	840	219,485	Initiated
Sorghum bicolor	Sorghum	10	760	203,575	Initiated
Zea mays cv B73	Maize	10	2300	1,014,701	Initiated

[a]Arabidopsis Genome Initiative, 2000.
[b]Yu *et al.*, 2002.
[c]Tuskan *et al.*, 2006.

Of the more than 25,000 genes predicted to be present in *Arabidopsis*, only 9% had been characterized experimentally by 2001 (Bevan *et al.*, 2001), and ~30% could not be assigned to any putative functional category based on sequence identity alone. These 25,000 genes (more recently updated to closer to 29,000) represented over 11,000 different protein types, and 35% of the predicted proteins occurred only once in the genome. The US National Science Foundation initiated a program, the "2010 Program," with the aim of understanding the functions of all the genes in the *Arabidopsis* genome by the year 2010 (Chory *et al.*, 2000). Awardees study specific gene families through combined approaches such as expression and analysis of recombinant proteins and characterization of the phenotypes of knockout mutations in the target genes. There have been regular reports of overall progress, which has been significant (Ausubel, 2002; Chory *et al.*, 2000), but it appears unlikely that the functions of all the genes will be understood within the next 4 years.

The initial *Arabidopsis* genome sequence yielded several surprises. For example, at least 47 expressed genes that encoded a wide variety of different protein types were found within the highly repetitive centromeric regions, gene families containing two or more members arranged in tandem arrays were common, and ~60% of the genome was present in 24 duplicated segments, each of more than 100 kb, suggesting that *Arabidopsis* may have had a tetraploid ancestor. Generation of full-length cDNA resources and application of DNA tiling array technology (Section III.B.1) has revealed a significant number of previously unsuspected genes in the *Arabidopsis* genome, many of which are transcribed but do not appear to code for proteins (Yamada *et al.*, 2003).

C. Rice

After the success in sequencing *Arabidopsis*, it was fitting that the next plant genome to be sequenced should be a monocot, and furthermore the world's major staple crop. The draft sequence of the rice genome was reported in 2002, for both the *indica* and *japonica* subspecies (Goff *et al.*, 2002; Yu *et al.*, 2002). The euchromatic portion of the rice genome was estimated to be 430 Mb, some 3.7 times larger than that of *Arabidopsis*. Similar to *Arabidopsis*, an apparent whole-genome duplication has occurred in rice, in this case 40–50 million years ago. The high degree of synteny among grass genomes (Freeling, 2001; Gale and Devos, 1998; Goff *et al.*, 2002), coupled with the ease of rice transformation (Tyagi *et al.*, 1999), excellent physical and genetic maps, and availability of mutant resources (Hirochika *et al.*, 2004), make rice an excellent model for other monocot crops. Indica rice is predicted to contain ~54,000 genes, of which only about

20% could be given a functional classification based on sequence alone (Yu *et al.*, 2002).

Approximately 85% of the predicted *Arabidopsis* proteins have significant homologues in rice, with overall mean identity of about 50%. Nevertheless, a significant number of *Arabidopsis* genes, most without functional annotation, are not present in rice and may represent dicot-specific genes. In contrast, most cereal genes discovered to date have very close homologues in rice; homologues of 98% of the maize, wheat and barley protein-coding genes known in 2002 were found in the rice genome (Goff *et al.*, 2002). This observation, coupled with the close synteny among cereal genomes, makes rice a valuable "scaffold" or nodal species for assembly of other economically important cereal genomes such as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), corn (*Zea mays*), and sorghum (*Sorghum bicolor*) (Devos and Gale, 2000; Goff *et al.*, 2002).

D. Poplar

The publication of a draft sequence of the poplar (*Populus trichocarpa*) genome in 2006 provided the first insights into the genomic organization of a tree species (Tuskan *et al.*, 2006). As with rice, the approach taken, by an international consortium, involved whole-genome shotgun sequencing and assembly, integrated with detailed genetic mapping. The *Populus* genome size is estimated to be ~485 Mbp, of which about 70% appears to be euchromatic. The ~75 Mbp of heterochromatic DNA remained unassembled. A significant proportion of the *Populus* genome appears to have arisen from a major genome duplication event. Poplar contains more than 45,000 putative coding genes, with similar frequencies of protein domains to those found in *Arabidopsis*, but a higher number of *Populus* homologues for each *Arabidopsis* gene. This is particularly apparent for genes involved in cell wall (lignocellulose) biosynthesis and defense (Tuskan *et al.*, 2006).

E. *Medicago truncatula* and *Lotus japonicus*

As a family, legumes are unique in their ability to fix atmospheric nitrogen through their association, in root nodules, with nitrogen-fixing bacteria (Rhizobia; Downie, 1997; Shanmugam *et al.*, 1978; Stacey *et al.*, 2006). From an agronomic perspective, legumes crops can be divided into the grain legumes, such as soybean, bean, and pea, and the forage legumes, such as alfalfa and clover. There was considerable debate in the late 1990s as to the best model species for legume genomics. Although considerable genetic resources were available in the above-mentioned grain legumes, the

sizes of their genomes, coupled with recalcitrance to genetic transformation, argued against their adoption. Eventually, one forage legume, *Medicago truncatula* (Cook, 1999; May and Dixon, 2004; Oldroyd and Geurts, 2001), and one leguminous weed, *Lotus japonicus* (Kawasaki and Murakami, 2000; Udvardi *et al.*, 2005), were selected as model species. As with *Arabidopsis*, a small diploid genome (Table I), self-fertility, rapid generation time, and availability of genetic transformation (although at nothing like the frequency achievable using the floral dip method with *Arabidopsis*) were the factors driving these choices. *M. truncatula* is very closely related to alfalfa (*Medicago sativa*), the world's major forage legume, which is, however, not itself useful as a model species, being an outcrossing autotetraploid. Initially, progress with legume genomics was primarily in the area of EST sequencing (Asamizu *et al.*, 2000), although whole-genome projects are now well underway in both *M. truncatula* (in the United States and Europe) and *L. japonicus* (in Japan) (VandenBosch and Stacey, 2003), and reports of the full draft sequences are expected within the next 12 months. Soybean (*Glycine max*) has also joined the list of legume species for which genome projects are underway (Jackson *et al.*, 2006). Extensive EST resources are also available for soybean, and for other legume species that are subjects of more modest genomics projects (VandenBosch and Stacey, 2003; Table II). In some cases, more limited EST projects have targeted specific metabolic processes, such as the biosynthesis of storage polysaccharides in guar (*Cyamopsis tetragonoloba*; Dhugga *et al.*, 2004; Naoumkina *et al.*, 2007). Surprisingly, many of the World's most important crop legume species lack substantial levels of EST resources (Table II).

Syntenic relationships exist among legume genomes. For example, linkage group V of *M. truncatula* exhibits macrosynteny with linkage groups V and I

Table II
Crop Legume EST Totals in GenBank, as of January 2006

Common name	Species	EST totals
Soybean	*Glycine max*	356,808
Common bean	*Phaseolus vulgaris*	21,377
Alfalfa	*Medicago sativa*	6613
Pea	*Pisum sativum*	3035
Peanut	*Arachis hypogaea*	2171
White lupin	*Lupinus albus*	2128
Chickpea	*Cicer arietinum*	724
Pigeon pea	*Cajanus cajan*	55
White clover	*Trifolium repens*	31
Lentil	*Lens culinaris*	8
Broad bean	*Vicia faba*	1

of alfalfa and pea, respectively (Cook, 1999). However, this does not extend to comparisons between *Medicago* and *Arabidopsis*, which are nevertheless quite closely related within the dicot subclass Rosidae. In spite of the lack of macrosynteny between *Medicago* and *Arabidopsis*, marker colinearity is often observed over small genetic distances (Zhu *et al.*, 2003). *Medicago*, *Lotus*, and soybean share a genome duplication event that occurred ~54 million years ago (Mudge *et al.*, 2005; Pfeil *et al.*, 2005). This genome duplication occurred after the divergence of the *Populus* lineage from legumes, but before the divergence of *Medicago*/*Lotus* and soybean (Cannon *et al.*, 2006).

F. Genetic Resources for Functional Genomics

The value of *Arabidopsis* as a model system comes from the ability to combine genomic sequence with genetic resources, and in this respect *Arabidopsis* is probably the best model plant system. Most importantly, the gene space has been nearly saturated with over 225,000 random *Agrobacterium* transferred DNA (T-DNA) insertion events, and the precise locations of the insertions in more than 20,000 of the *Arabidopsis* genes determined (Alonso *et al.*, 2003). Thus, loss-of-function mutants can be readily found for most of the *Arabidopsis* genes. Furthermore, if a gene from another species (e.g., an important crop) has a close orthologue in *Arabidopsis*, its function can initially be deduced by study of the phenotype of the corresponding *Arabidopsis* knockout, or by complementation of the *Arabidopsis* mutant with the gene from the crop species. Among the many examples of this approach are confirmation of function of a rice ethylene-signaling component (Mao *et al.*, 2006), a soybean jasmonate signaling component (Wang *et al.*, 2005b), a maize cell division regulator (Lim *et al.*, 2005), and a maize ABA signaling gene (Suzuki *et al.*, 2001).

Gain-of-function mutants are also available in *Arabidopsis*, at a lower frequency than the knockouts, from activation tagging projects in which the T-DNA insert contains multiple 35S enhancer sequences at the right border. Integration of the T-DNA construct within a gene can lead to a knockout, but integration near to a gene can result in the overexpression of that gene, irrespective of the orientation of the enhancer sequences relative to the transcription start site of the proximal gene (Weigel *et al.*, 2000). A good example of this approach is the discovery of the *p*roducer of *a*nthocyanin *p*igment (PAP1) mutant in which a MYB transcription factor (TF) controlling anthocyanin pigment formation is ectopically expressed as a result of the integration of proximal enhancer sequences (Borevitz *et al.*, 2000). The value of this discovery for the development of a "bloat-safe" alfalfa is described in Section VI.C.

The *Medicago* research community has developed a number of genetic resources to assist in gene discovery and functional annotation of legume genes. These include various populations of mutants. The first were produced by classical chemical mutagenesis of a polymorphic ecotype (A17; Penmetsa and Cook, 2000). Use has been made of fast neutrons to generate chromosomal deletions, and large populations of fast neutron deletion lines are now being generated (Wang *et al.*, 2006). In addition, following the demonstration that the transgenically inserted tobacco retrotransposon *Tnt1* could be activated in *M. truncatula* following tissue culture, and therefore be used for insertional mutagenesis (d'Erfurth *et al.*, 2003), efforts have been put in place to develop transposon-tagged populations of *Medicago* (Tadege *et al.*, 2005). With both approaches, the goal is to facilitate forward and reverse genetic screens. Although fast neutron deletions are very easy to generate and access through forward genetic screens for altered phenotype, cloning the deleted genes is less straightforward than in transposon- or T-DNA-tagged lines, and until recently has required map-based cloning. Success has now been reported for cloning genes based on comparisons of transcript abundance between wild-type and mutant lines using microarrays (Mitra *et al.*, 2004), which opens up possibilities for efficient gene identification through readily generated genetic resources.

Targeting induced local lesions in genomes (TILLING) is a new genetic tool for identifying genetic variation at the single base pair level (Henikoff *et al.*, 2004). It is a nontransgenic reverse genetics approach for identifying novel genetic variations. Reverse genetic screens using mutant populations have utilized TILLING in *L. japonicus* (Perry *et al.*, 2003), and a similar approach is being taken in *M. truncatula* (VandenBosch and Stacey, 2003). Without any prior knowledge of gene products, TILLING can investigate functions of a gene of interest in potentially any crop, and thus it is a useful tool for functional genomics. TILLING uses DNA pools from chemically mutagenized plants, and relies on the ability of an endonuclease (CEL1 from celery) to detect mismatches in heteroduplexes formed between wild-type and mutant PCR products of a specific sequence (McCallum *et al.*, 2000). Note that TILLING requires the generation of a mutant population, but does not require additional resources such as a sequenced genome or DNA arrays. PCR-based approaches may also be adaptable for the rapid reverse genetic screening of pooled fast neutron deletion populations to provide a rapid route to identification of individual plants harboring deletions in specific genes (Wang *et al.*, 2006). Similarly, a database of transposon-flanking sequences should be developed to provide a reverse genetic resource based on transposon insertion lines (Tadege *et al.*, 2005). Hopefully, all these resources will be in place for *Medicago* by the time the genome sequence is completed, thereby facilitating functional annotation of a legume genome.

III. TRANSCRIPTOMICS, PROTEOMICS, AND METABOLOMICS

A. Introduction

Global transcript profiling has become one of the most popular tools for analysis of plant gene expression, and this revolution has been driven primarily through the development of DNA microarray technology. The transcriptomes of plant species without genomics resources can also be interrogated, on a hitherto unprecedented scale, through the use of differential display or serial sequencing procedures. Being able to determine how an external stimulus, or endogenous developmental factor, regulates gene expression at the scale of the whole-genome provides a powerful tool for gene discovery and for understanding transcriptional networks. Technically speaking, transcript-profiling approaches are easier and higher throughput than massively parallel analysis of proteins (proteomics) or metabolites (metabolomics), and this explains their popularity and preponderance as analytical tools. This does not, however, mean that analyzing the transcriptome is more informative than proteomics or metabolomics. Indeed, changes in transcript levels are often quite transient, whereas the longer half-lives of proteins and metabolites give a more balanced and integrated "readout" of the biochemical phenotype of an organism.

B. Approaches for Transcript Profiling

1. For Sequenced Genomes

ESTs are rapidly generated, single pass sequences of cDNAs. Many sources of EST sequence information for plants are available online. Table III summarizes available resources for legumes, and also includes information on transcriptome and proteome information. Table II provides an indication of the numbers of ESTs sequenced for various legume species as of January 2006. The various "Gene Indices" available through the Institute for Genome Research (TIGR; http://www.tigr.org/tdb/tdb.html) [now the Dana-Farber Cancer Institute (DFCI)] are among the most extensive and user-friendly sources of EST information (Quackenbush *et al.*, 2000). Simply mining these data online can provide a rapid, first-pass analysis of the expression profile of a particular gene of interest. This is because ESTs are derived from transcripts sequenced from cDNA libraries that represent a particular tissue type, or tissue subjected to a specific biotic or abiotic stress. As an example, the current *M. truncatula* Gene Index (MtGI) contains over

Table III
Legume Genomic Resources

Database (URL)	Data types	Represented organism(s)
The Legume Information System (comparative-legumes.org/)	EST, genome, QTL, and comparative maps	*Glycine*, *Medicago*, *Lotus*, and *Phaseolus*
The Institute for Genomics Research (www.tigr.org)	EST, genome, repeat sequence, and pathways	*Glycine*, *Medicago*, and *Lotus*
NCBI (www.ncbi.nlm.nih.gov)	EST, genome, and expression	*Glycine*, *Medicago*, *M. sativa*, *Lotus*, and *Phaseolus*
MtDB–CCGB (www.medicago.org/MtDB/)	EST and genome	*Medicago*
Medicago EST Navigation System (MENS) (http://medicago.toulouse.inra.fr/Mt/EST/)	EST and pathways	*Medicago*
OpenSputnik Comparative genomics platform (http://sputnik.btk.fi/ests)	EST, BLAST, and SNP	*Glycine*, *Medicago*, *Lotus*, and *Phaseolus*
PlantGDB (www.plantgdb.org)	EST and BLAST	*Glycine*, *Medicago*, *M. sativa*, *Pisum*, *Arachis*, and *Phaseolus*
SoyBase (http://soybase.ncgr.org)	EST, genome, QTL and genetic maps, pathways, germplasm	*Glycine*
Sequencing *M. truncatula*, University of Oklahoma (http://www.genome.ou.edu/medicago.html)	Genome and BLAST	*M. truncatula*
Medicago Genome Database (http://mips.gsf.de/projects/medicago)	MIPS genome	*M. truncatula*
Kazusa *Lotus japonicus* (www.kazusa.or.jp/lotus/)	EST, genome, and genetic map	*Lotus*
M. truncatula Consortium (www.medicago.org/genome/)	Linkage maps, BAC overlap and clone/marker data	*M. truncatula*
Soybean Functional Genomics (Vodkin) (http://soybeangenomics.cropsci.uiuc.edu/)	Transcriptomics	*Glycine*
Soybean Genomics and Microarray Database (http://psi081.ba.ars.usda.gov/SGMD/Default.htm)	Transcriptomics	*Glycine*
Noble Foundation (Sumner) (www.noble.org/2DPage/Search.asp)	Proteomics	*Medicago*
M. truncatula Functional Genomics and Bioinformatics (http://medicago.vbi.vt.edu/)	Transcriptomics, proteomics, metabolomics, pathways, and literature	*Medicago*
Mt Proteomics (http://www.mtproteomics.fr.st/)	Proteomics	*Medicago* and *Sinorhizobium meliloti*

Table III (*continued*)

Database (URL)	Data types	Represented organism(s)
Australian National University 2D-PAGE Database (http://semele.anu.edu.au/)	Proteomics	*Medicago*
AlfaGenes (http://ukcrop.net/perl/ace/search/AlfaGenes)	EST, genetic map, and pathways	*M. sativa*
BeanGenes (http://beangenes.cws.ndsu.nodak.edu/)	Genetic map, gene classification, pathology, and cultivar data	*Phaseolus* and *Vigna*
CoolGenes (http://ukcrop.net/perl/ace/search/CoolGenes)	Genetic map	*Cicer* and *Lens*

36,878 tentative consensus sequences (TCs) or unigenes representing transcripts from over 61 different cDNA libraries. Figure 1 and the inset in Fig. 6 provide examples of how such data can be mined to provide a first indication of expression pattern to assist in identification of gene function (Section VI.B.1 below). Specific examples relating to the functional annotation of genes involved in the synthesis of phenylpropanoid-derived natural products have been reviewed (Costa *et al.*, 2003; Dixon *et al.*, 2002). Although simple, analysis of EST frequency in sequenced libraries requires some caution. First, there are problems associated with potential sequencing errors affecting the assembly of ESTs into the "contigs" known as TCs in the case of MtGI (Fig. 1; Rudd, 2003). The word "tentative" is important, as the assemblies can change when additional EST sequence information becomes available. In view of this, the TIGR EST databases keep track of all previous TC numbers for each contig, and these will sometimes split or coalesce until final confirmation is obtained from whole-genome sequence data. Second, because the selection phase of EST sequencing simply involves random picking of colonies, statistical issues affect the reliability of EST frequency within a library, particularly when considering libraries with low numbers of sequenced ESTs. With these limitations in mind, the increasing number of EST resources for important crop plants (Kuenne *et al.*, 2005) nevertheless provides an excellent starting point for selection of target genes, preliminary expression profiling, and development of molecular markers (Section IV.C below).

DNA microarrays provide an adaptable and rapid approach to transcript profiling. However, because they rely on previously determined gene or EST sequences, they represent a targeted profiling technology unless the arrays contain a complete unigene set for a particular organism. Several articles summarize the most important factors associated with the production, hybridization, and analysis of microarrays, and their applications for plant science (Baldwin *et al.*, 1999; Brazma *et al.*, 2001; Kehoe *et al.*, 1999; Robinson *et al.*, 2004;

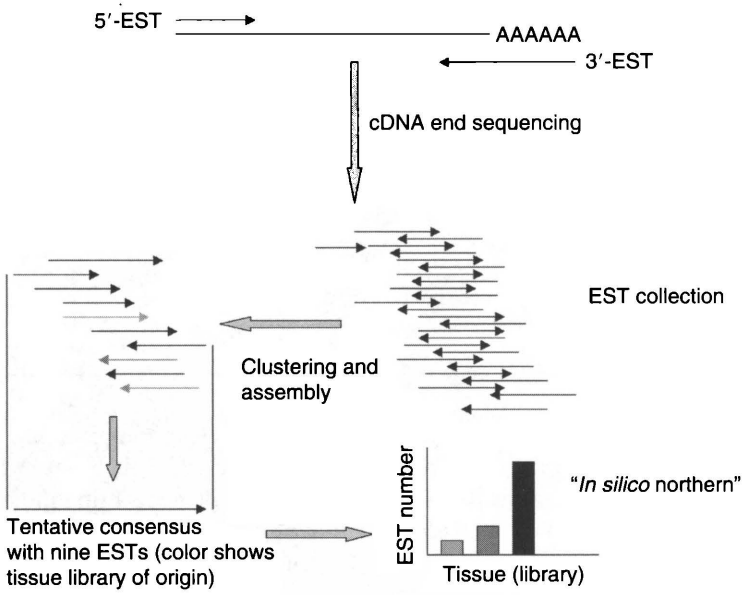


Figure 1 Generation and analysis of ESTs. A population of transcripts is converted to cDNA, which is then cloned and randomly sequenced. Each sequence run provides an EST. ESTs are clustered and aligned into TCs; each TC theoretically represents the transcript from one gene, complete only as far as the random sequence information allows. The number of occurrences of a particular EST in a particular cDNA library represents the relative transcript level of the corresponding TC in the biological material from which the library was constructed. (See Color Insert.)

van de Peppel *et al.*, 2003; Wu *et al.*, 2001). There are two major types of microarrays, representing different technology platforms for generation and analysis of the arrays. Spotted arrays consist of a large number of DNA species arrayed as a grid on a glass slide. The DNA may be from a cDNA clone, particularly in the case of "custom arrays" made by one laboratory for analyzing a specific set of transcripts, although better results are often obtained if all the spots contain DNA fragments of the same size, optimized for hybridization characteristics, as in the case of oligonucleotide arrays. However, this requires significant informatics input, which can be provided by commercial providers, such as Qiagen Operon, Alameda, CA, who will then make the required oligonucleotide set for in-house spotting, or Agilent Technologies, Palo Alto, CA, who will provide DNA arrays to you. Affymetrix DNA arrays utilize photolithographic masking methods and combinatorial chemistry to synthesize large numbers of unique probes on each array. Each annotated open reading frame is represented by around 11–13 pairs of oligonucleotides. Each pair is composed of a perfect match and one-base pair mismatched oligonucleotide (Barnett *et al.*, 2004).

As an example of the development of plant microarray resources, the early arrays generated for *M. truncatula* first consisted of spotted EST clones [2K, 6K, and 8K (the latter represented 6300 nonredundant genes); Firnhaber *et al.*, 2005], followed by a commercial 16K oligonucleotide array (Aziz *et al.*, 2005; Suzuki *et al.*, 2005), a custom Affymetrix array containing 10,000 *M. truncatula* probe sets and the complete genome of *Medicago*'s rhizobial symbiont *Sinorhizobium meliloti* (Barnett *et al.*, 2004), and finally a commercial Affymetrix array containing 32,167 *M. truncatula* EST/mRNA-based and chloroplast gene-based probe sets, 18,733 *M. truncatula* IMGAG and phase 2/3 BAC prediction-based probe sets from the whole-genome project, 1896 alfalfa EST/mRNA-based probe sets (primarily from trichome ESTs), and 8305 *S. meliloti* gene prediction-based probe sets. Further iterations of this array will occur as the *Medicago* genome attains completion.

Examples of gene expression profiling using legume gene chips cover many aspects of legume biology, from determining those transcripts that are specifically associated with the nodulation process (Barnett *et al.*, 2004; Colebatch *et al.*, 2004) and flower and pod development (Firnhaber *et al.*, 2005), to identifying genes of secondary metabolite biosynthesis activated in response to microbial elicitors or wound signals (Suzuki *et al.*, 2005) or expressed in glandular trichomes (Aziz *et al.*, 2005). The 16K *Medicago* oligonucleotide arrays have also been used to determine the "substantial equivalence" of transgenic plants expressing an engineered natural product pathway for isoflavone formation compared to plants not expressing the new pathway (Deavours and Dixon, 2005). An equally wide number of applications of microarray technology have been reported in *Arabidopsis*, ranging from studies on ethylene signaling (De Paepe *et al.*, 2004) and response to UV light (Casati and Walbot, 2003) to methyl jasmonate (MeJA) signaling (Sasaki-Sekimoto *et al.*, 2005). Work is currently in progress to generate a publicly available "gene expression atlas" for *Medicago* through microarray analysis (using the commercial Affymetrix arrays) of RNA samples from multiple tissues and different physiological treatments (M. Udvardi, personal communciation). These will supplement the large sets of *Arabidopsis* microarray data that are already publicly available (http://affymetrix.Arabidopsis.info/narrays/experimentbrowse.pl; http://www.genevestigator.ethz.ch/), and the emerging whole-genome expression profiles for important crops such as rice, soybean, barley, and tomato (Rensink and Buell, 2005). Some databases contain combinations of microarray and EST abundance data (Fei *et al.*, 2006).

The value of an EST collection or a microarray experiment for giving a readily accessible picture of the transcriptome of a particular tissue type is further enhanced if the degree of resolution can be increased from the organ to the cellular level. Microarray analysis was, in the past, limited by the relatively large amount of RNA required for hybridization (generally in the region of 50–200 μg of total RNA per hybridization, equating to 50–100 mg

of plant tissue). Such a requirement precluded the technology from taking advantage of the increasing refinement of tissue isolation procedures, such as laser capture microdissection (Kerek *et al.*, 2003), or methods for physical isolation of appendages such as trichomes (Lange *et al.*, 2000; Wagner, 1991). Methods have been developed for amplification of target RNA samples so that microarray analysis is now possible with as little as 100 ng of total RNA (Hertzberg *et al.*, 2001), and PCR-amplification methods have allowed for the generation of EST libraries from very small tissue samples such as isolated glandular trichomes (Aziz *et al.*, 2005). Trichomes are a particularly attractive target tissue for EST (in the absence of a sequenced genome) or microarray analysis, since they often show high specialization for the synthesis and secretion of species-specific bioactive secondary metabolites that confer insect and pest resistance (Georgieva, 1998; Lovinger *et al.*, 2000; Maluf *et al.*, 2001). Several examples illustrate the insights that can be gained into this specialized biochemistry by randomly sequencing, annotating, and functionally characterizing the biosynthetic enzyme gene transcripts that are often abundantly expressed in trichomes (Aziz *et al.*, 2005; Fridman *et al.*, 2005; Gang *et al.*, 2001; Wagner *et al.*, 2004).

Availability of sufficient RNA for hybridization is not the only factor that can limit the scope and reproducibility of a microarray experiment. Some genes, particularly regulatory genes, such as TFs, are expressed at very low levels, and the absolute signal strength of their hybridization approaches the noise level on the array (Czechowski *et al.*, 2004). In such cases, alternative profiling methods may be necessary. Sets of oligonucleotide probes have been developed for profiling the complete TF complement of *Arabidopsis* by quantitative real-time polymerase chain reaction (qRT-PCR; Czechowski *et al.*, 2004), and the technique has been used to study both developmental and wound/pathogen defensive TF gene expression (Czechowski *et al.*, 2004; McGrath *et al.*, 2005). An application of this method to profile glycosyltransferase gene expression in *Medicago* has shown it to be highly sensitive, reproducible, and to correlate well with parallel analysis by Affymetrix microarray analysis (Modolo *et al.*, 2007). It may therefore prove a popular technology for profiling transcript levels if highly accurate quantification is necessary.

Classical DNA microarrays are assembled from a set of unigenes that generally represent protein-coding transcripts. However, several types of noncoding RNAs have been discovered. The small noncoding RNAs play important roles in gene regulation (Bartel and Bartel, 2003), are generally in the region of 21–30 nucleotides in length, and fall into at least three distinct classes; microRNAs (miRNAs), small interfering RNAs (siRNAs), and repeat-associated small interfering RNAs (rasiRNAs) (Zamore and Haley, 2005). It is becoming increasingly clear that a significant proportion of the RNA transcripts in human do not encode proteins (Claverie, 2005), and the

search is now on for their function using targeted genetic screens (Mattick, 2005).

Global analysis of plant genomes, in addition to studies specifically targeting miRNAs and siRNAs (Bartel and Bartel, 2003), also supports the importance of noncoding RNAs in plants. Initially, EST sequences that were either short or did not appear to encode an open reading frame were ignored as artifacts, and for this reason such sequences did not generally appear on microarrays. With the advent of whole-genome tiling arrays, it has become possible to verify, and determine the expression pattern, of noncoding RNAs on a global scale. An excellent example of this approach in plants used a combination of full-length cDNA discovery and hybridization of RNA populations to whole-genome arrays to define the transcriptional units of all *Arabidopsis* genes (Yamada *et al.*, 2003). The tiling array consisted of 12 individual slides, each containing around 834,000 ordered 25-mer oligonucleotides that together represented about 94% of the *Arabidopsis* genome. This seminal work resulted in a full transcriptional annotation of the *Arabidopsis* genome in terms of genes that were (1) annotated and expressed, (2) annotated but not expressed, and (3) not annotated but expressed. The latter class were discovered in what had been thought to be intergenic regions. Surprising findings were the relatively large number of antisense transcripts, the high transcriptional activity of the centromeric regions, and the transcription of many genes previously classed as "pseudogenes," suggesting that these might serve a regulatory function.

2. For Species Lacking Genomics Resources

It is possible to carry out global scale transcript profiling in crop plants for which neither genomic nor extensive EST sequence information is available. Where a crop plant is closely related to a model species, it may be possible to utilize microarray resources from the model species. Examples include the use of *M. truncatula* microarrays to profile transcripts in alfalfa (Aziz *et al.*, 2005; Deavours and Dixon, 2005), and tomato arrays for profiling transcripts from other *Solanaceous* species such as pepper and eggplant (Moore *et al.*, 2005). Among genes represented in the alfalfa glandular trichome ESTs as TCs that have orthologues in *M. truncatula*, 66.5% had 100% sequence identity to the corresponding *M. truncatula* orthologue (Aziz *et al.*, 2005) and, of the total 5647 alfalfa trichome ESTs sequenced, 4804 had *M. truncatula* orthologues with E values of -20 and below. Because of this very high degree of sequence similarity between alfalfa genes and their *M. truncatula* orthologues, significant signal was observed for most of the 16,086 genes represented on the oligonucleotide arrays when hybridized with alfalfa total RNA (Aziz *et al.*, 2005). The overall number of features with signal more

than 300 pixels above background was the same as observed when the same arrays were hybridized with labeled RNA from *M. truncatula* stems.

cDNA amplified fragment length polymorphism (cDNA-AFLP) and SAGE are nontargeted transcript-profiling techniques that can theoretically be applied to any living organism. The basic principles behind the two methods are outlined in Fig. 2. cDNA-AFLP is an RNA fingerprinting approach which involves cDNA synthesis from the RNA transcripts to be analyzed, restriction digestion of the primary templates and ligation of anchors to their termini, preamplification with anchor-specific primers, and selective amplification of the restriction fragments with primers extended with one or more specific bases (Bachem *et al.*, 1998). The method allows for the simultaneous analysis of multiple samples. Examples of its use in

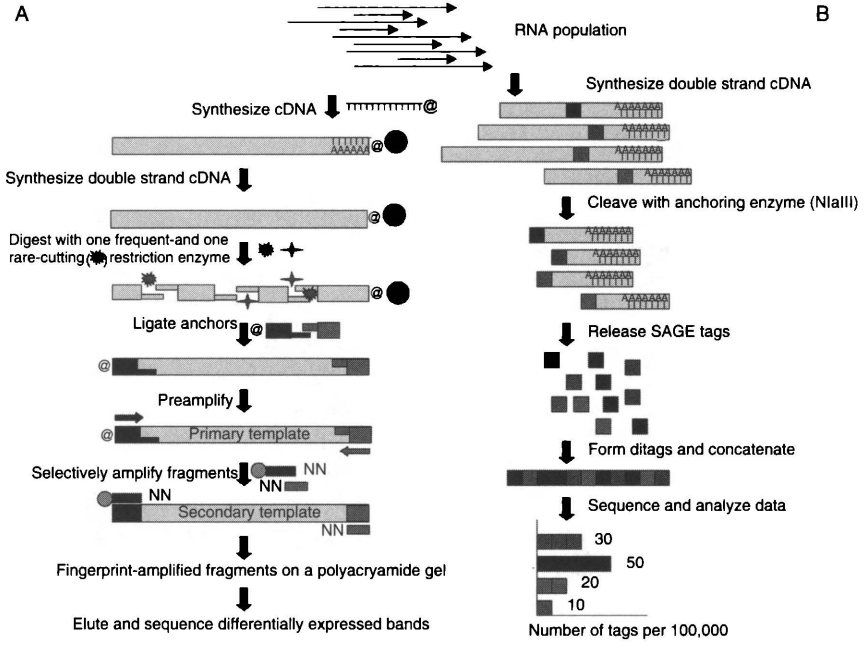


Figure 2 cDNA-AFLP and SAGE, techniques for nonbiased transcript profiling. (A) Procedure for cDNA-AFLP analysis. @ = biotin group, black circle = streptavidin bead. The rare-cutting site anchors and primers are shown in black, and the red circle symbol represents a ^{32}P label. See Bachem *et al.* (1998) for further details. (B) Procedure for SAGE analysis. After cutting the cDNAs with a frequent cutting enzyme (usually *Nla*III), linkers are ligated to the 5′-ends; these linkers contain a site for a type IIs restriction enzyme (*Bsm*FI) which cuts a 15-bp fragment (SAGE tag) of the cDNA (joined to the linker). These fragments are ligated tail to tail to form ditags, which are then amplified, concatenated, and sequenced. Special software determines the frequency of the SAGE tags among the sequenced DNA. (See Color Insert.)

plant biology include the profiling of transcripts responding to ethylene in *Arabidopsis* (De Paepe *et al.*, 2004) and to MeJA in tobacco BY2 cell suspension cultures (Goossens *et al.*, 2003). In the latter example, the authors pointed out the importance of a nontargeted approach for discovering natural product biosynthetic genes from the many unrelated medicinal plants, none of which currently has genome resources.

