Gene discovery for crop improvement

Gregory B Martin

Future improvements of crop plants will benefit from the isolation and characterization of genes that underlie both simply-inherited and polygenically-controlled traits. The molecular isolation of economically important plant genes has been facilitated by the construction and application of genetic maps, transposon-based gene tagging, protein–protein interaction cloning, and the development and analysis of large collections of cDNA sequences.

Addresses
Department of Agronomy, Purdue University, West Lafayette, IN 47907-1150, USA; e-mail: gmartin@dept.agry.purdue.edu

Current Opinion in Biotechnology 1998, **9**:220–226

http://biomednet.com/elecref/0958166900900220

Abbreviations

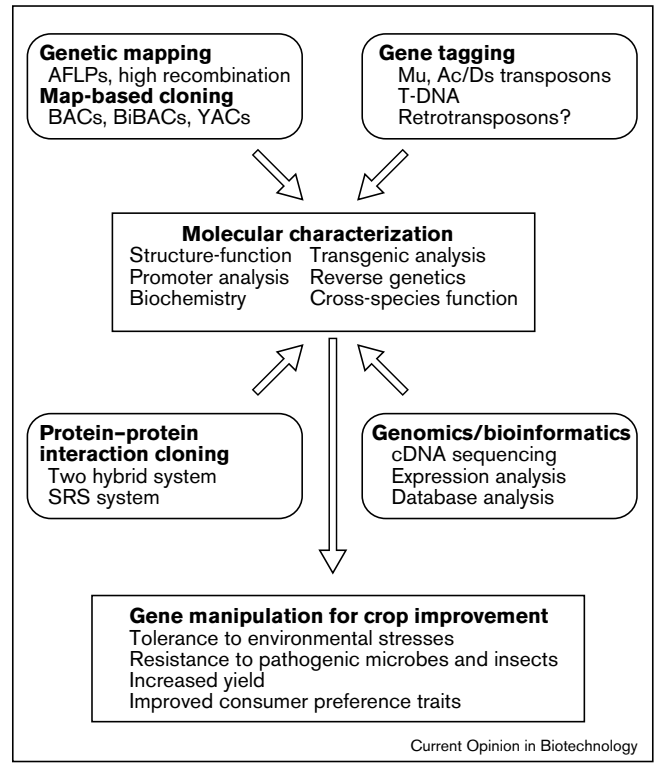
Ac/Ds	activator/dissociation
AFLP	amplified fragment length polymorphism
AIMS	amplification of mutagenized sites
BAC	bacterial artificial chromosome
EST	expressed sequence tag
GEF	GDP-GTP exchange factor
Mu	Mutator
QTL	quantitative trait loci
R	disease resistance
T-DNA	*Agrobacterium* transfer DNA
YAC	yeast artificial chromosome

Introduction

Plant breeding has been remarkably successful in the improvement of both qualitative and quantitative traits that affect agronomic performance and consumer preferences. Our inadequate knowledge about underlying biological processes continues, however, to impede breeding progress for quantitatively-inherited traits such as environmental stress tolerance, yield, and even some simply-inherited traits such as resistance to pathogens and insects. Future improvement of these traits will be facilitated by a deeper understanding of complex metabolic and signal transduction pathways that are involved in their expression. The discovery of new genes provides an important starting point to dissect these fundamental plant processes and to manipulate them for the improvement of crop plants.

There has been much progress in the development of strategies to discover new plant genes. In large part, these developments derive from four experimental approaches: firstly, genetic mapping in plants and the associated ability to use map-based gene isolation strategies; secondly, transposon/T-DNA tagging which allows the direct isolation of a gene; thirdly, protein–protein interaction cloning that permits the isolation of multiple genes contributing to a single pathway or metabolic process; and finally, bioinformatics/genomics — specifically the development and use of large EST databases, search software, and methods to quickly analyze gene expression in response to environmental stresses or during specific developmental stages. Recent progress in these four areas is permitting the isolation of many new plant genes that will have a major impact on crop improvement (Figure 1).

Figure 1



Gene discovery in plants. BiBACs, binary vector/bacterial artificial chromosome; SRS, Sos-recruitment system.

Genetic mapping and map-based cloning

Genetic maps based upon large numbers of cloned DNA markers are now available for all major crop plants and several model plant species (http://probe.nalusda.gov). These maps have revealed that related species often have highly conserved gene order along extensive segments of their chromosomes [1–4]. Such synteny allows cross-referencing among plant genomes, some of which have many mapped agronomically important loci [2–4], and this will expedite the identification of related genes from several species [5••,6••]. The first clear example of synteny was demonstrated between tomato and potato but conservation of gene order is perhaps most compelling among the grasses, where synteny has been found among

segments of the rice, corn, wheat, sorghum, oats and barley genomes [1–4]. In one particularly striking report, it was shown that gene order in small regions of chromosomes may be conserved even over the large evolutionary distance encompassed by the monocot-dicot division of plants [7••]. Cross-genome referencing may thus prove to be a widely applicable gene discovery strategy for linking well-studied model genomes such as *Arabidopsis* with crop plants.

