Plant Genome Analysis: The State of the Art

Christiane Gebhardt,* Renate Schmidt,† and Katharina Schneider‡
*Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany
†Max Planck Institute for Molecular Plant Physiology, 14476 Golm, Germany
‡GSF-National Research Center for Environment and Health, 85764 Neuherberg, Germany

Plants are the basis for the survival of all ''higher'' organisms on Earth. Development of molecular genetics tools has allowed analysis of the structure, evolution, and function of whole plant genomes, rather than individual genes. DNA-based markers were instrumental in constructing detailed genetic maps of model plants and all major crop species. These molecular maps were the basis of physical maps and the first plant whole genome sequences. Comparative analysis based on genetic, cytogenetic, and physical maps and DNA sequence information provided new insights into the evolution of plant nuclear and organellar genomes. Mapping factors controlling Mendelian and quantitative traits made possible the cloning and functional characterization of novel genes, which function in plant development, adaptation to biotic and abiotic stress, or in the formation of other agronomic characters. The parallel analysis of all transcripts, proteins, and metabolites present in plant cells or tissues has generated information that may lead to a better integrated understanding of genome function. Postfunctional analysis of natural variation of gene function and its effects on phenotype is envisaged to provide new diagnostic and therapeutic molecular tools for applications in plant breeding, adaptation, and ecology.

KEY WORDS: Plants, Genome, Linkage maps, Synteny, Gene function, Cytogenetics, Genome sequence, Genomics.

0074-7696/05 $35.00
DOI: 10.1016/S0074-7696(05)47005-9

I. Introduction

The term ''genome analysis'' circumscribes scientific approaches and methodologies that aim at elucidation of the structure, evolution, and function of whole genomes rather than individual genes. These approaches and methodologies are similar for all eukaryotic genomes such as yeast, plants, animals, and humans. The state of the art of genome analysis varies greatly, however, for individual species, depending on genome complexity, suitability for the addressing of questions of fundamental or applied science, and economic importance. Unique features of plant genomes compared with other eukaryotes are the chloroplast genome, which is essential for photosynthesis, and the polyploidy of many plant species, including important crops such as bread wheat (*Triticum aestivum*), potato (*Solanum tuberosum*), oilseed rape (*Brassica napus*), and sugar cane (*Saccharum* spp.).

Analysis of plant genomes began with the cytogenetic description and identification of the 10 chromosomes of maize (*Zea mays*) (McClintock, 1929) followed a few years later by the construction of the first genetic linkage maps, using as landmarks morphological traits that followed simple Mendelian inheritance patterns (Emerson *et al.*, 1935). To date, cytogenetic and linkage maps are still, or again, important tools for the structural analysis of plant genomes. This is because their analytical power has dramatically improved with the introduction of molecular genetics tools. Molecular cytogenetics and linkage mapping can be applied to a wide range of plant species at comparatively modest input of human and financial resources. Physical maps and whole genome sequences represent the ultimate precision in structural genome analysis (*Arabidopsis* Genome Initiative, 2000), but are restricted to only a few plant species so far, because of the high investments required. Nevertheless, it is expected that over the next 10 years the number of fully sequenced plant genomes will increase.

Our understanding of plant genome evolution has greatly advanced by comparative mapping and genomic sequencing in different species that cannot be hybridized by sexual crosses. Genome-wide DNA polymorphisms also provide accurate descriptors of genetic diversity and are being used to address questions of plant evolution, taxonomy, and crop domestication.

The functional analysis of plant genomes involves the localization of genetic factors, which control Mendelian or complex phenotypic traits, on molecular maps. This is the first step toward the identification and positional cloning of the underlying genes. Large, comprehensive mutant collections, which are the basis for functional analysis of the wild-type genes, and the genome-wide evaluation of gene expression and its consequences in different tissues and in response to environmental parameters are rapidly evolving tools with which to elucidate functional networks. The analysis of natural

variation of gene function in plant populations reaches beyond the analysis of gene function per se. Linkage disequilibrium studies are becoming an important area of plant genome research, as they can elucidate the molecular basis of complex phenotypes such as crop yield and adaptation to biotic and abiotic stress.

In a review on such a broad topic as "plant genome analysis," it is not possible to give credit to all the good and extensive work that has been done on a large number of different plant species in laboratories, large and small, all over the world. Instead, we try to summarize concepts, technologies, results, and trends on the basis of illustrative examples and to put the "state of the art of plant genome analysis" in a historical context.

II. Molecular Linkage Maps of Plant Species and Their Applications

Molecular linkage maps are constructed by estimating the frequency of recombination between DNA-based markers, which jointly segregate in experimental mapping populations (Gebhardt and Salamini, 1992). All DNA-based markers result from natural DNA variants that distinguish individuals of the same species. The molecular basis of this variation consists of point mutations (single-nucleotide polymorphisms, SNPs), insertions, deletions (InDels), or inversions of DNA fragments in one allele versus another. In contrast to induced mutations, which often have severe phenotypic effects, these DNA polymorphisms have survived through evolutionary time because they did not compromise the viability and competitiveness of the individuals that carried them. The basic advantage of DNA-based markers is their unlimited number and, at least for the majority of them, phenotypic neutrality, which allows the construction of detailed linkage maps with thousands of loci (Harushima *et al*., 1998; Isidore *et al*., 2003). The methods to detect DNA polymorphisms coevolved with molecular genetics tools. The first molecular linkage maps of plant species were constructed on the basis of restriction fragment length polymorphisms (RFLPs) (Bernatzky and Tanksley 1986; Bonierbale *et al*., 1988; Botstein *et al*., 1980; Chang *et al*., 1988 Helentjaris *et al*., 1986; Landry *et al*., 1987; McCouch *et al*., 1988).

With the invention of the polymerase chain reaction (PCR, Saiki *et al*., 1988) new, PCR-based marker technologies were developed, which facilitated linkage mapping in plants and largely replaced RFLP technology. Random-amplified polymorphic DNA (RAPD; Welsh and McClelland, 1990; Williams *et al*., 1990) and amplified fragment length polymorphism (AFLP; Vos *et al*., 1995) markers are generated with arbitrary primers that do not require a priori sequence information. Several hundred markers can

be generated *de novo* within a short time from less than 1 μg of genomic DNA. These marker types may be used to construct molecular linkage maps in plant species that were not previously accessible to genetic analysis, such as trees (Binelli and Bucci, 1994; Grattapaglia and Sedoroff, 1994; Tulsieram *et al.*, 1992), or to saturate low-density molecular maps (Menendez *et al.*, 1997; van Eck *et al.*, 1995). AFLP markers, in particular, are suitable for the genome-wide assessment of genetic similarity between accessions of the same species or between closely related species. Questions on species boundaries (Van den Berg *et al.*, 2002), taxonomic relationships (Dehmer and Hammer, 2004; Mace *et al.*, 1999), and crop domestication (Badr *et al.*, 2000; Heun *et al.*, 1997) have been addressed in this way. Microsatellites or simple sequence repeats (SSRs) (Tautz and Renz, 1984; Tautz *et al.*, 1986) are PCR-based markers that require sequence information for specific primer design. Because of their high degree of polymorphism and heterozygosity (Powell *et al.*, 1996), SSRs are today's most popular markers in plant genetics, particularly in crop species, such as wheat (Röder *et al.*, 1998) and soybean (Akkaya *et al.*, 1995), that present notorious difficulties in the detection of intraspecific DNA variation. In the genome-sequencing era, the availability of DNA sequence information is no longer limiting for marker and map development. PCR-based markers can be designed and mapped for virtually any plant gene for which sequence information is accessible in the literature and in databases (Chen *et al.*, 2001). Gene-based markers are most valuable in the biological context, as the DNA polymorphisms present in functional genes may be causal for natural phenotypic variation in plant populations (Fridman *et al.*, 2004, Thornsberry *et al.*, 2001). The most direct and precise method to detect DNA variation is comparative sequence analysis in different individuals of the same species. The ultimate DNA-based markers are therefore SNPs. With the reduction in sequencing costs, SNPs will gain more and more importance in plant genome analysis. At present, SNP discovery and analysis on the genome-wide scale is limited to a few plant species (Buckler *et al.*, 2001; Cho *et al.*, 1999; Rickert *et al.*, 2003; Schneider *et al.*, 2001).

On the basis of RFLP, RAPD, AFLP, and SSR markers, molecular linkage maps have been constructed for more than 60 plant species, spanning the alphabet from A (alfalfa) to W (wheat), including cereals, legumes, as well as root, tuber, and oil crops, fruit and drug species, vegetables, trees, and, of course, model plants (Riera-Lizarazu *et al.*, 2001).

A. Comparative Mapping of Plant Genomes

Before the advent of DNA-based markers, genetic analysis of a given plant species was restricted to individuals of the same species and its closest

relatives, because of the limitations imposed by feasibility of sexual hybridization. Genetic systems were developed for one species, for example, maize, or a single plant family, for example, the cereals, which were isolated from the rest of the plant kingdom. The RFLP assay is based on nucleic acid hybridization between a labeled marker probe and a membrane-bound genomic target sequence. Hence, depending on the experimental conditions used, cross-hybridization is detected between DNA sequences that are not identical but similar and RFLP markers originating from one species can be used for the construction of linkage maps in related species irrespective of their cross-compatibility. The observation of conserved genetic linkage between marker loci that share sequence similarity in different species indicates structural similarity between the different genomes (macrosynteny). RFLP markers made it possible to compare the structure of plant genomes over a wider taxonomic range than was previously possible. The earliest examples of this approach were comparisons between the genomes of potato and tomato, both members of the Solanaceae family (Bonierbale *et al*., 1988), between the A, B, and D genomes of hexaploid bread wheat (AABBDD) (Chao *et al*., 1989), and between maize and sorghum of the Poaceae family (Hulbert *et al*., 1990). Since then, numerous comparative mapping studies have been performed, mostly between different members of the same taxonomic family (Paterson *et al*., 2000). In all comparisons of this kind, structural conservation or synteny was observed, although to various degrees.

For example, the closely related solanaceous species potato (*Solanum tuberosum*) and tomato (*Solanum lycopersicum*) have the same basic number of 12 chromosomes and similar DNA content (Bennett and Smith, 1991). The RFLP linkage maps of the 12 potato and tomato chromosomes show the same linear order of marker loci detected with the same probe, except for 5 map segments having an inverted marker order, which are interpreted as paracentric inversions of chromosome arms (Tanksley *et al*., 1992). The comparison of tomato/potato with solanaceous pepper (*Capsicum annuum*), which has a much higher DNA content but the same basic chromosome number, revealed many more rearrangements, including translocations and inversions of several large chromosomal segments (Livingstone *et al*., 1999). Wheat, barley, rye, rice, maize, sorghum, sugarcane, and many other important crop species are members of the Poaceae family and their genomes vary about 100-fold in size (Bennett and Smith, 1991). Not surprisingly, members of this family were most extensively studied by comparative mapping (Devos and Gale, 2000). These comparisons revealed a degree of structural conservation that was expected for cereals known to be closely related, such as wheat, barley, and rye, but was not anticipated for comparisons between wheat, rice, and maize (Ahn *et al*., 1993). This finding soon led to the concept that genes of interest in species with large genomes such as wheat, maize, or sugarcane may be cloned on the basis of colinearity with smaller sized

genomes from the same family such as rice or sorghum (Bennetzen and Freeling, 1993; Moore *et al.*, 1993). Subsequent comparative sequencing of homologous regions in different grass species revealed cases in which colinearity was retained at the DNA sequence level (microsynteny) (Chen *et al.*, 1997; Feuillet and Keller, 1999). In a growing number of other cases, however, deletions or insertions, duplications, inversions, or translocations of genes have occurred during evolution (Brunner *et al.*, 2003; Feuillet and Keller, 2002; Kilian *et al.*, 1997; Tarchini *et al.*, 2000). Violation of sequence colinearity was even observed between alleles of two inbred lines of maize (Fu and Dooner, 2002).