In SAGE analysis, a method that takes advantage of the target sequence recognition properties of class-II DNA restriction enzymes that cut a short distance away from the enzyme's DNA recognition site, transcripts are reduced to short tags which are then concatenated and sequenced (Velculescu, 1999) (Fig. 2). SAGE is therefore a nontargeted or "open" system. In contrast to DNA arrays which are considered a "closed system," nontargeted transcript-profiling methods such as SAGE allow for the identification and analyses of previously undescribed RNAs (e.g., antisense RNAs). Comparisons of SAGE and microarray analysis using the same RNA samples show quite good correlations (Ishii *et al.*, 2000), and SAGE has become a popular transcript-profiling technique for plants, from loblolly pine (Lorenz and Dean, 2002) to rice (Matsumura *et al.*, 1999). Adaptations have been made to the technique to make it applicable to the analysis of transcripts from microdissected cells and other small samples (Velculescu *et al.*, 2000).

MPSS (Brenner *et al.*, 2000) identifies short sequence signatures (20 bp) generated from a position immediately adjacent to the *DpnII* restriction enzyme site nearest to the poly-A tail of an mRNA transcript. The relative abundance of these signatures in a given mRNA sample (library) represents a quantitative estimate of expression of that gene. MPSS and now clonal single molecule arrays (CSMA™) technologies were developed by Solexa, Inc. (www.solexa.com). Solexa has discontinued providing MPSS as a service and now exclusively offers CSMA™. Although the data output from MPSS and CSMA™ are essentially the same, CSMA™ is based on a high-density, eight distinct channel flow-cell array format whereas MPSS is a bead-based technology. Sequence data generated by the CSMA™ platform is based on sequencing-by-synthesis (SBS) and reversible terminator chemistry, and leverages massively parallel sequencing of cDNA fragments to generate data from millions of fragments simultaneously. Solexa's SBS approach is anticipated to generate up to one billion bases of data per run at costs more economical than MPSS.

C. Proteomics

Chemically, DNA is a relatively simple polymer with only four building blocks, in contrast to the 20 different protein amino acids and the many thousands of primary and secondary metabolites found in plants. Profiling

the proteome and metabolome therefore poses significant technical challenges compared to transcript profiling. In fact, few if any studies have been able to profile the complete proteome or metabolome of a complex organism; this is as much a problem of initial separation of complex mixtures as it is one of final detection. Because both proteomics and metabolomics require specific chemical determination of molecules with quite different structures, mass spectrometry has become the detection/analytical method of preference, and is capable of both extreme sensitivity and mass discrimination.

Classically, proteomics has relied on two-dimensional isoelectric focusing SDS-polyacrylamide gel electrophoresis for initial protein separation, and this approach can routinely resolve around 1000 different plant proteins (Lei *et al.*, 2005; Watson *et al.*, 2003; Yan *et al.*, 2006). This is nevertheless, only a small fraction of the predicted proteins in a tissue based on the numbers of genes known to be expressed. This lack of penetration represents less of a problem if the technique is applied to specific subcellular fractions (Majeran *et al.*, 2005; Nelson *et al.*, 2006; Taylor *et al.*, 2005; Ytterberg *et al.*, 2006), tissues with a preponderance of a specific protein type under study (e.g., seed storage proteins; Thiellement *et al.*, 1999), or specialized structures such as trichomes (Wienkoop *et al.*, 2004). Approaches to profiling proteins have used shotgun methods without gel fractionation, relying instead on more rapid separation methods, protein tagging, and the versatility of modern mass spectrometers (Aebersold and Mann, 2003; Chen *et al.*, 2006a; Hass *et al.*, 2006; Shen *et al.*, 2005; Wolff *et al.*, 2006).

Proteomic approaches are being applied to address many of the same types of questions currently investigated by transcript profiling; these include genetic diversity, phylogenetic relationships, characterization of mutants, studying responses to abiotic stresses such as UV light and cold, and understanding seed development (Agrawal and Thelen, 2006; Casati *et al.*, 2005; Thiellement *et al.*, 1999; Yan *et al.*, 2006). However, because of the high cost of the required mass spectrometers and current limitations to the depth of profiling, proteomics will likely remain, at least for the time being, less utilized than transcriptomics as regards applications to agronomy and plant breeding.

D. Metabolomic Analysis

1. Introduction

Levels of plant metabolites are controlled by both genetic and environmental factors, and the metabolome is often referred to as the functional manifestation of gene expression. Metabolite profiling can be classified into three approaches, targeted profiling, fingerprinting, or true metabolomics

(in depth and unbiased). Several articles provide reviews of the technology and its challenges for these various types of approach (Dixon *et al.*, 2006; Fiehn, 2002; Fuell *et al.*, 2004; Hall *et al.*, 2002; Sumner *et al.*, 2002, 2003). Early studies in plant metabolomics utilized gas chromatography-mass spectrometry (GC-MS) to profile mainly hydrophilic primary metabolites (Roessner *et al.*, 2000, 2001). Importantly, this work demonstrated that metabolite profiling was of value for providing genetic, as well as chemical, understanding of plant systems. Thus, analysis of GC-MS profiles of extracts from different potato genotypes, when compared to profiles from transgenic potato lines modified in sucrose catabolism, revealed how metabolite profile analysis clearly shows the way in which environmental factors can lead to metabolic phenotypes linked to specific genetic changes (Roessner *et al.*, 2001). This work defined both the strengths and potential difficulties of the approach. One major strength is that data mining tools such as hierarchical cluster analysis and principle component analysis (PCA) allow for clear visualization of factors that relate or distinguish different metabolite profiles, thus making the profile a very rich source of information for comparative genetic analysis (Fiehn, 2002; Roessner *et al.*, 2001). These, and other informatic approaches, have been reviewed (Sumner *et al.*, 2003). As with proteomics, one weakness is that most metabolic profiling technologies only sample a proportion of the total metabolome, Thus, the early GC-MS analysis of potato tissues only resolved about 80 different compounds, whereas it is estimated that a simple plant such as *Arabidopsis* contains in excess of 5000 different metabolites. Improvements in technology, for example, by the use of raid scanning time of flight mass spectrometry, have increased the number of metabolites detectable in crude plant extracts to around 1000 (Hall *et al.*, 2002).

2. Targeted Profiling

The large numbers of secondary metabolites produced by plants, perhaps in excess of 200,000 throughout the plant kingdom, present the biggest problem for nontargeted metabolomics. These compounds are chemically very diverse, often species specific, and the physical details of most are not present in chemical databases. Of course, such metabolites do not have to be actually identified in initial metabolomics experiments; an "unknown" can be treated exactly the same way as a "known" during clustering and statistical analysis, and can be treated as a genetic marker in the absence of its chemical identity. However, a greater problem for inclusion of secondary metabolites in the high throughput profiling necessary for the technology to be used in genetic mapping and breeding is the chemical diversity of these compounds. This necessitates specific extraction and sometimes separation protocols for

specific classes of secondary metabolites. Simple GC-MS is seldom used because of problems with derivatization and subsequent identification of the derivatives. Rather, most studies on secondary metabolite profiling have used a targeted approach designed to address a single class of compound, and one of the most commonly used analytical approaches has been high performance liquid chromatography (HPLC) coupled with mass spectrometry and/or UV/visible spectrophotometric detection (Sumner *et al.*, 2003).

Targeted metabolite profiling is an old technology, and good methods for many different classes of compounds have been available for some time. Examples include methods for profiling flavonoids and isoflavonoids (Graham, 1991; Liu and Dixon, 2001), phenylpropanoids (Chen *et al.*, 2003), triterpenes (Huhman and Sumner, 2002), carotenoids (Fraser *et al.*, 2000), various classes of alkaloids (He, 2000; Kalén *et al.*, 1992; Stashenko *et al.*, 2000), and acyl CoAs (Larson and Graham, 2001). These and related targeted profiling approaches have been applied to understanding the genetic basis of metabolite abundance via quantitative trait locus (QTL) analysis (Morrell *et al.*, 2006) (Section IV.D below), and determining phenotypic effects of transgenic modification of plants for improved quality traits (Deavours and Dixon, 2005; Morreel *et al.*, 2004; Xie *et al.*, 2006). Figure 3 provides an example of targeted profiling of flavonoid compounds in alfalfa. The extraction and HPLC method used favored the extraction and separation of (iso)flavonoids and their glycosides (Deavours and Dixon, 2005). By this approach, it was possible to show that constitutive expression of an isoflavone synthase transgene in alfalfa led to accumulation of isoflavone glucosides in the leaves, whereas the endogenous flavonoids found in the leaves were glucuronic acid conjugates. Plants constitutively expressing the isoflavone synthase produced higher levels of potentially defensive isoflavonoid metabolites following exposure to biotic or abiotic stress (Deavours and Dixon, 2005).

3. Metabolic Fingerprinting

As suggested above, it is not necessary to know the exact chemical nature of the components of a metabolic profile to be able to use the profile as a genetic and phenotypic tool. Because of this, more rapid analytical methods that provide a "metabolic fingerprint" rather than a detailed profile of individually separated molecules are being applied in the field of molecular agriculture. These include nuclear magnetic resonance (NMR) and near infrared (NIR) spectroscopy. NMR and NIR profiles can be subjected to the same types of statistical analysis as GC or HPLC elution profiles, and regions of the profiles that exhibit the greatest variation between samples/treatments can be correlated with genotype, environment, or expression of

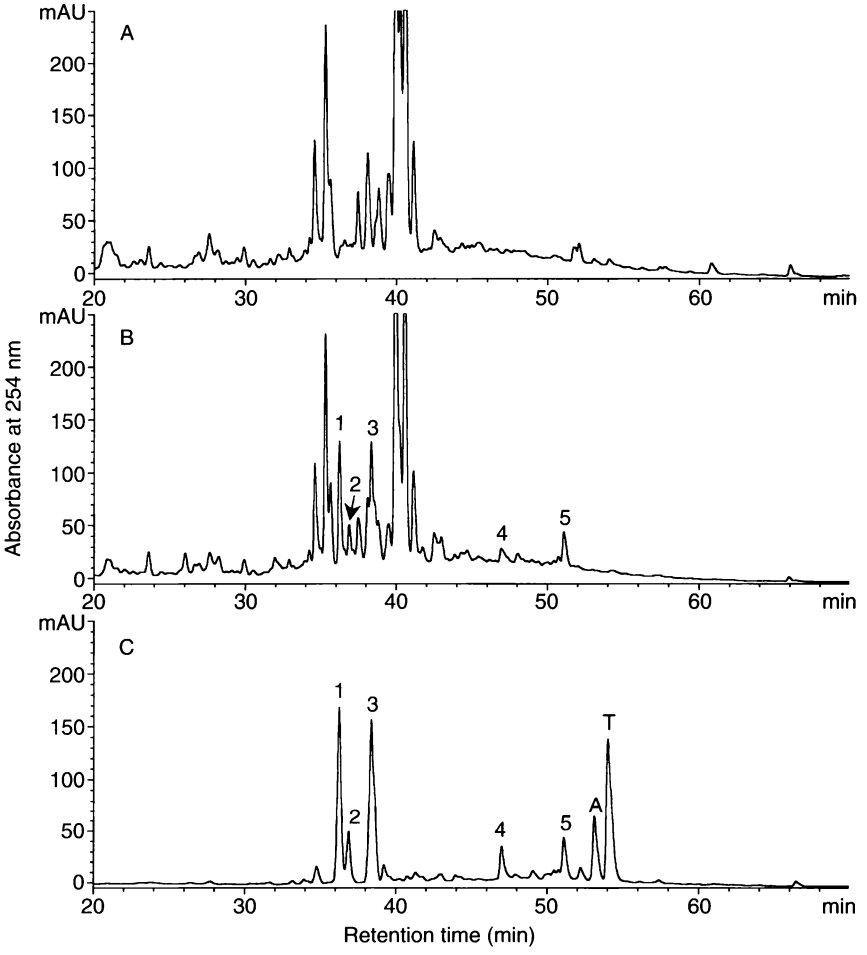


Figure 3 Use of targeted metabolite profiling to show production of isoflavone glycosides in leaves of transgenic alfalfa plants expressing an *M. truncatula* isoflavone synthase gene (Mt*IFS1*) under control of the constitutive cauliflower mosaic virus 35S promoter. HPLC traces show unhydrolyzed leaf extracts of (A) an empty vector control line and (B) an Mt*IFS1*-expressing line. Peaks with UV spectra similar to the isoflavone genistein that are not present in the control extracts are numbered 1–5. Peaks 1 and 4 were identified by LC/MS as the isoflavone glucosides genistin and sissotrin, respectively. Panel (C) shows an HPLC trace of a leaf extract of the Mt*IFS1*-expressing line after digestion with β-glucuronidase. This converts the major endogenous leaf flavonoids, glucoronides of the flavones apigenin and tricin, to their corresponding aglycones (peaks A and T, respectively), better revealing the isoflavone glycosides. HPLC was carried out on an ODS2 reverse-phase column (5-μm particle size, 4.6 × 250 mm^2) and eluted in 1% (v/v) phosphoric acid with an increasing gradient of acetonitrile (0–5 min, 5%; 5–10 min, 5–10%; 10–25 min, 10–17%; 25–30 min, 17–23%; 30–65 min, 23–50%; 65–69 min, 50–100%) at a flow rate of 1 ml min^{-1} (Deavours and Dixon, 2005).

a transgene. NMR fingerprinting has begun to find uses in functional genomics, the study of genetic diversity, the evaluation of the safety of transgenic crops, and determining responses of plants to infection (Charlton *et al.*, 2004; Ward and Beale, 2006). In addition to providing a standard method for assessment of plant quality traits, such as fiber and digestibility (Jung, 1997), NIR techniques have been developed that allow for rapid estimation of specific metabolites, such as ergot alkaloids in endophyte-infected tall fescue (Roberts *et al.*, 2005).

4. Nonbiased Metabolomics

Although it essentially targets those classes of molecules that are subject to the applied extraction protocol(s) and separation methods, the standard GC-MS approach to plant metabolomics is largely viewed as a nontargeted approach. Its effectiveness depends on the availablilty of mass spectral libraries to assist the identification of both known and unknown components (Kopka, 2006). In addition to the examples provided in Section D.I above, GC-MS profiling has become an important tool for functional genomics and analysis of biotic stress responses in legumes (Broeckling *et al.*, 2005; Debrosses *et al.*, 2005).

The remarkably high mass resolution power of Fourier transform ion cyclotron mass spectrometry (FT-MS) allows metabolite profiling without the need for preseparation of metabolites. Complex mixtures can be injected directly into the mass spectrometer, and the components are essentially resolved via the mass discrimination of the instrument (Aharoni *et al.*, 2002). Identification is based on absolute mass measurement. This is a very powerful approach and has been applied to gene discovery associated with nutritional stress in *Arabidopsis* (Hirai *et al.*, 2004). However, the method can not discriminate between isomers.

Extensive metabolomic analysis is facilitated by metabolic pathway databases for the plant species of interest. Several such databases have been developed, and some incorporate features for simultaneous display of gene expression data from microarrays or other formats directly onto the metabolic pathway maps (Krieger *et al.*, 2004; Lange and Ghassemian, 2005; Thimm *et al.*, 2004). The version for *M. truncatula* is called MedicCyc (Urbanczyk-Wochniak and Sumner, 2007) and features more than 250 metabolic pathways with associated genes, enzymes, and metabolites. More challenging to construct are databases that actually store the raw data obtained from different types of "omics" approaches. *D*atabase of "*ome*s" (DOME) is an early example of such an approach, initially constructed to house transcript, protein, and metabolite data from *M. truncatula* cell lines responding to biotic and abiotic elicitors (Mehrotra and Mendes, 2006).

5. Integrating Transcriptomic and Metabolomic Datasets

Several examples demonstrate the advantages of being able to simultaneously profile the transcriptome and the metabolome. This is a powerful new approach to the analysis of biological systems because it allows statistical analysis tools to be used to analyze, and therefore correlate, both genotype and phenotype (the metabolome; Phelps *et al.*, 2002). However, such an approach presents some problems in whole plant systems, because the existence of different cell types in a complex organism makes it difficult to correlate transcripts and metabolites that might not be expressed or synthesized in the same cells. Because of this, the first examples of the approach were reported for prokaryotic systems (Phelps *et al.*, 2002), which also have the advantage of being more readily amenable than plants to flux analysis through labeling with isotopic precursors followed by MS or NMR analysis. Likewise, to take advantage of more homogeneous cell populations, the first examples of integrated transcriptomics/metabolomics in plants utilized cell suspension cultures. In a study aimed at elucidating the genes of secondary metabolism, particularly nicotine alkaloid biosynthesis, expressed in tobacco BY2 cell suspension cultures in response to the wound signal MeJA, targeted metabolite profiling by GC-MS was coupled with cDNA-AFLP analysis of differentially induced transcripts (Goossens *et al.*, 2003). This approach led to the identification of a number of candidate genes for involvement in the nicotine biosynthesis pathway itself, or in its transcriptional regulation. A similar approach, but using oligonucleotide microarrays for transcript profiling and LC-MS for analysis of secondary metabolites, was used to identify genes encoding glycosyltransferases involved in MeJA-induced accumulation of triterpene saponins in *M. truncatula* cell suspension cultures (Achnine *et al.*, 2005).

A study of the relationships between the transcriptome and primary and secondary metabolism in *Arabidopsis* seedlings under conditions of sulfur or nitrogen deprivation revealed the power of this integrated approach for gene discovery, especially when combined with powerful informatic analysis (Hirai *et al.*, 2004). The transcriptome analysis used a macroarray that contained EST clones corresponding to around 9000 *Arabidopsis* genes, and the metabolome was profiled using extraction in three solvent systems of different polarities followed by FT-MS analysis. General responses to sulfur and nitrogen deficiency were identified through mathematical analysis of transcriptome and metabolome datasets using PCA and batch-learning self-organizing map analysis (Hirai *et al.*, 2004). Using the same techniques, detailed metabolite and transcript profiling of the *Arabidopsis* PAP1 mutant, which overexpresses anthocyanins (Sections II.F and VI.B.2), revealed the presence of eight novel anthocyanins, and, among the 32 genes that were shown to be specifically upregulated by PAP1, two were identified as specific

flavonoid- and anthyocyanin-glycosyltransferases (Tohge *et al.*, 2005). Such approaches represent a powerful tool for functional annotation of genes, particularly those that are members of large families that encode enzymes with promiscuous and overlapping *in vitro* substrate specificities such as glycosyltransferases (Bowles *et al.*, 2006). For such enzymes, correlation of gene and metabolite expression patterns might be the deciding approach for confirming *in vivo* function.

6. Profiling Technologies and "Substantial Equivalence"

During the regulatory process for approval of transgenic plants for commercial use, it is necessary to demonstrate that the regulated product is "substantially equivalent," from a compositional viewpoint, to its unmodified parent material. Generally, such equivalence is taken as meaning that the nutritional properties of the plant are not altered and that no potentially toxic compounds have been introduced. In a study with herbicide-resistant alfalfa, field-grown material was analyzed for fiber content, amino acid and mineral composition, and levels of the potentially estrogenic isoflavonoid coumestrol (McCann *et al.*, 2006) using classical analytical procedures for each parameter. Although the increasing ability to perform more global analysis of transcripts, proteins, and metabolites suggests that "omics" approaches may become standard for demonstration of substantial equivalence, there are also arguments against this. In some respects, "omics" approaches are too sensitive, and it is sometimes the case that variations between different tissues, varieties, or environmental conditions are greater than the changes observed following introduction of a transgene. This is illustrated by the PCA analysis of soluble phenolic compound profiles in control and transgenic alfalfa lines modified in lignin content and composition (Chen *et al.*, 2003). PCA analysis could resolve profiles from transgenics from those from controls for stem extracts, but not for leaf extracts. At the same time, the method resolved differences between two nontransgenic cultivars when considering extracts from either leaf or stem tissue (Chen *et al.*, 2003).

Nevertheless, metabolite and transcript profiling and fingerprinting have been performed to establish, or refute, substantial equivalence. In one study with peas, NMR fingerprinting showed that expression of a transgene did indeed affect the metabolite profile, but that this effect was masked by changes induced by environmental factors such as drought (Charlton *et al.*, 2004). In a study of alfalfa plants ectopically producing isoflavones in the leaves, microarray analysis failed to demonstrate significant changes in transcript levels, other than for the expressed transgene, in pair-wise comparisons of controls and trangenics, although there was significant interplant variation (Deavours and Dixon, 2005). If the new "omics" technology is to be

applied to the assessment of substantial equivalence, it is important that a public consensus is reached as to which analytes are of significance for human and animal health considerations and that tolerance intervals are defined that encompass the variations found in commercial populations (Dixon *et al.*, 2006; Ridley *et al.*, 2002).

IV. MOLECULAR MARKERS

A. Marker Types

Following the segregation of Mendelian genetic markers is the most powerful method to understand hereditary transmission (Beckmann and Soller, 1993). The advent of agriculture and domestication began with selection of superior genotypes/lines. Classical plant breeding techniques were mainly based on phenotypic selection (PS) where traits of interest were tagged with markers like seed color, leaf size, and flower color, which could distinguish between genotypes. However, morphological markers are influenced by the environment, may be linked to undesirable traits, and their use for selection is time consuming, requiring large population sizes and space for testing. In the early molecular era, isozyme and protein markers were used to select genotypes in plant breeding programs. These biochemical markers are, however, characterized by low polymorphism, especially between similar or related cultivars. The advent of rapid DNA sequencing led to the discovery of DNA-based markers (molecular markers) and these have became the marker class of choice. Molecular markers are based on DNA polymorphism as a result of mutation and are highly heritable. Their main advantage is that they are much more numerous and polymorphic than morphological or biochemical markers. The genomes of most plant species contain between 10^8 and 10^{10} nucleotides, and thus even a small proportion of polymorphic sites can yield a large number of potential markers (Paterson *et al.*, 1991).

Early identification of DNA markers relied on restriction fragment length polymorphisms or RFLPs (Tanksley *et al.*, 1989). RFLP markers segregate as codominant alleles capable of identifying all three morphs, thus being highly informative. The polymerase chain reaction (PCR) technique revolutionized molecular marker technology. Rapid amplification of discrete DNA fragments by PCR enables quick identification of DNA polymorphisms within a genome. The rapid identification of such markers linked to important loci facilitates their integration into plant breeding programs. Randomly amplified polymorphic DNA (RAPD) markers were the first markers of this kind to be developed (Williams *et al.*, 1990). In the past two decades, many

different molecular marker systems have been developed to serve specific needs, many of which have relied on genome and EST-sequencing projects. The marker systems currently employed include RFLPs, single strand conformation polymorphisms (SSCPs), sequence-tagged microsatellite sites (STMSs), RAPDs, sequence characterized amplified regions (SCARs), ESTs, microsatellites, or simple sequence repeats (SSRs), amplified fragment length polymorphisms (AFLPs), sequence-tagged sites (STSs), cleaved amplified polymorphic sequences (CAPSs), single nucleotide polymorphisms (SNPs), and heteroduplex markers. The ideal marker class should provide more markers per unit DNA, be stable, easily detectable, safe, and cost-effective, and have a high degree of polymorphism. Molecular markers are now an indispensable tool for cultivar identification and parentage analysis (Dudley *et al.*, 1992; Sefc *et al.*, 2000), genetic diversity analysis (Mian *et al.*, 2005a; Smith and Smith, 1992), genome mapping, and the tagging of genetically important traits (Cardinal *et al.*, 2003).

B. Molecular Genetic Maps

Linkage maps are constructed by following the segregation pattern of molecular markers in a population. Markers are placed in linear order based on pair-wise recombination frequencies between the markers. High marker polymorphism in a population is the key for successful linkage analysis. Backcross, F_2, recombinant inbred lines (RILs) and doubled haploids are the most commonly used populations for molecular mapping of self-pollinated crops (Chen *et al.*, 2001; Eujayl *et al.*, 1998). The pseudo F_2 cross (between two heterozygous parents) is the most frequently used population in mapping cross-pollinated crops (Saha *et al.*, 2005; Van Eck *et al.*, 1993). All of the above populations have both advantages and disadvantages. The F_2 and backcross populations show higher segregation, but are not available for subsequent studies. RIL populations can be permanently propagated and offer unique advantages in quantitative trait loci (QTL) mapping (Burr and Burr, 1991). However, development of RILs is time consuming and very difficult in self-incompatible species. In doubled haploid populations, homozygocity for a particular locus can be obtained quickly but segregation distortion is a major problem (Cloutier *et al.*, 1995).

Molecular markers are commonly used to generate genetic linkage maps, and have provided a major contribution to the genetic knowledge of many cultivated plants. Over the past two decades, genetic linkage maps have been developed for most of the agriculturally important plant species (Alm *et al.*, 2003; Chen *et al.*, 2001; Eujayl *et al.*, 1998; Gebhardt *et al.*, 1991; Jacobs *et al.*, 1995; Jones *et al.*, 2002; Kuhl *et al.*, 2001; Perez *et al.*, 1999;

Saha *et al.*, 2005; Tanksley *et al.*, 1992; Warnke *et al.*, 2004; Xu *et al.*, 1995). Such molecular maps have been used to map major genes (Van Eck *et al.*, 1993) and to identify the genetic components of polygenic traits (Bonierbale *et al.*, 1994; Qiu *et al.*, 2006). Genetic linkage maps have been used successfully for the identification of markers linked to a gene of interest (Xu *et al.*, 1999). The availability of genetic maps opened the door for comparative mapping, which allows the comparison of genome organization and orientation of one species to that of a closely or distantly related species through common markers between maps. Comparative mapping has revealed that gene content and order are generally conserved among closely related species (Alm *et al.*, 2003; Jones *et al.*, 2002; Van Deynze *et al.*, 1995). It has also been used for extending genetic information from model organisms to genetically more complex species (Paterson *et al.*, 1995).

C. Genomics for Generation of Molecular Markers

Among the different marker classes, SSRs have become the marker class of choice due to their manifold advantages over other marker systems. Single nucleotide polymorphism (SNP) markers are also becoming more popular as genome sequences for agriculturally important crops are becoming available and SNPs are detected at high frequencies. TILLING is also an attractive system for genome analysis.

1. Simple Sequence Repeats

SSRs, also known as microsatellites, are short stretches of DNA that are repeated many times. The di-, tri-, tetra-, or pentanucleotide tandem repeats are often multialleleic, chromosome-specific and dispersed throughout the genome (Weber and May, 1989). The basis of polymorphism is due to variation in the number of tandemly repeated nucleotide motifs, and this is thought to arise from slippage of the DNA polymerase during DNA replication. Although these SSRs are highly abundant in animal and plant genomes (Hamada *et al.*, 1982), the dinucleotide repeats are more common in genomic SSRs (Lee *et al.*, 2004), and trinucleotide motifs are the most abundant in EST-SSRs (Saha *et al.*, 2004). SSR markers are inherited in a Mendelian fashion and are mostly codominant in nature (Saghai-Maroof *et al.*, 1994).

Genomic or EST libraries can be screened for sequences that contain microsatellite motifs in order to develop primers (Panaud *et al.*, 1996). In the early 1990s, SSR markers were mainly developed from genomic libraries, an expensive and inefficient procedure (Squirrell *et al.*, 2003). The availability of large numbers of ESTs and other DNA sequence data made

SSR marker development efficient and cost effective for many plant species. The development of SSRs from ESTs has been reported in various crop species, including rice (Cho *et al.*, 2000), durum wheat (Eujayl *et al.*, 2002), barley (Thiel *et al.*, 2003), rye (Hackauf and Wehling, 2002), *M. truncatula* (Eujayl *et al.*, 2004), and tall fescue (Saha *et al.*, 2004). The frequency of SSRs in the EST databases of cereal crops (rice, wheat, maize, barley, and sorghum) varies from 1.5% (maize) to 4.7% (rice) (Kantety *et al.*, 2002), while in tall fescue it is only 1.3% (Saha *et al.*, 2004). The rate of polymorphism of EST-SSRs is comparatively lower than that of genomic-SSRs (Cho *et al.*, 2000; Thiel *et al.*, 2003).

SSRs have been used for the construction of linkage maps in a number of species including *Arabidopsis* (Bell and Ecker, 1994), maize (Senior *et al.*, 1996), wheat (Röder *et al.*, 1995), rice (Panaud *et al.*, 1996), barley (Liu *et al.*, 1996), and soybean (Akkaya *et al.*, 1995). Allelic profiles of genotypes have been studied using SSR markers for the purpose of genotype identification in potato (Schneider and Douches, 1997), soybean (Maughan *et al.*, 1995), grape (Thomas and Scott, 1993), and rapeseed (Kresovich *et al.*, 1995). Selection of agronomic traits was also accomplished using SSR markers (Yu *et al.*, 1994).

Map alignment through common markers is important for making mapping studies universally useful within a species (Powell *et al.*, 1996). Specific SSR primers from one species can be used to amplify DNA from another related species. As EST-SSR markers are derived from transcribed regions of DNA, they are expected to be more conserved and have a higher rate of transferability than genomic SSR markers (Scott *et al.*, 2000). For instance, tomato SSR sequences generated polymorphic and useful alleles in potato (Provan *et al.*, 1996). SSR loci have high rates of transferability across species (>50%) within a genus (Eujayl *et al.*, 2004; Gaitán-Solís *et al.*, 2002; Thiel *et al.*, 2003). However, the transferability of SSR loci across genera and beyond seems to be low (Roa *et al.*, 2000; Thiel *et al.*, 2003; White and Powell, 1997).

2. Single Nucleotide Polymorphisms

SNP has emerged as an important molecular marker system. The utility of SNPs in answering a large range of biological questions in a variety of fields is now beyond question. SNPs greatly expedite the understanding of many diseases and genetic variations in humans. SNPs associated with different human conditions, such as risk of cardiovascular disease, and susceptibility to Alzheimer's, susceptibility to hip osteoarthritis (Mototani *et al.*, 2005), and risk of thrombosis (Ridker *et al.*, 1995), have been identified. In plants, many SNPs have been shown to be associated with useful traits. For example in rice, SNPs for the fragrance trait and starch gelatinization temperature have been identified (Bradbury *et al.*, 2005).