Genetic linkage maps on which agronomically important loci have been located are also proving useful for mapping expressed sequence tags (ESTs) or PCR products encoding motifs that are characteristic of certain classes of gene products. Recently, this strategy was used to isolate and map DNA sequences with homology to disease resistance (*R*) genes from plants [8–10,11••]. *R* genes encode various motifs including leucine-rich repeats, nucleotide binding sites, protein kinase domains, interleukin-1 receptor homologies, and other short conserved sequences with unknown function. To isolate homologs of *R* genes, these features were targeted by degenerate PCR or EST database searches. The resulting sequences were placed on genetic maps to determine whether they are located near known *R* loci. This approach successfully identified over 100 *R* gene homologs from potato, soybean and *Arabidopsis* [8–10,11••]. It is important to note that close linkage or even co-segregation with a particular *R* locus is not sufficient evidence that the candidate gene is responsible for disease resistance; confirmation requires analysis of mutant alleles or complementation of the mutant phenotype by transformation of the cloned gene into the mutant plant. Nevertheless, genetic mapping of candidate genes should prove generally useful for isolating many genes whose products display highly conserved motifs.

Probably the most important development in genetic mapping over the past two years has been the demonstration that quantitative trait loci (QTL) can be localized near cloned DNA markers with very high resolution [12,13,14••,15]. QTL often underlie traits of great economic importance and are typically difficult to manipulate in plant breeding programs. Genetic maps based on DNA markers have allowed the dissection of some quantitative traits into single component loci which contribute substantial percentages of the phenotypic variation for a trait. For example, the tomato *fw2.2* locus accounted for up to 47% of the phenotypic variance in fruit weight, and the *fs8.1* locus accounted for up to 37% of the phenotypic variance in fruit shape [13,14••]. Although such associations have been observed in many crops for different QTL, it has now been demonstrated that QTL can be localized very precisely near specific DNA markers using nearly-isogenic lines [13,14••]. A surprising result from QTL mapping involving germplasm of wild species of crop plants is that the allele contributing the positive effect is often derived from the parent that is not especially notable for the trait under study [15,16,17••]. For example, a QTL that contributes to dark red fruit color in tomato was identified in a wild tomato species with fruit that does not even turn red upon ripening [16]. These observations indicate that many useful genes are present in germplasm collections of wild species and primitive varieties of crops. Plant breeders frequently ignore such collections because they do not show obviously valuable phenotypes, they are difficult to screen, and often are not adapted to temperate environments. Identification of DNA markers linked to specific QTL offers the possibility of marker-assisted selection for these traits, and high resolution mapping of QTL will allow map-based cloning of the genes responsible.

Map-based cloning involves the identification of closely linked DNA markers and use of these markers to isolate a genomic clone that spans the target locus. The genomic clone is used to identify candidate genes by sequencing, complementation of a mutant plant with subclones, or by use of the clone as a hybridization probe to isolate corresponding cDNA clones. Key developments in this strategy have been the realization that saturation of the target region with markers and the development of very high resolution maps (with less than 1 cM between the target locus and linked marker) can greatly expedite gene cloning [18]. It is relatively easy in plants to analyze large segregating populations, and the development of high-throughput DNA markers has expedited high resolution mapping of many loci. As a result, map-based cloning, although still not routine, has been successful in plants ranging from *Arabidopsis* (with a genome size of 145 Mb), to rice, sugar beet, tomato and barley (430 Mb, 760 Mb, 950 Mb, and 5000 Mb, respectively; [19–21,22••]).

DNA markers are used to identify DNA polymorphisms among different crop varieties. They have proven useful for following the inheritance of genetic loci in segregating populations. Originally, DNA markers consisted of RFLPs (restriction fragment length polymorphisms detected by hybridization of genomic DNA with either random genomic clones or cDNAs), or PCR-based markers such as random amplified polymorphic DNAs (RAPDs). Recently, other PCR-based strategies have been developed, such as amplification from small DNA linkers attached to restriction fragments (AFLPs; [23]), amplification of simple-sequence repeats (SSR; [24]) or amplification of the spacer regions between these repeats (inter simple-sequence repeats [ISSR]; [25]). These high-throughput marker strategies in combination with near-isogenic lines or bulk segregant analysis now allow a specific region of the genome to be saturated with DNA makers with much less effort than previously. AFLPs in particular have proven integral to the isolation of several *R* genes, including the *Mlo* gene from barley and the *Rx* gene from potato [22••,26].