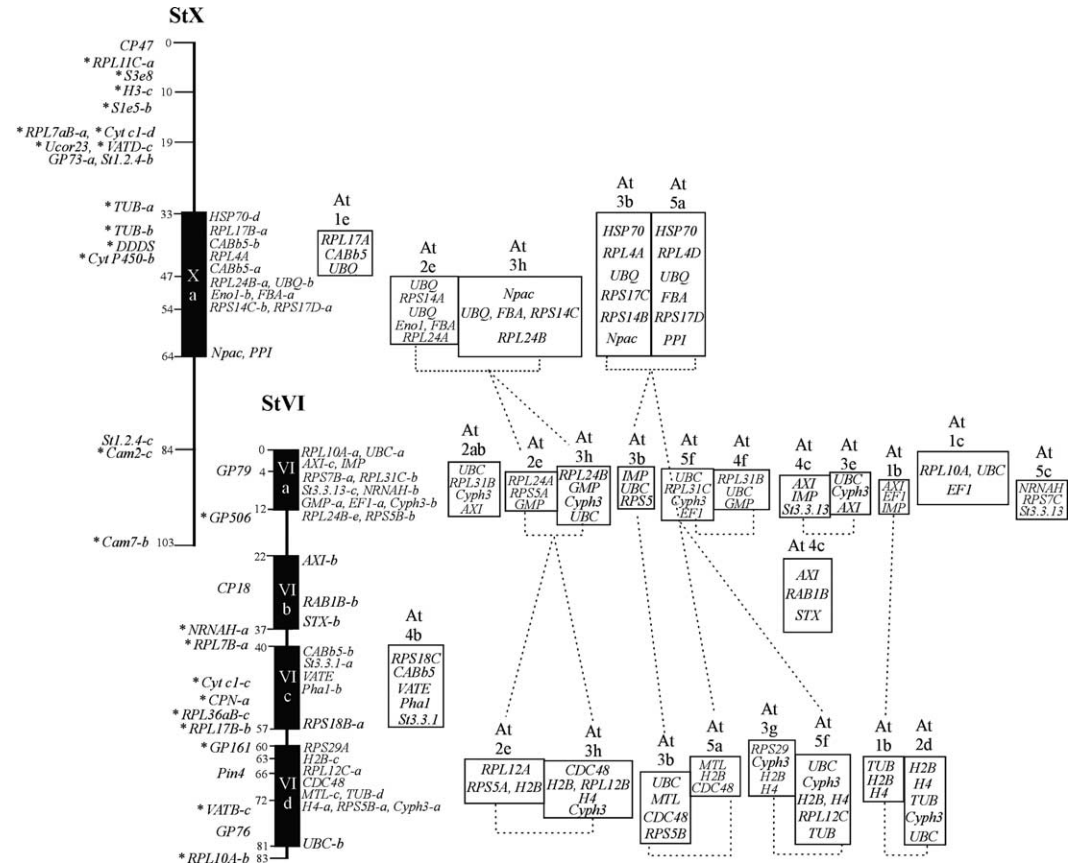
With increasing evolutionary distance between plant species, hybridization-based comparative mapping becomes more difficult and is restricted to "housekeeping" genes such as histone or ribosomal genes, which are highly conserved across plant families and even kingdoms. With the first plant genome sequences of *Arabidopsis thaliana* and rice completed and becoming available in public databases (see Section IV.A), comparisons between plant species, for which only genetic maps are available, and sequenced plant genomes become possible. In principle, marker sequences, primarily of expressed sequence tags (ESTs), which have been localized on the genetic map of a given plant species, are compared with the *Arabidopsis* or rice genome sequence. After defining a threshold for sequence similarity, homologous loci are identified *in silico* on the physical map of the sequenced plant genome. Statistical methods are then used to define threshold criteria for accepting the hypothesis that the observed conservation of genetic and physical linkage between a given number of sequence-related markers results from structural conservation of genome segments (syntenic blocks) between the two species. This approach has been used to compare the genome structure of the dicotyledonous species sugar beet, sunflower, *Prunus* spp., potato, cotton, the legumes soybean and *Medicago truncatula*, and species of the Brassicaceae such as *Brassica oleracea* and *Capsella rubella* with *Arabidopsis thaliana* (Boivin *et al.*, 2004; Bowers *et al.*, 2003, Dominguez *et al.*, 2003; Gebhardt *et al.*, 2003; Lee *et al.*, 2001; Lukens *et al.*, 2003; Zhu *et al.*, 2003). Similarly, genetic maps of the monocotyledonous species maize and wheat were compared with the rice genome sequence (Salse *et al.*, 2004; Sorrells *et al.*, 2003). In all cases studied so far, genome segments with structural conservation were identified, even between distantly related plant species, although fragmented and covering only part of the genomes compared. The types of syntenic relationships revealed when comparing molecular linkage maps of distantly related, dicotyledonous crop plants with the *Arabidopsis* physical map are exemplified in Fig. 1, based on potato chromosomes VI and X (Gebhardt *et al.*, 2003). A major observation is the redundancy of syntenic relationships. Most syntenic blocks are related to several

Arabidopsis genome segments and vice versa. Syntenic relationships of different potato genome segments to the same *Arabidopsis* genome segment indicated that potato genome evolution included ancient intra- and interchromosomal duplication events, which today are largely "hidden" because of subsequent divergence. Similar analysis of other plant species (Dominguez *et al*., 2003) and comparative analysis of the whole genome sequences of *Arabidopsis* and rice, which are now possible, produce similar conclusions (Salse *et al*., 2004; Simillion *et al*., 2002), suggesting that polyploidization and segmental duplications are general mechanisms of plant genome evolution. Remains of an ancestral genome of dicotyledonous plants may have been "excavated" by comparative mapping between sugar beet, sunflower, potato, and *Prunus*, which are all members of different taxonomic families, using as the common point of reference the *Arabidopsis* genome sequence (Dominguez *et al*., 2003). Map segments were independently found in all four crop species, which shared structural similarity with the same *Arabidopsis* genome segment. These structurally conserved genome segments may encode core functions of a plant genome.

The exploitation of genome colinearity for positional gene cloning succeeded in some cases but failed in others (Delseny, 2004). The main reason for failure is the lack of gene-by-gene colinearity at the sequence level, even between closely related species. Despite good overall structural conservation, the gene targeted in a crop plant may be deleted in the corresponding region of the model genome (Brown *et al*., 2003; Brunner *et al*., 2003; Desloire *et al*., 2003; Dunford *et al*., 2002). The redundancy of plant genomes also makes it difficult to identify truly orthologous regions. The extrapolation of gene function from a model genome such as *Arabidopsis thaliana* to other, distantly related plant genomes may be limited to single, orthologous genes (Peng *et al*., 2000) rather than relying on conserved genome structure.

B. Plant Function Maps

Molecular linkage maps provide the positional coordinates for plant genomes, equivalent to geographical longitude and latitude. These molecular coordinates are used to localize the genes that control phenotypic traits of interest on plant chromosomes. Saturated molecular linkage maps allowed for the first time the molecular cloning of novel plant genes on the basis of map position alone, without prior knowledge of the encoded protein (Arondel *et al*., 1992; Giraudat *et al*., 1992; Martin *et al*., 1993). The mapping of a phenotypic trait involves the construction of an experimental plant population (Gebhardt and Salamini, 1992), which segregates for both trait alleles and a sufficient number of markers. The same population is then


StX
CP47
* RPL11C-a
* S3e8
* H3-c
* S1e5-b
* RPL7aB-a, * Cyt c1-d
* Ucor23, * VATD-c
GP73-a, St1.2.4-b
* TUB-a
* TUB-b
* DDDS
* Cyt P450-b
Xa
HSP70-d
RPL17B-a
CABb5-b
RPL4A
CABb5-a
RPL24B-a, UBQ-b
Eno1-b, FBA-a
RPS14C-b, RPS17D-a
Npac, PPI
St1.2.4-c
* Cam2-c
* Cam7-b
StVI
GP79
* GP506
VIa
RPL10A-a, UBC-a
AXI-c, IMP
RPS7B-a, RPL31C-b
St3.3.13-c, NRNAH-b
GMP-a, EF1-a, Cyph3-b
RPL24B-e, RPS5B-b
CP18
Pin4
VIb
AXI-b
RAB1B-b
STX-b
VIc
* NRNAH-a
* RPL7B-a
CABb5-b
St3.3.1-a
VATE
Pha1-b
* Cyt c1-c
* CPN-a
* RPL36aB-c
* RPL17B-b
RPS18B-a
VId
* GP161
H2B-c
RPL12C-a
CDC48
MTL-c, TUB-d
H4-a, RPS5B-a, Cyph3-a
* VATB-c
GP76
UBC-b
* RPL10A-b
At 1e
RPL17A CABb5 UBQ
At 2e
UBQ RPS14A UBQ Eno1, FBA RPL24A
At 3h
Npac UBQ, FBA, RPS14C RPL24B
At 3b
HSP70 RPL4A UBQ RPS17C RPS14B Npac
At 5a
HSP70 RPL4D UBQ FBA RPS17D PPI
At 2ab
UBC RPL31B Cyph3 AXI
At 2e
RPL24A RPS5A GMP
At 3h
RPL24B GMP Cyph3 UBC
At 3b
IMP UBC RPS5
At 5f
UBC RPL31C Cyph3 EF1
At 4f
RPL31B UBC GMP
At 4c
AXI IMP St3.3.13
At 3e
UBC Cyph3 AXI
At 1b
AXI EF1 IMP
At 1c
RPL10A, UBC EF1
At 5c
NRNAH RPS7C St3.3.13
At 4c
AXI RAB1B STX
At 4b
RPS18C CABb5 VATE Pha1 St3.3.1
At 2e
RPL12A RPS5A, H2B
At 3h
CDC48 H2B, RPL12B H4 Cyph3
At 3b
UBC MTL CDC48 RPS5B
At 5a
MTL H2B CDC48
At 3g
RPS29 Cyph3 H2B H4
At 5f
UBC Cyph3 H2B, H4 RPL12C TUB
At 1b
TUB H2B H4
At 2d
H2B H4 TUB Cyph3 UBC

accurately phenotyped for the trait and genotyped with DNA-based markers. Genetic linkage analysis is carried out between the markers and the trait alleles. Detection of linkage between markers of known map position and trait alleles reveals the position of the genes that control the phenotype of interest. This approach is applicable to the mapping of Mendelian traits controlled by single genes as well as complex or quantitative traits, which are controlled by a number of genes and environmental factors. Over the last 15 years, many heritable characters were genetically dissected in this way in a range of plant species including the major crops, for example, mutant phenotypes (Castiglioni *et al.*, 1998), resistance to pests and pathogens (Barone *et al.*, 1990; Calenge *et al.*, 2004; Leonards-Schippers *et al.*, 1994; Schön *et al.*, 1993), adaptation and resistance to abiotic stress such as drought and cold (Saranga *et al.*, 2001; Wheeler *et al.*, 2005), morphological and developmental traits (Börner *et al.*, 2002), crop yield and heterosis (Bezant *et al.*, 1997; Stuber *et al.*, 1992; Xiao *et al.*, 1995), and quality traits (Causse *et al.*, 2004; Menendez *et al.*, 2002; Thomas *et al.*, 1996). Quantitative trait locus (QTL) mapping was the enabling, initial step toward the map-based cloning of QTL for fruit size of tomato (Frary *et al.*, 2000; Grandillo *et al.*, 1999), for tomato sugar content (Eshed and Zamir, 1995; Fridman *et al.*, 2000), and a QTL for photoperiod sensitivity in rice (Yano *et al.*, 2000).

FIG. 1 Synteny between the molecular genetic maps of potato chromosomes VI and X and the sequenced *Arabidopsis* genome (modified from Gebhardt *et al.*, 2003). Sequence similarity between potato EST loci and *Arabidopsis* genes is indicated by similar names. For gene identification and putative function assignment see Gebhardt *et al.* (2003). Potato RFLP loci defining the syntenic blocks are shown on the right of the linkage groups. Loci in the same line cosegregated. RFLP loci shown on the left of the linkage groups were not involved in the syntenic blocks. RFLP loci preceded by an asterisk (*) have sequence similarity with *Arabidopsis* genes but do not participate in syntenic blocks. The remaining loci to the left of linkage groups do not have hits in the *Arabidopsis* genome. For definitions of thresholds for sequence similarity see Gebhardt *et al.* (2003). Lower case letter extensions indicate duplicated RFLP loci that were detected with the same marker probe. The potato map segments showing synteny with *Arabidopsis* are indicated as black bars. The position, in centimorgans, of the borders of the syntenic blocks and some internal positions are shown to the left of the linkage groups. *Arabidopsis* syntenic blocks based on three or more genes are shown on the right of each linkage group and are identified by chromosome number and a lower case letter for the chromosome segment. The height of each block corresponds to the size of the syntenic segment, in centimorgans, on the potato linkage group. The vertical order of the loci in the *Arabidopsis* syntenic blocks corresponds to their order on the physical map. Loci in the same line were located in the same or overlapping *Arabidopsis* BAC clones. Pairs of *Arabidopsis* syntenic blocks that correspond to segmental genome duplications in *Arabidopsis* (*Arabidopsis* Genome Initiative, 2000) are positioned next to each other and are connected by a dotted line. Putative intra- and interchromosomal duplications of potato genome segments are indicated by dotted lines connecting *Arabidopsis* syntenic blocks.