The advantage of SNP markers is that they occur at a high frequency in genomes of all organisms. However, the frequency is highly dependent on the type of DNA surveyed, for example, coding versus noncoding sequences, genes of choice, and species investigated. In genomic DNA of maize inbred lines, one SNP was identified per 83 bases, while in the barley intronless *Isa* gene one SNP occurs every 27 bases (Bundock and Henry, 2004), and in sugarcane ESTs one SNP is found per 50 bases (Cordeiro *et al.*, 2006). Significant improvements have been made in SNP detection protocols, including dCAPS (Michaels and Amasino, 1998) and mass spectrometry using MALDI-TOF MS (Stoerker *et al.*, 2000). There have also been advances in fluorescence-based technologies, for example Amplifluor® (Serological Corporation), TaqMan®, SnaPshot®, and SNPlex® (Applied Biosysteems), and Illumina® (Illumina, Inc.), to detect SNPs. Chip-based technologies, for example Genechips (Affymetrix), and microarray technology (Wang *et al.*, 2005a) have also been used for SNP detection. Dot-blot-SNP analysis was described for application in plant breeding and cultivar identification in rice (Shirasawa *et al.*, 2006). A comparison of three SNP genotyping methods including GOOD (Sauer *et al.*, 2000), Amplifluor®, and TaqMan® for three different herbicide resistance genes from *A. thaliana* found the best results with TaqMan® for PCR specificity, flexibility in primer design, and success rate (Giancola *et al.*, 2006). However, all three genotyping techniques were successful in discriminating alleles in various plant species.

SNPs are very useful as genetic markers for population studies, germplasm fingerprinting and cultivar identification, molecular mapping, genotype/phenotype association, and for positional-cloning of specific genes. They have practical utility in identifying mutant lines developed from an original cultivar where most of the other marker systems are ineffective (Shirasawa *et al.*, 2006). The addition of SNP markers significantly increased the overall map length and marker density in sunflower (*Helianthus annuus* L.; Lai *et al.*, 2005). SNP markers are considered useful for gene mapping using populations derived from crosses between closely related lines; molecular markers like AFLPs and SSRs are found to be less polymorphic in these populations. A new breeding method named "DNA-selection breeding" has been proposed whereby genes associated with different agronomically important traits are selected by SNP analyses and used for selecting superior genotypes (Shirasawa *et al.*, 2006).

3. Tilling

TILLING (Section II.F) is a high-throughput, sensitive, cost-effective, and rapid means of finding genetic variation in a population. TILLING is effective in small or large genomes, diploid or hexaploids, and has great

potential to identify both induced and naturally occurring variation in many species. Thousands of plants or animals can be screened to identify any single base change as well as small indels (insertions and/or deletions) in any gene or genomic regions (Comai *et al.*, 2004). A million base pairs of genomic DNA can be screened per single assay, which makes TILLING a high-throughput technique (Slade and Knauf, 2005). This technique combines traditional chemical mutagenesis and modern high-throughput genotyping. DNA from eight mutant lines can be combined in one PCR tube, thus a 96-well PCR plate can screen 768 genotypes. During electrophoresis, mutant lines reveal polymorphic fragments relative to wild-type lines. TILLING is suitable for SNP discovery because it is sensitive enough to detect rare SNPs.

D. Metabolomic-Based "Markers"

The term "genetical metabolomics" was defined to describe the use of metabolite profiling in QTL mapping (Morrell *et al.*, 2006). If levels of specific metabolites can be used as quantitative traits to define metabolic QTLs (mQTLs) that control levels of specific metabolites, the nature of the genes underlying the mQTLs might be more readily obvious that in classical genetic QTL mapping, since the (probably) known structure of the metabolic pathway under study might suggest regulatory control points. As an example of this approach, flavonoid profiles (from targeted HPLC analysis) of apical tissues were used for mQTL mapping in two full-sib families of poplar, and three mQTLs tentatively shown to map to enzymes of the flavonoid pathway (Morrell *et al.*, 2006).

E. Advantages of Marker-Assisted Breeding

Marker-assisted selection (MAS) is a complementary technology which expedites the conventional methods of genetic selection for plant and animal improvement. In classical plant breeding systems, many cycles of selection and backcrossing are required to obtain a desirable genetic gain. Besides, classical breeding is mostly successful for dominant traits which are easily inherited in subsequent generations. However, genetic gains from classical breeding methods in major crop species have reached an apparent plateau. The use of molecular markers associated with qualitative and quantitative traits has been successfully used for the indirect selection of genes of interest. The advantages of MAS include ability to reveal sites of variation in a DNA sequence, and accelerated progress by shortening the breeding cycle. MAS not only gives larger genetic responses but also dramatically increases the frequencies of superior genotypes as compared to PS (Liu *et al.*, 2004). It is

particularly promising when dominant alleles are present and linked in coupling phase (Berloo and Stam, 1998). Molecular markers can alleviate complications of phenotype-based selection, provided they cosegregate with the gene of interest.

Crop production is significantly affected by a number of biotic and abiotic stresses, responses to which are mostly controlled by many genes. Resistant cultivar development is the practical solution for many quantitatively inherited traits. Polygenic control, along with large environmental influences, largely limits the effectiveness of PS for these traits. MAS provides an efficient way to accelerate development of resistant varieties (Frisch *et al.*, 1999). For example, the submergence tolerance gene *Sub1* of rice was tagged with two microsatellite markers, RM219 and RM464A, and several lines were identified that were homozygous for these loci and were genetically similar to the parent M-202 (Xu *et al.*, 2004). To validate the major QTL for scab resistance in rice, the associated SSR markers were analyzed in the $F_{2:3}$ lines of one population and in the $F_{3:4}$ lines of the other (Zhou *et al.*, 2003b). Markers from the original population were also closely associated with scab resistance in both validation populations. MAS was found to be more effective than PS. However, the most effective selection strategy was MAS during the seedling stage followed by PS after flowering. In another example, the eating and cooking quality of Zhenshan 97, an elite parent of hybrid rice, was developed by introgressing the Waxy gene region of Minghui63 through MAS breeding (Zhou *et al.*, 2003a). MAS was likewise used for pyramiding three bacterial blight resistance genes (*Xa5*, *Xa13*, and *Xa21*) into indica rice cultivar PR106 (Singh *et al.*, 2001). The gene combination provided a wide spectrum of resistance to the pathogen population that consisted of 23 different *Xanthomonas oryzae* isolates.

In major cereals such as rice, wheat, maize, and barley, molecular markers associated with different qualitative and quantitative traits have been identified and used for MAS. Substantial use of MAS in maize, with a slower pace of uptake in wheat and rice breeding, has been observed. Large-scale genotyping and MAS programs have been initiated through Rice CAP and Wheat CAP projects with funding from USDA, CSREES. Application of MAS to breeding programs depends on its relative cost and expected economic return. The best prospect for MAS is in multiple-trait improvement. Excluding costs, multiple-trait MAS can be used to increase the aggregate breeding values in quantitative characters and is expected to be more effective than conventional selection or single-trait MAS (Xie and Xu, 1998).

Gains from MAS and PS were compared for quantitative traits in sweet corn (Yousef and Juvik, 2001). A total of 52 paired comparisons were made between MAS and PS composite populations. MAS led to significantly higher gain in 38% of the paired comparisons compared to only 4% for PS. The average gains from MAS and PS were 10.9% and 6.1%, respectively.

It was also observed that MAS was most appropriate when traits are difficult and costly to measure and that the higher gain from MAS compensated for the higher costs. It was concluded that "incorporating DNA markers to traditional breeding programs could expedite selection progress and be cost-effective." A RAPD marker associated with common bacterial blight resistance in a common bean population (PC50/XAN159) was transformed into a SCAR marker and used for screening a different population (Yu *et al.*, 2000). The SCAR marker was 94.2% accurate in recognizing the resistant genotype. Cost comparison of MAS with greenhouse screening indicated that MAS was about one-third less expensive.

V. TRANSGENESIS

A. Transgenesis as a Tool for Functional Genomics

Transgenesis refers to the introduction of heterologous or homologous DNA into a plant genome resulting in its stable integration and expression. The technology has played a critical role in defining the *in vivo* functions of plant genes. In recent years, with the rapid increase in gene sequence information, systematic transgenic approaches have been taken to characterize large numbers of genes in both reverse and forward genetic studies, particularly in model systems. Predictions of gene function based on sequence homology alone do not necessarily provide information on the exact biological role of the gene *in planta* (van Enckevort *et al.*, 2005). After completion of the *Arabidopsis* genome sequence, at least 40% of the initial gene predictions based on computational annotation were subsequently found to be erroneous (Alonso and Ecker, 2006). As one of the key experimental methods in functional genomics, transgenesis has the advantage of revealing the direct link between gene sequence and function; such results not only further the understanding of basic biological questions, but also facilitate exploitation of genomic information for crop improvement.

Transgenesis has been widely used for loss-of-function and gain-of-function analyses of plant genes. Insertional mutagenesis using T-DNA is one of the major tools for functional analysis that can provide a phenotype as a clue to gene function (Xu *et al.*, 2005). T-DNA mutant collections are commonly produced by *Agrobacterium*-mediated transformation using a simple Ti plasmid carrying a selectable marker gene. If the T-DNA inserts within the boundaries of a gene, it can alter or abolish the function of the gene. Because of the disruptive nature of randomly inserted T-DNA, this type of mutagenesis is commonly associated with loss-of-function of endogenous genes. Compared with mutagenesis caused by chemical agents

[e.g., ethyl methanesulphonate (EMS)] or physical agents (e.g., fast neutrons, γ-radiation), the use of T-DNA as a mutagen offers the advantage of easy identification of the mutated gene. The T-DNA not only disrupts the expression of the gene into which it is inserted but also acts as a marker for subsequent identification of the mutation (Krysan *et al.*, 1999). When coupled with transposons, the introduction of the transposon containing T-DNA into the plant genome allows for the simultaneous disruption of different loci (Tadege *et al.*, 2005). Large numbers of T-DNA insertional lines have been produced in *Arabidopsis* and rice (Alonso and Ecker, 2006; Krysan *et al.*, 1999; Walden, 2002; Xu *et al.*, 2005). So far, more than 360,000 insertion sites have been mapped in the *Arabidopsis* genome, covering ~90% of the genes. One of the most exciting uses of the near complete collection of gene-indexed *Arabidopsis* mutations is the ability to carry out genome-wide forward genetic screens (Alonso and Ecker, 2006).

T-DNA mutagenesis has its limitation in analyzing the function of redundant genes (Xu *et al.*, 2005). Gain-of-function approaches such as gene overexpression and T-DNA activation tagging are straightforward and powerful approaches for elucidating gene function. Transgenic expression of all the cDNAs found in *Arabidopsis* resulted in the identification of many genes conferring interesting phenotypes. Efforts were also made to overexpress all the TF genes in *Arabidopsis*. Because of the unique characteristics and modes of action of TFs, this overexpression strategy is considered particularly effective in revealing gene function (Zhang, 2003). Sometimes the same gene can be found by different approaches. For example, the identification of WIN1, an *Arabidopsis* ethylene response factor-type TF that can activate wax deposition, was achieved by systematic overexpression of all gene sequences predicted to encode proteins sharing conserved domains with cognate TFs (Broun *et al.*, 2004). In an independent study, the *SHN1* gene, which shares the same sequence as *WIN1*, was obtained by screening a collection of 2000 transposon activation-tagged lines (Aharoni *et al.*, 2004).

Functional genomics has been broadly defined to include many endeavors on a genome-wide scale, such as transcriptional profiling to determine gene expression patterns, sequence alignment-based comparisons to identify homologues between and within organisms, and the use of virus-induced gene silencing to rapidly detect phenotypic effects (Xu *et al.*, 2005). Transgenesis studies are generally required to confirm the functions of the genes identified by these methods. While extremely useful, most other approaches to gene function are correlative and do not necessarily prove a causal relationship between gene sequence and function (Krysan *et al.*, 1999). Sometimes unexpected results have been obtained by transgenic analysis. The overexpression of *COL9*, a member of the *CONSTANS-LIKE* gene family, resulted in delayed flowering in *Arabidopsis*, which is opposite to the role that

the *CONSTANS* (*CO*) gene plays. Further analysis revealed that *COL9* delays flowering possibly by antagonistically repressing the expression of *CO*, and concomitantly reducing *FLOWERING LOCUS T* (*FT*) expression (Cheng and Wang, 2005). In *M. truncatula*, a TF gene (*WXP1*) related to wax biosynthesis was identified; overexpression of *WXP1* resulted in improved drought tolerance in alfalfa (Zhang *et al.*, 2005). Sequence comparison with *WXP1* revealed its homologue in *M. truncatula*, designated *WXP2*. Transgenic expression of both *WXP1* and *WXP2* in *Arabidopsis* resulted in improved drought tolerance; however, the transgenic plants were opposite in their freezing tolerance, with *WXP1* plants more tolerant and *WXP2* plants more sensitive to freezing stress (Zhang *et al.*, 2007).

B. Current Approaches to the Generation of Transgenic Plants

There are many variations of gene transfer methods to introduce transgenes into the plant genome. The most widely used methods are *Agrobacterium*-mediated gene transfer and biolistic transformation. Both have been applied to legume transformation, although the *Agrobacterium*-mediated approach has been the most popular (Somers *et al.*, 2003).

A. tumefaciens is a soilborne bacterium that, in nature, is capable of inserting a discrete portion of its DNA into the genome of a wide range of dicotyledonous plants (Valentine, 2003). Most of the machinery necessary for the gene transfer resides on a tumor-inducing (Ti) plasmid that carries two important genetic components: the T-DNA delimited by two 25-bp direct repeats at its ends and the virulence region (Tzfira and Citovsky, 2006). *Agrobacterium*-mediated transformation systems take advantage of this natural gene transfer mechanism in plants. Two key advances, the development of binary Ti vectors and of a range of disarmed *Agrobacterium* strains, have made *Agrobacterium* transformation the first option in engineering transgenic plants (Hellens and Mullineaux, 2000). *Agrobacterium*-mediated gene transfer offers the following advantages: (1) a significant portion of the transformants contains single copy transgenes, (2) *in planta* transformation without the need of tissue culture is possible in *Arabidopsis*, (3) numerous vector systems are now available, and (4) it is possible to transfer large DNA fragments, including bacterial artificial chromosomes (Herrera-Estrella *et al.*, 2005).

The biolistic method was developed as a necessity to transform species initially considered recalcitrant to *Agrobacterium* transformation (Herrera-Estrella *et al.*, 2005). Biolistics, or microprojectile bombardment, employs high-velocity gold or tungsten particles to deliver DNA into living cells for stable transformation (Christou, 1992; Sanford, 1988). Gene delivery to plant cells and tissues by microprojectiles has led to the production of

transgenic plants from many species, particularly monocots. Because biolistic transformation is a physical process and therefore involves only one biological system, it is a fairly reproducible procedure that can be easily adapted from one laboratory to another. Biolistic transformation is the only reliable method for chloroplast transformation. The main disadvantage of this method is the frequently occurring multiple copy integration.

Since the successful creation of transgenic *Nicotiana* and *Petunia* in the early 1980s, *Agrobacterium*-mediated transformation has been the method of choice in producing transgenic plants in a wide range of dicot species. Although it was initially considered impossible to transform monocot species with *Agrobacterium*, transgenic plants have been obtained with many monocot crops since the mid 1990s, including major cereals like rice, maize, wheat, barley, and a number of forage and turf grasses (Cheng *et al.*, 1997, 2004; Hiei *et al.*, 1994; Ishida *et al.*, 1996; Tingay *et al.*, 1997; Wang and Ge, 2006).

The cost associated with meeting regulatory requirements is a substantial impediment for the commercialization of transgenic crops (Bradford *et al.*, 2005). In recent years, public concern about the extent to which transgenic crops might differ from their traditionally bred counterparts has resulted in revised molecular strategies and choices of genes (Rommens *et al.*, 2004). Due to the nature of popularly used promoters, vectors, and selectable markers for plant transformation, most transgenic plants contain DNA from multiple organisms. It has been proposed to categorize GMOs into different classes based on the genetic distance between the target organism and the source of the transgenes (Nielsen, 2003). In an attempt to address some of the public perception issues relating to introduction of foreign DNA into plants, it has been shown that certain host plant DNA sequences can function in the same way as the *Agrobacterium* T-DNA border sequences (Rommens *et al.*, 2004). By incorporating such sequences to guide integration of the inserted transgene, and linking a positive selection for temporary expression of the selectable marker with a negative selection against its integration, it was possible to produce transgenic potato plants with reduced expression of tuber-specific polyphenol oxidase that contain no foreign DNA (Rommens *et al.*, 2004).

Driven by the complexity of intellectual property issues that limit the use of *Agrobacterium* in both public and private sectors, several species of bacteria outside the *Agrobacterium* genus have been modified to mediate gene transfer to different plant species (Broothaerts *et al.*, 2005). These plant-associated symbiotic bacteria, including *Rhizobium* species NGR234, *S. meliloti*, and *Mesorhizobium loti*, were made competent for gene transfer by acquisition of both a disarmed Ti plasmid and a suitable binary vector. Tobacco, *Arabidopsis* and rice were infected by these bacteria and transgenic plants were obtained. Of the bacteria used, at least *S. meliloti* is competent to transfer genes into both dicot and monocot plants and into a range of tissues, including leaf tissue, undifferentiated calli, and immature ovules

(Broothaerts *et al.*, 2005). The results suggest that non-*Agrobacterium* species are capable of the full range of genetic transformation mechanisms shown by their *Agrobacterium* counterparts. This alternative approach to the *Agrobacterium*-mediated gene transfer method, in addition to affording an "open source" platform for plant biotechnology, may lead to new uses of natural bacteria–plant interactions for crop improvement.

Although in general it is desirable to have the transgene integrated into the nuclear DNA, in some cases the plastid genome may be an appropriate target for transformation. The advantage of plastid transformation includes high transgene expression levels, increased biosafety because of maternal inheritance of cytoplasmic genomes in most crops, and lack of gene silencing and position effects (Bock and Khan, 2004; Maliga, 2004). Transplastomic lines have been mostly produced by biolistic transformation, although direct gene transfer to protoplasts has also been utilized. The expression of a *Bacillus thuringiensis* (Bt) toxin gene in the tobacco plastid genome yielded high levels of the Bt toxin protein (3–5% of the total soluble protein) and produced plants with high-level resistance to herbivorous insects (McBride *et al.*, 1995). The expression and accumulation of the human growth hormone somatotropin in transgenic tobacco plastids reached 7% of total soluble protein, and demonstrated the capacity of chloroplasts to allow correct formation of disulfide bonds in a protein of eukaryotic origin (Staub *et al.*, 2000). Although several successful examples of plastid engineering have set a foundation for various future applications, the adaptation of plastid transformation protocols for major food crops has proved significantly more difficult than initially anticipated (Grevich and Daniell, 2005; Maliga, 2003). After the first successful transformation in tobacco, it took almost 10 years before plastid transformation was achieved in two other Solanaceae species, potato (Sidorov *et al.*, 1999) and tomato (Ruf *et al.*, 2001). Engineering of plastids offers great promise for the production of edible vaccines, antibodies, and other pharmaceutical proteins in plants.

C. Strategies for Overcoming Recalcitrance of Crop Species to Genetic Transformation

To date, most transformation procedures involve certain tissue culture steps, particularly callus culture. It is well known that callus induction and plant regeneration from the induced callus is not only time consuming and laborious but also causes somaclonal variation (Bregitzer *et al.*, 1998; Goldman *et al.*, 2004; Spangenberg *et al.*, 1998). Tissue culture-based methods also generally require considerable training of the practitioner to develop the skills needed to generate sufficient numbers of transgenic plants (Somers *et al.*, 2003). In addition, transformation frequency varies significantly with

the genotype used, even in extensively studied species such as wheat and maize. In many cases, commercial cultivars may be difficult to transform, and crossing the initial transgenic with an elite line, followed by significant backcrossing, will be needed for cultivar development.

Tremendous efforts have been made to overcome recalcitrance of crop species to genetic transformation. There have been numerous reports and significant progress on optimizing tissue culture and transformation parameters such as modifying media composition, growth regulators, and culture conditions, and identifying or manipulating more highly virulent *Agrobacterium* strains. Because the usefulness of the results is often limited to the species or even the genotype tested, it is impractical to summarize such optimization work here. However, it is worth noting that transformation efficacy can be significantly improved by minimizing tissue culture steps. The use of cotyledonary explants in white clover, soybean, and *M. truncatula* bypassed the callus formation phase and allowed direct regeneration from the infected explants; such a procedure at least partially solved the genotype dependence problem and allowed rapid production of transgenics (Larkin *et al.*, 1996; Olhoft *et al.*, 2003; Wright *et al.*, 2006). The use of stolons as explants in some grass species also bypassed callus formation and accelerated the process of plant regeneration (Ge *et al.*, 2006; Wang and Ge, 2005).

Another strategy to increase transformation frequency is to improve the tissue culture response. In the model grass plant *Lolium temulentum*, screening of a large number of genotypes revealed a few lines with relatively better callus induction frequency (Wang *et al.*, 2002). Crosses were made between the selected lines, and a significant improvement in tissue culture response of *L. temulentum* was achieved by the production of haploid and double haploid lines from anthers of F_2 plants of the crosses (Wang *et al.*, 2005c). By using the highly tissue culture responsive doubled haploid line, a large number of fertile transgenic *L. temulentum* plants were produced by *Agrobacterium*-mediated transformation (Ge *et al.*, 2007).

The most successful story in plant transformation is the development of the nontissue culture approach for *Arabidopsis*. Generation of transgenic lines by *in planta* transformation is simple and routine (Bent, 2000; Clough and Bent, 1998). The impact of the high throughput method on *Arabidopsis* research has been truly remarkable. Studies on the mechanism of transformation revealed that ovules are the primary target for *Arabidopsis in planta* transformation by the floral dip method (Bechtold *et al.*, 2000; Desfeux *et al.*, 2000; Ye *et al.*, 1999). On the basis of the lack of success to date, much time and effort will likely be needed to develop similar transformation methods for other species.

A different approach to overcoming recalcitrance to transforamtion is to understand in detail the molecular basis of the T-DNA transfer process from *Agrobacterium* to the plant genome. Although outside the scope of this

article, much progress has been made in this area, especially as regards the proteins that interact with the T-DNA during its transfer (Anand and Mysore, 2005; Gelvin, 2003a; Tzfira and Citovsky, 2006). Forward genetic screens have been performed in *Arabidopsis* to identify T-DNA-tagged lines that are resistant to *Agrobacterium*-mediated transformation (Gelvin, 2003a). As a result, several plant genes have been identified which, if over-expressed, increase transformation frequency. These include histone H2A-1, VIP1 (necessary from nuclear import of T-DNA), and a protein that interacts with the *Agrobacterium* VirB2 (T-pilus) protein. More details are provided elsewhere (Gelvin, 2003a,b; Tzfira and Citovsky, 2006). It has also been shown that the yeast Rad54 protein, which is involved in chromatin remodeling, improves transformation efficiency when expressed in *Arabidopsis* (Shaked *et al.*, 2005). These exciting results hold promise for the generation of high transformation efficient plant lines of many species that currently exhibit recalcitrance to transformation.

D. Transgenesis for Trait Integration and Commercialization

Over the last decade, transformed plants have moved from laboratory to the field, where new transgenic cultivars are grown in large acreages throughout the world. The adoption of transgenic crops has experienced double-digit growth rates every single year since biotech crops were first commercialized in 1996, with the number of biotech countries increasing from 6 to 21 in the same period (James, 2005). The global biotech crop area has seen a remarkable increase of more than 50-fold in the first decade of commercialization, with 90 million hectares planted in 2005. The accumulated global biotech crop area in its first decade was 475 million ha or 1.17 billion acres (James, 2005).

The United States has been the biggest adopter of transgenic crops, with 49.8 million hectares planted in 2005, which represent 55% of the global biotech area. By 2005, herbicide-tolerant soybeans accounted for 87% of total US soybean acreage, herbicide-tolerant cotton was planted on 60% of total cotton acreage, insect-resistant cotton accounted for 52% of cotton acreage, and insect-resistant corn was planted on 35% of the total acreage (Fernandez-Cornejo and Caswell, 2006).

Farmers continued to choose biotech crops due to significant benefits, including enhanced crop yields, improved insurance against pest problems, reduced pest management costs, decreased pesticide use, and overall increase in grower returns (Sankula, 2006). Planted acreage has mainly concentrated in the following trait–crop combinations: herbicide-resistant alfalfa, canola, corn, cotton, and soybean; insect-resistant corn, cotton, rice, and sweet corn; virus-resistant squash and papaya.

Obviously, in the first-generation transgenic crops, herbicide tolerance has consistently been the dominant trait, followed by insect resistance and virus

resistance. Some new cultivars have stacked genes for herbicide tolerance and insect resistance. The initial strategies for introducing single gene traits have been very successful in developing novel transgenic cultivars.

Current strategies for plant improvement have moved toward engineering more complicated traits, such as stress tolerance, yield potential, and growth rate. With rapid advances in functional genomics, many new genes have been discovered and functionally tested. Different approaches have been developed to manipulate complex traits or engineer metabolic pathways. In some cases, overexpression or disruption of a single gene can lead to the required phenotypic change; for example, drought or cold tolerance can be improved by the expression of a single TF gene (Gilmour *et al.*, 2000; Kasuga *et al.*, 2004; Zhang *et al.*, 2005), or lignin biosynthesis can be modified by down-regulation of a gene coding for one of the key enzymes in the pathway (Guo *et al.*, 2001a; Reddy *et al.*, 2005; Section VI.B.I). Other cases, such as the production of β-carotene in rice, require the introduction of multiple genes (Paine *et al.*, 2005; Ye *et al.*, 2000). The need for multigene transformations has long been cited as a negative factor for the development of metabolically engineered plants. However, this technical obstacle is gradually being overcome. Up to nine transgenes have now been incorporated simulatenously, into rice by a cotransforamtion strategy (Wu *et al.*, 2002), and into *Brassica juncea* by both stepwise engineering and through an actual nine-gene construct (Wu *et al.*, 2005). In the latter case, the genes formed a complete biosynthetic pathway to polyunsaturated fatty acids. In cases where constitutive expression of a particular transgene throughout plant growth and development has deleterious effects, a number of promoter systems are now available for chemically induced transgene expression (Tang *et al.*, 2004).

The number of encouraging scientific reports and the range of transgenic materials currently undergoing field testing are truly extensive. Transgenic technologies have proven utility for improving disease resistance, yield potential, abiotic stress (drought, cold, salinity, aluminum) tolerance, nutrient use efficiency, feed quality for animals, processing properties of biofuel crops, and nutritional quality (increased protein and oil content), for delaying ripening, for modifying starch content, and for producing nutraceuticals (vitamins, iron, β-carotene, flavonoids) and pharmaceuticals.

E. Virus-Induced Gene Silencing as an Alternative to Stable Transformation for Functional Genomics

The discovery of the importance of nonprotein-coding RNAs (ncRNAs) in the regulation of cellular processes has been one of the most important breakthroughs in genetics. One aspect of this field was recognized by the award of the 2006 Nobel Prize for Medicine to Craig Mello and Andrew Fire, for their pioneering work on gene silencing by RNA interference (RNAi),

a process which involves small RNA molecules to target the destruction of transcripts destined for turnover. In fact, it was research on plant systems that provided some of the first evidence for gene silencing (Jorgensen, 1995; Mueller *et al.*, 1995).

These discoveries are encapsulated in two practical approaches for the downregulation of genes: RNAi and virus-induced gene silencing (VIGS). As used in plants, RNAi usually involves stable transformation with a gene construct that, when expressed, produces a small double-stranded RNA homologous to a portion of the target gene sequence. This is usually generated via an inverted repeat of the short target sequence interrupted by a plant intron sequence (Wesley *et al.*, 2001). This approach has been widely used for modifying a number of plant traits through targeted downregulation of a specific gene or genes. Examples include engineering altered flower color (Ono *et al.*, 2006) and nutritional quality (Davuluri *et al.*, 2005).

VIGS takes advantage of an endogenous plant defense mechanism against virus infection which, in plants simply infected by the virus, targets the viral genome for degradation (Lu *et al.*, 2003). Virus-based vectors have been designed in which a small portion of the target gene sequence is included; the gene silencing process is then targeted against the corresponding host mRNAs. Although mechanistically similar to RNAi, VIGS has two major advantages over stable RNAi transformation in its high throughput and speed of the response. The one disadvantage is lack of universal application due to species specificity of suitable viral vectors. Most studies with VIGS have used *Nicotiana benthamiana* as host with a tobacco rattle virus (TRV)-based vector. *N. benthamiana* is universally susceptible to most viruses, and this may be because it lacks one component of the pathway for generating the small silencing RNA molecules (Yang *et al.*, 2004). *N. benthamiana* is a good model for other *Solanaceous* species, particularly tomato, although VIGS also works quite well in this species (Ryu *et al.*, 2004). A *Brome mosaic virus* (BMV)-based vector has been developed for VIGS applications in grasses such as tall fescue and rice (Ding *et al.*, 2006). This was based on a strain of BMV that was serendipitously discovered in a tall fescue breeding population (Mian *et al.*, 2005b).

In classical VIGS, the virus is simply physically inoculated onto the leaves. It is sometimes more efficient to introduce the virus vector via an *Agrobacterium*-based binary vector, and this process, called "agroinoculation," is usually performed through leaf infiltration with *Agrobacterium* harboring the necessary constructs (Dinesh-Kumar *et al.*, 2003). It was shown that the physical inoculation step can be avoided, and the soil adjacent to the plant roots is simply drenched with the *Agrobacterium* suspension containing the TRV-based VIGS vector (Ryu *et al.*, 2004). This "Agrodrench" technique provides a rapid approach for high throughput and large-scale analysis of gene function, but is currently limited to *Solanaceous* species (Ryu *et al.*, 2004).

F. TILLING as an Alternative to Transgenesis for Gene Knockdowns

TILLING has the unique advantage of allowing the generation of an allelic series for potentially any target gene. It can significantly expedite the crop improvement process. As it is a nontransgenic approach, resulting crop varieties are not subject to the strict regulatory approval process for transgenic crops. It has been used for improving the oil and protein content of soybean with the ultimate target of making allergen-free soybeans (Comis, 2005). The applicability of TILLING to soybean, maize, romaine and iceberg lettuce, tomato, rice, peanut, bread and durum wheat, and castor has been successfully demonstrated (Slade and Knauf, 2005). However, if a transformation system is available for a crop and only a few genes are targeted for knockout, RNAi is still the method of choice. Besides, RNAi has the advantages of knocking down the expression of multiple related genes with one construct (Lawrence and Pikaard, 2003).

VI. CASE STUDIES FOR ALFALFA IMPROVEMENT

A. Introduction

Alfalfa is the most widely used forage legume crop in the world today due to its high biomass yield (the record is over 18,000 kg ha^{-1} of forage); high protein, energy, vitamin, and mineral feed quality for livestock; ability to fix atmospheric nitrogen; wide adaptation to various environments; improvement of soil composition when used as a rotation crop in sustainable agricultural systems; utility as a model system for genetic studies of autotetraploid species; and ease of use with the new biotechnologies (Bouton, 2001).

The primary center of origin for the genus *Medicago* is found in the Caucasus, northwestern Iran, and northeastern Turkey. *M. sativa* is a complex of several perennial subspecies, both diploids and tetraploids that are interfertile and possess a similar karyotype (Quiros and Bauchan, 1988). *M. sativa* ssp. *sativa*, *M. sativa* ssp. *falcata*, and *M. sativa* ssp. *glutinosa* are tetraploid subspecies while *M. sativa* ssp. *coerulea* (progenitor to cultivated alfalfa) and some *M. sativa* ssp. *falcata* are diploids. Cultivated alfalfa (ssp. *sativa*) is an autotetraploid with $2n = 4x = 32$ (Stanford, 1951).

Alfalfa genetics are complex because of the plant's autotetraploid nature (Stanford, 1951) and an allogamous breeding system that does not tolerate inbreeding. The development and use of molecular markers is limited due to the difficulty of resolving allele dosage and linkage phases in autotetraploids. For this reason in the past, genetic linkage maps have been developed in

diploid forms of the *M. sativa* species complex (Brummer *et al.*, 1993; Echt *et al.*, 1993; Kalo *et al.*, 2000; Kiss *et al.*, 1993; Tavoletti *et al.*, 1996). However, the utility of such diploid genetic maps in the breeding of tetraploid alfalfa depend on high synteny across the ploidy levels. In addition to serving as an important framework genetic linkage map in *Medicago* species, the diploid maps are also useful in transferring unique genes from the diploid level to the cultivated, tetraploid level. The widespread natural occurrence of restitutional 2*n* gametes (i.e., gametes with the somatic chromosome number) in the *Medicago* sp. (Bingham, 1968; Harlan and deWet, 1975; Stanford *et al.*, 1972) provided early support for the hypothesis that gene flow across different ploidy levels occurs continuously and naturally via 2*n* pollen. Such gene transfer via the restitutional gametes can aid in transferring valuable traits from diploid relatives into cultivated alfalfa in a breeding program (Bingham, 1980).