The identification of very closely linked markers potentiates the direct isolation of a large-insert genomic clone spanning the target locus without the need for chromosome walking [18]. Such chromosome 'landing' greatly simplifies gene isolation by eliminating the difficult procedure of isolating overlapping clones spanning a region. The availability of extensive physical maps of rice and *Arabidopsis* is further simplifying this step of map-based cloning in these species [27–29]. An impediment to map-based cloning in the past has been the relative difficulty of constructing and manipulating yeast artificial chromosome (YAC) clone libraries. The development of bacterial artificial chromosome (BAC) vectors has allowed the construction of large BAC libraries for many crop species [28,30,31]. It is easier to isolate large amounts of DNA from BAC clones than YAC clones; thus, subsequent isolation of end-clones and subclones for complementation is expedited. Complementation of a mutant phenotype by transformation of a clone carrying the putative target locus, or comparison with multiple mutant alleles, remains the definitive proof that the target gene has been cloned. The development of a BAC vector and a YAC system that can be used directly in *Agrobacterium* transformation promises to facilitate this often time-consuming step [32•,33•].

Map-based cloning strategies usually culminate in the need to identify the target gene within a large region of genomic DNA (30–400 kb). This process can be simplified if the gene is 'marked' by a deletion as sometimes results from fast-neutron or gamma irradiation [34]. Such deletions are often detectable by using the candidate gene as a probe on genomic DNA or mRNA blots. In some cases, a simply scorable phenotype, such as resistance or susceptibility to a disease will permit identification of plants that have an intragenic recombination between mutant alleles giving rise to the wild type trait [22••]. Other promising approaches include the capture of cDNAs by using hybridization to the BAC/YAC DNA bound to nitrocellulose or latex particles, or sequencing of the entire BAC/YAC insert followed by computer analysis for regions with high probability of protein-encoding sequences [22••,35••]. In the syntenic cereal genomes where there is little conservation of repeated sequences, it appears feasible to identify potential genes by hybridization of a BAC/YAC from a related species to a cDNA or genomic library of the target species [5••,6••].

Transposons/T-DNAs

Gene discovery based on gene inactivation by insertion of a transposon or T-DNA into an open reading frame or regulatory sequence continues to be successful in many plant species [36–38,39•]. A particularly powerful approach (and one which can also shed light on gene function) is the use of large maize populations that contain active Mutator (Mu) transposons [36]. Mu transposons are especially amenable to gene discovery because they cause high mutation rates, insert randomly throughout the genome rather than preferentially to linked sites, and are stable once inserted. Several groups have developed populations that contain thousands of Mu-insertions for determining the function of genes that have already been cloned. However, such a 'site-selected insertion'strategy requires knowledge of the gene sequence in order to develop PCR primers.

A related strategy has recently been developed to aid in the discovery of previously uncharacterized genes that are known to carry a Mu insertion [40•]. Previously, such Mu-tagged alleles were identified by the co-segregation of a Mu-hybridizing restriction fragment with the mutation. Determining co-segregation is often difficult, however, because Mu can be present in many copies in the genome complicating analysis by hybridization. A new approach termed 'amplification of mutagenized sites' (AIMS) incorporates the AFLP approach of adding DNA linkers to restriction fragments followed by PCR with linker-targeted primers and Mu-specific primers [40•]. Mu-tagged fragments that differ by as little as a few base pairs are visualized by this method, and products that co-segregate with the mutation are used as probes for further mapping and cDNA isolation. This clever combination of techniques promises to greatly simplify identification of tagged alleles and should be applicable in any plant species where routine insertional activation is possible.

The maize activator/dissociation (Ac/Ds) transposon system is also functional in species such as *Arabidopsis*, tobacco, petunia and tomato. As in maize, Ac/Ds elements preferentially transpose to linked sites on the chromosome in these species. Thus, for gene identification by transposon inactivation it is an advantage if an Ac/Ds insertion is present near a target gene. *Arabidopsis* and tomato populations are being developed in which transposon insertions have already been mapped to facilitate targeted gene tagging [37,38,41]. Potential limitations to heterologous transposon tagging are the relatively low efficiency of transformation in some plant species which makes generation of large transposon-tagged populations difficult and, in *Arabidopsis*, the observation that Ac shows reduced levels of activity. The recent identification of *Arabidopsis* mutants that show increased Ac excision (although low levels of re-insertion) may lead to a clearer understanding of host factors involved in controlling transposon activity and may improve their utility in gene discovery [42].

The integration of *Agrobacterium* transfer DNA (T-DNA) into plant genomes (primarily *Arabidopsis*) has also proved useful for tagging genes [39•]. Populations of over 40,000 T-DNA-tagged lines are now available in *Arabidopsis* and are being used to isolate large numbers of new genes. Estimates with these populations indicate that as many as 35–40% of the mutant lines carry a T-DNA insertion within a gene. T-DNA tagging combined with a gene trapping approach may increase this number to 80% [43•]. Based on mapping studies, it appears that T-DNA

insertions occur randomly within genes and throughout the genome. Thus, it should be feasible to use this approach to saturate the *Arabidopsis* genome. By using a combination of a site-selected insertion (or AIMS-style) PCR strategy with multiplexing of DNA pools derived from these lines it should soon be possible to identify an insertion in practically any gene in the *Arabidopsis* genome.