Molecular linkage maps gain additional information value when they are constructed or enriched with functional gene markers. Such markers are derived from or are highly similar to DNA sequences of genes of known function, encoding, for example, enzymes of biochemical pathways, transcriptional regulators, or components of signal transduction chains (Causse *et al.*, 2004; Chen *et al.*, 2001; Leister *et al.*, 1996; Ramalingam *et al.*, 2003; Wheeler *et al.*, 2005). Another source of functional markers is ESTs that show differential transcript levels in genotypes with contrasting phenotypes (resistant/susceptible, wild type/mutant) or under contrasting environmental conditions (high/low nutrient or water supply, light, temperature). The map positions of functional genes can be compared with map positions of related phenotypic traits (see previous discussion), either by mapping both in the same experimental population or by linkage to the same markers mapped in different populations. There are two reasons why functional candidate genes and related phenotypic trait loci colocalize: either the candidate gene and the trait locus are linked but functionally unrelated, or alleles present at the candidate locus are, in fact, responsible for the trait variation. Further experiments such as complementation analysis (Paal *et al.*, 2004) or linkage disequilibrium mapping (see Section II.C) are needed to validate the causal role of a candidate gene in the production of a phenotypic trait. Integrated maps of gene functions and traits provide criteria to prioritize among a large number of functional candidates those genes that might be the most relevant for natural trait variation (Pflieger *et al.*, 2001). For example, integrated function maps for resistance to pests and pathogens have been compiled from several mapping experiments for maize (McMullen and Simcox, 1995) and for the solanaceous species potato, tomato, and pepper (Gebhardt and Valkonen, 2001; Grube *et al.*, 2000; Leister *et al.*, 1996). On the basis of this, a picture emerges for the organization of genes for pathogen resistance in plant genomes, which—in combination with the structural and functional characterization of plant genes for pathogen recognition, signal transduction, and defense responses (reviewed in Martin *et al.*, 2003)—leads to models for the molecular basis of complex plant disease resistance (Gebhardt and Valkonen, 2001). Genes for resistance to different types of plant pathogens (viruses, bacteria, fungi) often cluster in the same map segments, and single genes for resistance (*R* genes) sometimes colocalize with quantitative resistance loci (QRL) for the same or other pathogens. This suggests a common molecular origin for different disease resistance phenotypes, which probably evolved by local gene duplications and subsequent functional divergence. The genome-wide organization of *Arabidopsis* genes, which encode function *R* genes or are sequence-related to *R* genes of other plant species (Holub, 2001), corroborate the model derived from function maps for pathogen resistance in crop plants.

C. Linkage Disequilibrium and Marker–Phenotype Associations in Plant Populations

Whereas linkage mapping is based on estimating the recombination frequency in one meiotic generation between the alleles at physically linked loci, linkage disequilibrium mapping evaluates the extent of coinheritance of alleles over several meiotic generations in populations of individuals related by descent. In principle, the frequency of coinheritance of physically linked alleles decreases with increasing recombination frequency between them and increasing number of meiotic generations, which separate a founder population from a contemporary population studied. Linkage disequilibrium (LD) is highly variable across the genome and depends on the population history (Flint-Garcia *et al*., 2003), which is certainly different for crop plant populations as compared with human populations. Many plant species, including major crops, reproduce by selfing or vegetative propagation. This is expected to increase LD when compared with outcrossing species, because the opportunity or the consequences of meiotic recombination for LD are reduced. Natural interspecific hybridization and man-made introgression of wild species into crop genomes can generate plant populations with allele admixtures from different sources, which may have founder effects and affect population structure. Particularly in crop species, the breeding history and the ''unnatural'' selection by human individuals are relevant parameters to be considered when studying LD.

In human population genetics, studies of LD and marker–phenotype associations are considered the strategy to elucidate the molecular basis of complex inherited diseases. The association between a molecular haplotype and disease phenotype will allow genetic risk assessment and, eventually, personalized therapy (Risch, 2000; Schafer and Hawkins, 1998). In plants, the association between DNA markers or haplotypes and natural phenotypic variation may allow the assessment of genetic potential in populations of varieties, breeding lines and wild species individuals that are related by descent. Markers associated with crop performance will provide excellent diagnostic tools for marker-assisted breeding and will facilitate identification and molecular cloning of the genes that control complex agronomic characters. In populations of wild plant species, haplotype–trait associations will be highly valuable to address at the molecular level questions of ecology, adaptation, and speciation.

The analysis of LD and marker–phenotype associations in plant populations is in its infancy at present but will certainly gain importance in plant genome analysis, as it now becomes more and more approachable, thanks to the availability of molecular maps, function maps, whole genome sequences, and functional genomics tools. In studies of LD in plants, molecular maps, a limited number of DNA-based markers, and small population sizes

compared with LD studies in human populations were used to explore LD in sugarcane cultivars (Jannoo *et al.*, 1999), sugar beet breeding lines (Kraft *et al.*, 2000), maize inbred lines (Remington *et al.*, 2001; Tenaillon *et al.*, 2001), and barley cultivars (Kraakman *et al.*, 2004). The whole genome sequence was the basis for a first analysis of LD in the wildflower *Arabidopsis thaliana* (Nordborg *et al.*, 2002). In all studies conducted so far, LD between DNA marker alleles was observed and ranged from a few hundred base pairs in maize to 50 centimorgans (cM) in *Arabidopsis* (reviewed in Flint-Garcia *et al.*, 2003).

Genome-wide marker–trait association studies require the genotyping of sufficiently large, phenotyped populations with a large number of SNP markers, which is costly and causes statistical problems due to multiple testing (Hirschhorn and Daly, 2005). The genome-wide approach aims at systematic, full coverage of the physical map. It is unbiased but currently rather prohibitive to be performed in most plant species. However, even when genotyping small populations with low to modest numbers of genome-wide AFLP or SSR markers, it was possible to identify marker–trait associations in *Beta vulgaris* ssp. *maritima* for the bolting gene *B* controlling growth habit (Hansen *et al.*, 2001), in *Hordeum spontaneum* for water stress tolerance (Ivandic *et al.*, 2003), and in spring barley (*Hordeum vulgare*) for yield components (Kraakman *et al.*, 2004). This might be due to inbreeding and/or a limited number of meiotic generations separating the individuals in the analyzed populations.

The alternative to the whole genome approach is the candidate gene approach to find marker–trait associations. Function maps, knowledge of biochemical pathways, and functional genomics results can be used to select candidate genes for a trait of interest (see Section II.B). Allelic variation of one or more candidate genes can then be evaluated for association with trait variation in suitable plant populations. Although biased, the candidate gene approach is a valid shortcut to the elucidation of the molecular basis of complex, polygenic traits as has been shown in plants. In a population of 92 maize inbred lines, DNA polymorphisms in the *Dwarf-8* gene, which controls plant height (Peng *et al.*, 1999), were associated with a QTL for flowering time (Thornsberry *et al.*, 2001). Similarly, haplotypes at the *CRY2* flowering time locus in *Arabidopsis* were associated with natural variation of this trait (Olsen *et al.*, 2004). In populations of potato cultivars, markers derived from or tightly linked to single genes for resistance to the oomycete *Phytophthora infestans* causing the late blight disease, or to the fungus *Verticillium dahliae* were associated with quantitative resistance to the same pathogens (Gebhardt *et al.*, 2004; Simko *et al.*, 2004). On the basis of, at present, limited genome coverage, LD seems to extend across considerable physical distance in potato (*Solanum tuberosum*), which is a tetraploid, vegetatively propagated and outcrossing species.

A fine example of how synteny, function maps, map-based QTL cloning, and the candidate gene approach to marker–trait associations can complement each other toward better understanding of natural variation of complex plant characters is a syntenic locus encoding the enzyme invertase on chromosome 9 of both tomato and potato. Invertases are ubiquitous enzymes that hydrolyze sucrose into glucose and fructose. Plant invertases occur in several isoforms and play important roles in source–sink relationships, defense responses to pathogens, growth, and other developmental processes (Tymowska-Lalanne and Kreis, 1998). In tomato (*Solanum lycopersicum*), an introgressed wild species allele at the invertase locus on chromosome 9 was identified by map-based cloning as being causal for a QTL for tomato fruit sugar content (Fridman *et al.*, 2000, 2004). In potato (*Solanum tuberosum*), molecular mapping of cloned invertase genes identified the orthologous locus on chromosome IX (Chen *et al.*, 2001). The potato invertase locus colocalized with a QTL for tuber sugar content, a trait that plays an important role for the culinary quality of processed products such as chips, crisps, and french fries (Menendez *et al.*, 2002). Subsequent analysis of DNA variation at the invertase locus in a population of potato breeding clones identified a natural invertase allele that is associated with better chip quality (Li *et al.*, 2005). This demonstrates that natural variation of the same, well-known gene can be responsible for functionally related but physiologically quite different QTL in related plant species.

III. Plant Molecular Cytogenetics and Physical Maps

A. Molecular Cytogenetics

The description of the karyotype of maize (*Zea mays*) based on the microscopic observation of mitotic chromosome spreads is the first example of a plant genome analysis (McClintock, 1929). Cytogenetics entered the molecular era with the development of the *in situ* hybridization technique (Gall and Pardue, 1969; John *et al.*, 1969), in which a labeled nucleic acid probe is hybridized to denatured DNA of chromosomes spread on a microscope slide. Since then, radioactive labeling methods have been replaced by nonisotopic *in situ* hybridization techniques using biotin, digoxigenin, or fluorescently labeled probes. Fluorescent probes are visualized directly by fluorescence microscopy; other probes are visualized indirectly by the use of fluorescent antibodies (Jiang and Gill, 1994; Schwarzacher, 2003). Because of its sensitivity and spatial resolution, fluorescence *in situ* hybridization (FISH) is now state of the art for the cytogenetic analysis of mitotic, meiotic, and interphase plant nuclei (Schwarzacher, 2003). FISH also allows the simultaneous use

and detection of several probes, each one labeled with a different fluorochrome. Advances in fluorescence microscopy with digital image analysis further increased the number of target sequences, allowing the "painting" of plant chromosomes (e.g., the five chromosomes of *Arabidopsis thaliana*) each in a different color (Lysak *et al.*, 2003). The spatial resolution of FISH performed on highly condensed somatic metaphase chromosomes is in the range of 1 to 3 Mbp (Schwarzacher, 2003). This resolution can be increased to 50–100 kb when probing meiotic pachytene chromosomes, which are 10 to 20 times longer than metaphase chromosomes and easily accessible in pollen cells (Xu and Earle, 1996, Zhong *et al.*, 1999). The highest resolution, down to 1 kb or even less, is reached when probing extended genomic DNA fibers from interphase nuclei (Fransz *et al.*, 1996).

Up to 90% of plant genomes consists of repetitive sequences, some of which are highly conserved whereas others are species specific (Heslop-Harrison, 2000). Total genomic DNA of a given plant species can therefore be specifically labeled with a fluorochrome and used as a probe in genomic *in situ* hybridization (GISH). GISH is being used to detect alien plant chromosomes and chromosome segments in interspecific hybrids and to trace them in the breeding process (Dong *et al.*, 2001; Forsström *et al.*, 2002; Schwarzacher *et al.*, 1992). Various classes of highly repetitive DNA sequences such as telomeric repeats, satellite repeats, ribosomal RNA genes, and retroelements have been cloned and characterized from plant species (Lapitan, 1992). When used as FISH probes, they are powerful descriptors of the large-scale organization of chromosomes and chromatin in various stages of the cell cycle (Harper and Cande, 2000; Heslop-Harrison, 2000; van Driel and Fransz, 2004). The abundance and distribution patterns of rRNA genes and satellite repeats, as detected by FISH, allow, for example, the identification of individual chromosomes (Doudrick *et al.*, 1995), the comparative analysis of the chromosome complements of related species, and the detection of evolutionary changes (Nagaki *et al.*, 1998; Taketa *et al.*, 1999). Structure and dynamic behavior of plant telomeres, centromeres, and heterochromatic regions have been studied by FISH and appropriate repetitive sequence probes (Cheng *et al.*, 2002; Dong and Jiang, 1998; Fransz *et al.*, 2002; Kilian *et al.*, 1995; Murata, 2002).

The sensitivity of FISH is insufficient to detect low copy number targets in large genomes with probes that are a few kilobases in length. This limitation was overcome when genomic libraries with large insertions from 10 kb to more than 100 kb became available (Section III.B). Pachytene FISH using bacterial artificial chromosome (BAC) clones as probes has gained an important role in plant genome analysis as it complements genetic and physical maps. BACs selected from libraries with markers that are linked on the genetic map can be physically mapped by pachytene FISH to the same chromosome. Measuring their physical distance establishes the correlation

to the genetic distance between the markers. Physical contigs (contiguous sequences) consisting of overlapping BACs (Section III.B) can be anchored by FISH analysis to chromosomes and genetic maps, thereby providing important information for whole genome sequencing projects (Section IV.A). Pachytene and fiber FISH have been used, for example, to construct an integrated genetic and physical map of rice chromosome 10 (Cheng *et al.*, 2001) and to compare the structure of plant nuclear and chloroplast genomes (Jackson *et al.*, 2000; Lilly *et al.*, 2001).