The complexities of alfalfa genetics are of less concern for initial trait insertion through transgenic approaches, but do impact segregation frequencies during subsequent introgression into elite cultivars.

B. Improvement of Aluminum Tolerance

1. Introduction

The use of alfalfa is mainly confined to the temperate areas of the world and not the tropics. There are several reasons for this, but the main one is an inability to tolerate acid, aluminum toxic soils that are widespread throughout the tropics (Bouton, 2001).

Acidity is common in soils where rainfall is high enough to leach appreciable amounts of exchangeable bases from the soil surface layers (Brady, 1974). This leaching effectively removes the buffering capacity of the soil and causes a drop in pH. Leaching also encourages acidity by allowing percolation of organic acids derived from naturally decomposing organic matter into the soil profile to replace the bases which are then removed by the drainage water.

Under very acid conditions, Al becomes soluble in soil and is present in the toxic Al^{3+} or $Al(OH)^{2+}$ forms (Brady, 1974). These then become adsorbed, even in preference to hydrogen ions, to clay minerals, with the adsorbed Al coming into equilibrium with the Al ions in the soil solution. The latter also contribute to overall soil acidity. When soil pH is moved toward neutrality with liming, the toxicity of Al is suppressed by changing to less toxic forms such as Al(OH).

Al toxicity occurs by definition when the ratio of extractable Al (found in the toxic forms at low pH) to extractable Al plus exchangeable Ca, Mg, and

K is greater than 60% within 50 cm of the soil surface. On the basis of this definition, Al toxicity is estimated to be present in 56% of the soils in the humid tropics (Buol and Eswaran, 1993). Some general level of Al tolerance will be necessary in most crops if these extensive areas are to be brought into some level of productivity.

The most common effect of Al on plant growth is the reduction of root elongation and proliferation, thereby leading to poor water and nutrient extraction (Buol and Eswaran, 1993). Exposure of plants to Al-toxic conditions causes inhibition of cell division at the root apex resulting in stunting of primary roots and inhibition of lateral root formation (Ryan *et al.*, 1993; Sivaguru and Horst, 1998). Al-sensitive plants are thus impaired in nutrient and water uptake, and tend to be drought susceptible with reduced crop yield and quality. Application of expensive soil amendments such as lime and organic acids raises the pH and converts Al into less toxic forms. However, even where liming is practiced, subsoils remain acid and Al-toxic. A cost-effective alternative is growing Al-tolerant cultivars in problem soils combined with soil amendments (Foy, 1988).

2. Marker-Assisted Breeding for Al Tolerance in Alfalfa

Screening and selecting cultivated alfalfa for tolerance to acidic, Al-containing soil has been reported (Baligar *et al.*, 1989; Bouton, 1996; Dall'Agnol *et al.*, 1996). Nevertheless, there is no *M. sativa* subsp. *sativa* cultivar or plant introduction that does not suffer a decline in performance under acid conditions. Conventional breeding to develop Al-tolerant germplasm in alfalfa is also limited (Bouton and Parrott, 1997). However, two genomic regions associated with Al tolerance were identified in a diploid (*M. sativa* subsp. *coerulea*) genotype using RFLP markers in conjunction with a callus growth bioassay by single marker analysis (Sledge *et al.*, 2002). A study was conducted to identify SSR markers that flank these QTLs using *M. truncatula* EST-SSR markers (Section IV.C.1 above) and also to identify additional Al-tolerance QTLs in a backcross population derived from the cross between Al-sensitive and Al-tolerant genotypes of *M. sativa* subsp. *coerulea* (Narasimhamoorthy *et al.*, 2007).

The ultimate goal underlying QTL mapping is often to identify the specific genes responsible for phenotypic variation. One method for doing this is the placement of candidate genes associated with a desirable phenotype from other species on to genetic maps to look for coincidence of map position. The *M. truncatula* EST and genome databases were mined to identify DNA sequences with high homology to Al tolerance genes identified in other plant species, to be used as candidate genes for genetic mapping in diploid alfalfa (Narasimhamoorthy *et al.*, 2007). Fifteen candidate genes selected for

candidate gene mapping included those that coded for proteins responsible for organic acid synthesis, genes involved in signal transduction, and genes that code for enzymes that alleviate oxidative stress. Evidence from other crop species supports Al-activated release from roots of carboxylates such as citric and malic acids as a major resistance mechanism (Kochian *et al.*, 2005). An intron-targeted mapping strategy was adopted for two specific genes involved in Al-activated root carboxylate release, namely citrate synthase (CS) and malate dehydrogenase. Six candidate gene markers designed from *M. truncatula* ESTs that showed homology to known Al-tolerance genes identified in other plant species were placed on the QTL map. Three putative QTLs on linkage groups LG I, LG II, and LG III, explaining 38%, 16%, and 27% of the phenotypic variation, respectively, were identified. A marker designed from a candidate gene involved in malic acid release mapped near a marginally significant QTL on LG I.

In order to move the Al tolerance QTL from the diploid (2x) *M. sativa* subsp. *coerulea* genotype to the cultivated tetraploid (4x) *M. sativa* subsp. *sativa*, 2x-4x crosses were made, using the diploid as the seed parent, and the tetraploid as the pollen parent. Alfalfa is essentially a bivalent forming autotetraploid where regular meiotic stages predominate in normal plants. Whereas the predominating gametes are mostly normal, restitutional $2n$ gametes occur at a low frequency in diploid alfalfa plants and are usually functional in fertilization events that involve tetraploid forms of cultivated alfalfa. The 2x-4x hybrids were genotyped to confirm the presence of markers linked to Al tolerance QTLs and phenotyped for tolerance to Al stress for further confirmation of their tolerance (Fig. 4A). These tetraploid hybrids were backcrossed to selected genotypes from the nondormant alfalfa cultivar "CUF 101." The BC_1F_1 plants were genotyped to select plants carrying the markers for Al tolerance QTLs which can be further backcrossed to elite clones. In addition to the markers linked to Al tolerance QTLs, a genome scan approach that randomly selects genetic markers spread over each linkage group to select for the cultivated alfalfa background should reduce the time involved in rigorous PS for cultivated alfalfa phenotypes during further backcrossing to introgress the Al tolerance QTLs. A synthetic population can be developed after 3–4 backcrosses to select plants that carry the Al tolerance QTL from the diploid alfalfa into the elite tetraploid cultivar background. These can then be tested in both greenhouse and field for acid soil tolerance.

3. A Transgenic Approach to Al Tolerance in Alfalfa

As outlined earlier, Al-induced secretion of organic acids from the roots has been proposed as a mechanism for Al tolerance (Delhaize and Ryan, 1995; Kochian, 1995; Ma, 2000). Two general patterns of Al-stimulated efflux of organic acids have been reported. In Pattern I, no discernible

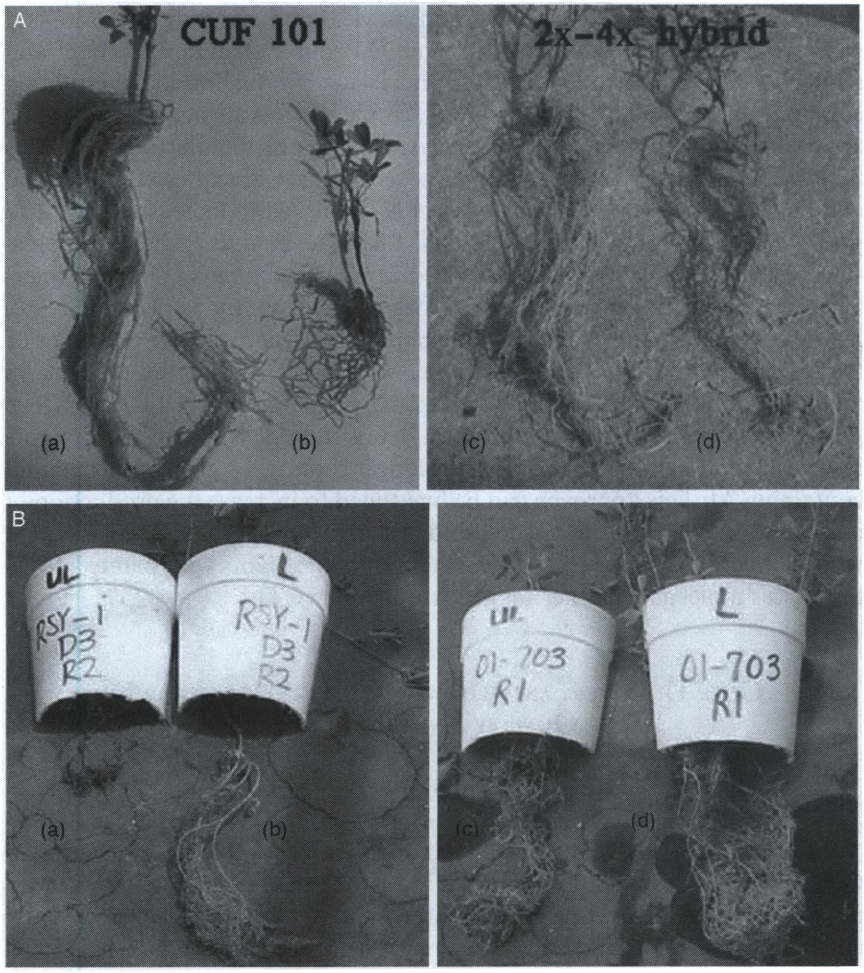


Figure 4 Improvement of aluminum tolerance in alfalfa. (A) The root systems of 8-week-old plants of different alfalfa clones obtained through marker-assisted breeding. (a) CUF 101 in limed soil, (b) CUF 101 in unlimed soil, (c) 2x-4x hybrid in limed soil, and (d) 2x-4x hybrid in unlimed soil. (B) Root systems of control and transgenic alfalfa lines after 8-week growth. (a) Regen-SY (nontransformed) in unlimed soil, (b) Regen-SY in limed soil, (c) Regen-SY transformed with a *CS* gene in unlimed soil, and (d) Regen-SY transformed with a *CS* gene in limed soil. (See Color Insert.)

delay is observed between the addition of Al and the onset of organic acid release, suggesting that Al activates a preexisting mechanism without a need for induction of novel proteins. In Pattern II, organic acid secretion is delayed for several hours after exposure to Al^{3+}, indicating that protein induction is required (Ma *et al.*, 2001).

Alfalfa is very sensitive to Al^{3+} and its yield and stand duration in acid soils are compromised due to both inhibited root system development and reduced symbiotic nitrogen fixation. Increasing production of Al-chelating compounds, particularly organic acids, in plant roots through a transgenic approach could enhance tolerance to Al toxicity. In the first report of this approach, a *CS* gene from *Pseudomonas aeruginosa* was expressed in tobacco and papaya plants (da la Fuente *et al.*, 1997); this led to increased citric acid production and tolerance to Al toxicity. Likewise, in alfalfa, the overexpression of malate dehydrogenase (MDH) in transgenic plants enhanced tolerance to Al toxicity through increased organic acid synthesis (Tesfaye *et al.*, 2001). In a similar approach, the *Arabidopsis ACT2* constitutive promoter or the tobacco *RB7* root-specific promoter were used to drive the *CS* gene in alfalfa (Rosellini *et al.*, 2003). Transgenic plants expressing the *CS* gene possessed better root growth and total dry matter yield than control plants in Al-toxic soils (Fig. 4B), and also had longer roots when grown in a medium containing Al. This approach can be fine-tuned to increase the production of citric acid in specific root tissues, and the *CS* transgene can be pyramided with Al tolerance QTLs by crossing the transgenic plants with plants known to possess molecular marker-tagged Al tolerance QTLs.

C. Gene Discovery and Metabolic Engineering for Forage Quality Enhancement

1. Introduction

Forage quality is a major but complex trait for plant improvement. Generally speaking, quality decreases as plants mature and enter the flowering stage, primarily as a result of lignification of secondary cell walls (Jung and Vogel, 1986). Other quality traits include protein content and amino acid composition, protein and nutrient bioavailabilty, presence of chemical antifeedants, and potential for causing pasture bloat. We here describe studies designed to better understand, and improve, forage quality traits in alfalfa. This provides another illustration of how some of the postgenomics technologies and resources described earlier, in this case those derived for/from *Arabidopsis* and *M. truncatula*, can be applied for variety improvement.

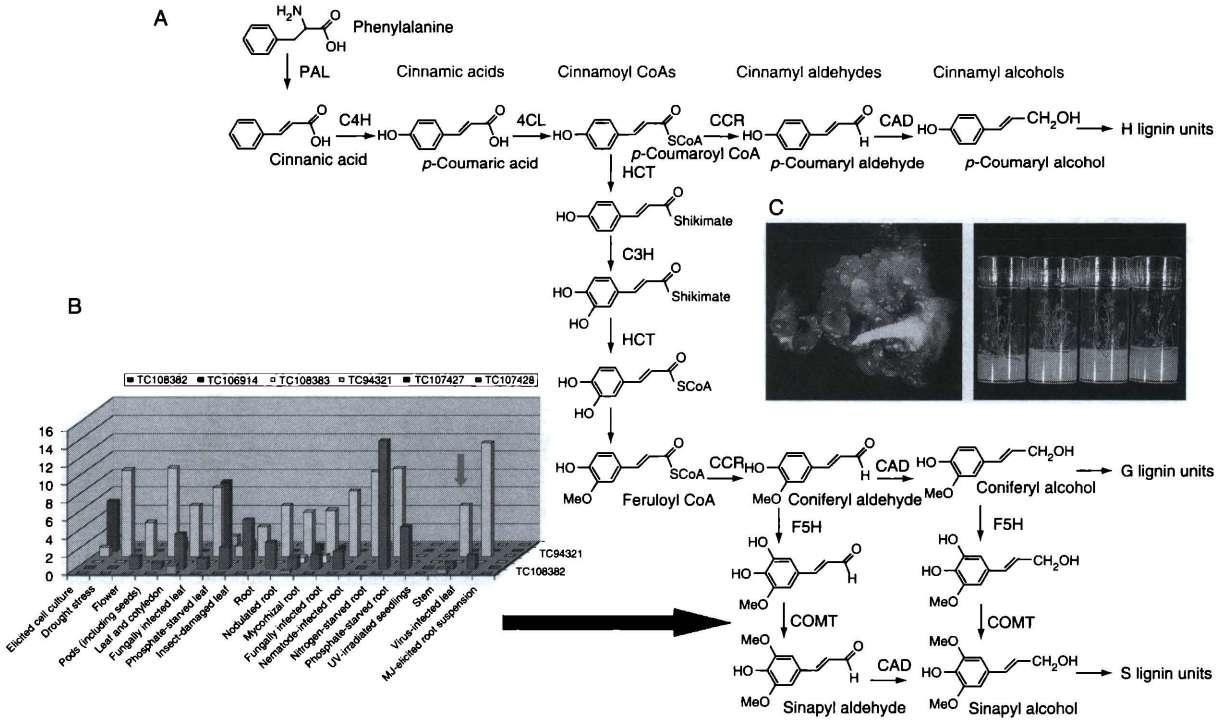
2. Improved Forage Digestibility

Feeding and grazing studies have shown that only small changes in forage digestibility can have significant effects on animal performance (Casler and Vogel, 1999). Improving digestibility is therefore an important goal of forage

breeding programs, and has, in the past, been addressed primarily through standard approaches of crossing and selection (Jung *et al.*, 1994). At the same time, this research has led to an understanding of relationships between various forage quality parameters, such as neutral detergent fiber (NDF), acid detergent fiber (ADF) and acid detergent lignin (ADL), and forage quality (Jung, 1997; Jung *et al.*, 1997). There is a general consensus that high lignin levels reduce digestibility (Jung and Deetz, 1993). However, most of the studies from which this conclusion is based used materials with different lignin contents and/or compositions as a result of divergent selection for forage quality traits (Méchin *et al.*, 2000), natural genetic variation of plant accessions (Casler, 1987), delignification (Jung *et al.*, 1992), or different maturity of plant tissues (Reeves, 1987). The results of such studies will be complicated by the many uncontrolled developmental and genetic variables that could potentially affect digestibility (Titgemayer *et al.*, 1996). Because alfalfa is an outbreeding autotetraploid, inbred lines are not available. Generating isogenic transgenic lines in which lignin content or composition are modified by altering expression of a target gene in the lignin pathway provides both a new approach to trait improvement and a strategy for better elucidating the lignin/digestibility relationship in alfalfa. The value of such an approach, for both basic and applied research, becomes more apparent when considering the potential variations that can exist in lignin structure.

Lignin is a polymer of hydroxylated and methoxylated phenylpropane units (monolignols) linked via oxidative coupling (Boudet *et al.*, 1995). There is a vigorous ongoing debate as to the extent, or lack, of orderliness in the polymer (Davin and Lewis, 2005; Ralph *et al.*, 2007), with the prevailing view being that lignin is assembled by a relatively random free radical-mediated process, that is nevertheless under developmental control at the level of substrate supply (Boerjan *et al.*, 2003). Angiosperm lignin contains two major monolignols, mono-methoxylated guaiacyl (G) and di-methoxylated syringyl (S) units, polymerized through at least five different linkage types (Boerjan *et al.*, 2003). It also contains low levels of *p*-hydroxyphenyl, or H units (Fig. 5). In many forage crops, lignin content and S/G ratio increase with stem maturity (Buxton and Russell, 1988; Jung and Vogel, 1986), and both content and S/G ratio therefore correlate negatively with forage digestibility in ruminant animals (Albrecht *et al.*, 1987; Buxton and Russell, 1988; Grabber *et al.*, 1992; Jung *et al.*, 1997; Sewalt *et al.*, 1996). However, the relationship between lignin composition and digestibility is far from clear, since the amount of G lignin has also been linked with reduced cell wall degradability in forages (Jung and Deetz, 1993), and studies with synthetic lignins (Grabber *et al.*, 1997) have yielded results that question effects of lignin composition alone on forage digestibility.

The biosynthesis of the monolignol building blocks of lignin is believed to proceed essentially according to the pathway in Fig. 5 (Hoffmann *et al.*, 2003;


A
Phenylalanine
PAL
Cinnamic acids
Cinnamoyl CoAs
Cinnamyl aldehydes
Cinnamyl alcohols
C4H
4CL
CCR
CAD
Cinnanic acid
p-Coumaric acid
p-Coumaroyl CoA
p-Coumaryl aldehyde
p-Coumaryl alcohol
H lignin units
HCT
Shikimate
C3H
Shikimate
HCT
SCoA
Feruloyl CoA
CCR
CAD
Coniferyl aldehyde
Coniferyl alcohol
G lignin units
F5H
F5H
COMT
COMT
CAD
Sinapyl aldehyde
Sinapyl alcohol
S lignin units
B
TC108382
TC106914
TC108383
TC94321
TC107427
TC107428
16
14
12
10
8
6
4
2
0
TC94321
TC108382
Elicited cell culture
Drought stress
Flower
Pods (including seeds)
Leaf and cotyledon
Fungally infected leaf
Phosphate-starved leaf
Insect-damaged leaf
Root
Nodulated root
Mycorrhizal root
Fungally infected root
Nematode-infected root
Nitrogen-starved root
Phosphate-starved root
UV-irradiated seedlings
Stem
Virus-infected leaf
MJ-elicited root suspension
C

Humphreys and Chapple, 2002; Humphreys *et al.*, 1999; Nair *et al.*, 2004; Schoch *et al.*, 2001), although it is possible that there are variations between species, especially as regards pathway regulation and whether linear or parallel pathways exist for G and S lignin synthesis (Chen *et al.*, 2006b; Parvathi *et al.*, 2001). As a result of the genome-sequencing projects outlined in Section II, and a number of EST programs in other species, genes encoding all of the enzymes in Fig. 5 have been identified from representative monocots and dicots. Most important for forage species are the EST collections from *Medicago*, tall fescue, and perennial ryegrass.

Because of the very high sequence identity between orthologous genes in *M. truncatula* and *M. sativa* (Aziz *et al.*, 2005), the genetic resources from *M. truncatula* have proven very useful for identifying lignin pathway gene sequences that can be applied directly for genetic modification of lignin in alfalfa. An example is provided in Fig. 5. Searching the DFCI MtGI (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=medicago) reveals five TC sequences whose BLAST annotation suggests that the gene might encode a caffeic acid 3-*O*-methyltransferase (COMT), the enzyme that carries out the final methylation step in the formation of S lignin (Fig. 5). By counting the number of individual ESTs corresponding to each TC in each of the more than 60 cDNA libraries sequenced to date, it is possible to obtain an estimate of the relative degree of expression of each of the TCs in different tissues (such an approach is called an "in silico Northern," after the "Northern" blot hybridization technique for measuring transcript levels). Similar information could be obtained from microarray analysis of specific tissue types. It is clear from Fig. 5B that only TC 94321 is strongly represented in stem tissue. This TC is therefore most likely the true COMT involved in lignification. The same approach was taken to identify *M. truncatula* genes encoding the three cytochrome P450 enzymes of the lignin pathway, namely cinnamate 4-hydroxylase (C4H), coumaroyl shikimate 3-hydroxylase [also known as coumarate 3-hydroxylase (C3H)] and coniferaldehyde 5-hydroxylase [also known as ferulate 5-hydroxylase (F5H)] (Reddy *et al.*, 2005). These three

Figure 5 Application of genomics/transgenesis to lignin modification in alfalfa. (A) Currently accepted model of the lignin biosynthetic pathway. Enzymes are: PAL, L-phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CCR, cinnamoyl CoA reductase; CAD, cinnamyl alcohol dehydrogenase; HCT, hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyl transferase; C3H, "coumarate 3-hydroxylase;" CCoAOMT, caffeoyl CoA 3-*O*-methyltransferase; F5H, "ferulate 5-hydroxylase;" COMT, "caffeic acid 3-*O*-methyltransferase." (B) cDNA library-specific EST counts for all TCs annotated as encoding caffeic acid *O*-methyltransferase. Note that only one TC is strongly expressed in stems, the major site of lignification. This sequence was therefore chosen for antisense and RNAi-mediated downregulation, and the corresponding vectors introduced into alfalfa by *Agrobacterium*-mediated transformation and regeneration via somatic embryogenesis (C) (Chen *et al.*, 2006b). (See Color Insert.)

enzymes catalyze strategically placed reactions in the formation of all monolignols, G-units, and S-units, respectively (Fig. 5).

The close sequence identity between orthologous genes in the two closely related *Medicago* species makes it possible to make gene constructs using *M. truncatula* sequences to target alfalfa genes for downregulation. Downregulation of C4H, C3H, or F5H in alfalfa therefore used the corresponding *M. truncatula* sequences expressed in the antisense orientation (Reddy *et al.*, 2005). The transgenes were driven by the bean phenylalanine ammonia-lyase PAL2 promoter (Liang *et al.*, 1989), which is expressed in most of the vascular tissues of alfalfa (Guo *et al.*, 2001a). Transformation used efficient methods based on cocultivation of leaf discs with *Agrobacterium* followed by regeneration via somatic embryogenesis (Samac and Temple, 2006) (Fig. 5C). Plants with validated reductions in target transcript and/or enzyme activity levels had either reduced lignin levels with relatively normal lignin composition (C4H transgenics), lignin rich in *p*-hydroxyphenyl (H) units (C3H transgenics), or lignin rich in G units with reduced S content (F5H transgenics) (Reddy *et al.*, 2005). Previous studies had used similar antisense technology to downregulate (COMT) and/or caffeoyl CoA 3-*O*-methyltransferase (CCoAOMT) in the same alfalfa genetic background (Guo *et al.*, 2001a). COMT downregulation reduced both lignin content and S/G ratio, whereas lignin content was reduced, but S lignin levels remained unaltered, in CCoAOMT downregulated plants (Guo *et al.*, 2001a).

The availability of sets of transgenic alfalfa plants with various combinations of altered lignin content and composition allowed a determination to be made of the relative importance for forage digestibility of altered lignin content or composition in the same genetic background. Plants were grown to the early bud stage, harvested, and analyzed for a number of forage quality parameters, including *in situ* digestibility in the rumens of fistulated steers (Guo *et al.*, 2001b; Reddy *et al.*, 2005). These studies clearly indicated that lignin content, rather than composition, impacted digestibility, with the greatest improvement in digestibility (up to 15% in plants downregulated in C3H) being observed with the plants with the most reduced lignin levels (Reddy *et al.*, 2005). Earlier studies, in which transgenic forage had been analyzed in fistulated sheep, also indicated that downregulation of CAD improved digestibility in alfalfa (Baucher *et al.*, 1999), although to a lesser extent than with the transgenes described earlier.

Although downregulation of C3H, and the enzyme preceding it (HCT), gives the largest digestibility improvements in alfalfa, there is no free lunch here, as these particular transgenic plants suffer from yield depression (Reddy *et al.*, 2005; Shadle *et al.*, 2007). The reason for this is not totally clear, although distorted vascular tissues are observed in the HCT lines (Shadle *et al.*, 2007), suggesting that water relations may be disturbed. The digestibility increases in COMT and CCoAOMT lines of around 5% through

a single transformation event, but not associated with negative agronomic performance, represent an economically beneficial transgenic improvement. Such lines have now been extensively field tested with a view to commercialization (Temple *et al.*, 2004). Additional research may lead to an understanding of the reduced growth phenotype of plants with strong lignin downregulation, such that this can be ameliorated either by targeting the lignin reduction to a more narrow set of cell types, or through introduction of compensatory mechanisms. In both cases, it is likely that postgenomics approaches will hold the key, by providing either promoters with a more specific cell type expression or high throughput platforms for trait identification.

3. Bloat Resistance

Although this may appear paradoxical in view of the above discussion of improving biomass digestibility in alfalfa, the high protein content of alfalfa can cause problems for ruminant animals because it is too rapidly digested by rumen microogransims (Marshall *et al.*, 1980). This leads to: (1) excess methane production in the rumen, exacerbated by foaming caused by a combination of the high protein content and presence of other agents such as saponins and leading to the condition known as pasture bloat; (2) increased urinary nitrogen excretion; and (3) reduced levels of "by-pass" protein not exiting the rumen and therefore not contributing to the nitrogen nutrition of the animal. Designing a "bloat safe alfalfa," with these other additional benefits, has been a major goal for alfalfa breeders (Coulman *et al.*, 2000). Studies in sheep have demonstrated significant improvements in performance and reduction in bloat, if the animals are fed forages contain reasonable levels of flavonoid polymers known as condensed tannins [CTs, also called proanthocyanidins (Aerts *et al.*, 1999)]. CTs bind to proteins and reduce their rate of microbial degradation.

In laboratory studies, treatment of feed proteins with modest amounts of CTs (around 2–4% of dry matter) reduced both proteolysis during ensiling and rumen fermentation. In studies with sheep, increasing dietary CTs (from only trace amounts to 4% of dry matter) increased by-pass protein, and a diet containing only 2% CT increased absorption of essential amino acids by the small intestine (Douglas *et al.*, 1999). Low concentrations of CTs also help counter intestinal parasites in lambs, and, as described earlier, confer bloat safety (Aerts *et al.*, 1999). Levels of CTs for bloat reduction are at the lower end of the range needed to significantly improve the nitrogen nutrition of the animal (Li *et al.*, 1996). The above properties of CTs are the main driving force behind efforts to genetically introduce the CT pathway into forage crops (Aerts *et al.*, 1999; Reed, 1995). However, high concentrations of

CTs (from 6% to 12% of dry matter) reduce the palatability of forages, and can negatively impact nutritive value, including digestibility, by directly binding bacterial enzymes or forming complexes with cell wall polysaccharides and thereby reducing their accessibility to degrading enzymes (Aerts *et al.*, 1999; Reed, 1995; Smulikowska *et al.*, 2001).

Alfalfa is a bloat-causing forage because its aerial portions do not contain measurable levels of CTs. These compounds do, however, accumulate to quite high levels in the seed coat (Koupai-Abyazani *et al.*, 1993). Thus, alfalfa contains all the genes necessary for CT biosynthesis. The trick is to express these genes ectopically in the aerial portions of the plant. To identify the necessary genes for engineering CTs in alfalfa, it is first important to consider the chemical structures of these compounds.

CTs are oligomeric and polymeric end products of the flavonoid biosynthetic pathway (Fig. 6). The past 3 years have seen important breakthroughs

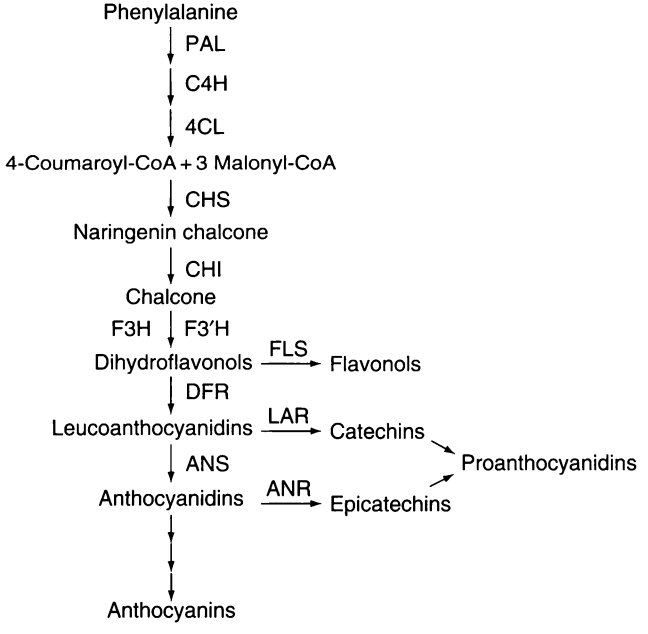


Figure 6 An overview of the biosynthesis of anthocyanins and condensed tannins. Enzymes are: ANR, anthocyanidin reductase or BANYULS; ANS, anthocyanidin synthase (LDOX leucoanthocyanidin dioxygenase); CHI, chalcone isomerase; CHS, chalcone synthase; 4CL, 4-coumarate:CoA ligase; C4H, cinnamate 4-hydroxylase; DFR, dihydroflavonol reductase; F3H, flavonoid-3-hydroxylase; F3′H, flavonoid-3′-hydroxylase; FLS, flavonol synthase; LAR, leucoanthocyanidin reductase; PAL, phenylalanine ammonia-lyase. The PAP1 myb TF controls expression of the genes encoding the enzymes of the anthocyanin pathway from PAL to ANS in *Arabidopsis*.

in our understanding of the biosynthesis of the building blocks of CTs, the flavan-3-ols (+)-catechin, and (−)-epicatechin (Fig. 6; Dixon *et al.*, 2005; Tanner *et al.*, 2003; Xie *et al.*, 2003). However, virtually nothing is known about the ways in which these units are assembled into the corresponding oligomers *in vivo* (Xie and Dixon, 2005). Molecular genetic approaches are leading to an understanding of the regulatory genes that control CT biosynthesis, and this information, together with the increased knowledge of the enzymes specific for the pathway, will facilitate the genetic engineering of plants for introduction of value added forage quality traits.

The major discovery engine for the genes specific for CT biosynthesis and its control has been the use of forward genetics in *Arabidopsis*. Interruption of CT biosynthesis at any stage results in the formation of a transparent testa (tt) phenotype. Many *tt* mutants of *Arabidopsis* have now been characterized, and the cloned genes that had been disrupted comprise biosynthetic enzymes of the CT pathway, TFs controlling both the pathway and endothelial cell development, transporters, a proton pump, and an oxidase (Lepiniec *et al.*, 2006).

On the basis of our results, at least three genes appear necessary for introducing the CT pathway into tissues that do not naturally make these compounds: these encode MYB family TFs functionally orthologous to *Arabidopsis* PAP1, which, when ectopically expressed, leads to massive accumulation of anthocyanin pigments (Borevitz *et al.*, 2000); the MYB family TF TT2 (Nesi *et al.*, 2001), which appears to regulate genes encoding late steps in CT biosynthesis; and the enzyme anthocyanidin reductase (ANR), which is encoded by the *BANYULS* gene of *Arabidopsis*, and converts anthocyanidins into their corresponding 2,3-*cis*-flavan-3-ols (e.g., cyandin to (−)-epicatechin, Fig. 6; Xie *et al.*, 2003, 2004). PAP1 was found by a T-DNA activation tagging approach (Borevitz *et al.*, 2000), and was one of the first genes to be discovered in this way in view of the obvious purple-red phenotype of plants overexpressing this gene. Discovery of the function of the *Arabidopsis* BANYULS gene, by analysis of the catalytic activity of the recombinant protein expressed *in vitro*, was soon followed by the isolation of the functional orthologue from *Medicago*, by utilizing EST information from a cDNA library representing transcripts from developing seeds (Xie *et al.*, 2004).