Finally, retrotransposons (which transpose by using an RNA intermediate) have been proposed as a possible tool for gene discovery in rice [44]. Although these elements are typically inactive during normal growth conditions, certain subclasses of the elements are active during tissue culture. One rice retrotransposon, Tos17, appears to be a good candidate as it can be activated by tissue culture, induces mutations at high frequency, and its copy number is quite low in some rice cultivars making it easier to identify new insertions [44].

Protein–protein interaction cloning

Gene isolation methods often rely on the ability to detect a mutant phenotype in the plant; however, genes that are indispensable or that give no discernible phenotype when mutated may be overlooked by such approaches. For example, sequencing of the yeast genome revealed that, despite the intensive study of this organism, 35% of the open reading frames had not been uncovered by genetic analysis. One approach to overcome this limitation is to use protein–protein interaction as a basis for gene isolation [45]. This approach relies on the fact that protein–protein interactions are fundamental to many cellular processes, including signal transduction, transcription, and metabolism. Thus, to identify genes contributing to a particular phenotype, a single gene is first identified that plays a critical role in a biological process, and the gene is then used as a 'bait' to capture other genes whose products physically interact with the protein encoded by the bait gene. In the past, assays of protein–protein interactions relied on biochemical techniques. Recently, genetic approaches, such as the yeast two-hybrid system, have proven extremely powerful in isolating new plant genes [45,46].

The yeast two-hybrid system relies on the physical interaction of a 'bait' protein fused to a DNA binding domain with a 'prey' protein fused to a transcription activation domain. Interaction of the bait and prey proteins activates expression of a reporter gene by localizing the transcriptional activation domain near the cognate promoter [46]. The two-hybrid system has been successful in testing interactions between specific plant proteins that were predicted to interact [47,48]. In addition, new genes have been discovered by using 'interaction hunts' involving a characterized bait protein and a cDNA library fused to the transcription activation domain. Using this approach, genes have been isolated that play a role in signal transduction [49,50•,51], transcription [52] and a metabolic pathway [53].

Despite this progress, the yeast two-hybrid system has several shortcomings. Particularly large or hydrophobic proteins often do not make good bait proteins, presumably because they do not fold properly or have difficulty entering the yeast nucleus. In addition, proteins with regions rich in acidic amino acids are sometimes found to activate the reporter genes autonomously when fused to the DNA-binding domain precluding their use as bait proteins. Introduction of a transcriptional repressor domain in-frame with the bait protein has been shown to suppress autoactivation of a reporter gene by the p53 protein [54] and may prove useful in identifying interactors with plant transcription factors and other 'autoactivators'. It might also be possible to surmount reporter gene autoactivation by truncating the bait protein, although this could interfere with the identification of interacting proteins that require the complete protein for interaction. To circumvent some of these limitations, new protein–protein interaction systems have been developed that do not rely on transcriptional read-out [55•,56].

The Sos-recruitment system avoids the requirement for reporter gene transcription by relying on protein–protein interactions directly in the yeast cytoplasm [55•]. The system depends on restoration of growth to a temperature-sensitive yeast GDP-GTP exchange factor (GEF) mutant, by expression of a truncated mammalian GEF (hSos) that must be localized to the plasma membrane. A bait protein is fused to hSos and the prey protein (or cDNA library) is fused to a myristylation targeting sequence. When the two fusion proteins are expressed in the yeast GEF mutant, cells grow only if the bait protein physically interacts with the myristylated prey fusion protein, thus recruiting hSos to the plasma membrane. The Sos-recruitment system was tested using the previously known interaction of c-Jun and c-Fos, and was also used to isolate new c-Jun interacting proteins [55•]. A potential drawback of the system is the large size of the truncated hSos protein (115 kDa) which, in fusion with large bait proteins, might lead to low expression or hinder physical interactions.

Another promising approach for detection of protein–protein interactions in the cytoplasm uses intracistronic complementation of the *lacZ* gene [56]. The system relies on the creation of bait and prey fusion proteins with mutant α and ω domains of the β-galactosidase protein that are active yet show reduced affinity for one another. Interaction of the bait protein with a prey forces interaction of the β-gal peptides and reconstitution of β-galactosidase activity. The system was developed using mammalian cells but it seems likely that it might be adapted for screening cDNA libraries against a specific bait protein in yeast. A similar approach for testing protein–protein interactions directly in plant cells would be especially useful, but the high

endogenous β-galactosidase activity in plant cells would probably preclude the direct application of this system.

Finally, because protein–protein interaction screens often give rise to many candidate interactors, it is important that strict criteria be developed for testing the biological significance of putative partner proteins. Such criteria include testing specificity of the interaction by substituting a closely related but non-functional partner protein in the assay [49,50•] and demonstrating interaction *in vivo* by co-immunoprecipitation or other biochemical techniques [53]. Ultimately, the most convincing evidence for biological significance is to show that perturbation of the candidate gene expression by mutagenesis, antisense, or overexpression alters the phenotype under study [49].