B. Physical Maps

A physical map provides a scaffold, measured in physical units of base pairs, for a genomic region or a whole genome. It is based on overlapping genomic DNA fragments represented by large-insert clones. This feature distinguishes it from a genetic map that is based on recombination units measured in centimorgans. Recombination events are unequally distributed over the chromosome with recombination hotspots in distal, gene-rich regions of chromosome arms, whereas the base pair counts of a physical map are an absolute distance measure that is independent from position in the genome. In this respect, physical maps are similar to cytogenetic maps, in which loci are positioned on the basis of their absolute or relative distance from microscopically visual chromosomal landmarks such as telomeres, centromeres, nucleolus-organizing regions (NORs), or others (Harper and Cande, 2000). In most cases, a physical map does not cover whole chromosomes because of technical difficulties in dealing with repetitive DNA (Section III.A), so the focus is on local absolute distances. In this respect the resolution of physical maps exceeds by far that of cytogenetic and genetic maps. Physical maps help do the following:

- Clone by position genes identified by QTL analysis or simple Mendelian inheritance (Section II.B);
- Analyze the "genomic environment," for example, the regulatory regions of candidate genes;
- Develop further genetic markers from a genomic region of interest for marker-assisted breeding;
- Analyze the structural dynamics of the genome, for example, by assessing regions for recombination frequencies and linkage disequilibrium in combination with genetic mapping data and phenotypic data (Section II.C);
- Analyze the genomes of related species on the basis of the synteny concept (Section II.A); and
- Provide a tiling path of minimally overlapping clones as a backbone for efficiently sequencing the whole genome (Section IV.A).

1. Tools and Strategies for Physical Mapping

Appropriate biological source material and efficient methods to physically dissect and assemble genomic regions are needed to construct a physical map. Before the advent of cloning protocols for large DNA fragments and the routine use of restriction endonucleases, physical mapping was restricted to total genomic DNA. Carle and Olson (1984) reported 11 bands of yeast DNA separated by an early type of pulsed-field gel electrophoresis. By DNA–DNA hybridization with specific probes they could associate three of them with chromosomes as a first electrophoretic karyogram spanning more than 900 kb. In a next step, λ phages and cosmids became available as cloning vectors for DNA fragments in the size range from 15 to 40 kb, and the first strategies for DNA fingerprinting by restriction digests were established in *Caenorhabditis elegans* and yeast (Coulson *et al.*, 1986; Olson *et al.*, 1986). Restriction fragment patterns of clones were visualized and compared in order to retrieve overlaps between clones of contiguous sequence, so-called contigs, some of which exceeded 100 kb. The potential of the regional physical maps was already recognized at that time, not only to determine the distance between linked genes, but also to be extended to the whole genome level as a common tool for the research community.

A major breakthrough for the generation of physical maps was the cloning of 300- to 1000-kbp DNA inserts into yeast artificial chromosome (YAC) vectors (Burke *et al.*, 1987). YAC inserts were often found to be chimeric, and clones carrying tandemly repeated sequences exhibited a high degree of instability (Schmidt *et al.*, 1994). Still, YAC libraries were constructed and successfully used for plant species such as *Arabidopsis thaliana* (Grill and Somerville, 1991; Schmidt *et al.*, 1995; Zachgo *et al.*, 1996), maize (Edwards *et al.*, 1992), and rice (Saji *et al.*, 2001). A further improvement was the development of the bacterial artificial chromosome (BAC) cloning system (Shizuya *et al.*, 1992), which allowed the stable cloning of 100- to 200-kbp DNA fragments. This cloning system proved versatile for high-throughput technologies including DNA sequencing and is the most popular at present in plant and animal genome projects. Plant BAC libraries are listed under http://www.chori.org and http://www.genome.clemson.edu. In addition, bacteriophage P1-derived artificial chromosomes (PACs; Ioannou *et al.*, 1994) proved to be reliable cloning vehicles for large genomic DNA inserts to construct physical maps.

Strategies to analyze and assemble DNA fragments had to be refined because of the increased number and size of the genomic DNA inserts in physical mapping projects. Thanks to their higher resolution, DNA-sequencing gels and capillary electrophoresis replaced, to a great extent, agarose gels for the separation of restricted, large genomic DNA inserts. Within a

project to develop a sorghum physical map, an efficient high-throughput AFLP-based method was developed to estimate clone overlaps (Klein *et al.*, 2000). Band visualization is facilitated by either radioactive or fluorescently labeled nucleotides. Banding patterns are scanned and then scored and interpreted by the fingerprint contig (FPC) program developed by Soderlund *et al.* (1997) for contig assembly. The technologies available for DNA fingerprinting, assembly strategies, and their evaluation for physical map construction were reviewed by Meyers *et al.* (2004a). As an alternative to restriction fragment-based strategies, protocols for high-throughput hybridization were developed to construct physical maps. In *Arabidopsis thaliana*, a BAC library was hybridized against both ends of 500 BAC clones in an iterative "sampling without replacement" strategy (Mozo *et al.*, 1998). A minimum of 220 contigs covering 41.8% of the library, which contains 7 genome equivalents, illustrated the feasibility of the approach for a species with a low content of repetitive DNA such as *Arabidopsis*. In most higher plants, however, repeat sequences cover 50–90% of the genome (Heslop-Harrison, 2000). To obtain specific results in these cases, hybridization probes must be scrutinized for uniqueness by testing them against repeat libraries. Gene coding sequences are frequently targeted by hybridization-based approaches because of their low redundancy compared with other genomic regions. To construct a physical map of maize that can be integrated with the genetic map, nearly 10,000 unique 40-mer oligonucleotides were deduced from expressed sequence-tagged unigenes and used for two-dimensional "overgo" hybridizations (Gardiner *et al.*, 2004). Overgo probes are designed to bind to nonrepeat regions and consist of two overlapping, complementary oligonucleotides, which allow radiolabeling to high specific activity by the Klenow enzyme. As an additional benefit, markers developed from expressed sequences have the potential to tag candidate genes for traits of interest. Genetic mapping of the same markers not only serves to integrate both maps, but also corroborates the genotype–phenotype linkage (Section II.B).

A multiparallel oligo-fingerprinting method used 35-mer oligonucleotides derived from EST and BAC end sequences as hybridization probes on a filter-arrayed BAC library for constructing a physical map of the medaka fish (Khorasani *et al.*, 2004). The same method is now being applied to physical mapping of the sugar beet genome (http://www.gabi.de). The advantages of hybridization-based strategies are as follows: first, contigs are directly anchored to the genetic map in case the map position of the cross-hybridizing oligonucleotide is known, whereas contigs from restriction-based approaches are largely anonymous; second, even small overlaps, which are likely to be eliminated by the stringency thresholds in the restriction-based approaches, are detected by hybridizations; and third, the hybridization-based strategy is less susceptible to SNPs, thereby allowing the use of DNA from different

genotypes for the assembly, which is problematic with the restriction-based approach.

Apart from cloning genomic DNA fragments into vectors and transforming them into microbial recipients such as yeast and bacteria, there is also the possibility to immortalize the donor DNA by fusing or integrating it with another, unrelated plant genome and to propagate this chimeric material in the form of radiation hybrid (RH) or addition lines. Whereas addition lines result from interspecific crosses and tissue culture procedures to select lines with additional alien chromosomes, radiation hybrid lines are produced by lethal irradiation of the donor cells to induce random chromosomal breaks followed by subsequent fusion with a suitable recipient cell line from a different species (Barrett, 1992). These strategies are most popular in mammalian genome research. The feasibility of constructing chromosome addition and radiation hybrid lines in maize and barley and the potential of this approach for covering the whole genome have been successfully demonstrated (Kynast *et al*., 2004; Wardrop *et al*., 2002). Genomic regions defined by their breakpoints can be analyzed in a different genetic background devoid of the rest of the donor genome, which might cause problems resulting from duplicated segments and repetitive sequences. This option is also useful if the genomic region of interest proves unclonable by the previously described strategies, and long-range accuracy and map integrity need to be secured. A further tool for physical mapping still waiting to be exploited in plant genome research is the HAPPY mapping technique (Dear and Cook, 1993). It is based on cosegregation analysis of markers amplified in pools of randomly sheared genomic DNA fragments whereby cosegregation is understood as the presence of both marker fragments in the same pool without the requirements of meiotic mapping involving parental polymorphisms (Waugh *et al*., 2002).

2. Local Versus Whole Genome Physical Maps: Comparison of Requirements and Objectives

If interest is focused on a particular genome region, for example, in a positional cloning project, then a local physical map covering the locus of interest is sufficient. In this case the strategy involves construction of a large-insert library from the genotype that contains the desired allele. In particular, the cloning of genes for resistance to plant pathogens, which are often introgressed from wild species, necessitates the generation of a large-insert library from the respective donor in each case. Cloning of the *R1* gene for resistance to late blight in potato (Ballvora *et al*., 2002) is an illustrative example. After genetic fine mapping, BAC clones were produced from the genotype harboring the *R1* resistance gene. A contig was built around the *R1*

locus and screened for candidate resistance genes. In this way, a candidate for the *R1* gene was identified, subcloned, and shown to encode the functional gene by complementation of a susceptible potato genotype. Because of the versatility of molecular cloning techniques, most protocols can be transferred between genotypes of one species and between species with relative ease. Therefore, local physical maps can be constructed for almost any locus irrespective of the plant species, provided that a dense framework of genetic markers within the region of interest is available for chromosome walking toward the locus of interest.

The situation is more critical if a whole genome physical map is intended. The complexity of plant genomes, due to duplications, rearrangements and deletions, the often large genome size, and the high degree of repetitive sequences, is the greatest obstacle for the generation of physical maps. For maize, which has an estimated genome size of 2500 Mbp (Arumuganathan and Earle, 1991), representing about 80% of the human genome in terms of size, a physical map based on three different BAC libraries is under construction (Coe *et al*., 2002; Cone *et al*., 2002; Yim *et al*., 2002). The number of contigs was reduced to 760 (http://www.genome.arizona.edu/fpc/maize/). Meanwhile, a massive sequencing effort of aligned BAC ends has partly elucidated sequence composition and genome organization and confirmed a repeat content of 58% in the maize genome (Messing *et al*., 2004). Additional criteria for selecting a species for a whole genome physical map construction include the potential to serve as model species and the expected increase in economic value with regard to crop improvement by molecular breeding strategies. Almost complete physical maps have been published for *Arabidopsis thaliana* (Chang *et al*., 2001; Marra *et al*., 1999; Mozo *et al*., 1999), rice (Chen *et al*., 2002; Tao *et al*., 2001), sorghum (Klein *et al*., 2000), and soybean (Wu *et al*., 2004b). To achieve complete genome coverage, large-insert libraries containing at least six to eight genome equivalents have proved to be a prerequisite (Zhang and Wu, 2001). The cloning efficiency of large inserts was found to be dependent on the restriction enzyme used because the distribution of restriction sites is uneven within a genome. Therefore the use of different restriction enzymes for cloning large fragments is necessary to fill gaps and to improve genomic coverage (Tao *et al*., 2001; C. C. Wu *et al*., 2004a; Yim *et al*., 2002). Projects to generate whole genome physical maps are inspired mostly by the desire to sequence entire genomes (Section IV.A). For this purpose the tiling path of minimally overlapping clones to be sequenced is deduced from the physical maps, which indeed were essential tools for the sequencing of both the *Arabidopsis* and rice genomes (*Arabidopsis* Genome Initiative, 2000; Feng *et al*., 2002; Rice Chromosome 10 Sequencing Consortium, 2003; Sasaki *et al*., 2002).

IV. Plant Genomes

A. Whole Genome Sequences

A catalog of transcribed sequences can be assembled for any species in a cost-efficient manner by generating thousands of single-pass sequencing reads of randomly chosen cDNA clones (expressed sequence tags, ESTs) (Section IV.C.1). For an EST project, cDNA clones are isolated either from a library made from mRNA isolated from a mixture of tissues and conditions (Newman *et al.*, 1994) or from different cDNA libraries made from a variety of tissues and cultured cells (Höfte *et al.*, 1993; Uchimiya *et al.*, 1992). It is possible to reconstruct consensus sequences of ESTs that are often longer than individual ESTs because many genes are represented multiple times in a particular EST collection (Quackenbush *et al.*, 2001; Rounsley *et al.*, 1996). Extensive EST collections have been established for many plant species. For example, more than 100,000 ESTs each are available in public databases for 14 different higher plant species (*Triticum aestivum, Zea mays, Arabidopsis thaliana, Oryza sativa, Glycine max, Saccharum officinarum, Pinus taeda, Medicago truncatula, Sorghum bicolor, Solanum tuberosum, Solanum lycopersicum, Malus domestica, Vitis vinifera*, and *Lotus japonicus*; http://www.ncbi.nlm.nih.gov/dbEST/). However, even these large data sets do not represent the entire gene repertoires of these organisms because scarce transcripts are poorly represented in cDNA libraries. Another drawback of EST sequencing is that this approach fails to provide information about elements of importance for gene regulation as these are located mostly outside transcribed regions. In contrast, a comprehensive description of the gene and repeat repertoire of a particular species is possible if the entire genome of a species is analyzed at the sequence level. Such an effort also gives an insight into chromosomal architecture.