Expressing PAP1, TT2, and ANR together appeared insufficient to allow for constitutive accumulation of CTs in *Arabidopsis* leaves and stems (Sharma and Dixon, 2005). In contrast, coexpression of *Arabidopsis* PAP1 and *Arabidopsis* or *Medicago* ANR leads to production of CTs in tobacco leaves and flowers, at levels that would be protective for bloat if they were in alfalfa (Xie *et al.*, 2006).

One limitation to moving this technology directly to alfalfa is the finding that *Arabidopsis* PAP1 does not appear to function well in legumes (G. J. Peel

and R. A. Dixon, unpublished results). This problem can be partially circumvented in *M. truncatula*, the leaves of which contain a central red spot rich in anthocyanins; simply expressing the *ANR* gene in these leaves therefore results in a detectable accumulation of CTs (Xie *et al.*, 2006). We have identified two novel MYB TFs by informatic analysis of the *M. truncatula* genome sequence, and these confer a strong purple pigment phenotype when transformed into alfalfa, *M. truncatula*, or clover (G. J. Peel, E. Wright, Z. Y. Wang, and R. A. Dixon, unpublished results). Thus, we believe that the development of a bloat-safe alfalfa by tannin engineering will soon become a reality. Previous studies have hinted at the potential for accumulating tannins in alfalfa foliage after transformation with flavonoid pathway TFs, but this accumulation required that the plants were placed under stress conditions such as cold or high light intensity (Ray *et al.*, 2003).

D. Issues for Molecular Development of Alfalfa

In self-pollinated crops, introgressing desirable exotic alleles from wild to cultivated backgrounds with the aid of molecular markers is a straightforward process. However, in tetraploid alfalfa with no extant inbred lines and an autoteraploid nature, such an effort presents a much more complicated challenge. Although QTLs and transgenes for Al tolerance, and transgene sets for lignin modification and introduction of condensed tannins, are currently being introgressed into elite cultivar backgrounds, their deployment requires estimation of the QTL or transgene effects in commercial breeding populations. In particular, the marker and QTL effects must be estimated on a regular basis to improve accuracy and to guard against unfavorable associations with other traits and against epistatic effects with the background genome or environment. The deployment of transgenes also requires further research to ensure stable expression and understand the effect of the transgenes in combination with the QTLs. These considerations are, of course, apart from the evolving web of patent and regulatory issues.

VII. THE FUTURE: BRIDGING THE GAP FROM MODELS TO CROPS

A long-range goal of translational genomics is to utilize bioinformatics to leverage genomic information from model and reference organisms for the economic benefit of important crop species. Given the abundant genomic

information available for model and crops species, it is perhaps surprising that a significant gap still exists between the wealth of available knowledge about genome structure and content and its utilization by plant breeders whose programs are directed toward agronomic improvement of crop species. For example, an informal survey at the 2006 International Plant Breeding Symposium revealed that very few of the breeders polled access plant genomics resources. What are required are intuitive, web-based tools that enable plant breeders to retrieve genomic, genetic, and phenotypic data and information that is relevant to them. An example of such an online database is the Soybean Breeder's Toolbox (http://soybeanbreederstoolbox.org/) that allows exploration of the genomic resources through easily retrieved information. The toolbox provides information about molecular markers on genetic maps, diseases and pests that damage soybean crops, and data associated with soybean quantitative traits such as the resistance of different soybean genotypes to biotic and abiotic stresses.

Using comparative genomics, information from model plant species can accelerate the discovery of genes responsible for disease and pest resistance, tolerance to plant stresses such as drought, and enhanced nutritional value including production of antioxidants and anticancer compounds. A sequenced and annotated genome can accelerate the identification of candidate genetic loci underlying phenotypes of interest. Because sequence and function of genes are largely conserved among related species, comparative genomics can leverage information and knowledge gained from a sequenced model, or reference species, to make hypotheses about the relationship between genotype and phenotype for related species. With the exception of *Medicago*, *Lotus*, and now soybean, crop legumes have not been sequenced because their large and complex polyploid genomes make genome sequencing endeavors cost-prohibitive. In regions of the genome where syntenic relationships exist between *Medicago* or *Lotus* and a crop legume, the annotated genomic sequence from these species can be leveraged to identify candidate genes of interest in other legumes.

The following is an example of how the Legume Information System (LIS; http://www.comparative-legumes.org) (Gonzales *et al.*, 2005) can be used to find candidate genes for sudden death syndrome (SDS) in soybean via the *Medicago* genome. SDS, caused by *Fusarium solani f.* sp. *glycines*, creates toxins in the roots resulting in root rot and leaf scorch that severely reduces soybean production each year. SDS is a major concern and has become the focus for breeders and scientists interested in producing a more resistant soybean plant. QTLs for SDS in soybean have been previously identified and mapped (Njiti *et al.*, 2002). Once QTL regions have been located, the actual genetic elements responsible for the phenotype may perhaps be identified. Specifically, genetic maps, physical maps, and annotated TC and EST

sequences from soybean and *M. truncatula* can be compared. In addition, the recently published *M. truncatula* genomic sequences can be used to identify *M. truncatula* candidate genes in a genomic region syntenic to a QTL region for SDS in soybean. Genomic sequences of candidate genes from *M. truncatula* can then be used to identify ESTs with sequence similarities from soybean for DNA marker development and cloning of potential soybean disease-causing alleles.

By using the CMap module of LIS to query and display soybean SDS QTLs on genetic maps, the soybean linkage maps are compared to *M. truncatula* maps to identify syntenic regions containing SDS QTLs. Once genomic markers in *M. truncatula* have been identified as syntenic to the SDS QTL region, the *M. truncatula* physical maps are used to identify the sequenced genomic clones within comparable regions. These genomic sequences within the physical region are then analyzed for candidate genes using annotations displayed in the LIS Comparative Functional Genomics Browser (Fig. 7). Finally, consensus sequences aligned to genomic sequence can be analyzed using the existing annotations to isolate candidate soybean EST sequences that may confer SDS resistance in soybean.

VIII. THE FUTURE TECHNOLOGIES, OPPORTUNITIES, AND CHALLENGES

"Science is built up with facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house."
Jules Henri Poincaré (1854–1912)
La Science et l'hypothése

Plant biologists who focus on crop species typically rely on integrative and comparative analyses using model and reference species. However, these scientists are increasingly challenged in translating diverse genetic, genomic (read "-omics"), and phenotypic information to address their long-term research goals. These diverse data types are dispersed within a growing number of independently evolving, web-based information resources. Quan *et al.* (2003) explain it well: "...many barriers exist between [scientists] and their data, which is scattered over dozens of machines in incompatible data stores in a myriad of formats." In addition, the researcher is commonly faced with data from different resources that are essentially equivalent from the biologist's point of view. For instance, various unigene/gene index/cluster sets, produced using different protocols, are available to organize transcriptome information. Second, researchers find web-based information often requires careful data management practices on the client side (desktop).

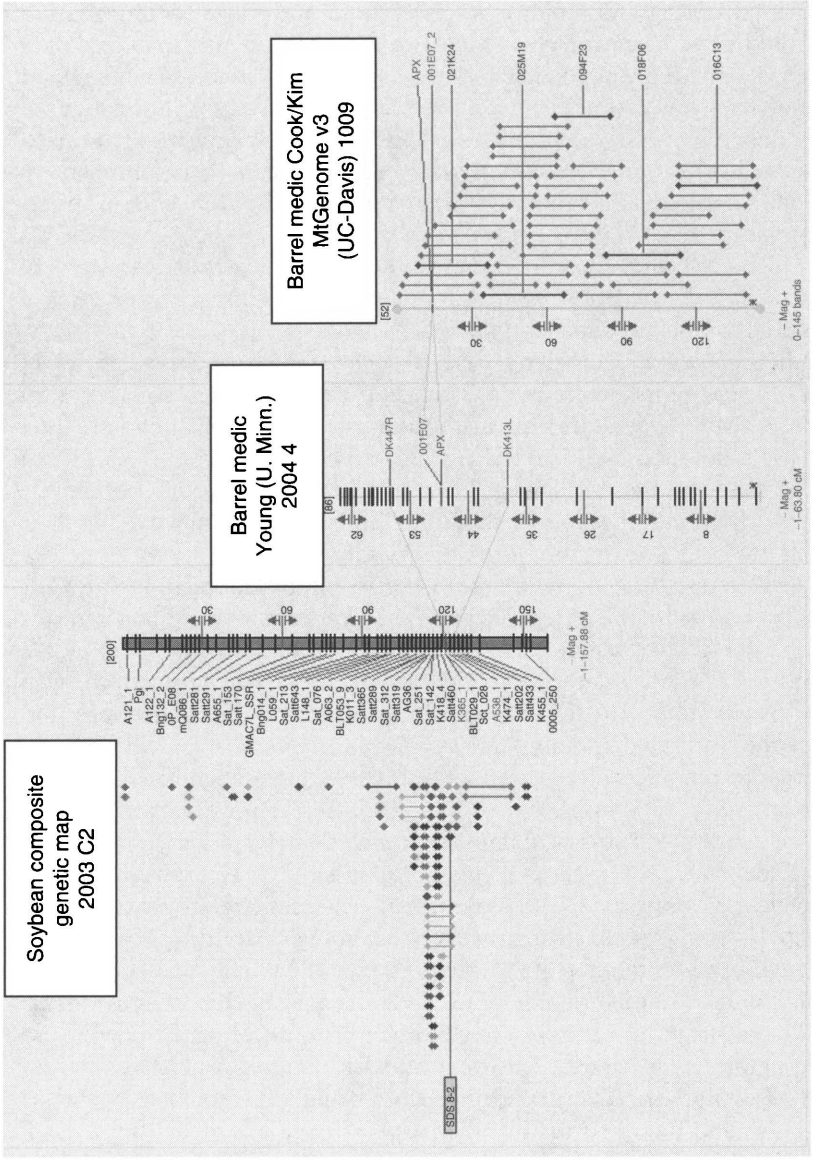


Figure 7 A candidate SDS resistance gene is identified in soybean using *Medicago* annotated genomic sequence in an area where syntenic relationships exist between *Medicago* and soybean. (See Color Insert.)

It is difficult to cross the boundaries between these resources beyond what is explicitly provided through links from one information resource to another, much less to download a combined set of data for follow-on analysis or experimental testing. As biology becomes more and more an information science, addressing the problems of integrating independently managed data resources is of extreme importance. Compounding this issue will be the flood of data as next-generation technologies, such as massively high throughput DNA sequence analysis, greatly increase the volume of data while dramatically reducing its unit cost. GenBank, for example, is a comprehensive nucleotide sequence public repository that contains DNA sequence information for more than 205,000 organisms with more than 3000 new species being added per month. GenBank has doubled in size approximately every 18 months since its inception. As of August 2006, GenBank warehouses more than 145 billion bases of nucleotide sequence.

Hendler makes a compelling case that "as modern science grows in complexity and scope, there is an increasing need for more collaboration between scientists at different institutions, in different subareas, and across scientific disciplines" (Hendler, 2003). The frustrations of biological researchers suggest that the current World Wide Web is not sufficient for the needs of collaboration across scientific disciplines or, ultimately, for their daily discovery activities. For example, a Google search using the term "gene" will return web pages for both the scientific journal *Gene* and for the Hollywood actor Gene Hackman. What are required are mechanisms that imply the correct semantic meaning to search terms as intended by the plant biologist. That said, the current distributed, dynamic nature of the web is particularly suited to the emerging, and ever-changing data types and needs of the research community.

Emerging semantic web and web services technologies (Schiltz *et al.*, 2004; Wilkinson *et al.*, 2003) appear to provide architecture for discovery of, and access to, distributed biological data sources and analysis services. Building on XML, semantic web service technologies underlie an emerging approach to provide not only data integration, but also data interoperability, of distributed web resources. Semantic web technologies are designed to scale and evolve using computer algorithms instead of human-developed "parsers" to identify, configure, compare, and combine data resources on the web. The specific requirements of a semantic web application include the use of geographically distributed information with diverse ownership (i.e., no control of evolution). The application should make use of heterogeneous information and data sources in other ways than intended by the original authors. Importantly, the application would adopt formal descriptions of the meaning of the information. Finally, the application should have a combination of static and dynamic knowledge. The Virtual Plant Information Network (VPIN; http://vpin.ncgr.org) is a National Science Foundation-funded

collaborative project to further develop the technology framework for a web-based, distributed, virtual plant network, into a single semantic web services (SWS) platform. This platform allows partners to share data and invoke web processes to operate on that data. As a SWS platform, VPIN allows partners to describe their data and services, and to find data and services, based on suitable definitions understood by both computers and humans.

Nucleic acid sequencing is a rapidly advancing field that has phenomenal potential to transform crop improvement strategies. Next-generation, non-electrophoretic DNA sequencing technologies generate data at more than 3000-fold the rate but at ~1/80th the cost of conventional capillary sequencing (Table IV). For example, Solexa's technology generates approximately one billion bases of DNA sequence per 3-day run with a reagent cost of 3500 US dollars. To put this into perspective, it would be possible to generate with 10 Solexa instruments the equivalent of all of the data currently in GenBank in a little more than 6 weeks!

Of these new technologies, 454 Life Sciences Corporation, Branford, CT, developed the first DNA sequencing platform to employ picoliter volumes in a highly multiplexed, flow-through array (Margulies *et al.*, 2005). Sequencing is performed on randomly fragmented cDNA using microbead-based pyrosequencing chemistry. This platform provides significant improvements in cost-effectiveness, ease of use and speed and has significant potential to fundamentally change DNA sequencing strategies for crop species. It has been used to sequence pooled RNA samples of *M. truncatula* (Cheung *et al.*, 2006).

DNA sequencing is used in four principal applications: (1) *de novo* genome sequencing to create a reference set of sequences that render a species genomically tractable; (2) gene or genome resequencing in which genes, genome segments, or entire genomes are sequenced in individuals within a population in order to undertake association studies; (3) RNA profiling or transcriptome sequencing, in which an RNA sample is converted to complementary DNA (cDNA) and sequenced in order to determine the sequence or abundance of transcripts (ESTs), for correlation with phenotypes; and

Table IV
Comparisons of Current and Next-Generation DNA Sequencing Technologies

	ABI 3730 × 1	ABI SOLiD	Solexa	454 Life sciences
Read length	400	2 × 20 (paired ends)	35	200
Reads per run	96	40 million	30 million	0.5 million
Bases per run	38,400	1 billion	1 billion	100 million
Run per day	72	1 per 5 days	1 per 3 days	1 per 2 days
Cost per base	0.125¢	TBD[a]	0.0004¢	0.012¢

[a]To be determined.

(4) metagenomics, in which DNA from mixed environmental samples is sequenced in order to evaluate correlation between environmental variables and species' abundances.

De novo genome sequencing has thus far been the focus of a substantial proportion of sequencing resources. Nonelectrophoretic DNA sequencing technologies offer cost and throughput advantages in *de novo* genome sequencing applications. By circumventing the need to propagate clone libraries in a living host, they avoid cloning bias associated with cloning artifacts. The emerging paradigm for use of nonelectrophoretic sequencing technologies in *de novo* genome-sequencing projects is a hybrid approach that uses a combination of capillary and pyrosequencing technologies. Hybrid assemblies of Sanger and pyrosequencing reads were shown to be feasible and cost-effective for development of draft and finished prokaryotic genome sequences (Goldberg *et al.*, 2006). A hybrid approach is likely to become the fundamental strategy for future analyses of crop species' genomes.

The goal of gene expression profiling experiments is typically to understand the dynamics of transcript abundance between states or temporal events in networks and pathways. Usually, this involves the identification of a set of transcripts whose expression differs as an external parameter is varied (e.g., developmental stage, genotype, stress). Several sequence-based transcript-profiling methods have been described that provide absolute counts of the number of times a transcript occurs in a sample (Kuo *et al.*, 2006; Mikkilineni *et al.*, 2004). These approaches appear to be extensible to nonelectrophoretic sequencing technologies (Mikkilineni *et al.*, 2004). An RNA profiling application where nonelectrophoretic sequencing instruments are clearly the platform of choice is in the identification and characterization of small RNA molecules. Several publications have demonstrated the broad utility of 454 pyrosequencing for identifying and profiling various classes of micro and small interfering RNA molecules (Henderson *et al.*, 2006; Lu *et al.*, 2006).

Next-generation sequencing approaches have considerable potential to impact crop species EST, genomic and resequencing efforts. Studies have provided an opportunity for benchmarking of a new paradigm in sequence technology for efficient and cost-effective genome analysis. In the case of DNA sequence analysis, the bottleneck is no longer data generation, but data management and integration for crop improvement.

ACKNOWLEDGMENTS

We thank Cuc Ly for assistance with artwork, and Dr. Yongzhen Pang for providing Fig. 7. Work described from the authors' laboratories was supported by the Samuel Roberts Noble Foundation, and by grants from the

NSF Plant Genome Program, US Department of Energy, Oklahoma Center for the Advancement of Science and Technology, Forage Genetics International and Halliburton Energy Services.

References

Achnine, L., Huhman, D. V., Farag, M. A., Sumner, L. W., Blount, J. W., and Dixon, R. A. (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J.* **41**, 875–887.

Aebersold, R. H., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* **422**, 198–207.

Aerts, R. J., Barry, T. N., and McNabb, W. C. (1999). Polyphenols and agriculture: Beneficial effects of proanthocyanidins in forages. *Agric. Ecosyst. Environ.* **75**, 1–12.

Agrawal, G. K., and Thelen, J. J. (2006). Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. *Mol. Cell. Proteomics* **5**, 2044–2059.

Aharoni, A., de Vos, C. H. R., Verhoeven, H. A., Maliepaard, C. A., Kruppa, G., Bino, R., and Goodenowe, D. B. (2002). Non-targeted metabolomic profiling using Fourier transform ion cyclotron mass spectrometry (FTMS). *OMICS J. Integrat. Biol.* **6**, 217–234.

Aharoni, A., Dixit, S., Jetter, R., Thoenes, E., van Arkel, G., and Pereira, A. (2004). The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis*. *Plant Cell* **16**, 2463–2480.

Akkaya, M. S., Shoemaker, R. C., Specht, J. E., Bhagwat, A. A., and Cregan, P. B. (1995). Integration of simple sequence repeat (SSR) DNA markers into a soybean linkage map. *Crop Sci.* **35**, 1439–1445.

Albrecht, K. A., Wedin, W. F., and Buxton, D. R. (1987). Cell-wall composition and digestibility of alfalfa stems and leaves. *Crop Sci.* **27**, 735–741.

Alm, V., Fang, C., Busso, C. S., Devos, K. M., Vollan, K., Grieg, Z., and Rognli, O. A. (2003). A linkage map of meadow fescue (*Festuca pratensis* Huds.) and comparative mapping with other *Poaceae* species. *Theor. Appl. Genet.* **108**, 25–40.

Alonso, J. M., and Ecker, J. R. (2006). Moving forward in reverse: Genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. *Nat. Rev. Genet.* **7**, 524–536.

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., *et al.* (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657.

Anand, A., and Mysore, K. S. (2005). Current advances in *Agrobacterium*-plant interactions and their implications in agricultural biotechnology. *In* "Advances in Plant Physiology" (A. Hemantaranjan, Ed.), pp. 221–242. Scientific Publishers, Jodhpur, India.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.

Asamizu, E., Watanabe, M., and Tabata, S. (2000). Large scale structural analysis of cDNAs in the model legume, *Lotus japonicus*. *J. Plant Res.* **113**, 451–455.

Ausubel, F. M. (2002). Summaries of National Science Foundation-sponsored *Arabidopsis* 2010 projects and National Science Foundation-sponsored plant genome projects that are generating *Arabidopsis* resources for the community. *Plant Physiol.* **129**, 394–437.

Aziz, N., Paiva, N. L., May, G. D., and Dixon, R. A. (2005). Profiling the transcriptome of alfalfa glandular trichomes. *Planta* **221**, 28–38.

Azpiroz-Leehan, R., and Feldmann, K. A. (1997). T-DNA insertion mutagenesis in *Arabidopsis*: Going back and forth. *Trends Genet.* **13**, 152–156.

Bachem, C. W. B., Oomen, R. J. E. F., and Visser, R. G. F. (1998). Transcript imaging with cDNA-AFLP: A step-by-step protocol. *Plant Mol. Biol. Rep.* **16**, 157–173.

Baldwin, D., Crane, V., and Rice, D. (1999). A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. *Curr. Opin. Plant Biol.* **2**, 96–103.

Baligar, V. C., Elgin, J. H., and Foy, C. D. (1989). Variability in alfalfa for growth and mineral uptake and efficiency ratios under aluminum stress. *Agron. J.* **81**, 223–229.

Barnett, M. J., Toman, C. J., Fisher, R. F., and Long, S. R. (2004). A dual-genome symbiosis chop for coordinate study of singal exchange and development in a prokaryote-host interaction. *Proc. Natl. Acad. Sci. USA* **101**, 16636–16641.

Bartel, B., and Bartel, D. P. (2003). MicroRNAs: At the root of plant development? *Plant Physiol.* **132**, 709–717.

Baucher, M., BernardVailhe, M. A., Chabbert, B., Besle, J. M., Opsomer, C., VanMontagu, M., and Botterman, J. (1999). Down-regulation of cinnamyl alcohol dehydrogenase in transgenic alfalfa (*Medicago sativa* L.) and the effect on lignin composition and digestibility. *Plant Mol. Biol.* **39**, 437–447.

Bechtold, N., Jaudeau, B., Jolivet, S., Maba, B., Vezon, D., Voisin, R., and Pelletier, G. (2000). The maternal chromosome set is the target of the T-DNA in the in planta transformation of *Arabidopsis thaliana*. *Genetics* **155**, 1875–1887.

Beckmann, J. S., and Soller, M. (1993). Restriction fragment length polymorphisms in genetic improvement: Methodologies, mapping and costs. *Theor. Appl. Genet.* **67**, 35–43.

Bell, C. J., and Ecker, J. R. (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**, 137–144.

Bent, A. F. (2000). *Arabidopsis* in planta transformation. Uses, mechanisms, and prospects for transformation of other species. *Plant Physiol.* **124**, 1540–1547.

Berloo, R., and Stam, P. (1998). Marker-assisted selection in autogamous RIL populations: A simulation study. *Theor. Appl. Genet.* **96**, 147–154.

Bevan, M., Bancroft, I., Mewes, H. W., Martienssen, R., and McCombie, R. (1999). Clearing a path through the jungle: Progress in *Arabidopsis* genomics. *Bioessays* **21**, 110–120.

Bevan, M., Mayer, K., White, O., Eisen, J. A., Preuss, D., Bureau, T., Salzberg, S. L., and Mewes, H. W. (2001). Sequence and analysis of the *Arabidopsis* genome. *Curr. Opin. Plant Biol.* **4**, 105–110.

Bingham, E. T. (1968). Transfer of diploid *Medicago* spp. germplasm to tetraploid *M. sativa* L. in 4x-2x crosses. *Crop Sci.* **8**, 760–762.

Bingham, E. T. (1980). Maximizing heterozygosity in autopolyploids. *In* Polyploidy, Biological Relevance (W. H. Lewis, Ed.), pp. 471–490. Plenum Press, New York.

Bock, R., and Khan, M. S. (2004). Taming plastids for a green future. *Trends Biotechnol.* **22**, 311–318.

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* **54**, 519–546.

Bonierbale, M. W., Plaisted, R. L., Pineda, O., and Tanksley, S. D. (1994). QTL analysis of trichome-mediated insect resistance in potato. *Theor. Appl. Genet.* **87**, 973–987.

Borevitz, J., Xia, Y., Blount, J. W., Dixon, R. A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* **12**, 2383–2393.

Boudet, A. M., Lapierre, C., and Grima-Pettenati, J. (1995). Tansley Review No. 80. Biochemistry and molecular biology of lignification. *New Phytol.* **129**, 203–236.

Bouton, J. H. (1996). Screening the alfalfa core collection for acid soil tolerance. *Crop Sci.* **36**, 198–200.

Bouton, J. H. (2001). Alfalfa. *In* Proceedings of the XIX International Grassland Congress (J. A. Gomide, W. R. S. Mattos, and S. C. da Silva, Eds.), pp. 545–547. FEALQ, Piracicaba, Sao Pedro, Sao Paulo, Brazil.

Bouton, J. H., and Parrott, W. A. (1997). Salinity and aluminum stress. *In* "Biotechnology and the Improvement of Forage Legumes" (B. D. McKersie and D. C. W. Brown, Eds.), pp. 203–226. CAB International, Wallingford, New York.

Bowles, D., Lim, E., Poppenberger, B., and Vaistij, F. (2006). Glycosyltransferases of lipophilic small molecules. *Annu. Rev. Plant Biol.* **57**, 567–597.

Bradbury, L. M. T., Fitzgerald, T. L., Henry, R. J., Jin, Q. S., and Waters, D. L. E. (2005). The gene for fragrance in rice. *Plant Biotechnol. J.* **3**, 363–370.

Bradford, K. J., Van Deynze, A., Gutterson, N., Parrott, W., and Strauss, S. H. (2005). Regulating transgenic crops sensibly: Lessons from plant breeding, biotechnology and genomics. *Nat. Biotechnol.* **23**, 439–444.

Brady, N. C. (1974). "The Nature and Property of Soils." Macmillian Publishing Co. Inc., New York.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C. J. J., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., *et al.* (2001). Minimum information about a microarray experiment (MIAME): Toward standards for microarray data. *Nat. Genet.* **29**, 365–371.

Bregitzer, P., Halbert, S. E., and Lemaux, P. G. (1998). Somaclonal variation in the progeny of transgenic barley. *Theor. Appl. Genet.* **96**, 421–425.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., *et al.* (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**, 630–634.

Broeckling, C. D., Huhman, D. V., Farag, M., Smith, J. T., May, G. D., Mendes, P., Dixon, R. A., and Sumner, L. W. (2005). Metabolic profiling of *Medicago truncatula* cell cultures reveals effects of biotic and abiotic elicitors on primary metabolism. *J. Exp. Bot.* **56**, 323–336.

Broothaerts, W., Mitchell, H. J., Weir, B., Kaines, S., Smith, L. M. A., Wei, Y., Mayer, J. E., Roa-Rodríguez, C., and Jefferson, R. A. (2005). Gene transfer to plants by diverse species of bacteria. *Nature* **433**, 629–633.

Broun, P., Poindexter, P., Osborne, E., Jiang, C.-Z., and Riechmann, J. L. (2004). WIN1, a transcriptional activator of epidermal wax accumulation in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **101**, 4706–4711.

Brummer, E. C., Bouton, J. H., and Kochert, G. (1993). Development of an RFLP map in diploid alfalfa. *Theor. Appl. Genet.* **86**, 329–332.

Bundock, P. C., and Henry, R. J. (2004). Single nucleotide polymorphism haplotype diversity and recombination in the *Isa* gene of barley. *Theor. Appl. Genet.* **109**, 543–551.

Buol, S. W., and Eswaran, H. (1993). Assessment and conquest of poor soils. *In* "Adaptation of Plants to Soil Stresses" (J. W. Maranville, Ed.), pp. 17–27. Intsormil Publication No. 94–2, University of Nebraska, Lincoln.

Burr, B., and Burr, F. A. (1991). Recombinant inbreds for molecular mapping in maize. *Trends Genet.* **7**, 55–60.

Buxton, D. R., and Russell, J. R. (1988). Lignin constituents and cell-wall digestibility of grass and legume stems. *Crop Sci.* **28**, 553–558.

Cannon, S. B., Sterck, L., Rombauts, S., Sato, S., Cheung, F., Gouzy, J. P., Wang, X., Mudge, J., Vasdewani, J., Scheix, T., Spannagl, M., Nicholson, C., *et al.* (2006). Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proc. Natl. Acad. Sci. USA* **103**, 14959–14964.

Cardinal, J., Lee, M., and Moore, J. (2003). Genetic mapping and analysis of quantitative trait loci affecting fiber and lignin content in maize. *Theor. Appl. Genet.* **106**, 866–874.

Casati, P., and Walbot, V. (2003). Gene expression profiling in response to ultraviolet radiation in maize genotypes with varying flavonoid content. *Plant Physiol.* **132**, 1739–1754.

Casati, P., Zhang, X. C., Burlingame, A. L., and Walbot, V. (2005). Analysis of leaf proteome after UV-B irradiation in maize lines differing in sensitivity. *Mol. Cell. Proteomics* **4**, 1673–1685.

Casler, M. D. (1987). *In vitro* digestibility of dry matter and cell wall constituents of smooth bromegrass forage. *Crop Sci.* **27**, 931–934.

Casler, M. D., and Vogel, K. P. (1999). Accomplishments and impact from breeding for increased forage nutritional value. *Crop Sci.* **39**, 12–20.

Charlton, A., Godward, J., Oehlschlager, S., Ambrose, M., Bowen, M., Welham, T., Mullineaux, P., and Domoney, C. (2004). Investigation of the metabolome of leaves from a pea germplasm array and from transgenic pea plants exposed to environmental stress. *In* "Proceedings of the Fifth European Conference on Grain Legumes," pp. 189–190, Dijon, France.

Chen, E. I., Hewel, J., Felding-Habermann, B., and Yates, J. R., III (2006a). Large scale protein profiling by combination of protein fractionation and multidimensional protein identification technology (MudPIT). *Mol. Cell. Proteomics* **5**, 53–56.

Chen, F., Duran, A. L., Blount, J. W., Sumner, L. W., and Dixon, R. A. (2003). Profiling phenolic metabolites in transgenic alfalfa modified in lignin biosynthesis. *Phytochemistry* **64**, 1013–1021.

Chen, F., Reddy, M. S. S., Temple, S., Jackson, L., Shadle, G. L., and Dixon, R. A. (2006b). Multi-site genetic modulation of monolignol biosynthesis suggests new routes for the formation of syringyl lignin and wall-bound ferulic acid in alfalfa (*Medicago sativa* L.). *Plant J.* **48**, 113–124.

Chen, X., Salamini, F., and Gebhardt, C. (2001). A potato molecular-function map for carbohydrate metabolism and transport. *Theor. Appl. Genet.* **102**, 284–295.

Cheng, M., Fry, J. E., Pang, S., Zhou, H., Hironaka, C. M., Duncan, D. R., Conner, T. W., Wan, Y., Cheng, M., Pang, S. Z., Zhou, H. P., and Wan, Y. C. (1997). Genetic transformation of wheat mediated by *Agrobacterium tumefaciens*. *Plant Physiol.* **115**, 971–980.

Cheng, M., Lowe, B. A., Spencer, T. M., Ye, X. D., and Armstrong, C. L. (2004). Factors influencing *Agrobacterium*-mediated transformation of monocotyledonous species. *In Vitro Cell. Dev. Biol.-Plant* **40**, 31–45.

Cheng, X.-F., and Wang, Z.-Y. (2005). Overexpression of *COL9*, a *CONSTANS-LIKE* gene, delays flowering by reducing *CO* and *FT* expression in *Arabidopsis thaliana*. *Plant J.* **43**, 758–768.

Cheung, F., Haas, B. J., Goldberg, S. M. D., May, G. D., Xiao, Y., and Town, C. D. (2006). Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology. *BMC Genomics* **7**, 272–282.

Cho, Y. G., Ishii, T., Temnykh, S., Chen, X., Lopovich, L., McCouch, S. R., Park, W. D., Ayres, N., and Cartinhour, S. (2000). Diversity of microsatellites derived from genomic libraries. *Theor. Appl. Genet.* **100**, 713–722.