Genomics/bioinformatics

Bioinformatics — the application of high volume information technology to biological databases — has created a paradigm shift in gene discovery. Large-scale sequencing of cDNA and genomic clones and the analysis of databases that contain this information underlie many gene discovery approaches [57,58••,59]. A clear match to an anonymous cDNA or previously characterized gene in the database simplifies the identification of transcribed regions within large inserts of BAC or YAC clones, candidate transposon-tagged genes, or genes encoding specific protein-interactors. A recent effort to identify overlapping ESTs derived from the same gene and group them into cDNA contigs further increases the utility of the *Arabidopsis* database for gene identification [60•].

To date, large numbers of cDNA sequences are publicly available for rice and *Arabidopsis* and numerous EST projects are underway for other model organisms and economic species (see http://probe.nalusda.gov, and the dbEST database at http://www.ncbi.nlm.gov/). Large cDNA databases for corn, soybean, and probably other crops have been developed by private industry but, to date, these have not been released publicly. Refinements in optimizing the sequence information in the *Arabidopsis* database and in search software have accelerated progress in identifying large numbers of genes based on specific motifs [11••,59]. For example, it has recently been possible using a search of the EST database, to identify a large number of cDNA clones that encode proteins with similarity to bacterial cellulose synthase [59]. Previous to this work, cellulose synthase genes had been notoriously difficult to isolate and this 'cloning in silico' promises to facilitate research on plant cell walls.

The pattern of gene expression in different tissues and developmental stages, and in response to specific stresses provides insight into gene function and, thus, plays an important role in discovery of new genes. In the past year, there has been exciting progress in the development of new techniques to analyze gene expression on a genome-wide basis. Two especially promising methods are hybridization to microarrays of characterized cDNAs [61,62,63••], and serial analysis of gene expression [64••]. The power of cDNA microarray technology when applied to the full collection of yeast genes and a large collection of human genes has been recently demonstrated [62,63••]. *Arabidopsis* and rice, with their large cDNA collections and EST databases, are prime candidates for similar analyses. The serial analysis of gene expression technique has been used to analyze gene expression differences between cancerous and normal human cells [64••]. From an examination of over 300,000 transcripts, 500 were identified that were differentially expressed. A similar approach to examine gene expression in plants during pathogen infection or environmental stresses, for example, would be expected to yield large numbers of new genes which may play a role in plant response to these events. Genome-wide expression differences in plants carrying specific mutations or overexpressed transgenes should similarly aid in gene discovery as well as functional analysis of previously characterized genes.

Conclusion

Convergence of genetic and physical maps, transposon tagging, protein–protein interaction screens, and the burgeoning information from genomics projects all point to rapid progress in the identification of new genes for crop improvement. The challenge will be to elucidate the function of these genes at the molecular level and to manipulate the genes in crop plants to produce cultivars with enhanced disease resistance, stress tolerance, quality traits, and the potential for production of useful new products.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Bennetzen JL, Freeling M: **Grasses as a single genetic system: genome composition, colinearity, and compatibility.** *Trends Genet* 1993, **9**:259-261.

2. Havukkala IJ: **Cereal genome analysis using rice as a model.** *Curr Opin Genet Dev* 1996, **6**:711-714.

3. Lagercrantz U, Putterill J, Coupland G, Lydiate D: **Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome colinearity and congruence of genes controlling flowering time.** *Plant J* 1996, **9**:13-20.

4. Paterson AH, Lin Y-R, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu S-C, Stansel JW, Irvine JE: **Convergent domestication of cereal crops by independent mutations at corresponding genetic loci.** *Science* 1995, **269**:1714-1718.

5. •• Chen M, Sanmiguel P, De Oliveira AC, Woo S-S, Zhang H, Wing RA, Bennetzen JL: **Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes.** *Proc Natl Acad Sci USA* 1997, **94**:3431-3435.

This first detailed analysis of colinearity at the sub-centimorgan level among three grass genomes indicates that gene isolation from the larger grass genomes can be expedited, in some cases, by cross-referencing with the rice and sorghum genomes.

6. •• Avramova Z, Tikhonov A, Sanmiguel P, Jin Y-K, Liu C, Woo S-S, Wing RA, Bennetzen JL: **Gene identification in a complex chromosomal continuum by local genomic cross-referencing.** *Plant J* 1996, **10**:1163-1168.

A clear demonstration that gene identification can be facilitated within a syntenic region by relying on the DNA sequence conservation of genes when compared with divergent repetitive DNAs. This technique should be generalizable to other plant species and may also permit the identification of conserved sequences other than genes (e.g. centromeres).

7. •• Paterson AH, Feldmann KA, Wendel JF: **Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence.** *Nat Genet* 1996, **14**:380-383.