Plant genomes consist of three components located in the nucleus, the plastids, and the mitochondria. The plastid genomes of land plants are generally similar in size (120–220 kbp) (Palmer, 1992). Complete sequences of chloroplast genomes were described for the liverwort *Marchantia polymorpha* and tobacco in 1986 (Ohyama *et al.*, 1986; Shinozaki *et al.*, 1986). Similar orders of the densely packed genes are observed in species as distantly related as *Marchantia, Arabidopsis*, tobacco, and rice (Hiratsuka *et al.*, 1989; Sato *et al.*, 1999a). Cytogenetic analyses of the *Arabidopsis*, tobacco, and pea chloroplast genomes revealed, in addition to monomeric circular forms, linear and circular multimers as well as rearranged, partial genome units (Lilly *et al.*, 2001).

Mitochondrial genomes of land plants are more variable in size (180–2500 kb; Ward *et al.*, 1981) than plastid genomes. Homologous recombination

between repeated sequences present in the mitochondrial genome results in the simultaneous occurrence of a variable number of different mitochondrial genome molecules in the organelle (Palmer and Shields, 1984). Comparison of the complete sequences for the mitochondrial genomes of the liverwort *Marchantia polymorpha, Arabidopsis thaliana*, and *Oryza sativa* (Notsu *et al.*, 2002; Oda *et al.*, 1992; Unseld *et al.*, 1997) revealed a greater degree of variation in gene repertoire than is found for plastid genomes. Sequences of nuclear and plastid genome origin were found in the mitochondrial genomes of *Arabidopsis* and rice (Notsu *et al.*, 2002; Unseld *et al.*, 1997). Likewise, evidence of frequent, ongoing transfer of sequences from the organelles to the nuclear genomes has been reported (Timmis *et al.*, 2004).

Both organellar genomes together encode only a small proportion of the genes in a plant. The vast majority are found in the nuclear genome. For example, the organellar genomes of *Arabidopsis* contribute coding information for little more than 0.5% of all genes in this organism (*Arabidopsis* Genome Initiative, 2000). Nuclear genomes of angiosperms vary more than 1000-fold in genome size (Leitch *et al.*, 1998). Thus, species with comparatively small genomes, such as *Arabidopsis* and rice, were chosen for the first plant genome sequencing projects. More recently, approaches have been devised that are suitable for the analysis of large genomes containing a high proportion of repetitive sequences. Adopting a clone-by-clone sequencing approach, *Arabidopsis thaliana* was the first higher plant species for which the sequence of the nuclear genome was determined. Clone contig chromosome maps assembled from libraries of large-insert clones (Section III.B) were a prerequisite for the map-based sequencing strategy. A tiling path of clones with minimal overlap was chosen and for each of these large-insert clones libraries of randomly sheared DNA were prepared in high-copy plasmid vectors. The inserts of the plasmid clone libraries were sequenced and the resulting sequences were assembled. Finally, the assembled sequences of the different large-insert clones were used to reconstruct the sequences of the five chromosomes. Sequences of two of the five chromosomes of the *Arabidopsis* accession Columbia were reported in 1999 (European Union *Arabidopsis* Sequencing Consortium, Cold Spring Harbor Laboratory, Washington University St. Louis, and PE Biosystems, 1999; Lin *et al.*, 1999). In 2000 the sequence of the genome covering 115.4 Mbp in total was released at accuracy rates between 99.99 and 99.999% (*Arabidopsis* Genome Initiative, 2000; European Union Chromosome 3 *Arabidopsis* Sequencing Consortium, Institute for Genomic Research, and Kazusa DNA Research Institute, 2000; Kazusa DNA Research Institute, Cold Spring Harbor and Washington University in St. Louis Sequencing Consortium, and European Union *Arabidopsis* Genome Sequencing Consortium, 2000; Theologis *et al.*, 2000). Regions containing highly repetitive tandemly organized sequences such as the rDNA repeat regions and

the centromeric repeat regions were not completely sequenced. Thus, the sequenced portions of the chromosomes extended from either the telomeres or ribosomal DNA repeats into the centromeric repeat regions.

Gene structures were determined by a combination of appropriate computer algorithms and by taking into account similarities to known proteins and ESTs (*Arabidopsis* Genome Initiative, 2000). Annotation efforts are still ongoing, and substantial refinements of the exon/intron structures were achieved by integrating information derived from large sets of full-length cDNA sequences (Castelli *et al.*, 2004; Haas *et al.*, 2002; Seki *et al.*, 2002; Yamada *et al.*, 2003). Genome-wide mapping of transcriptional units was also performed by hybridizing RNA populations prepared from various tissues to a set of high-density oligonucleotide arrays spanning the entire genome (Yamada *et al.*, 2003).

The availability of the genome sequence for the Columbia accession facilitates strategies aimed at genome-wide detection of polymorphisms for other *Arabidopsis* accessions. Hybridization of genomic DNA of two *Arabidopsis thaliana* accessions to oligonucleotide arrays that had been initially designed for gene expression monitoring enabled the detection of almost 4000 single-feature polymorphisms between the Columbia and Landsberg *erecta* accessions (Borevitz *et al.*, 2003). An even higher number of polymorphisms between these two accessions was the result of a genome-sequencing effort adopting a whole-genome shotgun strategy for the Landsberg *erecta* genotype. Approximately 500,000 sequences corresponding to *Arabidopsis* nuclear DNA and amounting to 263 Mbp of good-quality raw sequence were obtained and assembled in 92.1 Mbp of sequence. Comparison of this data set with the genome sequence of the Columbia accession was performed for polymorphism prediction. On average, a single-nucleotide polymorphism (SNP) was found every 3.3 kbp and an insertion/deletion event (InDel) was found every 6.6 kbp. In total, more than 50,000 polymorphisms were predicted, hence greatly facilitating map-based cloning efforts in *Arabidopsis* (Jander *et al.*, 2002). The InDel polymorphisms ranged from 1 bp to 38 kbp; insertions greater than 250 bp in Columbia often matched transposon-related proteins, indicating that a proportion of the large InDels were explained by transposon insertion or excision events. However, comparison of the genomic sequences of the Columbia and Landsberg *erecta* accessions also revealed evidence of gene translocation events (*Arabidopsis* Genome Initiative, 2000).

The annotated sequence of the *Arabidopsis* nuclear genome offers unique opportunities for comparative sequence analyses with closely and distantly related species. The comparative study of small chromosome segments revealed similar gene arrangements in *Arabidopsis* and *Capsella rubella*, a species that diverged from *Arabidopsis* approximately 6 to 10 million years ago (Acarkan *et al.*, 2000; Rossberg *et al.*, 2001). Notably, conservation of

microstructure was also observed when small genome segments of species were compared that belong to different clades of the eudicots, *Arabidopsis* and tomato (Ku *et al*., 2000; Rossberg *et al*., 2001). However, it was also discovered that gene loss in duplicated regions of a genome is a particularly important process in plant genome evolution (Section II.A). Many but not all the observed deviations from colinearity were explained by this process (Ku *et al*., 2000; Rossberg *et al*., 2001). This was also seen when *Arabidopsis* genome segments were compared with corresponding regions in *Brassica* (O'Neill and Bancroft, 2000; Rana *et al*., 2004), a lineage that was subjected to polyploidization after divergence of the progenitors of *Arabidopsis* and *Brassica* approximately 20 million years ago.

For sequencing of the rice genome in the context of the International Rice Genome Sequencing Project, the clone-by-clone strategy was chosen. The sequencing consortium selected the 430-Mbp nuclear genome of the *Oryza sativa* subspecies *japonica* as a target for their sequencing efforts. So far, sequences of 3 of the 12 rice chromosomes have been reported to the same degree of accuracy as for the *Arabidopsis* genome project (Rice Chromosome 10 Sequencing Consortium, 2003; Feng *et al*., 2002; Sasaki *et al*., 2002). Owing to the high abundance of highly repetitive satellite DNA, centromeric regions of plants or animals resist mapping and sequencing efforts. However, the centromeric region of chromosome 8 of rice has an unusually low abundance of highly repetitive satellite DNA. This enabled mapping and sequencing of this centromeric region, thus allowing insight into the structural architecture of such a region that is of crucial importance for chromosome segregation and karyotype stability (Nagaki *et al*., 2004; Wu *et al*., 2004). A 750-kbp region underlying the kinetochore was identified and several active genes were found in this heterochromatic area (Nagaki *et al*., 2004).

If gene discovery is the main aim of a genome-sequencing project, the production of a draft sequence—a sequence that is not completely assembled—is an attractive option. Such a draft sequence was generated for the rice genome (Barry, 2001). Information derived from 3391 BAC clones added up to 399 Mbp of sequence. After assembly of redundant sequences into contigs, 259 Mbp of sequence was obtained, thus determining the sequence of approximately 60% of the rice genome.

Two rice genome-sequencing programs took advantage of the whole genome shotgun approach to produce draft sequences of the rice genome (Goff *et al*., 2002; Yu *et al*., 2002). For the whole genome shotgun strategy, libraries of randomly sheared DNA, ideally of different size ranges, are prepared and the resulting clones are sequenced from both ends. The resulting sequences are assembled. Moreover, the information that nonoverlapping sequences are generated from a pair of end sequences of a given clone is exploited to link sequence assemblies into scaffolds. The depth of the sequence coverage

determines not only the average length of the sequence assemblies but also the quality of the sequence. The draft sequence for the *indica* subspecies was generated at approximately fourfold coverage and a total of 361 Mbp was assembled for the 466-Mbp genome (Yu *et al.*, 2002). The more than sixfold coverage of the genome of the *japonica* subspecies resulted in 389 Mbp of assembled sequence for the 430-Mbp genome (Goff *et al.*, 2002). Data produced by the two different shotgun approaches were combined to aid the assembly. The high genome colinearity of the *indica* and *japonica* subspecies was exploited to construct scaffolds and superscaffolds by integrating data of the *indica* subspecies for assembly of the *japonica* subspecies and vice versa. This approach resulted in increased long-range contiguity: 140 mapped superscaffolds with an average size of 8.2 Mbp were established for the *indica* subspecies and 119 mapped superscaffolds of an average size 11.6 Mbp were established for the *japonica* subspecies (Yu *et al.*, 2005). By integrating data from large-scale EST and cDNA sequencing programs, the gene repertoire of both subspecies is now estimated at approximately 38,000 to 40,000 genes (Yu *et al.*, 2005).

Many of the important crop plants have genomes as large or even larger than the human genome. A high abundance of mobile element sequences and in particular retrotransposon sequences is characteristic for large-genome plant species. For example, repetitive sequences make up 80% of the 2500-Mbp maize genome. Given that amplification of retrotransposon sequences within the last 3 million years caused a doubling in genome size (SanMiguel *et al.*, 1998) it is evident that many of these sequences are closely related. Thus, not only the abundance of repetitive elements, but also the high degree of sequence similarity between elements that may be widely dispersed in the genome, causes problems for mapping and sequencing projects. Therefore, strategies had to be devised that enrich for genes in shotgun libraries. In a solution of sheared, heat-denatured DNA a specific sequence reassociates at a rate proportional to its copy number in the genome. On the basis of such an analysis genomic DNA can be fractionated into highly repetitive, moderately repetitive, and single- and/or low-copy DNA; cloned; and sequenced. Adopting this strategy for sorghum and maize revealed that libraries made from the single- and/or low-copy fraction may be an effective means to capture the sequence complexity of these genomes (Peterson *et al.*, 2002; Yuan *et al.*, 2003), thus providing a cost-effective alternative to a whole genome sequencing effort. Sequencing by methyl filtration provides an alternative strategy. This approach exploits the fact that most repetitive elements are heavily methylated when compared with gene sequences. Given that shotgun clone libraries are prepared in a methylation-restrictive *Escherichia coli* host strain, repeat sequences are largely excluded whereas unmethylated genic regions are preserved in these libraries (Rabinowicz *et al.*, 1999). The effectiveness of these gene-enrichment strategies was demonstrated by the first

pilot experiments. Raw coverage of 300 Mbp of methylation filtration sequence of the 735-Mbp sorghum genome resulted in the sequence tagging of 96% of the genes in this species with, on average, two-thirds coverage across their lengths (Bedell *et al.*, 2005). A sixfold reduction in effective genome size and a fourfold increase in gene identification rate were achieved by applying a combination of both described gene enrichment strategies for analysis of the maize genome when compared with a nonenriched library (Whitelaw *et al.*, 2003).

B. Model Systems

The analysis of model organisms is of special importance in eukaryotic genome research. Ideally, eukaryotic model organisms for genome research combine high amenability to genetic analyses with a small genome size. *Saccharomyces cerevisiae* was the first eukaryotic genome to be fully sequenced (Mewes *et al.*, 1997). The description of the gene complement in this organism led to a wealth of follow-on experiments aimed at the elucidation of gene function, the systematic study of protein localization, and a comprehensive analysis of protein–protein interactions.