Chory, J., Ecker, J. R., Briggs, S., Caboche, M., Coruzzi, G. M., Cook, D., Dangl, J., Grant, S., Guerinot, M. L., Henikoff, S., Martienssen, R., Okada, K., *et al.* (2000). National Science Foundation-Sponsored Workshop report: 'The 2010 Project': Functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. *Plant Physiol.* **123**, 423–425.

Christou, P. (1992). Genetic transformation of crop plants using microprojectile bombardment. *Plant J.* **2**, 275–281.

Claverie, J.-M. (2005). Fewer genes, more non-coding RNA. *Science* **309**, 1529–1530.

Clough, S. J., and Bent, A. F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.

Cloutier, S., Cappadocia, M., and Landry, B. S. (1995). Study of microspore-culture responsiveness in oilseed rape (*Brassica napus* L.) by comparative mapping of a F2 population and two microspore-derived populations. *Theor. Appl. Genet.* **91**, 841–847.

Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Kopka, J., and Udvardi, M. K. (2004). Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant J.* **39**, 487–512.

Comai, L., Young, K., Till, B. J., Reynolds, S. H., Greene, E. A., Codomo, C. A., Enns, L. C., Johnson, J., Burtner, C., and Oden, A. R. (2004). Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J.* **37**, 778–786.

Comis, D. (2005). TILLING genes to improve soybeans. *Agric. Res.* **53**, 4–5.

Cook, D. R. (1999). *Medicago truncatula*: A model in the making! *Curr. Opin. Plant Biol.* **2**, 301–304.

Cordeiro, G. M., Eliott, F., McIntyre, C. L., Casu, R. E., and Henry, R. J. (2006). Characterisation of single nucleotide polymorphisms in sugarcane ESTs. *Theor. Appl. Genet.* **113**, 331–343.

Costa, M. A., Collins, R. E., Anterola, A. M., Cochrane, F. C., Davin, L. B., and Lewis, N. G. (2003). An *in silico* assessment of gene function and organization of the phenylpropanoid pathway metabolic networks in *Arabidopsis thaliana* and limitations thereof. *Phytochemistry* **64**, 1097–1112.

Coulman, B., Goplen, B., Majak, W., McAllister, T., Cheng, K. J., Berg, B., Hall, J., McCartney, D., and Acharya, S. (2000). A review of the development of a bloat-reduced alfalfa cultivar. *Can. J. Plant Sci.* **80**, 487–491.

Czechowski, T., Bari, R. P., Stitt, M., Scheible, W. R., and Udvardi, M. K. (2004). Real-time RT-PCR profiling of over 1400 *Arabidopsis transcription* factors: Unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J.* **38**, 366–379.

da la Fuente, J. M., Ramirez-Rodriguez, V., Cabrera-Ponce, J. L., and Herrera-Estrella, L. (1997). Aluminum tolerance in transgenic plants by alteration of citrate synthesis. *Science* **276**, 1566–1568.

Dall'Agnol, M., Bouton, J. H., and Parrott, W. A. (1996). Screening methods to develop alfalfa germplasms tolerant of acid, aluminum toxic soils. *Crop Sci.* **36**, 64–70.

Davin, L. B., and Lewis, N. G. (2005). Dirigent phenoxy radical coupling: Advances and challenges. *Curr. Opin. Biotechnol.* **16**, 398–406.

Davuluri, G. R., van Tuinen, A., Fraser, P. D., Manfredonia, A., Newman, R., Burgess, D., Brummell, D. A., King, S. R., Palys, J., Uhlig, J., Bramley, P. M., Pennings, H. M. J., *et al.* (2005). Fruit-specific RNAi-mediated suppression of *DET1* enhances tomato nutritional quality. *Nat. Biotechnol.* **23**, 890–895.

De Paepe, A., Vuylsteke, M., Van Hummelen, P., Zabeau, M., and Van Der Straeten, D. (2004). Transcriptional profiling by cDNA-AFLP and microarray analysis reveals novel insights into the early response to ethylene in *Arabidopsis*. *Plant J.* **39**, 537–559.

Deavours, B. E., and Dixon, R. A. (2005). Metabolic engineering of isoflavonoid biosynthesis in alfalfa (*Medicago sativa* L.). *Plant Physiol.* **138**, 2245–2259.

Debrosses, G. G., Kopka, J., and Udvardi, M. K. (2005). *Lotus japonicus* metabolic profiling. Development of gas chromatography mass spectrometry resources for the study of plant-microbe interactions. *Plant Physiol.* **137**, 1302–1318.

Delhaize, E., and Ryan, P. R. (1995). Aluminum toxicity and tolerance in plants. *Plant Physiol.* **107**, 315–321.

d'Erfurth, I., Cosson, V., Eschstruth, A., Lucas, H., Kondorosi, A., and Ratet, P. (2003). Efficient transposition of the *Tnt1* tobacco retrotransposon in the model legume *Medicago truncatula*. *Plant J.* **34**, 95–106.

Desfeux, C., Clough, S. J., and Bent, A. F. (2000). Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the *Arabidopsis* floral-dip method. *Plant Physiol.* **123**, 895–904.

Devos, K. M., and Gale, M. D. (2000). Genome relationships: The grass model in current research. *Plant Cell* **12**, 636–646.

Dhugga, K. S., Barreiro, R., Whitten, B., Stecca, K., Hazebroek, J., Randhawa, G. S., Dolan, M., Kinney, A. J., Tomes, D., Nichols, S., and Anderson, P. (2004). Guar seed β-mannan synthase is a member of the cellulose synthase super gene family. *Science* **303**, 363–366.

Dinesh-Kumar, S. P., Anandalakshmi, R., Marathe, R., Schiff, M., and Liu, Y. (2003). Virus-induced gene silencing. *In* "Plant Functional Genomics" (E. Grotewold, Ed.), pp. 287–293. Humana Press Inc., Totowa, NJ.

Ding, X. S., Schneider, W. L., Chaluvadi, S. R., Mian, R., and Nelson, R. S. (2006). Characterization of a *Brome mosaic virus* strain and its use as a vector for gene silencing in monocotyledonous hosts. *Mol. Plant Microbe Interact.* **19**, 1229–1239.

Dixon, R. A., Achnine, L., Kota, P., Liu, C.-J., Reddy, M. S., and Wang, L. (2002). The phenylpropanoid pathway and plant defense: A genomics perspective. *Mol. Plant Pathol.* **3**, 371–390.

Dixon, R. A., Sharma, S. B., and Xie, D. (2005). Proanthocyanidins: A final frontier in flavonoid research? *New Phytologist* **165**, 9–28.

Dixon, R. A., Gang, D. R., Charlton, A. J., Fiehn, O., Kuiper, H. A., Reynolds, T. L., Tjeerdema, R. S., Jeffrey, E. H., German, J. B., Ridley, W. P., and Seiber, J. N. (2006). Applications of metabolomics to agriculture. *J. Agric. Food Chem.* **54**, 8984–8994.

Douglas, G. B., Stienezen, M., Waghorn, G. C., Foote, A. G., and Purchas, R. W. (1999). Effect of condensed tannins in birdsfoot trefoil (*Lotus corniculatus*) and sulla (*Hedysarum coronarium*) on body weight, carcass fat depth, and wool growth of lambs in New Zealand. *NZ J. Agric. Res.* **42**, 55–64.

Downie, A. (1997). Fixing a symbiotic circle. *Nature* **387**, 352–354.

Dudley, J. W., Saghai Maroof, A., and Rufener, G. K. (1992). Molecular marker information and selection of parents in corn breeding programs. *Crop Sci.* **32**, 301–304.

Echt, C. S., Kidwell, K. K., Knapp, S. J., Osborn, T. C., and McCoy, T. J. (1993). Linkage mapping in diploid alfalfa (*Medicago sativa*). *Genome* **37**, 61–71.

Eujayl, I., Baum, M., Powell, W., Erskine, W., and Pehu, E. (1998). A genetic linkage map of lentil (*Lens* sp.) based on RAPD and AFLP markers using recombinant inbred lines. *Theor. Appl. Genet.* **97**, 83–89.

Eujayl, I., Sorrells, M. E., Baum, M., Wolters, P., and Powell, W. (2002). Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theor. Appl. Genet.* **104**, 399–407.

Eujayl, I., Sledge, M. K., Wang, L., May, G. D., Chekhovskiy, K., Zwonitzer, J. C., and Mian, M. A. R. (2004). *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. *Theor. Appl. Genet.* **108**, 414–422.

Fei, Z., Tang, X., Alba, R., and Giovannoni, J. (2006). Tomato expression database (TED): A suite of data presentation and analysis tools. *Nucleic Acids Res.* **34**, D770–D776 (database issue).

Fernandez-Cornejo, J., and Caswell, M. (2006). The first decade of genetically engineered crops in the United States (USDA Economic Information Bulletin Number 11. www.ers.usda.gov).

Fiehn, O. (2002). Metabolomics: The link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155–171.

Firnhaber, C., Pühler, A., and Küster, H. (2005). EST sequencing and time course microarray hybridizations identify more than 700 *Medicago truncatula* genes with develomental expression regulation in flowers and pods. *Planta* **222**, 269–283.

Foy, C. D. (1988). Plant adaptation to acid, aluminum toxic soils. *Comm. Soil Sci. Plant Anal.* **19**, 958–987.

Fraser, P. D., Pinto, M. E. S., Holloway, D. E., and Bramley, P. M. (2000). Application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. *Plant J.* **24**, 551–558.

Freeling, M. (2001). Grasses as a single genetic system. Reassessment 2001. *Plant Physiol.* **125**, 1191–1197.

Fridman, E., Wang, J., Iijima, Y., Froehlich, J. E., Gang, D. R., Ohlrogge, J., and Pichersky, E. (2005). Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones. *Plant Cell* **17**, 1252–1267.

Frisch, M., Bohn, M., and Melchinger, A. E. (1999). Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Sci.* **39**, 1295–1301.

Fuell, H., Hardy, N., Beckmann, M., Draper, J., Smith, A. R., Taylor, T., Fiehn, O., Goodacre, R., Bino, R., Hall, R., Kopka, J., Lange, B. M., *et al.* (2004). A proposed framework for the description of plant metabolomics experiments and their results. *Nat. Biotechnol.* **22**, 1601–1606.

Gaitán-Solís, E., Duque, M. C., Edwards, K. J., and Tohme, J. (2002). Microsatellite repeats in common bean (*Phaseolus vulgaris*): Isolation, characterization, and cross-species amplification in *Phaseolus* ssp. *Crop Sci.* **42**, 2128–2136.

Gale, M. D., and Devos, K. M. (1998). Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* **95**, 1971–1974.

Gang, D. R., Wang, J., Dudareva, N., Nam, K. H., Simon, J. E., Lewinsohn, E., and Pichersky, E. (2001). An investigation of the storage and biosynthesis of phenylpropenes in sweet basil. *Plant Physiol.* **125**, 539–555.

Ge, Y., Norton, T., and Wang, Z.-Y. (2006). Transgenic zoysiagrass (*Zoysia japonica*) plants obtained by *Agrobacterium*-mediated transformation. *Plant Cell Rep.* **25**(8), 792–798.

Ge, Y., Cheng, X.-F., Hopkins, A., and Wang, Z.-Y. (2007). Generation of transgenic *Lolium temulentum* plants by *Agrobacterium tumefaciens*-mediated transformation. *Plant Cell Rep.* (in press).

Gebhardt, C., Ritter, E., Barone, A., Debener, T., Walkemeier, B., Schachtschabel, E., Kaufman, H., Thompson, R. D., Bonierbale, M. W., Ganal, M. W., Tanksley, S. D., and Salamini, F. (1991). RFLP maps of potato and their alignment with the homologous tomato genome. *Theor. Appl. Genet.* **83**, 49–57.

Gelvin, S. B. (2003a). *Agrobacterium*-mediated plant transformation: The biology behind the "gene-jockeying" tool. *Microbiol. Mol. Biol. Rev.* **67**, 16–37.

Gelvin, S. B. (2003b). Improving plant genetic engineering by manipulating the host. *Trends Biotechnol.* **21**, 95–98.

Georgieva, I. D. (1998). Relation between glandular trichome phenolic compounds in tomato and resistance to whitefly (*Trialeurodes vaporariorum* West). *Genet. Breed.* **29**, 61–66.

Giancola, S., McKhann, H. I., Bérard, A., Camilleri, C., Durand, S., Libeau, P., Roux, F., Reboud, X., Gut, I. G., and Brunel, D. (2006). Utilization of the three high-throughput SNP genotyping methods, the GOOD assay, Amplifluor and TaqMan, in diploid and polyploidy plants. *Theor. Appl. Genet.* **112**, 1115–1124.

Gilmour, S. J., Sebolt, A. M., Salazar, M. P., Everard, J. D., and Thomashow, M. F. (2000). Overexpression of the *Arabidopsis* CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. *Plant Physiol.* **124**, 1854–1865.

Goff, S. A., Ricke, D., Lan, T.-H., Presting, G., Wang, R. D., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100.

Goldberg, S. M., Johnson, J., Busam, D., Feldblyum, T., Ferriera, S., Friedman, R., Halpern, A., Khouri, H., Kravitz, S. A., Lauro, F. M., Li, K., Rogers, Y.-H., *et al.* (2006). A Sanger/

pyrosequencing *et al*hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc. Natl. Acad. Sci. USA* **103**, 11240–11245.

Goldman, J. J., Hanna, W. W., Fleming, G. H., and Ozias-Akins, P. (2004). Ploidy variation among herbicide-resistant bermudagrass plants of cv. TifEagle transformed with the *bar* gene. *Plant Cell Rep.* **22**, 553–560.

Gonzales, M. D., Archuleta, E., Farmer, A., Gajendran, K., Grant, D., Shoemaker, R., Beavis, W. D., and Waugh, M. E. (2005). The Legume Information System (LIS): An integrated information resource for comparative legume biology. *Nucleic Acids Res.* **33**, D660–D665.

Goossens, A., Häkkinen, S., Laakso, I., Seppänen-Laakso, T., Biondi, S., De Sutter, V., Lammertyn, F., Nuutila, A. M., Söderlund, H., Zabeau, M., Inzé, D., and Oksman-Caldentey, K. M. (2003). A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proc. Natl. Acad. Sci. USA* **100**, 8595–8600.

Grabber, J. H., Jung, G. A., Abrams, S. M., and Howard, D. B. (1992). Digestion kinetics of parenchyma and sclerenchyma cell walls isolated from orchardgrass and switchgrass. *Crop Sci.* **32**, 806–810.

Grabber, J. H., Ralph, J., Hatfield, R. D., and Quideau, S. (1997). *p*-Hydroxyphenyl, guaiacyl, and syringyl lignins have similar inhibitory effects on wall degradability. *J. Agric. Food Chem.* **45**, 2530–2532.

Graham, T. L. (1991). A rapid, high resolution high performance liquid chromatography profiling procedure for plant and microbial aromatic secondary metabolites. *Plant Physiol.* **95**, 584–593.

Grevich, J. J., and Daniell, H. (2005). Chloroplast genetic engineering: Recent advances and future perspectives. *Crit. Rev. Plant Sci.* **24**, 83–107.

Guo, D., Chen, F., Inoue, K., Blount, J. W., and Dixon, R. A. (2001a). Down-regulation of caffeic acid 3-*O*-methyltransferase and caffeoyl CoA 3-*O*-methyltransferase in transgenic alfalfa (*Medicago sativa* L.): Impacts on lignin structure and implications for the biosynthesis of G and S lignin. *Plant Cell* **13**, 73–88.

Guo, D., Chen, F., Wheeler, J., Winder, J., Selman, S., Peterson, M., and Dixon, R. A. (2001b). Improvement of in-rumen digestibility of alfalfa forage by genetic manipulation of lignin *O*-methyltransferases *Transgenic Res.* **10**, 457–464.

Hackauf, B., and Wehling, P. (2002). Identification of microsatellite polymorphisms in an expressed portion of the rye genome. *Plant Breed.* **121**, 17–25.

Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L. W., and Bino, R. (2002). Plant metabolomics as the missing link in functional genomics strategies. *Plant Cell* **14**, 1437–1440.

Hamada, H., Petrino, M. G., and Kakunaga, T. (1982). A novel repeated element with Z-DNA forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* **79**, 6465–6469.

Harlan, J. R., and deWet, J. M. J. (1975). The origins of polyploids. *Bot. Rev.* **41**, 361–390.

Hass, W., Faherty, B. K., Gerber, S. A., Elias, J. E., Beausoleil, S. A., Bakalarski, C. E., Li, X., Villen, J., and Gygi, S. P. (2006). Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol. Cell. Proteomics* **5**, 1326–1337.

He, X. G. (2000). On-line identification of phytochemical constituents in botanical extracts by combined high-performance liquid chromatographic-diode array detection-mass spectrometric techniques. *J. Chromatogr. A* **880**, 203–232.

Hellens, R., and Mullineaux, P. (2000). A guide to *Agrobacterium* binary Ti vectors. *Trends Plant Sci.* **5**, 446–451.

Henderson, I. R., Zhang, X., Lu, C., Johnson, L., Meyers, B. C., Green, P. J., and Jacobsen, S. E. (2006). Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* **38**, 721–725.

Hendler, J. (2003). Communication. Science and the semantic web. *Science* **299**, 520–521.

Henikoff, S., Till, B. J., and Comai, L. (2004). TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiol.* **135**, 630–636.

Herrera-Estrella, L., Simpson, J., and Martinez-Trujillo, M. (2005). Transgenic plants: An historical perspective. *In* "Transgenic Plants: Methods and Protocols" (L. Pena, Ed.), pp. 3–32. Humana Press, Totowa, NJ.

Hertzberg, M., Sievertzon, M., Aspeborg, H., Nilsson, P., Sandberg, G., and Lundeberg, J. (2001). cDNA microarray analysis of small plant tissue samples using a cDNA tag target amplification protocol. *Plant J.* **25**, 585–591.

Hiei, Y., Ohta, S., Komari, T., and Kumashiro, T. (1994). Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282.

Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T., and Saito, K. (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **101**, 10205–10210.

Hirochika, H., Guiderdoni, E., An, G., Hsing, Y. I., Eun, M. Y., Han, C. D., Upadhaya, N., Ramachandran, S., Zhang, Q., Pereira, A., Sundaresan, V., and Leung, H. (2004). Rice mutant resources for gene discovery. *Plant Mol. Biol.* **54**, 325–334.

Hoffmann, L., Maury, S., Martz, F., Geoffroy, P., and Legrand, M. (2003). Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J. Biol. Chem.* **278**, 95–103.

Huang, S. (2000). The practical problems of post-genomic biology. *Nat. Biotechnol.* **18**, 471–472.

Huhman, D. V., and Sumner, L. W. (2002). Metabolic profiling of saponin glycosides in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospary ion-trap mass spectrometer. *Phytochemistry* **59**, 347–360.

Humphreys, J. M., and Chapple, C. (2002). Rewriting the lignin roadmap. *Curr. Opin. Plant Biol.* **5**, 224–229.

Humphreys, J. M., Hemm, M. R., and Chapple, C. (1999). New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. *Proc. Natl. Acad. Sci. USA* **96**, 10045–10050.

Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T., and Kumashiro, T. (1996). High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*. *Nat. Biotechnol.* **14**, 745–750.

Ishii, M., Hashimoto, S., Tsutsumi, S., Wada, Y., Matsushima, K., Kodama, T., and Aburatani, H. (2000). Direct comparison of genechip and SAGE on the quantitative accuracy in transcript profiling analysis. *Genomics* **68**, 136–143.

Jackson, S., Rokhsar, D., Stacey, G., Shoemaker, S., Schmutz, J., and Grimwood, J. (2006). Toward a reference sequence of the soybean genome: A multiagency effort. *Crop Sci.* **46**, 55–61.

Jacobs, J. M. E., Van Eck, H. J., Arens, P., Verkerk-Bakker, B., Linel-Hekkert, B., Bastiaanssen, H. J. M., El-Kharbotly, A., Pereira, A., Jacobsen, E., and Stiekema, W. J. (1995). A genetic map of potato (*Solanum tuberosum*) integrating molecular markers, including transposons, and classical markers. *Theor. Appl. Genet.* **91**, 289–300.

James, C. (2005). Global Status of Commercialized Biotech/GM Crops: 2005 (The International Service for the Acquisition of Agri-biotech Applications (ISAAA). http://www.isaaa.org).

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

Jorgensen, R. A. (1995). Cosuppression, flower color patterns, and metastable gene expression states. *Science* **268**, 686–691.

Jung, H. G. (1997). Analysis of forage fiber and cell walls in ruminant nutrition. *J. Nutr.* **127**, S810–S813.

Jung, H. G., and Deetz, D. A. (1993). Cell wall lignification and degradability. *In* "Forage Cell Wall Structure and Digestibility" (H. G. Jung, D. R. Buxton, R. D. Hatfield, and J. Ralph, Eds.), pp. 315–346. ASA-CSSA-SSSA, Madison.

Jung, H. G., and Vogel, K. P. (1986). Influence of lignin on digestibility of forage cell wall material. *J. Anim. Sci.* **62**, 1703–1712.

Jung, H. G., Valdez, F. R., Hartfield, R. D., and Blanchette, R. A. (1992). Cell wall composition and degradability of forage stems following chemical and biological delignification. *J. Sci. Food. Agric.* **58**, 347–355.

Jung, H. G., Smith, R. R., and Endres, C. C. (1994). Cell wall composition and degradability of stem tissue from lucerne divergently selected for lignin and *in vitro* dry matter disappearance. *Grass Forage Sci.* **49**, 295–304.

Jung, H. G., Mertens, D. R., and Payne, A. J. (1997). Correlation of acid detergent lignin and Klason lignin with digestibility of forage dry matter and neutral detergent fiber. *J. Dairy Sci.* **80**, 1622–1628.

Kalén, G., Samuelsson, R., and Torma, E. (1992). A method for the determination of alkaloids in reed canarygrass (*Phalaris arundinaceae* L.) using HPLC and solid phase extraction. *Plant Sci.* **42**, 224–229.

Kalo, P., Endre, G., Zimanyi, L., Csandi, G., and Kiss, G. B. (2000). Construction of an improved linkage map of diploid alfalfa. *Theor. Appl. Genet.* **100**, 641–657.

Kantety, R. V., Rota, M. L., Matthews, D. E., and Sorrells, M. E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.* **48**, 501–510.

Kasuga, M., Miura, S., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2004). A combination of the *Arabidopsis* DREB1A gene and stress-inducible rd29A promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer. *Plant Cell Physiol.* **45**, 346–350.

Kawasaki, S., and Murakami, Y. (2000). Genome analysis of *Lotus japonicus*. *J. Plant Res.* **113**, 497–506.

Kehoe, D. M., Villand, P., and Somerville, S. (1999). DNA microarrays for studies of higher plants and other photosynthetic organisms. *Trends Plant Sci.* **4**, 38–41.

Kerek, N. M., Ceserani, T., Tausta, S. L., Sussex, I. M., and Nelson, T. M. (2003). Laser capture microdissection of cells from plant tissues. *Plant Physiol.* **132**, 27–35.

Kiss, G. B., Csanadi, G., Kalman, K., Kalo, P., and Okresz, L. (1993). Construction of a basic genetic map of *Medicago* using RFLP, RAPD, isozyme and morphological markers. *Mol. Gen. Genet.* **238**, 129–137.

Kochian, L. V. (1995). Cellular mechanism of aluminum toxicity and resistance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**, 237–260.

Kochian, L. V., Pineros, M. A., and Hoekenga, O. A. (2005). The physiology, genetics and molecular biology of plant aluminum resistance and toxicity. *Plant Soil* **274**, 175–195.

Koornneff, M., Alonso-Blanco, C., and Vreugdenhil, D. (2004). Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* **55**, 141–172.

Kopka, J. (2006). Gas chromatography mass spectrometry. *In* "Plant Metabolomics" (K. Saito, R. A. Dixon, and L. Willmitzer, Eds.), pp. 3–20. Springer-Verlag, Berlin, Heidelberg.

Koupai-Abyazani, M. R., McCallum, J., Muir, A. D., Lees, G. L., Bohm, B. A., Towers, G. H. N., and Gruber, M. Y. (1993). Purification and characterization of a proanthocyanidin polymer from seed of alfalfa (*Medicago sativa* cv. beaver). *J. Agric. Food Chem.* **41**, 565–569.

Kresovich, S., Szewe-McFadden, A. K., Bliek, S. M., and McFerson, J. R. (1995). Abundance and characterization of simple sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. *Theor. Appl. Genet.* **91**, 206–211.

Krieger, C. J., Zhang, P., Mueller, L. A., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S. Y., and Karp, P. D. (2004). MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* **32**, D438–D442.

Krysan, P. J., Young, J. C., and Sussman, M. R. (1999). T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* **11**, 2283–2290.

Kuenne, C., Lange, M., Funke, T., Miehe, H., Thiel, T., Grosse, I., and Scholz, U. (2005). CR-EST: A resource for crop ESTs. *Nucleic Acids Res.* **33**, D619–D621.

Kuhl, J. C., Hanneman, R. E., and Havey, M. J. (2001). Characterization and mapping of Rpi1, a late-blight resistance locus from diploid (1EBN) Mexican *Solanum pinnatisectum*. *Mol. Genet. Genomics* **265**, 977–985.

Kuo, W. P., Liu, F., Trimarchi, J., Punzo, C., Lombardi, M., Sarang, J., Whipple, M. E., Maysuria, M., Serikawa, K., Lee, S. Y., McCrann, D., Kang, J., *et al.* (2006). A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies. *Nat. Biotechnol.* **24**, 832–840.

Lai, Z., Livingstone, K., Zou, Y., Church, S. A., Knapp, S. J., Andrews, J., and Rieseberg, L. H. (2005). Identification and mapping of SNPs from ESTs in sunflower. *Theor. Appl. Genet.* **111**, 1532–1544.

Lange, B. M., and Ghassemian, M. (2005). Comprehensive post-genomic data analysis approaches integrating biochemical pathway maps. *Phytochemistry* **66**, 413–451.

Lange, B. M., Wildung, M. R., Stauber, E. J., Sanchez, C., Pouchnik, D., and Croteau, R. (2000). Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. *Proc. Natl. Acad. Sci. USA* **97**, 2934–2939.

Larkin, P. J., Gibson, J. M., Mathesius, U., Weinman, J. J., Gartner, E., Hall, E., Tanner, G. J., Rolfe, B. G., and Djordjevic, M. A. (1996). Transgenic white clover: Studies with the auxin-responsive promoter, GH3, in root gravitropism and lateral root development. *Transgenic Res.* **5**, 325–335.

Larson, T. R., and Graham, I. A. (2001). A novel technique for the sensitive quantification of acyl CoA esters from plant tissues. *Plant J.* **25**, 115–125.

Lawrence, R. J., and Pikaard, C. S. (2003). Transgene-induced RNA interference: A strategy for overcoming gene redundancy in polyploids to generate loss-of-function mutations. *Plant J.* **36**, 114–121.

Lee, J. M., Nahm, S. H., Kim, Y. M., and Kim, B. D. (2004). Characterization and molecular genetic mapping of microsatellite loci in pepper. *Theor. Appl. Genet.* **108**, 619–627.

Lei, Z., Elmer, A., Watson, B. S., Dixon, R. A., Mendes, P., and Sumner, L. W. (2005). A two dimensional electrophoresis proteomic reference map and systematic identification of 1367 proteins from a cell suspension culture of the model legume *Medicago truncatula*. *Mol. Cell. Proteomics* **4**, 1812–1825.

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N., and Caboche, M. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* **57**, 405–430.

Li, Y. G., Tanner, G., and Larkin, P. (1996). The DMACA-HCl protocol and the threshold proanthocyanidin content for bloat safety in forage legumes. *J. Sci. Food Agric.* **70**, 89–101.

Liang, X., Dron, M., Schmid, J., Dixon, R. A., and Lamb, C. J. (1989). Developmental and environmental regulation of a phenylalanine ammonia-lyase β-glucuronidase gene fusion in transgenic tobacco plants. *Proc. Natl. Acad. Sci. USA* **86**, 9284–9288.

Lim, J., Jung, J. W., Lim, C. E., Lee, M. H., Kim, B. J., Kim, M., Bruce, W. B., and Benfey, P. N. (2005). Conservation and diversification of SCARECROW in maize. *Plant Mol. Biol.* **59**, 619–630.

Liu, C.-J., and Dixon, R. A. (2001). Elicitor-induced association of isoflavone *O*-methyltransferase with endomembranes prevents formation and 7-*O*-methylation of daidzein during isoflavonoid phytoalexin biosynthesis. *Plant Cell* **13**, 2643–2658.

Liu, P., Zhu, J., and Lu, Y. (2004). Marker-assisted selection in segregating generations of self-fertilizing crops. *Theor. Appl. Genet.* **109**, 370–376.

Liu, Z. W., Biyashev, R. M., and Saghai-Maroof, M. A. (1996). Development of simple sequence repeat DNA markers and their integration into a barley linkage map. *Theor. Appl. Genet.* **93**, 869–876.

Lorenz, W. W., and Dean, J. F. D. (2002). SAGE profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*). *Tree Physiol.* **22**, 301–310.

Lovinger, A., Liewehr, D., and Lamp, W. O. (2000). Glandular trichomes on alfalfa impede searching behavior of the potato leafhopper parasitoid. *Biol. Control* **18**, 187–192.

Lu, C., Kulkarni, K., Souret, F. F., MuthuValliappan, R., Tej, S. S., Poethig, R. S., Henderson, I. R., Jacobsen, S. E., Wang, W., Green, P. J., and Meyers, B. C. (2006). MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. *Genome Res.* **16**, 1276–1288.

Lu, R., Martin, H. A. M., Peart, J. R., Malcuit, I., and Baulcombe, D. C. (2003). Virus-induced gene silencing in plants. *Methods* **30**, 296–303.

Ma, J. F. (2000). Role of organic acids in detoxification of Al in higher plants. *Plant Cell Physiol.* **44**, 383–390.

Ma, J. F., Ryan, P. R., and Delhaize, E. (2001). Aluminum tolerance in plants and the complexing role of organic acids. *Trends Plant Sci.* **6**, 273–278.

Majeran, W., Cai, Y., Sun, Q., and van Wijk, K. J. (2005). Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. *Plant Cell* **17**, 3111–3140.

Maliga, P. (2003). Progress towards commercialization of plastid transformation technology. *Trends Biotechnol.* **21**, 20–28.

Maliga, P. (2004). Plastid transformation in higher plants. *Annu. Rev. Plant Biol.* **55**, 289–313.

Maluf, W. R., Campos, G. A., and das Gracas, C. M. (2001). Relationships between trichome types and spider mite (*Tetranychus evansi*) repellence in tomatoes with respect to foliar zingiberene contents. *Euphytica* **121**, 73–80.

Mao, C., Wang, S., Jia, Q., and Wu, P. (2006). OsEIL1, a rice homolog of the *Arabidopsis* EIN3 regulates the ethylene response as a positive component. *Plant Mol. Biol.* **61**, 141–152.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Photosystem I reaction center polypeptides of spinach are synthesized on thylakoid-bound ribosomes. *Nature* **437**, 376–380.

Marshall, D. R., Broué, P., Grace, J., and Munday, J. (1980). Tannins in pasture legumes: The annual and perennial *Medicago* species. *Aust. J. Exp. Agric. Anim. Husb.* **21**, 47–50.

Matsumura, H., Nirasawa, S., and Terauchi, R. (1999). Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *Plant J.* **20**, 719–726.

Mattick, J. S. (2005). The functional genomics of noncoding RNA. *Science* **309**, 1527–1528.

Maughan, P. J., Saghai, M. A. M., and Buss, G. R. (1995). Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. *Genome* **38**, 715–723.

May, G. D., and Dixon, R. A. (2004). Quick Guide: *Medicago truncatula*. *Curr. Biol.* **14**, R180–R181.

McBride, K. E., Svab, Z., Schaaf, D. J., Hogan, P. S., Stalker, D. M., and Maliga, P. (1995). Amplification of a chimeric *Bacillus* gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. *Bio. Technol.* **13**, 362–365.

McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000). Targeting Induced Local Lesions IN Genomes (TILLING) for plant functional genomics. *Plant Physiol.* **123**, 339–442.

McCann, M. C., Rogan, G. J., Fitzpatrick, S., Trujillo, W. A., Sorbet, R., Hartnell, G. F., Riodan, S. G., and Nemeth, M. A. (2006). Glyphosate-tolerant alfalfa is compositionally equivalent to conventional alfalfa. *J. Agric. Food. Chem.* **54**, 7187–7192.