This compelling report of conserved gene order between Arabidopsis (a dicot) and sorghum (a monocot) indicates that genomic cross-referencing might be possible between well-characterized model plants and crop species where more agronomic traits have been mapped.

8. Leister D, Ballvora A, Salamini F, Gebhardt C: **A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants.** *Nat Genet* 1996, **14**:421-429.

9. Yu YG, Buss GR, Maroof MAS: **Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site.** *Proc Natl Acad Sci USA* 1996, **93**:11751-11756.

10. Kanazin V, Marek LF, Shoemaker RC: **Resistance gene anologs are conserved and clustered in soybean.** *Proc Natl Acad Sci USA* 1996, **93**:11746-11750.

11. •• Botella MA, Coleman MJ, Hughes DE, Nishimura MT, Jones JDG, Somerville SC: **Map positions of 47 *Arabidopsis* sequences with similarity to disease resistance genes.** *Plant J* 1997, **12:**1197-1211.

This comprehensive identification and mapping of many *R* gene homologs in *Arabidopsis*, (and in other species [8–10]), demonstrates the power of combining degenerate PCR and EST database searches with genetic and physical mapping. A similar strategy aplied to other gene products with conserved features (e.g. transcription factors) should greatly expedite gene identification when combined with mutant analysis.

12. Young ND: **QTL mapping and quantitative disease resistance in plants.** *Annu Rev Phytopathol* 1996, **34**:479-501.

13. Grandillo S, Ku H-M, Tanksley SD: **Characterization of fs8.1, a major QTL influencing fruit shape in tomato.** *Mol Breeding* 1996, **2**:251-260.

14. •• Alpert KB, Tanksley SD: **High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: a major fruit weight quantitative trait locus in tomato.** *Proc Natl Acad Sci USA* 1996, **93**:15503-15507.

This is the first example where very high resolution mapping has made feasible the map-based cloning of an economically important QTL. The fw2.2 gene will likely be the first QTL to be molecularly characterized and its study should lead to new insights into fruit development.

15. Eshed Y, Gera G, Zamir D: **A genome-wide search for wild-species alleles that increase horticulture yield of processing tomatoes.** *Theor Appl Genet* 1996, **93**:877-886.

16. Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD: **Advanced backcross QTL analysis of tomato. I. Identification of QTL for traits of agronomic importance from *Lycopersicon hirsutum*.** *Theor Appl Genet* 1998, in press.

17. •• Tanksley SD, McCouch SR: **Seed banks and molecular maps: unlocking genetic potential from the wild.** *Science* 1997, **277**:1063-1066.

This excellent review article discusses inherent limitations in the current approach to utilizing exotic germplasm and presents a new paradigm for molecular breeding based on QTL mapping. The advanced backcross QTL mapping method will have a major impact on germplasm utilization for plant breeding and gene isolation.

18. Tanksley SD, Ganal MW, Martin GB: **Chromosome landing: a new paradigm for map-based cloning in species with large genomes.** *Trends Genet* 1995, **11**:63-68.

19. Cao H, Glazebrook J, Clarke JD, Volko S, Dong X: **The *Arabidopsis* NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats.** *Cell* 1997, **88**:57-63.

20. Cai D, Kleine M, Kiflr S, Harloff H-J, Sandal NN, Marcker KA, Klein-Lankhorst RM, Salentijn EM, Lange W, Stiekema WJ *et al.*: **Positional cloning of a gene for nematode resistance in sugar beet.** *Science* 1997, **275**:832-834.

21. Ori N, Eshed Y, Paran I, Preseting G, Aviv D, Tanksley SD, Zamir D, Fluhr R: **The *I2C* family from the wilt disease resistance locus *I2* belongs to the nucleotide binding leucine-rich repeat superfamily of plant resistance genes.** *Plant Cell* 1997, **9**:521-532.

22. •• Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J *et al.*: **The barley *Mlo* gene: a novel control element of plant pathogen resistance.** *Cell* 1997, **88**:695-705.

This is an especially elegant example of map-based cloning and subsequent verification of an important *R* gene from a plant species with a genome almost twice the size of the human genome.

23. Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M: **AFLP: a new concept for DNA fingerprinting.** *Nucleic Acids Res* 1995, **23**:4407-4414.

24. Taramino G, Tingey S: **Simple sequence repeats for germplasm analysis and mapping in maize.** *Genome* 1996, **39**:277-287.

25. Kantety RV, Zeng X, Bennetzen JL, Zehr BE: **Assessment of genetic diversity in dent and popcorn (*Zea mays* L.) inbred lines using inter-simple sequence repeat (ISSR) amplification.** *Mol Breeding* 1995, **1**:365-373.

26. Bendahmane A, Kanyuka K, Baulcombe DC: **High-resolution genetical and physical mapping of the *Rx* gene for extreme resistance to potato virus X in tetraploid potato.** *Theor Appl Genet* 1997, **95**:153-162.