In plants, *Arabidopsis* and rice are considered model species for eudicots and monocotyledonous plants, respectively. The following section gives a short overview on genome organization in these species and the available resources for functional genomic analyses. Many aspects of plant biology that are common to all plants can be studied in the tractable model systems *Arabidopsis* and rice. However, these two model systems do not, by any means, encompass the vast diversity found in the plant kingdom at the whole plant, physiological, biochemical, genetic, or molecular level. Moreover, *Arabidopsis* and rice are not suitable to study certain biological phenomena that are of great importance in plant biology, such as symbiotic nitrogen fixation or wood formation. To address such issues, the legumes *Lotus japonicus* and *Medicago truncatula*, the tree *Populus trichocarpa*, and the moss *Physcomitrella patens* have also been selected for plant genome projects.

1. *Arabidopsis thaliana*

The small crucifer *Arabidopsis thaliana* has with approximately 125 Mbp one of the smallest known nuclear genomes in higher plants. The genome is characterized by a low amount of repetitive sequences and a high density of gene sequences. On average, a gene was found every 4.5 kbp in the nuclear genome. Only 35% of the predicted proteins are unique in the genome; analysis of duplicated gene sequences revealed that 17% of all genes are

arranged in tandem arrays. Approximately 60% of the genome was found in large segmental duplications (*Arabidopsis* Genome Initiative, 2000). High-precision genetic mapping defined regions containing centromere function for each of the five chromosomes (Copenhaver *et al.*, 1999). Pericentromeric regions containing numerous classes of repetitive sequences flank the centromere cores that consist of homogeneous arrays of 180-bp repeats and in some cases of long tandem arrays of 5S rDNA sequences (*Arabidopsis* Genome Initiative, 2000). A first insight into sequence determinants of importance for chromosome condensation was provided by the sequence of a heterochromatic island on the short arm of chromosome 4 (Cold Spring Harbor Laboratory, Washington University Genome Sequencing Center, and PE Biosystems *Arabidopsis* Sequencing Consortium, 2000).

The function of only a small fraction of the *Arabidopsis* genes has been characterized experimentally. Even assigning a putative function based on sufficient similarity to proteins of known function is not possible for approximately 30% of the predicted genes (*Arabidopsis* Genome Initiative, 2000). With the annotated genome sequence in hand, functional analysis of all *Arabidopsis* genes became the goal. To meet this objective, emphasis is currently on global approaches such as transcriptomics, proteomics, and metabolomics (Somerville and Dangl, 2000; Section IV.C). Furthermore, the excellent amenability of *Arabidopsis* to mutational and genetic analyses is an essential prerequisite for functional genomics approaches. The short generation time of less than 2 months, the small stature of fully grown plants, and the large number of offspring render *Arabidopsis* ideally suited for genetic and mutational studies. Extensive mutant collections have been generated by ionizing radiation and chemicals such as ethyl methanesulfonate (EMS) and many mutants have been placed on the genetic map (Koornneef *et al.*, 1983; Meinke *et al.*, 1998; Rhee *et al.*, 1998), often as a first step toward gene isolation by map-based cloning approaches. Dense molecular marker maps (Chang *et al.*, 1988; Lister and Dean, 1993; Meinke *et al.*, 1998; Rhee *et al.*, 1998) extensively cross-referenced with contig maps based on artificial chromosome clones (Kotani *et al.*, 1997; Marra *et al.*, 1999; Mozo *et al.*, 1999; Sato *et al.*, 1998; Schmidt *et al.*, 1995; Zachgo *et al.*, 1996) facilitated positional cloning of mutant loci and sequencing of the genome. A compilation shows that more than 100 genes with mutant phenotypes have been isolated by map-based strategies (Meinke *et al.*, 2003). Increasingly, the study of naturally occurring genetic variation in *Arabidopsis thaliana* is used in addition to the study of mutants to gain insight into the control of important plant processes (Koornneef *et al.*, 2004).

Arabidopsis is readily transformed by various methods (Damm *et al.*, 1989; Lloyd *et al.*, 1986). High-throughput transformation methods devoid of tissue culture steps are in place (Clough and Bent, 1998) enabling, for example, protein localization studies by introducing fusions between reporter

genes and cDNA or gene sequences (Cutler *et al*., 2000; Tian *et al*., 2004) or efficient transgene-induced RNA interference (McGinnis *et al*., 2005). Furthermore, the T-DNA of *Agrobacterium tumefaciens* and transposable elements from maize that had been introduced into *Arabidopsis* have been exploited to establish large populations of insertion mutants (Azpiroz-Leehan and Feldmann, 1997; Meissner *et al*., 1999). Cloning of mutant genes by tagging approaches has been widely adopted in *Arabidopsis*. Meinke *et al*. (2003) compiled more than 295 cases. The extensive mutant collections are also exploited for reverse genetic approaches. Point mutations in a gene of interest can be screened for in a high-throughput fashion by TILLING (Henikoff *et al*., 2004). Sequences flanking insertions of the T-DNA of *Agrobacterium tumefaciens* have been determined for more than 88,000 different events (Alonso *et al*., 2003) and the analysis of this mutant collection showed that mutants had been identified for approximately 75% of the predicted genes. A centralized entry point to information and research materials that have been assembled for this species is provided by the *Arabidopsis* Information Resource (TAIR, http://arabidopsis.org; Rhee *et al*., 2003). Moreover, this database offers many links to other databases and/or functional genomics projects.

2. *Oryza sativa*

Rice has been selected as a model organism because it combines several attributes such as a small genome size of approximately 430 Mbp (Arumuganathan and Earle, 1991), amenability to transformation (Christou, 1997; Hiei *et al*., 1997), and a wealth of genetic and molecular resources. Furthermore, it is a principal food crop. Rice belongs to the Poaceae and is related to species of agronomic importance such as maize and wheat, which have much larger genomes (Arumuganathan and Earle, 1991). Despite the fact that some Poaceae species diverged as long as 60 million years ago, a remarkable degree of genome conservation has been established for this plant family, using comparative genetic mapping experiments (Gale and Devos, 1998). Thus, information collected for the rice genome has an immediate impact on studies in many other crop plants.

Extensive germplasm collections have been assembled for rice. As many as 80,000 samples are, for example, stored at the International Rice Research Institute (IRRI; Jackson, 1997). The collection comprises landrace varieties, modern as well as obsolete varieties, special genetic stocks, and wild species. Comprehensive molecular marker maps (Causse *et al*., 1994; Harushima *et al*., 1998) enabled mapping of quantitative trait loci for many different traits in rice (McCouch and Doerge, 1995; Yano and Sasaki, 1997). Clone contig maps based on artificial chromosome clones have been established for all 12 rice chromosomes (Chen *et al*., 2002; Saji *et al*., 2001; Wu *et al*., 1998,

2002, 2003). These maps form an indispensable basis not only for the sequencing of the rice genome but also for positional cloning of quantitative trait loci (Takahashi *et al.*, 2001; Yano *et al.*, 2000).

The wealth of sequence information about the nuclear, chloroplast, and mitochondrial genomes of the two rice subspecies *indica* and *japonica* (Barry, 2001; Feng *et al.*, 2002; Goff *et al.*, 2002; Hiratsuka *et al.*, 1989; Notsu *et al.*, 2002; Rice Chromosome 10 Sequencing Consortium, 2003; Sasaki *et al.*, 2002; Tang *et al.*, 2004; Yu *et al.*, 2002, 2005) is complemented by extensive collections of ESTs (Sasaki *et al.*, 1994; Uchimiya *et al.*, 1992; Zhou *et al.*, 2003) and the availability of more than 28,000 full-length cDNA sequences (Rice Full-Length cDNA Consortium, 2003). Estimates of gene number in rice vary, but all gene counts suggest that rice contains more genes than *Arabidopsis* (Goff *et al.*, 2002; Yu *et al.*, 2002, 2005). Approximately 80% of the *Arabidopsis* genes have a homolog in rice, but only 50% of the predicted rice genes have a homolog in *Arabidopsis* (Yu *et al.*, 2002). In different species of the Gramineae (rice, maize, barley, and wheat) gradients in GC content, codon usage, and amino acid usage are observed in genes along the direction of transcription. In contrast, when genes were studied in several eudicot species, similar features were not found. These different characteristics of monocot and eudicot genes severely hamper the detection of protein homologs across the monocot–eudicot divide (Wong *et al.*, 2002).

Analysis of duplication patterns in the rice nuclear genome revealed evidence of a recent segmental duplication (Wu *et al.*, 1998; Yu *et al.*, 2005) and an ancient whole genome duplication before the divergence of the cereals from one another (Paterson *et al.*, 2004; Yu *et al.*, 2005). Individual gene duplications are particularly frequent events (Yu *et al.*, 2005), with a significant proportion of genes being present in tandem arrays (Sasaki *et al.*, 2002).

The *indica* and *japonica* genotypes vary little with respect to gene content. Alignments of *indica*- and *japonica*-derived sequences showed extensive colinearity, but deviations from colinearity due to InDels, especially in intergenic regions, are frequent (Yu *et al.*, 2005). Comparing homologous regions of approximately 2.3 Mbp in size revealed 9056 SNPs and 201 InDels between the *indica* and *japonica* genotypes. More than 35% of the InDels were longer than 1 kbp and at least some of them correspond to gene regions (Feng *et al.*, 2002). Comparative analysis of the genome sequences for the two subspecies has enabled genome-wide identification of single-nucleotide polymorphisms and single-base InDels between these genotypes (Feltus *et al.*, 2004). The vast majority of known maize, wheat, and barley genes have homologs in rice (Goff *et al.*, 2002). Moreover, a comparison of rice chromosome sequence maps with genetic maps of sorghum and maize corroborated the extensive synteny between rice and other cereal genomes (Goff *et al.*, 2002; Rice Chromosome 10 Sequencing Consortium, 2003). In contrast, little

conservation in gene order was found between rice and *Arabidopsis* (Feng *et al*., 2002; Goff *et al*., 2002).

Studies of gene function at a genome-wide level in rice can build on the availability of high-density arrays for gene expression profiling (Rensink and Buell, 2004) and a proteome database (Komatsu and Tanaka, 2005). Moreover, rice can be transformed by different methods (Christou, 1997; Hiei *et al*., 1997). The T-DNA of *Agrobacterium tumefaciens, Ac/Ds* and *En/Spm*, maize transposable elements that had been introduced into rice, and *Tos17*, an endogenous retrotransposon, have all been used to establish large populations of insertion mutants and in a number of cases gene tagging has resulted in the cloning of mutant genes (Hirochika *et al*., 2004). Screening of deletion (Li *et al*., 2001) or insertion (Enoki *et al*., 1999; Lee *et al*., 2003; Sato *et al*., 1999b) mutant populations by PCR approaches has been established and mutants for genes of interest have been isolated by reverse genetic approaches. The use of reverse genetic approaches in rice will be even more facile in future because programs have been initiated that aim at large-scale identification of rice sequences that flank insertions of the T-DNA of *Agrobacterium tumefaciens* (Sallaud *et al*., 2004), the *Ds* transposable element (Kolesnik *et al*., 2004), and the *Tos17* element (Miyao *et al*., 2003).

3. Legume Models: *Medicago truncatula* and *Lotus japonicus*

Legumes provide unique insights into the phenomenon of plant–microbe symbiosis. Therefore, two legume species, *Lotus japonicus* and *Medicago truncatula*, which are hosts for the nitrogen-fixing symbiotic bacteria *Sinorhizobium meliloti* and *Mesorhizobium meliloti*, respectively, have been selected as plant model species. Symbiotic relationships with arbuscular mycorrhizal fungi can also be studied in these species. *Lotus japonicus* and *Medicago truncatula* are self-fertilizing diploid species that have a generation time of approximately 3 months and are amenable to *Agrobacterium*-mediated transformation (Barker *et al*., 1990; Handberg and Stougaard, 1992). Screening mutant populations generated by insertion mutagenesis, EMS, or irradiation treatments led to the discovery of different classes of symbiotic mutants in *Medicago* and *Lotus* (Stougaard, 2001). Importantly, a set of common signal transduction components has been identified that is required to redirect the development of root cells so that the plants can engage in symbiotic relationships with bacteria and fungi (Kistner and Parniske, 2002). Map-based cloning of some of the genes involved has been carried out. Thus, the study of plant symbioses has clearly benefited from the resources that have been put in place by the ongoing clone-by-clone genome sequencing programs for these species (http://www.medicago.org/genome/; http://www.kazusa.or.jp/lotus/), which have, with 470 Mbp, relatively

compact genomes. Large-scale EST programs have enabled transcript profiling in *Medicago* and *Lotus*, and many genes were detected that were more highly expressed in nodules than in roots (Federova *et al.*, 2002; Colebatch *et al.*, 2004). Furthermore, genes of importance for arbuscular mycorrhizal symbiosis were identified in *Medicago* (Liu *et al.*, 2003). With the complete genome sequences of the symbiotic bacteria in hand (Galibert *et al.*, 2001; Kaneko *et al.*, 2000), transcriptional analyses of rhizobia are also likely to reveal new details concerning many different aspects of symbiosis.