McGrath, K., Dombrecht, B., Manners, J. M., Schenk, P. M., Edgar, C. I., Maclean, D. J., Scheible, W. R., Udvardi, M. K., and Kazan, K. (2005). Repressor- and activator-type ethylene response factors function in jasmonate signaling and disease resistance identified via a genome-wide screen of *Arabidopsis* transcription factor gene expression. *Plant Physiol.* **139**, 949–959.

Méchin, V., Argillier, O., Menanteau, V., Barriére, Y., Mila, I., Pollet, B., and Lapierre, C. (2000). Relationship of cell wall composition to *in vitro* cell wall digestibility of maize inbred line stems. *J. Sci. Food Agric.* **80**, 574–580.

Mehrotra, B., and Mendes, P. (2006). Bioinformatics approaches to integrate metabolomics and other systems biology data. *In* "Plant Metabolomics" (K. Saito, R. A. Dixon, and L. Willmitzer, Eds.), pp. 105–115. Springer-Verlag, Berlin, Heidelberg.

Mian, M. A. R., Zwonitzer, J. C., Chen, Y., Saha, M. C., and Hopkins, A. A. (2005a). AFLP diversity within and among hardinggrass populations. *Crop Sci.* **45**, 2591–2597.

Mian, M. A. R., Zwonitzer, J. C., Hopkins, A. A., Ding, X. S., and Nelson, R. S. (2005b). Response of tall fescue genotypes to a new strain of Brome mosaic virus. *Plant Dis.* **89**, 224–227.

Michaels, S. D., and Amasino, R. M. (1998). A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. *Plant J.* **14**, 381–385.

Mikkilineni, V. R., Mitra, D., Merritt, J., DiTonno, J. R., Church, G. M., Ogunnaike, B., and Edwards, J. S. (2004). Digital quantitative measurements of gene expression. *Biotechnol. Bioeng.* **86**, 117–124.

Mitra, R. M., Gleason, C. A., Edwards, A., Hadfield, J., Downie, J. A., Oldroyd, G. E. D., and Long, S. R. (2004). A Ca^{2+}/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. *Proc. Natl. Acad. Sci. USA* **101**, 4701–4705.

Modolo, L., Blount, J. W., Achnine, L., Naoumkina, M., Wang, X., and Dixon, R. A. (2007). A functional genomics approach to (iso)flavonoid glycosylation in the model legume *Medicago truncatula. Plant Mol. Biol.* (in press).

Moore, S., Payton, P., Wright, M., Tanksley, S., and Giovannoni, J. (2005). Utilization of tomato microarrays for comparative gene expression analysis in the Solanaceae. *J. Exp. Bot.* **56**, 2885–2895.

Morreel, K., Ralph, J., Lu, F., Goeminne, G., Busson, R., Herdewijn, P., Goeman, J. L., Van der Eycken, J., Boerjan, W., and Messens, E. (2004). Phenolic profiling of caffeic acid *O*-methyltransferase-deficient poplar reveals novel benzodioxane oligolignols. *Plant Physiol.* **136**, 4023–4036.

Morrell, K., Goeminne, G., Storme, V., Sterck, L., Ralph, J., Coppieters, W., Breyne, P., Steenackers, M., Georges, M., Messens, E., and Boerjan, W. (2006). Genetical metabolomics of flavonoid biosynthesis in *Populus*: A case study. *Plant J.* **47**, 224–237.

Mototani, H., Mabuchi, A., Saito, S., Fujioka, M., Iida, A., Takatori, Y., Kotani, A., Kubo, T., Nakamura, K., Sekine, A., Murakami, Y., Tsunoda, T., *et al.* (2005). A functional single nucleotide polymorphism in the core promoter region of CALM1 is associated with hip osteoarthritis in Japanese. *Hum. Mol. Genet.* **14**, 1009–1017.

Mudge, J., Cannon, S. B., Kalo, P., Oldroyd, G. E. D., Roe, B. A., Town, C. D., and Young, N. D. (2005). Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. *BMC Plant Biol.* **5**, 15–30.

Mueller, E., Gilbert, J., Davenport, G., Brigneti, G., and Baulcombe, D. C. (1995). Homology-dependent resistance: Transgenic virus resistance in plants related to homology-dependent gene silencing. *Plant J.* **7**, 1001–1013.

Nair, R. B., Bastress, K. L., Ruegger, M. O., Denault, J. W., and Chapple, C. (2004). The *Arabidopsis thaliana* reduced epidermal fluorescence1 gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell* **16**, 544–554.

Naoumkina, M., Torres-Jerez, I., Allen, S., He, J., Zhao, P., Dixon, R. A., and May, G. D. (2007). Analysis of cDNA libraries from developing seeds of guar (*Cyamopsis tetragonoloba* (L.) Taub). *BMC Plant Biol.* (in review).

Narasimhamoorthy, B., Bouton, J. H., Olsen, K. M., and Sledge, M. K. (2007). Quantitative trait loci and candidate gene mapping of aluminum tolerance in diploid alfalfa. *Theor. Appl. Genet.* **114**, 901–913.

Nelson, C. J., Hegeman, A. D., Harms, A. C., and Sussman, M. R. (2006). A quantitative analysis of *Arabidopsis* plasma membrane using trypsin-catalyzed 18O-labeling. *Mol. Cell. Proteomics* **5**, 1382–1395.

Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The *Arabidopsis* TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* **13**, 2099–2114.

Nielsen, K. M. (2003). Transgenic organisms—time for conceptual diversification? *Nat. Biotechnol.* **21**, 227–228.

Njiti, V. N., Meksem, K., Iqbal, M. J., Johnson, J. E., Kassem, M. A., Zobrist, K. F., Kilo, V. Y., and Lightfoot, D. A. (2002). Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. *Theor. Appl. Genet.* **104**, 294–300.

Oldroyd, G. E., and Geurts, R. (2001). *Medicago truncatula*, going where no plant has gone before. *Trends Plant Sci.* **6**, 552–554.

Olhoft, P. M., Flagel, L. E., Donovan, C. M., and Somers, D. A. (2003). Efficient soybean transformation using hygromycin B selection in the cotyledonary-node method. *Planta* **216**, 723–735.

Ono, E., Fukuchi-Mizutani, M., Nakamura, N., Fukui, Y., Yonekura-Sakakibara, K., Yamaguchi, M., Nakayama, T., Tanaka, T., Kusumi, T., and Tanaka, Y. (2006). Yellow flowers generated by expression of the aurone biosynthetic pathway. *Proc. Natl. Acad. Sci. USA* **103**, 11075–11080.

Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., Wright, S. Y., Hinchliffe, E., Adams, J. L., Silverstone, A. L., and Drake, R. (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* **23**, 482–487.

Panaud, O., Chen, X., and McCouch, S. R. (1996). Development of microsatellite markers and characterization of simple-sequence length polymorphism (SSLP) in rice (*Oryza sativa*). *Mol. Gen. Genet.* **252**, 597–607.

Parvathi, K., Chen, F., Guo, D., Blount, J. W., and Dixon, R. A. (2001). Substrate preferences of *O*-methyltransferases in alfalfa suggest new pathways for 3-*O*-methylation of monolignols. *Plant J.* **25**, 193–202.

Paterson, A. H., Damon, S., Hewitt, J. D., Zamir, D., Rabinowitch, H. D., Lincoln, S. E., Lander, E. S., and Tanksley, S. D. (1991). Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. *Genetics* **127**, 181–197.

Paterson, A. H., Lin, Y. R., Li, S., Schertz, K. F., Doebley, J. F., Pinson, S. R. M., Liu, S. C., Stansel, J. W., and Irvine, J. E. (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**, 1714–1717.

Penmetsa, R. V., and Cook, D. R. (2000). Production and characterization of diverse developmental mutants of *Medicago truncatula*. *Plant Physiol.* **123**, 1387–1397.

Perez, F., Menendez, A., Dehal, P., and Quiros, C. F. (1999). Genomic structural differentiation in *Solanum*: Comparative mapping of the A- and E-genomes. *Theor. Appl. Genet.* **98**, 1183–1193.

Perry, J. A., Wang, T. L., Welham, T. J., Gardner, S., Pike, J. M., Yoshida, S., and Parniske, M. (2003). A TILLING reverse gentics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiol.* **131**, 866–871.

Pfeil, B., Schlueter, J., Shoemaker, R. C., and Doyle, J. (2005). Placing palepolyploidy in relation to taxon divergence: A phylogenetic analysis in legumes using 39 gene families. *Syst. Biol.* **54**, 441–454.

Phelps, T. J., Palumbo, A. V., and Beliaev, A. S. (2002). Metabolomics and microarrays for improved understanding of phenotypic characteristics controlled by both genomics and environmental constraints. *Curr. Opin. Biotechnol.* **13**, 20–24.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., and Rafalski, A. (1996). The utility of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* **2**, 225–238.

Provan, J., Powell, W., and Waugh, R. (1996). Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). *Theor. Appl. Genet.* **92**, 1078–1084.

Qiu, D., Morgan, C., Shi, J., Long, Y., Liu, J., Li, R., Zhuang, X., Wang, Y., Tan, X., Dietrich, E., Weihmann, T., Everett, C., *et al.* (2006). A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. *Theor. Appl. Genet.* **114**, 67–80.

Quackenbush, J., Liang, F., Holt, I., Pertea, G., and Upton, J. (2000). The TIGR gene indices: Reconstruction and representation of expressed gene sequences. *Nucleic Acids Res.* **28**, 141–145.

Quan, D., Martin, S., and Grossman, D. (2003). Applying Semantic Web Techniques to Bioinformatics. http://theory.csail.mit.edu/~dquan/iswc2003-bioinformatics.pdf.

Quiros, C. F., and Bauchan, G. R. (1988). The genus *Medicago* and the origin of the *Medicago sativa* complex. *In* "Alfalfa and Alfalfa Improvement" (A. A. Hanson, D. K. Barnes, and R. R. Hill, Eds.), pp. 93–124. Agronomy Monographs, American Society of Agronomy, Madison, WI.

Ralph, J., Brunow, G., Harris, P., Dixon, R. A., and Boerjan, W. (2007). Lignification: Are lignins biosynthesized via simple combinatorial chemistry or via proteinaceous control and template replication? *In* "Recent Advances in Polyphenols Research" (F. Daayf and V. Lattanzio, Eds.) Blackwell Publishing, Malden, MA (in press).

Ray, H., Yu, M., Auser, P., Blahut-Beatty, L., McKersie, B., Bowley, S., Westcott, N., Coulman, B., Lloyd, A., and Gruber, M. Y. (2003). Expression of anthocyanins and proanthocyanidins after transformation of alfalfa with maize *Lc*. *Plant Physiol.* **132**, 1448–1463.

Reddy, M. S. S., Chen, F., Shadle, G. L., Jackson, L., Aljoe, H., and Dixon, R. A. (2005). Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). *Proc. Natl. Acad. Sci. USA* **102**, 16573–16578.

Reed, J. D. (1995). Nutritional toxicology of tannins and related polyphenols in forage lugumes. *J. Anim. Sci.* **73**, 1516–1528.

Reeves, J. B. (1987). Lignin and fiber compositional changes in forages over a growing season and their effects on *in vitro* digestibility. *J. Dairy Sci.* **70**, 1583–1594.

Rensink, W. A., and Buell, C. R. (2005). Microarray expression profiling resources for plant genomics. *Trends Plant Sci.* **10**, 603–609.

Ridker, P. M., Hennekens, D. H., Lindpainter, K., Stampfer, M. J., Eisenbeerg, P. R., and Miletich, J. P. (1995). Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. *N. Engl. J. Med.* **332**, 912–917.

Ridley, W. P., Sidhu, R. S., Pyla, P. D., Nemeth, M. A., Breeze, M. L., and Astwood, J. D. (2002). A comparison of the nutritional profile of Roundup Ready corn event NK603 to that of conventional corn (*Zea mays* L.). *J. Agric. Food Chem.* **50**, 7235–7243.

Roa, A. C., Chavarriaga-Aguirre, P., Duque, M. C., Maya, M. M., Bonierbale, M. W., Iglesias, C., and Tohme, J. (2000). Cross-species amplification of cassava (*Manihot esculenta*) (*Euphorbiaceae*) microsatellites: Allelic polymorphism and degree of relationship. *Am. J. Bot.* **87**, 1647–1655.

Roberts, C. A., Benedict, H. R., Hill, N. S., Kallenbach, R. L., and Rottinghaus, G. E. (2005). Determination of ergot alkaloid content in tall fescue by near-infrared spectroscopy. *Crop Sci.* **45**, 778–783.

Robinson, P. N., Wollstein, A., Böhme, U., and Beattie, B. (2004). Ontologizing gene-expression microarray data: Characterizing clusters with gene ontology. *Bioinformatics* **20**, 979–981.

Röder, M., Plaschke, J., Konig, S., Borner, A., Sorrells, M., Tanksley, S., and Ganal, M. (1995). Abundance, variability and chromosomal location of microsatellites in wheat. *Mol. Gen. Genet.* **246**, 327–333.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N., and Willmitzer, L. (2000). Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J.* **23**, 131–142.

Roessner, U., Leudemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., and Fernie, A. R. (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**, 11–29.

Rommens, C. M., Humara, J. M., Ye, J., Yan, H., Richael, C., Zhang, L., Perry, R., and Swords, K. (2004). Crop improvement through modification of the plant's own genome. *Plant Physiol.* **135**, 421–431.

Rosellini, D., Barone, P., Bouton, J. H., LaFayette, P., Sledge, M. K., Veronesi, F., and Parrott, W. A. (2003). Alfalfa for acid soils: A biotech approach. Czech. *J. Plant Breed.* **30**, 168–172.

Rudd, S. (2003). Expressed sequence tags: Alternative or complement to whole genome sequences? *Trends Plant Sci.* **8**, 321–329.

Ruf, S., Hermann, M., Berger, I. J., Carrer, H., and Bock, R. (2001). Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat. Biotechnol.* **19**, 870–875.

Ryan, P. R., DiTomaso, J. M., and Kochian, L. V. (1993). Aluminum toxicity in roots: An investigation of spatial sensitivity and the role of the root cap. *J. Exp. Bot.* **44**, 437–446.

Ryu, C. M., Anand, A., and Mysore, K. S. (2004). Agrodrench: A novel and effective agroinocualtion method for virus-induced gene silencing in roots and diverse *Solanaceous* species. *Plant J.* **40**, 322–331.

Saghai-Maroof, M. A., Biyashev, R. M., Yang, G. P., Zhang, Q., and Allard, R. W. (1994). Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. *Proc. Natl. Acad. Sci. USA* **91**, 5466–5470.

Saha, M. C., Mian, M. A. R., Eujayl, I., Zwonitzer, J. C., Wang, L., and May, G. D. (2004). Tall fescue EST-SSR markers with transferability across several grass species. *Theor. Appl. Genet.* **109**, 783–791.

Saha, M. C., Mian, M. A. R., Zwonitzer, J. C., Chekhovskiy, K., and Hopkins, A. A. (2005). An SSR- and AFLP-based genetic linkage map of tall fescue (*Festuca arundinacea* Schreb.). *Theor. Appl. Genet.* **110**, 323–336.

Samac, D. A., and Temple, S. J. (2006). Development and utilization of transformation in *Medicago* species. *In* "Genetically Modified Crops" (G. H. Liang and D. Z. Skinner, Eds.), pp. 165–202. The Haworth Press, Inc., Binghamton, NY.

Sanford, J. C. (1988). The biolistic process. *Trends Biotechnol.* **6**, 299–302.

Sankula, S. (2006). Quantification of the Impacts on US Agriculture of Biotechnology-Derived Crops Planted in 2005 (National Center for Food and Agricultural Policy. www.ncfap.org).

Sasaki-Sekimoto, Y., Taki, N., Obayashi, T., Aono, M., Matsumoto, F., Sakurai, N., Suzuki, H., Hirai, M. Y., Noji, M., Saito, K., Masuda, T., Takamiya, K., *et al.* (2005). Coordinated activation of metabolic pathways for antioxidants and defence compounds by jasmonates and their roles in stress tolerance in *Arabidopsis*. *Plant J.* **44**, 653–668.

Sauer, S., Lechner, D., Berlin, K., Plancon, C., Heuermann, A., Lehrach, H., and Gut, I. G. (2000). Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. *Nucleic Acids Res.* **28**, E100.

Schiltz, G., Gessler, D., and Stein, L. (2004). Semantic MOBY.Position Paper, W3C Workshop on Semantic Web for Life Sciences. http://lists.w3.org/Archives/Public/public-swls-ws/2004Sep/att-0036/smoby-w3c-sw-ls.pdf.

Schneider, K., and Douches, D. S. (1997). Assessment of PCR-based simple sequence repeats to fingerprint North American potato cultivars. *Am. Potato J.* **74**, 149–160.

Schoch, G., Goepfert, S., Morant, M., Hehn, A., Meyer, D., Ullmann, P., and Werck-Reichart, D. (2001). CYP98A3 from *Arabidopsis thaliana* is a 3′-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *J. Biol. Chem.* **276**, 36566–36574.

Scott, K. D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E. M., Lee, L. S., and Henry, R. J. (2000). Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.* **100**, 723–726.

Sefc, K. M., Lopes, M. S., Lefort, F., Botta, R., Roubelakis-Angelakis, K. A., Ibáñez, P., Wagner, H. W., Glössl, J., and Steinkellner, H. (2000). Microsatellite variability in grapevine cultivars from different European regions and evaluation of assignment testing to assess the geographic origin of cultivars. *Theor. Appl. Genet.* **100**, 498–505.

Senior, M. L., Chin, E. C. L., Lee, M., Smith, J. S. C., and Stuber, C. W. (1996). Simple sequence repeat markers developed from maize sequences found in the GenBank database: Map construction. *Crop Sci.* **36**, 1676–1683.

Sewalt, V. J. H., Glasser, W. G., Fontenot, J. P., and Allen, V. G. (1996). Lignin impact on fiber degradation. 1. Quinone methide intermediates formed from lignin during *in vitro* fermentation of corn stover. *J. Sci. Food Agric.* **71**, 195–203.

Shadle, G., Chen, F., Reddy, M. S. S., Jackson, L., Nakashima, J., and Dixon, R. A. (2007). Down-regulation of hydroxycinnamoyl CoA: Shikimate hydroxy cinnamoyl transferase in transgenic alfalfa impacts lignification, development and forage quality. *Phytochemistry* (in press).

Shaked, H., Melamed-Bessudo, C., and Levy, A. A. (2005). High frequency gene targeting in *Arabidopsis* plants expressing the yeast RAD54 gene. *Proc. Natl. Acad. Sci. USA* **102**, 12265–12269.

Shanmugam, K. T., O'Gara, F., Andersen, K., and Valentine, R. C. (1978). Biological nitrogen fixation. *Annu. Rev. Plant Physiol.* **29**, 263–276.

Sharma, S. B., and Dixon, R. A. (2005). Metabolic engineering of proanthocyanidins by ectopic expression of transcription factors in *Arabidopsis thaliana*. *Plant J.* **44**, 62–75.

Shen, Y., Zhang, R., Moore, R., Kim, J., Metz, T., Hixson, K., Zhao, R., Livesay, E., Udseth, H., and Smith, R. (2005). Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000–1500 and capabilities in proteomics and metabolomics. *Anal. Chem.* **77**, 3090–3100.

Shirasawa, K., Shiokai, S., Yamaguchi, M., Kishitani, S., and Nishio, T. (2006). Dot-blot-SNP analysis for practical plant breeding and cultivar identification in rice. *Theor. Appl. Genet.* **113**, 147–155.

Sidorov, V. A., Kasten, D., ShengZhi, P., Hajdukiewicz, P. T. J., Staub, J. M., and Nehra, N. S. (1999). Stable chloroplast transformation in potato: Use of green fluorescent protein as a plastid marker. *Plant J.* **19**, 209–216.

Singh, S., Sidhu, J. S., Huang, N., Vikal, Y., Li, Z., Brar, D. S., Dhaliwal, H. S., and Khush, G. S. (2001). Pyramiding three bacterial blight resistance genes (*Xa5, Xa13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. *Theor. Appl. Genet.* **102**, 1011–1015.

Sivaguru, M., and Horst, W. J. (1998). The distal part of the transition zone is the most aluminum-sensitive apical root zone of maize. *Plant Physiol.* **116**, 155–163.

Slade, A. J., and Knauf, V. C. (2005). TILLING moves beyond functional genomics into crop improvement. *Transgenic Res.* **14**, 109–115.

Sledge, M. K., Bouton, J. H., Dall'Agnoll, M., Parrott, W. A., and Kochert, G. (2002). Identification and confirmation of aluminum tolerance QTL in diploid *Medicago sativa* subsp. *coerulea*. *Crop Sci.* **42**, 1121–1128.

Smith, O. S., and Smith, J. S. C. (1992). Measurement of genetic diversity among maize hybrids: A comparison of isozymic, RFLP, pedigree, and heterosis data. *Maydica* **37**, 53–60.

Smulikowska, S., Pastuszewska, B., Swiech, E., Ochtabinska, A., Mieczkowska, A., Nguyen, V. C., and Buraczewska, L. (2001). Tannin content affects negatively nutritive value of pea for monogastrics. *J. Anim. Feed Sci.* **10**, 511–523.

Somers, D. A., Samac, D. A., and Olhoft, P. M. (2003). Recent advances in legume transformation. *Plant Physiol.* **131**, 892–899.

Spangenberg, G., Wang, Z.-Y., and Potrykus, I. (1998). "Biotechnology in Forage and Turf Grass Improvement." Springer, Berlin.

Squirrell, J., Hollingsworth, P. M., Woodhead, M., Russell, J., Lowe, A. J., Gibby, M., and Powell, W. (2003). How much effort is required to isolate nuclear microsatellites from plants? *Mol. Ecol.* **12**, 1339–1348.

Stacey, G., Libault, M., Brechenmacher, L., Wan, J., and May, G. D. (2006). Genetics and functional genomics of legume nodulation. *Curr. Opin. Plant Biol.* **9**, 110–121.

Stanford, E. H. (1951). Tetrasomic inheritance in alfalfa. *Agron. J.* **43**, 2222–2225.

Stanford, E. H., Clement, W. M., and Bingham, E. T. (1972). Cytology and evolution of the *Medicago sativa-coerulea-falcata* complex. *In* "Alfalfa Science and Technology" (C. H. Hanson, Ed.), pp. 87–102. American Society of Agronomy, Madison, WI.

Stashenko, E. E., Acosta, R., and Martinez, J. R. (2000). High-resolution gas-chromatographic analysis of the secondary metabolites obtained by subcritical-fluid extraction from Colombian rue (*Ruta graveolens* L.). *J. Biochem. Biophys. Methods* **43**, 379–390.

Staub, J. M., Garcia, B., Graves, J., Hajdukiewicz, P. T. J., Hunter, P., Nehra, N., Paradkar, V., Schlittler, M., Carroll, J. A., Spatola, L., Ward, D., Ye, G.-N., *et al.* (2000). High-yield production of a human therapeutic protein in tobacco chloroplasts. *Nat. Biotechnol.* **18**, 333–338.

Stoerker, J., Mayo, J. D., Tetzlaff, C. N., Sarracino, D. A., Schwope, I., and Richert, C. (2000). Rapid genotyping by MALDI-monitored nuclease selection from probe libraries. *Nat. Biotechnol.* **18**, 1213–1216.

Sumner, L. W., Duran, A. L., Huhman, D. V., and Smith, J. T. (2002). Metabolomics: A developing and integral component in functional genomic studies of *Medicago truncatula*. *In* "Phytochemistry in the Genomics and Post-Genomics Eras" (J. T. Romeo and R. A. Dixon, Eds.), pp. 31–61. Elsevier, Oxford.

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

Suzuki, H., Reddy, M. S. S., Naoumkina, M., Aziz, N., May, G. D., Huhman, D. V., Sumner, L. W., Blount, J. W., Mendes, P., and Dixon, R. A. (2005). Methyl jasmonate and yeast elicitor induce differential genetic and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula*. *Planta* **220**, 698–707.

Suzuki, M., Kao, C. Y., Cocciolone, S., and McCarty, D. R. (2001). Maize VP1 complements *Arabidopsis* abi3 and confers a novel ABA/auxin interaction in roots. *Plant J.* **28**, 409–418.

Tadege, M., Ratet, P., and Mysore, K. S. (2005). Insertional mutagenesis: A Swiss Army Knife for functional genomics of *Medicago truncatula*. *Trends Plant Sci.* **10**, 229–235.

Tang, W., Luo, X., and Samuels, V. (2004). Regulated gene expression with promoters responding to inducers. *Plant Sci.* **166**, 827–834.

Tanksley, S. D., Young, N. D., Paterson, A. H., and Bonierbale, M. W. (1989). RFLP mapping in plant breeding. New tools for and old science. *Biotechnology* **7**, 257–264.

Tanksley, S. D., Ganal, M. W., Prince, J. P., de Vicente, M. C., Bonierbale, M. W., Broun, P., Fulton, T. M., Giovannoni, J. J., Grandillo, S., Martin, G. B., Messeguer, R., Miller, J. C., *et al.* (1992). High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**, 1141–1160.

Tanner, G. J., Francki, K. T., Abrahams, S., Watson, J. M., Larkin, P. J., and Ashton, A. R. (2003). Proanthocyanidin biosynthesis in plants. Purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. *J. Biol. Chem.* **278**, 31647–31656.

Tavoletti, S., Veronesi, F., and Osborn, T. C. (1996). RFLP linkage map of an alfalfa meiotic mutant based on an F1 population. *J. Hered.* **87**, 167–170.

Taylor, N. L., Heazlewood, J. L., Day, D. A., and Millar, A. H. (2005). Differential impact of environmental stresses on the pea mitochondrial proteome. *Mol. Cell. Proteomics* **4**, 1122–1133.

Temple, S. J., Whalen, D., Drummond, B. J., Dixon, R. A., and McCaslin, M. (2004). Improving alfalfa digestibility: Field evaluation of transgenic alfalfa engineered for reduced lignin. Abstracts of the National Alfalfa Improvement Conference. Quebec City, Quebec.

Tesfaye, M., Temple, S. J., Allan, D. J., Vance, C. P., and Samac, D. A. (2001). Overexpression of malate dehydrogenase in transgenic alfalfa enhances organic acid synthesis and confers tolerance to aluminum. *Plant Physiol.* **127**, 1836–1844.

Thiel, T., Michalek, W., Varshney, R. K., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422.

Thiellement, H., Bahrman, N., Damerval, C., Plomion, C., Rossignol, M., Santoni, V., de Vienne, D., and Zivy, M. (1999). Proteomics for genetic and physioloigcal studies in plants. *Electrophoresis* **20**, 2013–2026.

Thimm, O., Blaesing, O., Gibon, Y., Nagel, A., Meyer, S., Krueger, P., Selbig, J., Mueller, L., Rhee, S. Y., and Stitt, M. (2004). MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939.

Thomas, M. R., and Scott, N. S. (1993). Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence tagged sites (STSs). *Theor. Appl. Genet.* **86**, 985–990.

Tingay, S., McElroy, D., Kalla, R., Fieg, S., Wang, M., Thornton, S., Brettell, R., and Wang, M. B. (1997). *Agrobacterium tumefaciens*-mediated barley transformation. *Plant J.* **11**, 1369–1376.

Titgemayer, E. C., Cochran, R. C., Towne, E. G., Armendariz, C. K., and Olson, K. C. (1996). Elucidation of factors associated with the maturity-related decline in degradability of big bluestem cell wall. *J. Anim. Sci.* **74**, 648–657.

Tohge, T., Nishiyama, Y., Hirai, M. Y., Yano, M., Nakajima, J.-I., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D. B., Kitayama, M., Noji, M., Yamazaki, M., *et al.* (2005). Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant J.* **42**, 218–235.

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., *et al.* (2006). The genome of black cottonwood *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604.

Tyagi, A. K., Mohanty, A., Bajaj, S., Chaudhury, A., and Maheshwari, S. C. (1999). Transgenic rice: A valuable monocot system for crop improvement and gene research. *Crit. Rev. Biotechnol.* **19**, 41–79.

Tzfira, T., and Citovsky, V. (2006). *Agrobacterium*-mediated genetic transformation of plants: Biology and biotechnology. *Curr. Opin. Biotechnol.* **17**, 147–154.

Udvardi, M. K., Tabata, S., Parniske, M., and Stougaard, J. (2005). *Lotus japonicus*: Legume research in the fast lane. *Trends Plant Sci.* **10**, 222–228.

Urbanczyk-Wochniak, E., and Sumner, L. W. (2007). MedicCyc: A biochemical pathway database for *Medicago truncatula*. *Bioinformatics* (in press).

Valentine, L. (2003). *Agrobacterium tumefaciens* and the plant: The David and Goliath of modern genetics. *Plant Physiol.* **133**, 948–955.

van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, M., and Holstege, F. C. P. (2003). Monitoring global messenger RNA changes in externally controlled microarray experiments. *EMBO J.* **4**, 387–393.

Van Deynze, A. E., Nelson, J. C., O'Donoughue, L. S., Ahn, S. N., Siripoonwiwat, W., Harrington, S. E., Yglesias, E. S., Braga, D. P., McCouch, S. R., and Sorrells, M. E. (1995). Comparative mapping in grasses. Oat relationships. *Mol. Gen. Genet.* **249**, 349–356.

Van Eck, H. J., Jacobs, J. M. E., van Dijk, J., Stiekema, W. J., and Jacobsen, E. (1993). Identification and mapping of three flower colour loci of potato (*Solanum tuberosum* L.) by RFLP analysis. *Theor. Appl. Genet.* **86**, 295–300.

van Enckevort, L. J. G., Droc, G., Piffanelli, P., Greco, R., Gagneur, C., Weber, C., González, V. M., Cabot, P., Fornara, F., Berri, S., Miro, B., Ping, L., *et al.* (2005). EU-OSTID: A collection of transposon insertional mutants for functional genomics in rice. *Plant Mol. Biol.* **59**, 99–110.

VandenBosch, K. A., and Stacey, G. (2003). Summaries of legume genomics projects from around the globe. Community resources for crops and models. *Plant Physiol.* **131**, 840–865.

Velculescu, V. E. (1999). Tantalizing transcriptomes—SAGE and its use in global gene expression analysis. *Science* **286**, 1491–1492.

Velculescu, V. E., Vogelstein, B., and Kinzler, K. W. (2000). Analysing uncharted transcriptomes with SAGE. *Trends Genet.* **16**, 423–425.

Wagner, G. J. (1991). Secreting glandular trichomes: More than just hairs. *Plant Physiol.* **96**, 675–679.

Wagner, G. J., Wang, E., and Shepherd, R. W. (2004). New approaches for studying and exploiting an old protuberance, the plant trichome. *Ann. Bot.* **93**, 3–11.

Walden, R. (2002). T-DNA tagging in a genomics era. *Crit. Rev. Plant Sci.* **21**, 143–165.

Wang, H., Li, G., and Chen, R. (2006). Fast neutron bombardment (FNB) induced deletion mutagenesis for forward and reverse genetic studies in plants. *In* "Floriculture, Ornamental and Plant Biotechnology: Advances and Topical Issues" (J. T. da Silva, Ed.), pp. 629–639. Global Science Books, London, UK.

Wang, H. Y., Luo, M., Tereshchenko, I. V., Frikker, D. M., Cui, X., Li, J. Y., Hu, G., Chu, Y., Azaro, M. A., Lin, Y., Shen, L., Yang, Q., *et al.* (2005a). A genotyping system capable of simultaneously analyzing >1000 single-nucleotide polymorphisms in a haploid genome. *Genome Res.* **15**, 276–283.

Wang, Z., Dai, L., Jiang, Z., Peng, W., Zhang, L., Wang, G., and Xie, D. (2005b). GmCOI1, a soybean F-box protein gene, shows ability to mediate jasmonate-regulated plant defense and fertility in *Arabidopsis*. *Mol. Plant Microbe Interact.* **18**, 1285–1295.

Wang, Z.-Y., and Ge, Y. (2005). Rapid and efficient production of transgenic bermudagrass and creeping bentgrass bypassing the callus formation phase. *Funct. Plant Biol.* **32**, 769–776.