27. Monna L, Miyao A, Zhong HS, Yano M, Iwamoto M, Umehara Y, Kurata N, Hayasaka H, Sasaki T: **Saturation mapping with subclones of YACs: DNA marker production targeting the rice blast disease resistance gene, Pi-b.** *Theor Appl Genet* 1997, **94**:170-176.

28. Nakamura S, Asakawa S, Ohmido N, Fukui K, Shimizu N, Kawasaki S: **Construction of an 800-kb contig in the near-centomeric region of the rice blast resistance gene Pi-ta2 using a highly representative rice BAC library.** *Mol Gen Genet* 1997, **254**:611-620.

29. Somerville C: **The physical map of an *Arabidopsis* chromosome.** *Trends Plant Sci* 1996, **1**:3.

30. Zhang HB, Choi SD, Woo SS, Li ZK, Wing RA: **Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population.** *Mol Breeding* 1996, **2**:11-24.

31. Choi S, Wing RA: **The construction of bacterial artificial chromosome (BAC) libraries.** In *Plant Molecular Biology Manual* Edited by Gelvin S, Schilperoot R. Drodrecht: Kluwer Academic Publishers 1998; in press.

32. • Hamilton CM: **A binary-BAC system for plant transformation with high-molecular-weight DNA.** *Gene* 1997, **200**:107-116.

Direct transformation of BAC clones using this novel vector system will expedite map-based cloning in plant species that are readily transformed by *Agrobacterium*.

33. • Adam G, Mullen JA, Kindle K: **Retrofitting YACs for direct DNA transfer into plant cells.** *Plant J* 1997, **11**:1349-1358.

Addition of plant selectable markers to the arms of YAC clones spanning a target locus allows direct transformation into plant cells and increases the usefulness of many plant YAC libraries.

34. Century KS, Shapiro AD, Repeti PP, Dahlbeck D, Holub E, Staskawicz BJ: **NDR1, a pathogen-induced component required for *Arabidopsis* disease resistance.** *Science* 1997, **278**:1963-1965.

35. •• Seki M, Hayashida N, Kato N, Yohda M, Shinozaki K: **Rapid construction of a transcription map for a cosmid contig of *Arabidopsis thaliana* genome using a novel cDNA selection method.** *Plant J* 1997, **12**:481-487.

This refinement of earlier cDNA selection methods provides a powerful approach to quickly identify transcribed regions within a target region to expedite map-based cloning.

36. Gray J, Close PS, Briggs SP, Johal GS: **A novel suppressor of cell death in plants encoded by the Lls1 gene of maize.** *Cell* 1997, **89**:25-31.

37. Bhatt AM, Page T, Lawson EJR, Lister C, Dean C: **Use of Ac as an insertional mutagen in *Arabidopsis*.** *Plant J* 1996, **9**:935-945.

38. Cooley MB, Goldsbrough AP, Still DW, Yoder JI: **Site-selected insertional mutagenesis of tomato with maize Ac and Ds elements.** *Mol Gen Genet* 1996, **252**:184-194.

39. • Azpiroz-Leehan R, Feldmann KA: **T-DNA insertion mutagenesis in *Arabidopsis*—going back and forth.** *Trends Genet* 1997, **13**:152-156.

An excellent review of T-DNA tagging for gene cloning that also presents a multiplex approach for reverse genetics.

40. • Frey M, Stettner C, Gierl A: **A general method for gene isolation in tagging approaches: amplification of insertion mutagenized sites (AIMS).** *Plant J* 1998, in press.
A novel approach to identify transposon or T-DNA tagged genes that utilizes AFLP-based technology to simplify the identification of DNA fragments which co-segregate with the mutant allele. Will be especially useful in maize but also in other systems where high copy number of transposons obscures co-segregation analysis.

41. Smith D, Yanai Y, Liu Y-G, Ishiguro S, Okada K, Shibita D, Whittier RF, Fedoroff NV: **Characterization and mapping of Ds-GUS-T-DNA lines for targeted insertional mutagenesis.** *Plant J* 1996, **10**:721-732.

42. Jarvis P, Belzile F, Pages T, Dean C: **Increased Ac excision (iae): *Arabidopsis thaliana* mutations affecting Ac transposition.** *Plant J* 1997, **11**:907-919.

43. • Babiychuk E, Fuanghthong M, Van Montagu M, Inze D, Kushnir S: **Efficient gene tagging in *Arabidopsis thaliana* using a gene trap approach.** *Proc Natl Acad Sci USA* 1997, **94**:12722-12727.
A T-DNA carrying a kanamycin resistance gene was constructed that generates a translational fusion when it integrates into a plant gene. This system increase the recovery of plants with T-DNA-tagged genes and is applicable to other plant species.

44. Hirochika H: **Retrotransposons of rice: their regulation and use for genome analysis.** *Plant Mol Biol* 1997, **35**:231-240.

45. Mcnabb DS, Guarente L: **Genetic and biochemical probes for protein–protein interactions.** *Curr Opin Biotechnol* 1996, **7**:557-559.