4. Tree Model: *Populus trichocarpa*

It may be difficult to extrapolate findings from annual, herbaceous model systems to organisms such as trees that are distinguished by a long life span, large size, extensive formation of secondary xylem, dormancy, and phase change from juvenile to mature state. This notion together with the economic and ecological importance of forest trees has fueled the idea to promote a tree species as a model for perennial plants. *Populus* combines many features of importance for a model organism. It has a relatively small genome of approximately 550 Mbp and a fast growth rate (Bradshaw *et al.*, 2000; Taylor, 2002). Vegetative propagation is easily possible and *Populus* can be readily transformed. An outcrossing mating system, the long generation time, and the large physical size of the trees prevent the generation of collections of recessive mutations. Therefore, transgenic strategies aimed at dominant changes through suppression or overexpression of particular genes are preferred for the study of gene–function relationships in trees (Brunner *et al.*, 2004). A large set of *Populus* EST sequences enabled a first comparison of the gene repertoires of *Populus* and *Arabidopsis*. The high similarity of gene content found possibly indicates that differences between annual and perennial angiosperms are primarily manifested in altered gene regulation rather than altered gene content (Sterky *et al.*, 2004). Much more refined analyses will be possible once the assembled and annotated genome sequence is available. The *Populus* genome sequence was generated by the DOE Joint Genome Institute, using a whole genome shotgun approach (http://www.jgi.doe.gov/sequencing/). Paired end-reads were obtained for clone libraries of different size classes and sequencing proceeded until 7.5-fold coverage was reached (Tuskan *et al.*, 2004).

5. Moss Model: *Physcomitrella patens*

All plant model systems described so far belong to the angiosperms. Mosses and flowering plants diverged approximately 450 million years ago. A comparison of a moss genome with that of a flowering plant may be a particularly fruitful way to learn more about plant genome evolution. In seed plants the

diploid sporophyte is the predominating generation; for most of their lives seed plants stay in the sporophytic generation. In contrast, the haploid gametophyte is the predominating generation in the life cycle of mosses. This characteristic offers unique advantages for mutagenesis because mutant phenotypes can be directly observed in the haploid gametophyte. The moss *Physcomitrella patens* combines many features that make it an attractive model system, for example, simple culture methods, cryopreservation, and a high regenerative capacity. Moreover, the study of cell polarity and pattern determination is facilitated by accessibility to direct observation of all cells in the protonemal tissue of the gametophyte stage (Cove, 2000). A major asset of this experimental system is the discovery that the efficiency of homologous recombination is high enough in this species to be exploited as an efficient tool for the study of gene–function relationships. Large mutant collections have been established by exploiting homologous recombination (Egener *et al*., 2002). In higher plants, this has so far not been possible because the efficiency of homologous recombination is several orders of magnitude lower than in *Physcomitrella* (Schaefer, 2001). Large-scale EST programs resulted in a set of more than 19,000 nonredundant open reading frames (ORFs). Only about 30% of these EST contigs were found to have a homolog in both rice and *Arabidopsis* (Rensing *et al*., 2005). Sequencing of the *Physcomitrella* genome, which encompasses 511 Mbp (Reski and Cove, 2004), is currently ongoing at the DOE Joint Genome Institute by means of a whole genome shotgun strategy (http://www.jgi.doe.gov/sequencing/).

C. Functional Genomics

Functional genomics integrates multiparallel approaches to assess the activities of plant cells and tissues at different levels. The analysis of automatically generated "high-throughput" experimental data by bioinformatics aims at a more holistic understanding of complex processes such as plant cell differentiation, organ development, or cell death. "Systems biology" has been coined as a new discipline uniting the different approaches in this area. In the following sections, the methodological strategies and concepts to analyze three main classes of cellular components, namely transcripts, proteins, and metabolites, for which plant research is most advanced, are described and an outlook on future developments is presented. Selected illustrative examples are given in each section.

1. Transcriptomics

The parallel analysis of large numbers of transcripts is called "transcriptomics." Transcripts are the prerequisite for and the primary product of gene

expression according to the central dogma of molecular biology (revisited in Crick, 1970). The discovery of viral RNA-dependent DNA polymerase or reverse transcriptase enabled the *in vitro* generation of cDNA (copy DNA) complementary to a ribonucleic acid template (Ross *et al.*, 1972). Nowadays, cDNA synthesis is still the first step in all transcriptomics research. Subsequent enzymatic steps lead to the molecular cloning of a population of cDNAs, which represent a snapshot of type and abundance of the transcripts present in a given tissue or organ at a given developmental stage. The resulting collection of cDNAs, the cDNA library, is an inventory of the genes expressed in the tissue sample. Mass sequencing of cDNAs generates expressed sequence tags (ESTs; Adams *et al.*, 1991), which typically have a length of 300–700 bp. For many different plant species extensive EST collections have been assembled (http://www.ncbi.nlm.nih.gov/dbEST/; Section IV.A).

Northern analysis used to be the technique of choice for monitoring the expression of a single gene in different tissues, which is based on the hybridization of a labeled DNA probe to electrophoretically separated mRNAs immobilized on a membrane support. With the onset of transcriptomics, array technologies were developed for the parallel monitoring of thousands of transcripts. Individual cDNAs are fixed at a known position to a solid support and then hybridized with a complex transcript sample converted into labeled cDNA. For so-called macroarrays, PCR-amplified cDNAs are spotted onto nylon membranes, which are then used for hybridization with radioactively labeled complex cDNA probes. This method requires relatively little equipment and produces robust results of satisfying sensitivity and reliability. Macroarrays are therefore the method of choice for transcriptomics research on plant species for which genomic information and tools are at present limited, such as barley (Potokina *et al.*, 2002) and sugar beet (Bellin *et al.*, 2002). Microarrays differ from macroarrays by miniaturization of the solid support carrier, which is usually a glass slide printed with the hybridization targets, and by using a hybridization protocol involving two cDNA samples labeled with different fluorescent dyes in the same experiment (Aharoni and Vorst, 2001; Brown and Botstein, 1999; Schena *et al.*, 1995). As an alternative to printed slides, photolithographic procedures allow the synthesis of up to 100,000 oligonucleotides of 25-mer length directly on a chip surface (Fodor *et al.*, 1991). Thus, all possible transcripts of the annotated *Arabidopsis thaliana* genome (transcriptome) can be accommodated on a single chip. This chip is now used to analyze at a genome-wide level the response of the *Arabidopsis* transcriptome to environmental factors such as nitrate supply (Scheible *et al.*, 2004; Wang *et al.*, 2003). To improve comparability and transferability of the large data sets produced by microarray experiments, common standards for experimental and data analysis procedures have been defined in the MIAME (minimum information about a

microarray experiment) protocol (Brazma *et al*., 2001), including documentation of plant cultivation, which is now widely accepted in the plant research community.

The array approach depends on the availability of suitable cDNA collections or the sequence information thereof. Only transcripts occurring in the interrogated pool can be assessed for their expression profile within the technical limits of the array. As an alternative to this ''closed'' system, several ''open'' systems have been developed, which are based on the generation of specific cDNA samples amplified by PCR and their independent fingerprinting. Among the first methods used was differential display (Liang and Pardee, 1992). However, because of nonstringent PCR conditions for short random primers, the results showed low reproducibility and were not sensitive enough. The cDNA AFLP method (Bachem *et al*., 1996) combines cDNA synthesis with the fidelity of AFLP fingerprinting, using two selective PCR amplification steps. Differentially expressed cDNA fragments are recovered and cloned for further analysis. The cDNA AFLP technique proved highly effective in different studies on potato, tomato, and *Zinnia elegans* (Bachem *et al*., 1996; Durrant *et al*., 2000; Milioni *et al*., 2002). This technique was upscaled by the use of quantitative PCR for cDNA amplification, resulting in precise transcript profiles for cell cycle-modulated genes in tobacco and genes expressed during rice seed development (Breyne *et al*., 2003; Suzuki *et al*., 2005). Suppression subtractive hybridization (SSH; Diatchenko *et al*., 1996) is a method to enrich for cDNAs that are specifically expressed in one of two samples. Safener-inducible transcripts in poplar (Rishi *et al*., 2004) or abscisic acid-inducible transcripts in the desiccation-tolerant plant *Craterostigma plantagineum* (Kirch *et al*., 2001) were selected by this method.

A second group of ''open system'' transcript profiling techniques is based on the enzymatic generation of short sequence signatures (tags) from each cDNA consisting of maximally 20 bp. Serial analysis of gene expression (SAGE; Velculescu *et al*., 1995) is an example of this strategy. It involves the restriction of a cDNA sample with two enzymes to generate the tags, which are then ligated, cloned, and mass sequenced. The abundance of each tag is a direct measurement for the transcript level of the corresponding gene. For plant genomes with large sequence data resources available, SAGE provides sufficient information for the reliable identification of the corresponding gene. SAGE was applied to analysis of the *Arabidopsis* root transcriptome, using the genome sequence as reference (Fizames *et al*., 2004). Using a specialized bioinformatics tool, 144,083 sequenced tags were found to represent at least 15,964 different mRNAs expressed in roots, some of which had not been identified previously. A comparison of tag frequencies in SAGE libraries generated from plants grown either on nitrate or ammonium nitrate as nitrogen source revealed 270 differentially expressed genes. The second technique to be mentioned in this context is massive parallel signature

sequencing (MPSS; Brenner *et al.*, 2000), which relies on the enzymatic generation of restriction fragments of 16–20 nucleotides and nonconventional sequencing of microbead-bound templates. MPSS libraries are now available for *Arabidopsis thaliana*, rice, and grape (http://mpss.udel.edu/). Nearly 37 million 17-bp sequence tags were identified from 14 MPSS libraries of different *Arabidopsis* samples alone (Meyers *et al.*, 2004b). Many of them identified previously uncharacterized transcripts.

Apart from improving the resolution of gene expression levels, there is a strong trend to refine the spatial pattern of plant gene expression by cell type-specific profiling (Schnable *et al.*, 2004). Transcripts of single cells are retrieved by microcapillary methods, protoplast preparation, and laser-capture microdissection before expression profiling. By protoplasting *Arabidopsis* root cells that were cell type specifically labeled with green fluorescent protein (GFP), and classifying them into three developmental stages, a three-dimensional gene expression map of the *Arabidopsis* root was constructed on the basis of a whole genome microarray (Birnbaum *et al.*, 2003). This approach enabled the correlation of the expression of groups of genes to specific cell fates.

2. Proteomics

Proteomics describes the science of analyzing the complement of proteins (the proteome) in a given sample. Proteins are the translation products of mRNAs, and some of them carry out the gene function triggered by transcription immediately. However, there is also regulation at the translational and posttranslational level to control gene activity. In yeast it was found that only a small number of transcripts and proteins followed similar profiles on carbon source perturbation whereas the majority showed quantitative as well as qualitative differences (Griffin *et al.*, 2002). Therefore, the profiling of gene expression at the transcriptome and proteome levels is complementary, and both levels are required for a "systems biology" approach. In addition, the protein complement is believed to be much more complex than the transcriptome. As estimated for humans, each gene encodes at least six proteins (Service, 2001) and 10^5–10^6 proteins are present in a eukaryotic cell (Patterson and Aebersold, 2003). Explanations for this complexity are alternative splicing of transcripts to generate proteins with different domains and a multitude of posttranslational modifications such as phosphorylation and glycosylation.

The challenges of proteomics in general, and plant proteomics in particular, are intact and comprehensive protein extraction, separation and identification, and, finally, data analysis and integration (Rose *et al.*, 2004). Because of the sensitive and reactive nature of proteins, elaborate extraction protocols using different solvents and physical treatments have been established

(Rabilloud, 1996). However, with most protocols only one protein fraction is extracted. It also must be noted that especially during plant development, metabolic changes such as cell wall synthesis occur that influence cross-reactivity with proteins to be extracted. This will necessitate adaptation of extraction protocols to generate correct protein profiles (Rose *et al*., 2004). Two-dimensional gel electrophoresis (2D-PAGE), originally introduced by O'Farrell (1975), has been adapted for protein separation and improved for reliable, large-scale proteomic analyses (Rabilloud, 2002). In the first dimension, the proteins are separated according to the isoelectric point depending on protein charge, and in the second dimension according to protein molecular mass. For protein visualization, classic stains such as Coomassie blue and silver, and covalent labeling of protein extracts with different fluorescent dyes, similar to those used in microarray transcriptome analysis, are used. The latter technique, called difference gel electrophoresis (DIGE; Unlu *et al*., 1997), allows the analysis of different protein samples in the same gel and thereby improves comparative studies. DIGE was applied to the analysis of photosynthetic proteins in mutant and wild-type plants (Kubis *et al*., 2003). For image processing and data analysis specialized software was developed (Rose *et al*., 2004). Gel-free approaches such as multidimensional protein identification technology (MudPIT), which is based on biphasic capillary liquid chromatography, were shown to complement 2D-PAGE efficiently in detecting proteins (Koller *et al*., 2002). In fact, with the exception of 165 proteins that were detectable only by 2D-PAGE, all of the approximately 2500 proteins identified in leaves, roots, and seeds of rice were discovered by MudPIT.