Wang, Z.-Y., and Ge, Y. (2006). Recent advances in genetic transformation of forage and turf grasses. *In Vitro Cell. Dev. Biol. Plant* **42**, 1–18.

Wang, Z.-Y., Hopkins, A., and Scott, M. (2002). Plant regeneration from embryogenic cell suspension cultures of *Lolium temulentum*. *In Vitro Cell. Dev. Biol. Plant* **38**, 446–450.

Wang, Z.-Y., Ge, Y., Mian, R., and Baker, J. (2005c). Development of highly tissue culture responsive lines of *Lolium temulentum* by anther culture. *Plant Sci.* **168**, 203–211.

Ward, J. L., and Beale, M. H. (2006). NMR spectroscopy in plant metabolomics. *In* "Plant Metabolomics" (K. Saito, R. A. Dixon, and L. Willmitzer, Eds.), pp. 81–91. Springer-Verlag, Berlin, Heidelberg.

Warnke, S. E., Barker, R. E., Jung, G., Sim, S.-C., Mian, M. A. R., Saha, M. C., Brilman, L. A., Dupal, M. P., and Forster, J. W. (2004). Genetic linkage mapping of an annual × perennial ryegrass population. *Theor. Appl. Genet.* **109**, 294–304.

Watson, B. S., Asirvatham, V. S., Wang, L., and Sumner, L. W. (2003). Mapping the proteome of barrel medic (*Medicato truncatula*). *Plant Physiol.* **131**, 1104–1123.

Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–396.

Weigel, D., Ahn, J. H., Blazquez, M. A., Borewitz, J., Christensen, S. K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malanchaurvil, E. J., Neff, M. M., Nguyen, J. T., Sato, S., *et al.* (2000). Activation tagging in *Arabidopsis*. *Plant Physiol.* **122**, 1003–1013.

Wesley, V. S., Helliwell, C. A., Smith, N. A., Wang, M. B., Rouse, D. T., Liu, Q., Gooding, P. S., Singh, S. P., Abbott, D., Stoutjesdijk, P. A., Robinson, S. P., Gleave, A. P., *et al.* (2001). Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J.* **27**, 581–590.

White, G., and Powell, W. (1997). Isolation and characterization of microsatellite loci in *Swietenia humilis* (*Meliaceae*): An endangered tropical hardwood species. *Mol. Ecol.* **6**, 851–860.

Wienkoop, S., Zoeller, D., Ebert, B., Simon-Rosin, U., Fisahn, J., Glinski, M., and Weckwerth, W. (2004). Cell-specific protein profiling in *Arabidopsis thaliana* trichomes: Identification of trichome-located proteins involved in sulfur metabolism and detoxification. *Phytochemistry* **65**, 1641–1649.

Wilkinson, M., Gessler, D. D., Farmer, A., and Stein, L. (2003). The biomoby project explores open-source, simple, extensible protocols for enabling biological database interoperability. *In* "Proceedings of the Virtual Conference on Genomics and Bioinformatics," pp. 16–26.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A., and Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**, 6531–6535.

Wolff, S., Otto, A., Albrecht, D., Zeng, J. S., Buttner, K., Gluckmann, M., Hecker, M., and Becher, D. (2006). Gel-free and gel-based proteomics in *Bacillus subtilis*. *Mol. Cell. Proteomics* **5**, 1183–1192.

Wright, E., Dixon, R. A., and Wang, Z.-Y. (2006). *Medicago truncatula* transformation using cotyledon explants. *In* "Agrobacterium Protocols" (K. Wang, Ed.), 2nd ed., pp. 129–135. Humana Press, Totowa, New Jersey.

Wu, G., Truksa, M., Datla, N., Vrinten, P., Bauer, J., Zank, T., Cirpus, P., Heinz, E., and Qiu, X. (2005). Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat. Biotechnol.* **23**, 1013–1017.

Wu, L., Nandi, S., Chen, L., Rodriguez, R. L., and Huang, N. (2002). Expression and inheritance of nine transgenes in rice. *Transgenic Res.* **11**, 533–541.

Wu, S. H., Ramonell, K., Gollub, J., and Somerville, S. (2001). Plant gene expression profiling with DNA microarrays. *Plant Physiol. Biochem.* **39**, 917–926.

Xie, C., and Xu, S. (1998). Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. *Heredity* **80**, 489–498.

Xie, D.-Y., and Dixon, R. A. (2005). Proanthocyanidin biosynthesis—still more questions than answers? *Phytochemistry* **66**, 2126–2143.

Xie, D. Y., Sharma, S. R., Paiva, N. L., Ferreira, D., and Dixon, R. A. (2003). Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. *Science* **299**, 396–399.

Xie, D.-Y., Sharma, S. B., and Dixon, R. A. (2004). Anthocyanidin reductases from *Medicago truncatula* and *Arabidopsis thaliana*. *Arch. Biochem. Biophys.* **422**, 91–102.

Xie, D.-Y., Sharma, S. B., Wright, E., Wang, Z.-Y., and Dixon, R. A. (2006). Engineering plants for introduction of health-beneficial proanthocyanidins. *Plant J.* **45**, 895–907.

Xu, K., Deb, R., and Mackill, D. J. (2004). A microsatellite marker and a codominant PCR-based marker for marker-assisted selection of submergence tolerance in rice. *Crop Sci.* **44**, 248–253.

Xu, M. L., Melchinger, A. E., Xia, X. C., and Lübberstedt, T. (1999). High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR, and AFLP markers. *Mol. Gen. Genet.* **261**, 574–581.

Xu, W. W., Sleper, D. A., and Chao, S. (1995). Genome mapping of tall fescue (*Festuca arundinacea* Schreb.) with RFLP markers. *Theor. Appl. Genet.* **91**, 947–955.

Xu, Y. B., McCouch, S. R., and Zhang, Q. (2005). How can we use genomics to improve cereals with rice as a reference genome? *Plant Mol. Biol.* **59**, 7–26.

Yamada, K., Lim, J., Dale, J. M., Chen, H., Shinn, P., Palm, C. J., Southwick, A. M., Wu, H. C., Kim, C., Nguyen, M., Pham, P., Cheuk, R., *et al.* (2003). Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **302**, 842–846.

Yan, S.-P., Zhang, Q.-Y., Tang, Z.-C., Su, W.-A., and Sun, W.-N. (2006). Comparative proteomic analysis provides new insights into chilling stress responses in rice. *Mol. Cell. Proteomics* **5**, 484–496.

Yang, S.-J., Carter, S. A., Cole, A. B., Cheng, N.-H., and Nelson, R. S. (2004). A natural variant of a host RNA-dependent RNA polymerase is associated with increased susceptibility to viruses by *Nicotiana benthamiana*. *Proc. Natl. Acad. Sci. USA* **101**, 6297–6302.

Yates, J. R., III (1998). Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **33**, 1–19.

Ye, G. N., Stone, D., Pang, S. Z., Creely, W., Gonzalez, K., and Hinchee, M. (1999). *Arabidopsis* ovule is the target for *Agrobacterium* in planta vacuum infiltration transformation. *Plant J.* **19**, 249–257.

Ye, X., Salim, A.-B., Klöti, A., Zhang, J., Lucca, P., Beyer, P., and Potrykus, I. (2000). Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287**, 303–305.

Yousef, G. G., and Juvik, J. A. (2001). Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. *Crop Sci.* **41**, 645–656.

Ytterberg, A. J., Peltier, J. B., and Van Wijk, K. J. (2006). Protein profiling of plastoglobules in chloroplasts and chromoplasts. A surprising site for differential accumulation of metabolic enzymes. *Plant Physiol.* **140**, 984–997.

Yu, J., Hu, S., Wang, J., Wong, G. K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92.

Yu, K., Park, S. J., and Poysa, V. (2000). Marker-assisted selection of common beans for resistance to common bacterial blight: Efficiency and economics. *Plant Breed.* **119**, 411–415.

Yu, Y. G., Saghai-Maroof, M. A., Buss, G. R., Maughan, P. J., and Tolin, S. A. (1994). RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. *Phytopathology* **84**, 60–64.

Zamore, P. D., and Haley, B. (2005). Ribo-gnome: The big world of small RNAs. *Science* **309**, 1519–1524.

Zhang, J.-Y., Broeckling, C. D., Blancaflor, E. B., Sledge, M., Sumner, L. W., and Wang, Z.-Y. (2005). Overexpression of *WXP1*, a putative *Medicago truncatula* AP2 domain-containing

transcription factor gene, increases cuticular wax accumulation and enhances drought tolerance in transgenic alfalfa (*Medicago sativa*). *Plant J.* **42**, 689–707.

Zhang, J.-Y., Broeckling, C., Sumner, L. W., and Wang, Z.-Y. (2007). Heterologous expression of two putative *Medicago truncatula* ERF transcription factor genes, *WXP1* and *WXP2*, in *Arabidopsis* led to increased leaf wax accumulation and improved drought tolerance, but differential response in freezing tolerance. *Plant Mol. Biol.* **64**, 265–278.

Zhang, J. Z. (2003). Overexpression analysis of plant transcription factors. *Curr. Opin. Plant Biol.* **6**, 430–440.

Zhou, P. H., Tan, Y. F., He, Y. Q., Xu, C. G., and Zhang, Q. (2003a). Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. *Theor. Appl. Genet.* **106**, 326–331.

Zhou, W. C., Kolb, F. L., Bai, G. H., Domier, L. L., Boze, L. K., and Smith, N. J. (2003b). Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant Breed.* **122**, 40–46.

Zhu, H., Kim, D. J., Baek, J.-M., Choi, H.-K., Ellis, L. C., Kuester, H., McCombie, W. R., Peng, H.-M., and Cook, D. R. (2003). Syntenic relationships between *Medicago truncatula* and *Arabidopsis* reveal extensive divergence of genome organization. *Plant Physiol.* **131**, 1018–1026.

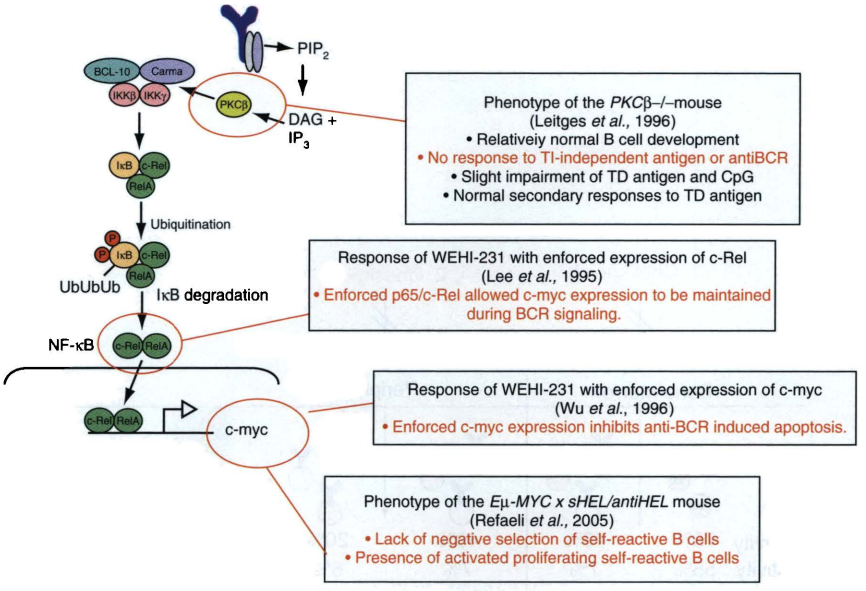


Plate 1

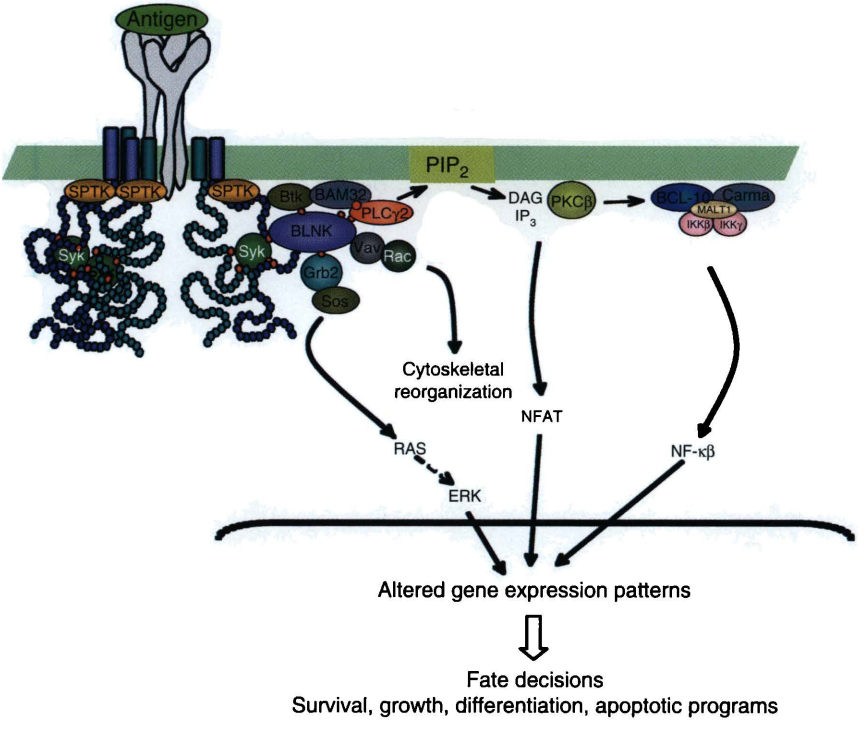


Plate 2

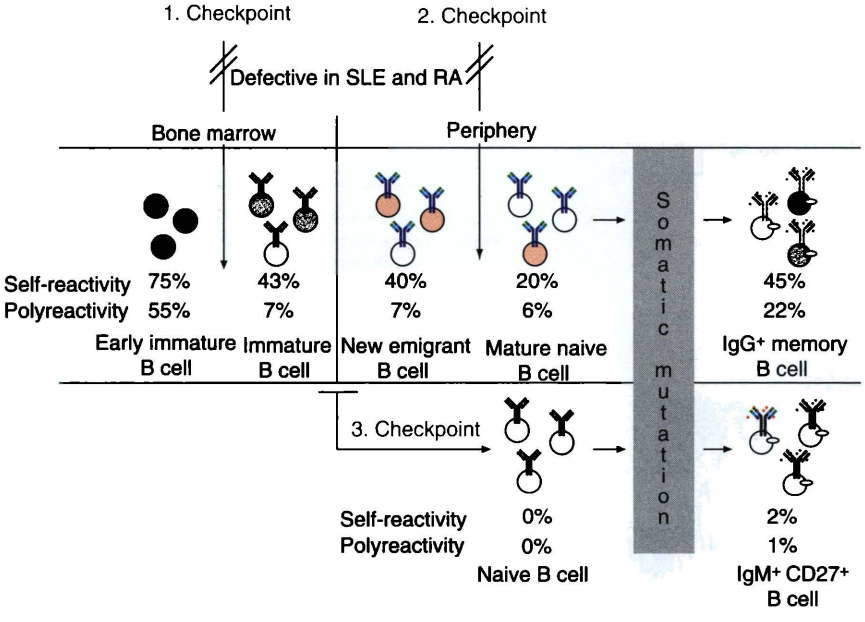

1. Checkpoint
2. Checkpoint
Defective in SLE and RA
Bone marrow
Periphery
Self-reactivity
Polyreactivity
75%
55%
43%
7%
40%
7%
20%
6%
45%
22%
Early immature B cell
Immature B cell
New emigrant B cell
Mature naive B cell
IgG+ memory B cell
Somatic mutation
3. Checkpoint
Self-reactivity
Polyreactivity
0%
0%
Naive B cell
2%
1%
IgM+ CD27+ B cell

Plate 3

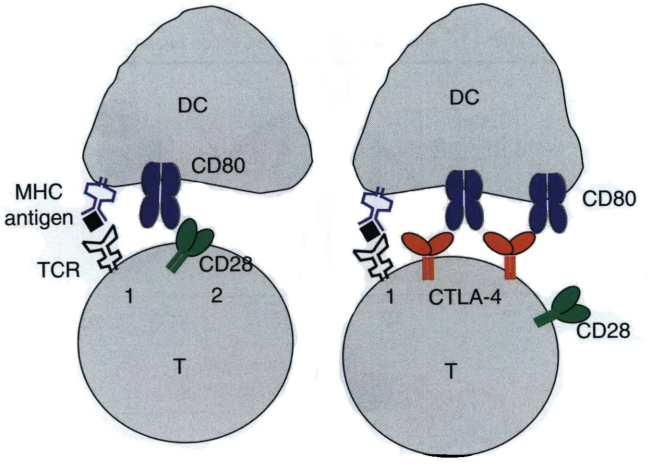


Plate 4

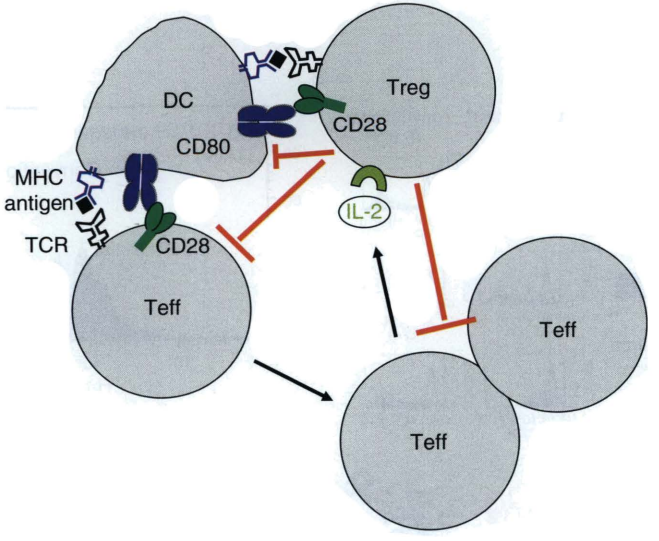


Plate 5

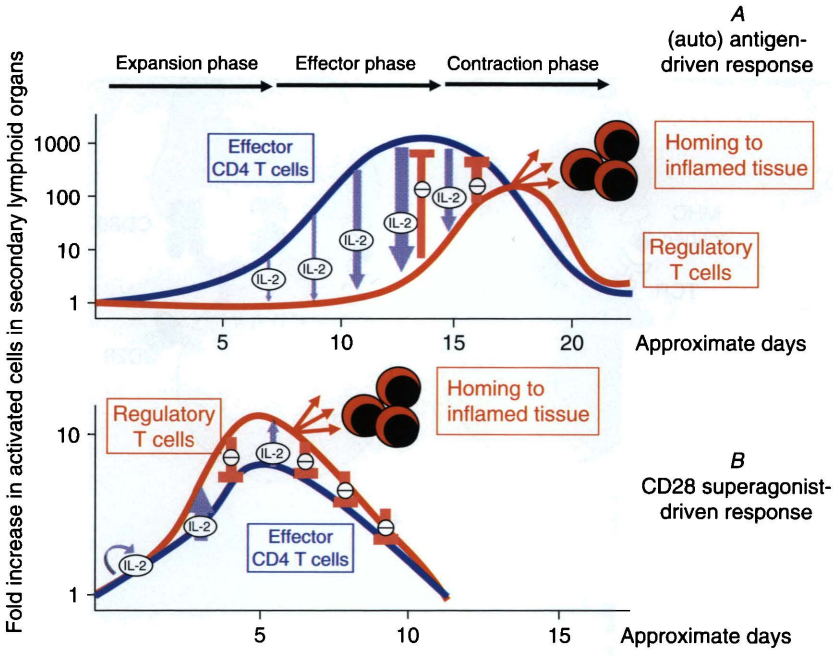


Plate 6

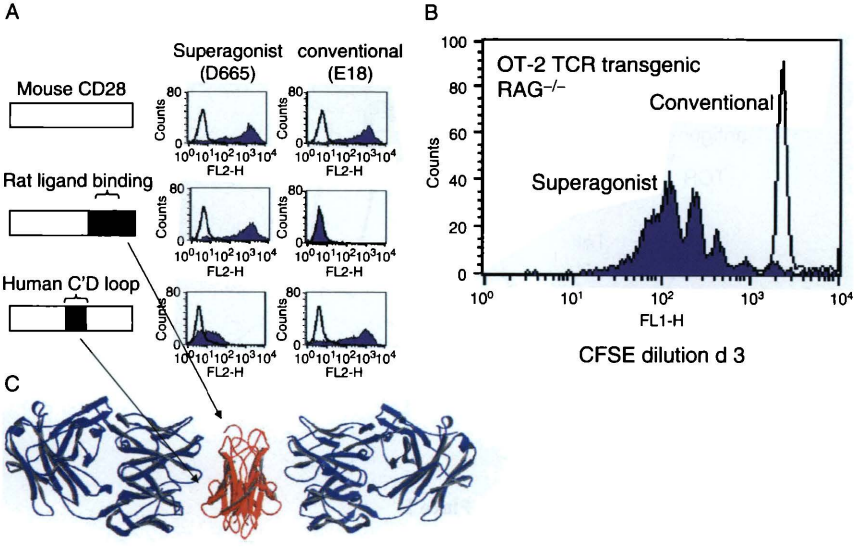


Plate 7

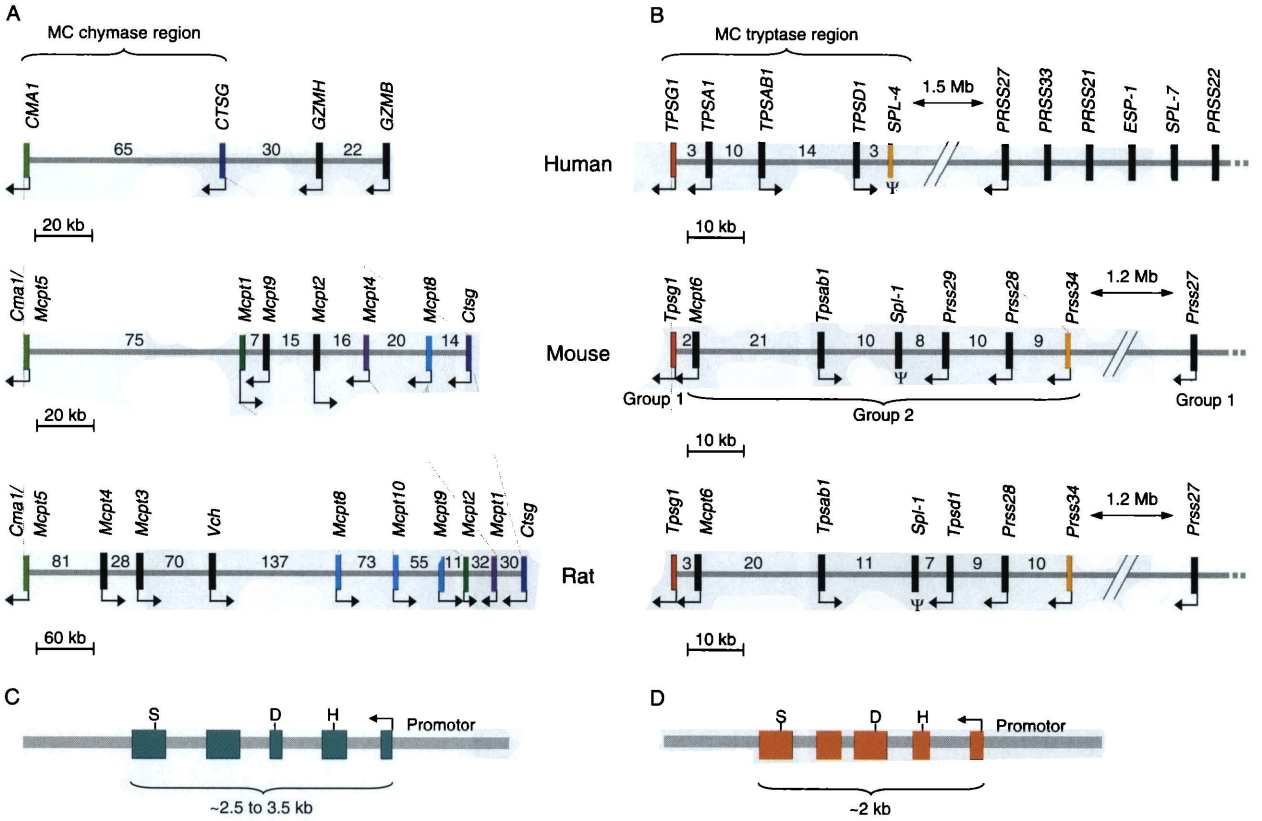
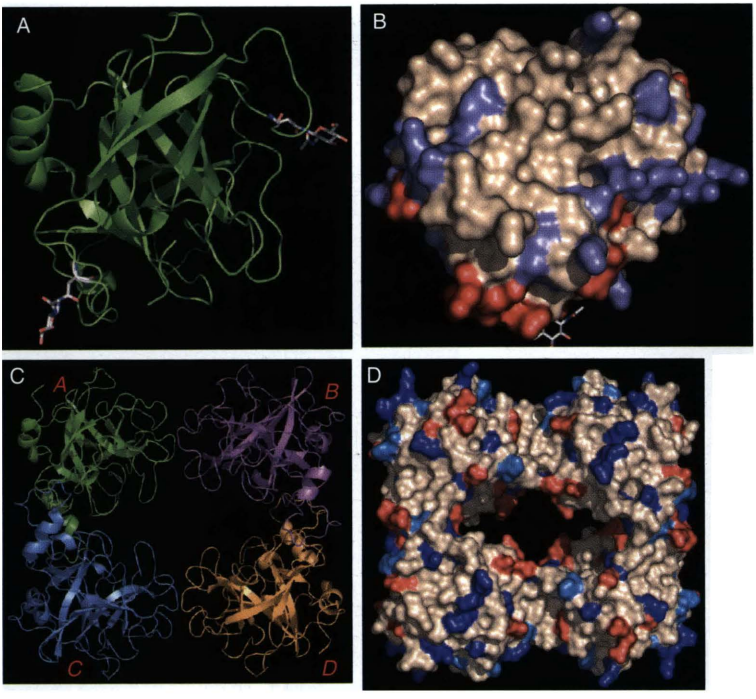

A
MC chymase region
CMA1
CTSG
GZMH
GZMB
65
30
22
20 kb
Cma1/
Mcpt5
Mcpt1
Mcpt9
Mcpt2
Mcpt4
Mcpt8
Ctsg
75
7
15
16
20
14
20 kb
Cma1/
Mcpt5
Mcpt4
Mcpt3
Vch
Mcpt8
Mcpt10
Mcpt9
Mcpt2
Mcpt1
Ctsg
81
28
70
137
73
55
11
32
30
60 kb
B
MC tryptase region
Human
TPSG1
TPSA1
TPSAB1
TPSD1
SPL-4
1.5 Mb
PRSS27
PRSS33
PRSS21
ESP-1
SPL-7
PRSS22
3
10
14
3
Ψ
10 kb
Mouse
Tpsg1
Mcpt6
Tpsab1
Spl-1
Prss29
Prss28
Prss34
1.2 Mb
Prss27
2
21
10
8
10
9
Ψ
Group 1
Group 2
Group 1
10 kb
Rat
Tpsg1
Mcpt6
Tpsab1
Spl-1
Tpsd1
Prss28
Prss34
1.2 Mb
Prss27
3
20
11
7
9
10
Ψ
10 kb
C
S
D
H
Promotor
~2.5 to 3.5 kb
D
S
D
H
Promotor
~2 kb

Plate 8


A
B
C
A
B
C
D
D

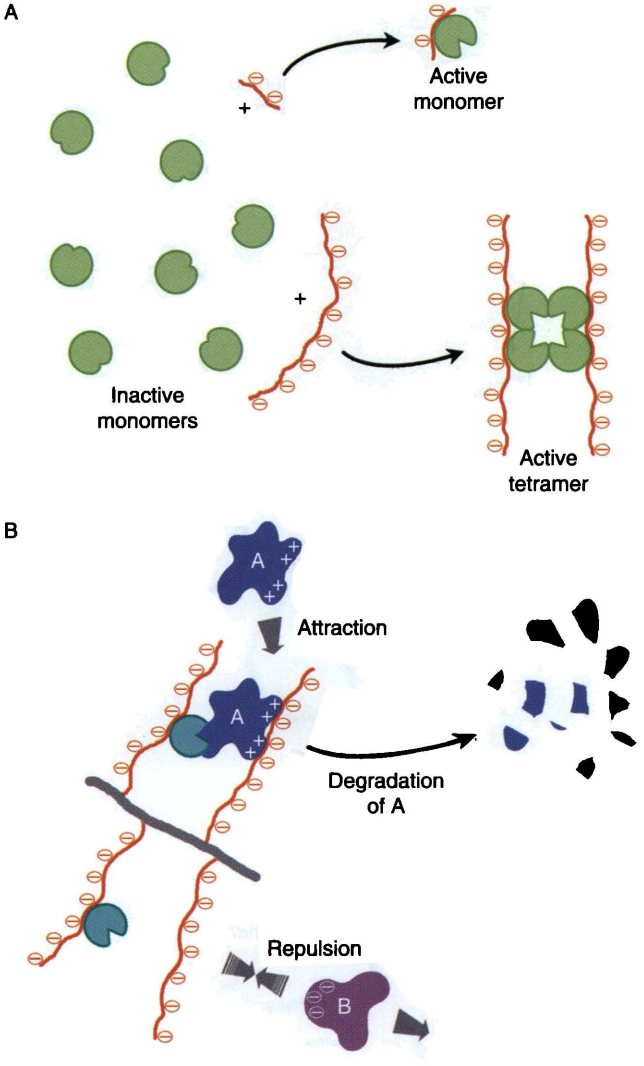

A
Active
monomer
+
Inactive
monomers
+
Active
tetramer
B
A
Attraction
A
Degradation
of A
Repulsion
B

Plate 10

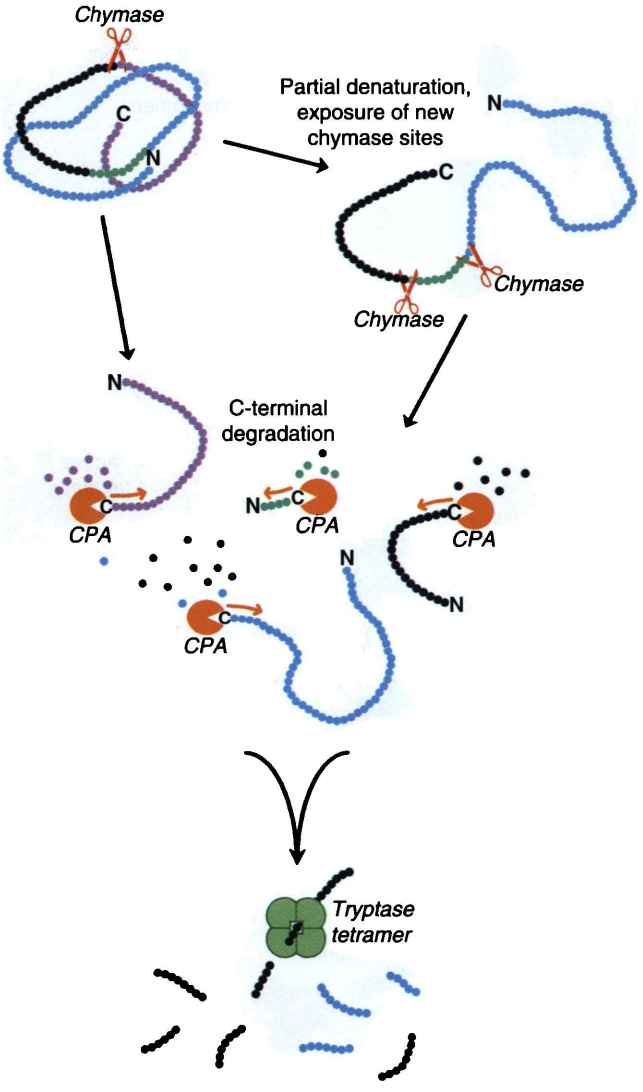

Chymase
C
N
Partial denaturation,
exposure of new
chymase sites
N
C
Chymase
Chymase
N
C-terminal
degradation
C
CPA
N
C
CPA
C
CPA
N
C
CPA
N
Tryptase
tetramer

Plate 11