46. Fields S, Sternglanz R: **The two-hybrid system: an assay for protein–protein interactions.** *Trends Genet* 1994, **10**:286-292.

47. Tang X, Frederick RD, Zhou J, Halterman DA, Jia Y, Martin GB: **Initiation of plant disease resistance by physical interaction of AvrPto and the Pto kinase.** *Science* 1996, **274**:2060-2063.

48. Scofield SR, Tobias CM, Rathjen JP, Chang JH, Lavelle DT, Michelmore RW, Staskawicz BJ: **Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato.** *Science* 1996, **274**:2063-2065.

49. Zhou J, Loh Y-T, Bressan RA, Martin GB: **The tomato gene Pti1 encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response.** *Cell* 1995, **83**:925-935.

50. • Zhou J, Tang X, Martin GB: **The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a cis-element of pathogenesis-related genes.** *EMBO J* 1997, **16**:3207-3218.
This paper and that of Zhou *et al.* 1995 [49] are early reports of using the yeast two-hybrid system to isolate genes involved in plant signal transduction and defense responses.

51. De Veylder L, Segers G, Glab N, Casteels P, Van Montagu M, Inze D: **The Arabidopsis Cks1At protein binds the cyclin-dependent kinases Cd2aAt and Cdc2bAt.** *FEBS Lett* 1997, **412**:446-452.

52. Schultz TF, Quatrano RS: **Characterization and expression of a rice RAD23 gene.** *Plant Mol Biol* 1997, **34**:557-562.

53. Wedel N, Soll J, Paap B: **CP12 provides a new mode of light regulation of Calvin cycle activity in higher plants.** *Proc Natl Acad Sci USA* 1997, **94**:10479-10484.

54. Cormack RS, Somssich IE: **Dampening of bait proteins in the two-hybrid system.** *Anal Biochem* 1997, **248**:184-186.

55. • Aronheim A, Zandi E, Hennemann H, Elledge SJ, Karin M: **Isolation of an Ap-1 repressor by a novel method for detecting protein–protein interactions.** *Mol Cell Biol* 1997, **17**:3094-3102.
This new genetic screening system detects protein–protein interactions in the yeast cytoplasm rather than relying on transcription of reporter genes. It should provide a useful alternative for bait proteins that are troublesome in the standard two-hybrid system.

56. Rossi F, Charlton CA, Blau HM: **Monitoring protein–protein interactions in intact eukaryotic cells by β-glacatosidase complementation.** *Proc Natl Acad Sci USA* 1997, **94**:8405-8410.

57. Delseny M, Cooke R, Raynal M, Grellet F: **The *Arabidopsis thaliana* cDNA sequencing projects.** *FEBS Lett* 1997, **403**:221-224.

58. •• Somerville C, Flanders D, Cherry JM: **Plant biology in the post-Gutenberg era.** *Plant Physiol* 1997, **113**:1015-1022.
A comprehensive and very clear introduction to the vast amount of information on plant biology available on the World Wide Web. Web sites covering all aspects of plant biology from taxonomy to DNA sequence databases are discussed making this a useful guide which saves much random browsing time.

59. Cutler S, Somerville C: **Cellulose synthesis: cloning in silico.** *Curr Biol* 1997, **7**:108-111.

60. • Rounsley SD, Glodek A, Sutton G, Adams MD, Somerville CR, Venter JG, Kerlavage AR: **The construction of *Arabidopsis* expressed sequence tag assemblies.** *Plant Physiol* 1996, **112**:1177-1183.
Redundant ESTs in the Arabidopsis database were grouped using computer programs optimized on the human EST collection to develop a minimal set of cDNA contigs. This approach will prove generally useful for all plant EST projects and will facilitate the development of cDNA microarray chips.

61. Schena M: **Genome analysis with gene expression microarrays.** *Bioessays* 1996, **18**:427-431.

62. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW: **Parallel human genome analysis: microarray-based expression monitoring of 1000 genes.** *Proc Natl Acad Sci USA* 1996, **93**:10614-10619.

63. •• Derisi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science* 1997, **278**:680-686.
This paper and Zhang *et al.* 1997 [64••] have used the same experimental method to analyze gene expression. They both provide excellent examples of the tremendous potential for genome-wide expression analysis. In both papers, large numbers of genes were shown to be differentially expressed as a result of environmental stimuli or specific mutations in the cell lines. Application of these approaches to plants should be straightforward and will provide unprecedented insight into the molecular basis of economically important traits.

64. •• Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW: **Gene expression profiles in normal and cancer cells.** *Science* 1997, **276**:1268-1272.
This report and Derisi *et al.* 1997 [63••] are excellent examples of the tremendous potential for genome-wide expression analysis. In both papers, large numbers of genes were shown to be differentially expressed as a result of environmental stimuli or specific mutations in the cell lines. Application of these approaches to plants should be straightforward and will provide unprecedented insight into the molecular basis of economically important traits.