The final step is the identification of the proteins, their association with a sequence and a gene. First, proteins are digested in-gel by specific proteases to yield defined peptide fragments. Thanks to improvement and adaptation, mass spectroscopy is now the method of choice for the identification of peptide mass fingerprints by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry or the *de novo* sequencing of peptide fragments by electrospray ionization (ESI)–tandem mass spectrometry (MS/MS) (Aebersold and Mann, 2003). Apart from the high costs associated particularly with the latter approach, the success of a proteomics project largely depends on the availability of cDNA or genomic sequence data to compare the experimental with the predicted peptide fragments for identification. The aim of large-scale identification of protein populations is at present feasible only with two plant species, *Arabidopsis thaliana* and rice, for which the genome sequence is known (see Section IV). The complexity of plant proteomes puts severe constraints on interspecies comparisons (Saravanan and Rose, 2004). The 2D gels of complex protein samples from tomato and pepper fruits showed little congruence, whereas comparative transcriptomics of fruit ripening have already yielded first results (Fei *et al*.,

2004). More reports are available on intraspecies analyses of genetic variability of proteins and their use as genetic markers (Canovas *et al.*, 2004; Thiellement *et al.*, 1999). The 2D-PAGE approach appears also to be suited for studies of phylogenetic relationships between different Triticeae species, based on the computational analysis of shared and nonshared protein spots (Zivy *et al.*, 1995). However, at that time the identity of proteins could not be verified by sequence analysis.

The resolution of 2D gels is limited to identify not more than 2000 plant proteins per sample (Tsugita and Kamo, 1999). This technical limitation generated the idea that analyzing several, less complex subproteomes separately helps in assembling the entire proteome. Large-scale studies on the proteome of chloroplasts and mitochondria have been carried out (Millar *et al.*, 2005; Peltier *et al.*, 2002). To ensure organelle purity, isotope-coded affinity tags (ICATs) were used to specifically label the proteins of one cell compartment (Dunkley *et al.*, 2004). Other studies focused on membrane proteins (Borner *et al.*, 2005) or cell wall-associated proteins (Chivasa *et al.*, 2002). These analyses led to the discovery of proteins that were not previously identified or predicted. The association of an unknown protein with a location in the cell and with other known proteins in the same compartment provides a first clue to its function.

Most proteins interact with one or several other proteins in complexes or networks to execute their function. The analysis of protein interaction *in vitro* was greatly facilitated by the yeast two-hybrid (Y2H) system (Fields and Song, 1989). The system was adapted to large-scale plant projects (Soellick and Uhrig, 2001), and numerous analyses were performed in plants to characterize groups of interacting proteins such as transcription factors (Ouellet *et al.*, 2001; Zimmermann *et al.*, 2004).

The first results from plant protein arrays have been reported (Kramer *et al.*, 2004). Clones from a cDNA expression library of filial tissue of developing barley seeds were selected for expression of the encoded proteins and fixed on a chip. This array was subjected to a kinase assay using a specific barley protein kinase and [γ-^{33}P]ATP as substrate. The identification of target proteins for posttranslational modification by autoradiography and phosphor imaging indicates the power of this emerging technology in plant functional genomics.

3. Metabolomics

Metabolomics is the analysis of all low molecular weight molecules present in cells in a particular physiological or developmental state. It is a nontargeted, open and unbiased approach toward the complete characterization of the metabolome, which is estimated to comprise between 90,000 and 200,000 different compounds (Fiehn, 2002). At the level of metabolites, the

components of the metabolic networks set up by gene expression and regulated by, but also regulating, protein activity are directly captured. Despite the high number of different metabolites it must be seen that there is also redundancy in their use by the cell. The same metabolite might play a role in different pathways in the same compartment or in different cell compartments, in which the metabolite concentration might be quite different. An example is malate, which takes part in the tricarboxylic acid pathway and in amino acid biosynthesis, and shuttles between the cytosol and organelles. In such cases, the information associated with the identification and quantification of the metabolite in a tissue sample is limited.

At present, the complex chemical nature of the molecules, which differ in their composition, configuration, charge, polarity, mass, and susceptibility to environmental conditions, restricts the number of individual compounds that can be analyzed in a single extraction and detection protocol. Therefore, rapid methods for tissue harvesting, extraction, and sample preparation were developed to minimize artifacts and to increase the number of detectable metabolites (Fiehn, 2002). To obtain a better resolution, complex mixtures are first subjected to a chromatographic step such as gas or liquid chromatography. The separated compounds are then analyzed by different types of mass spectrometers for qualitative analysis. The chromatographic retention indices and the mass-to-charge ratios of the metabolites as well as the mass spectral fragmentation patterns serve to identify each substance. The identification of new metabolites is still laborious (Fiehn *et al*., 2000b). The relatively small number of cataloged metabolites accounts for the fact that only 77 of 150 metabolites detected in potato tubers were identified (Roessner *et al*., 2000), as were only 90 of 400 metabolites detected in phloem exudates of *Cucurbita maxima* (Fiehn, 2003), and only 75 of the 325 metabolites that increased with cold acclimation of *Arabidopsis thaliana* (Cook *et al*., 2004). Specific mass spectral data collections are now being developed by plant laboratories in addition to commercially available resources (Bino *et al*., 2004). To help the interpretation of primary data with respect to metabolic pathways, databases and data analysis tools are accessible by internet (Bino *et al*., 2004). Biological replicates were found to have the largest effect on variability, causing 40% standard deviation for metabolites in *Arabidopsis* leaves (Fiehn *et al*., 2000a), whereas the analytical variability contributed less than 10%. This high biological variability reflects the dynamic state of the metabolome and stresses the necessity of controlled environment for standardized analyses.

A further challenge is the handling of large data sets in comparative metabolomics because the metabolites identified must be grouped with respect to occurrence and abundance in different samples. To extract this information, statistical data-mining procedures for multivariate clustering such as hierarchical component analysis (HCA) and principal component

analysis (PCA) are performed (Fiehn *et al.*, 2000a). Metabolic phenotypes are defined, related to genotype, and the distances between two samples are evaluated. Metabolomics research currently aims at having a framework of standards to be able to link the results between different laboratories and to transfer procedures (Jenkins *et al.*, 2004).

Because of its complexity, metabolomics is at present the greatest challenge in plant functional genomics, but also the most promising, particularly in combination with other approaches such as transcriptomics and proteomics. Integrated protocols have already been developed (Weckwerth *et al.*, 2004), and first results based on integration of transcriptomics and metabolomics to analyze responses to nitrogen and/or sulfur deficiencies in *Arabidopsis* have been reported (Hirai *et al.*, 2004). In this case, coordinated modulation of genes involved in glucosinolate metabolism was revealed, and the potential of such analyses for the genetic engineering of crop plants with modified metabolite content was highlighted.

V. Conclusion and Outlook

This chapter has shown that plant genome analysis is a rapidly evolving and growing part of genome research. Fruitful collaborations with specialists in disciplines such as nanotechnology and engineering have helped to upscale experiments such as nucleic acid detection, sequencing, and protein as well as metabolite identification. There is a strong trend toward miniaturization of the biological experiments in order to increase throughput for systematic genome-wide analyses. Second, genome analysis aims at increasing spatial and temporal resolution when studying biological processes. Looking at many single components in parallel at the same time, rather than at complex mixtures or fractions separated by long intervals of time, helps to better understand the dynamics of biological processes. At the same time it is now possible to pinpoint single nucleotides in whole chromosomes and to study their particular function. Because of the special features of plant cells, plant genome research faces great challenges for the adaptation of the technologies addressed. The new developments lead to an unprecedented size of data sets which can be handled and interpreted only with adequate bioinformatics software, another area of research. Thus, in modern plant molecular genetics the strength will come from an integration of resources to complement each other. This is the aim that plant systems biology is trying to achieve, and the efforts are motivated by our need for healthy plants in a healthy environment.

Plants are the basis for the survival of all "higher" organisms on Earth, and the manipulation of crop plants in agriculture is the material basis of

most human civil societies. The results of research in chemistry, physics, plant biology, and genetics have changed agricultural practice and plant breeding in the nineteenth and twentieth centuries. These changes had a dramatic impact on the present way of life in industrialized countries, as it reduced the input of labor required to ensure adequate food supply for a growing population to a fraction. Molecular biology and genetics have changed plant science completely. What impact, besides satisfying our curiosity, will modern plant genome research have, hopefully for the good of mankind in the twenty-first century? We envision the same two options as seen for genome research in humans and animal models: new therapeutics and new diagnostics.

- New therapeutics: Knowledge of the molecular basis of plant development, morphology, and chemical composition can lead to the discovery of new targets for crop protection or to the generation of novel crops by transgenic approaches.
- New diagnostics: The same knowledge can also be used to diagnose more efficiently the effects and consequences of natural variation, not only on crop performance but also on plant ecology and adaptation. As in medicine, correct diagnosis is the first prerequisite for effective treatment.

References

Acarkan, A., Rossberg, M., Koch, M., and Schmidt, R. (2000). Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis thaliana* and *Capsella rubella*. *Plant J.* **23,** 55–62.

Adams, M. A., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. (1991). Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* **252,** 1651–1656.

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

Aharoni, A., and Vorst, O. (2001). DNA microarrays for functional plant genomics. *Plant Mol. Biol.* **48,** 99–118.

Ahn, S., Anderson, J. A., Sorrells, M. E., and Tanksley, S. D. (1993). Homoeologous relationships of rice, wheat and maize chromosomes. *Mol. Gen. Genet.* **241,** 483–490.

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

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., Jeske, A., Koesema, E., Meyers, C. C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D. E., Marchand, T., Risseeuw, E.,

Brogden, D., Zeko, A., Crosby, W. L., Berry, C. C., and Ecker, J. R. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301,** 653–657.

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

Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H. M., and Somerville, C. R. (1992). Map-based cloning of a gene controlling ω-3 fatty acid desaturation in *Arabidopsis*. *Science* **258,** 1353–1355.

Arumuganathan, K., and Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9,** 208–218.

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., van der Hoeven, R. S., de Bruijn, S. M., Vreugdenhil, D., Zabeau, M., and Visser, R. G. (1996). Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. *Plant J.* **9,** 745–753.

Badr, A., Müller, K., Schäfer-Pregl, R., El Rabey, H., Effgen, S., Ibrahim, H. H., Pozzi, C., Rohde, W., and Salamini, F. (2000). On the origin and domestication history of barley (*Hordeum vulgare*). *Mol. Biol. Evol.* **17,** 499–510.

Ballvora, A., Ercolano, M. R., Weiß, J., Meksem, K., Bormann, C., Oberhagemann, P., Salamini, F., and Gebhardt, C. (2002). The *R1* gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. *Plant J.* **30,** 361–371.

Barker, D. G., Bianchi, S., Blondon, F., Dattée, Y., Duc, G., Essad, S., Flament, P., Gallusci, P., Génier, G., Guy, P., Muel, X., Tourneur, J., Dénarié, J., and Huguet, T. (1990). *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*–legume symbiosis. *Plant Mol. Biol. Rep.* **8,** 40–49.

Barone, A., Ritter, E., Schachtschabel, U., Debener, T., Salamini, F., and Gebhardt, C. (1990). Localization by restriction fragment length polymorphism mapping in potato of a major dominant gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Mol. Gen. Genet.* **224,** 177–182.

Barrett, J. H. (1992). Genetic mapping based on radiation hybrid data. *Genomics* **13,** 95–103.

Barry, G. F. (2001). The use of the Monsanto draft rice genome sequence in research. *Plant Physiol.* **125,** 1164–1165.

Bedell, J. A., Budiman, M. A., Nunberg, A., Citek, R. W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K., McMenamy, J., Smith, M., Holeman, H., Roe, B. A., Wiley, G., Korf, I. F., Rabinowicz, P. D., Lakey, N., McCombie, W. R., Jeddeloh, J. A., and Martienssen, R. A. (2005). Sorghum genome sequencing by methylation filtration. *PLoS Biol.* **3,** e13.

Bellin, D., Werber, M., Theis, T., Weisshaar, B., and Schneider, K. (2002). EST sequencing, annotation and macroarray transcriptome analysis identify preferentially root-expressed genes in sugar beet. *Plant Biol.* **4,** 700–710.

Bennett, M. D., and Smith, J. B. (1991). Nuclear DNA amounts in angiosperms. *Phil. Trans. R. Soc. Lond. B* **334,** 309–345.

Bennetzen, J. L., and Freeling, M. (1993). Grasses as a single genetic system: Genome composition, collinearity and compatibility. *Trends Genet.* **9,** 259–261.

Bernatzky, R., and Tanksley, S. D. (1986). Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* **112,** 887–898.

Bezant, J., Laurie, D., Pratchett, N., Chojecki, J., and Kearsey, M. (1997). Mapping QTL controlling yield and yield components in a spring barley (*Hordeum vulgare* L.) cross using marker regression. *Mol. Breed.* **3,** 29–38.

Binelli, G., and Bucci, G. (1994). A genetic linkage map of *Picea abies* Karts., based on RAPD markers, as tool in population genetics. *Gene. Theor. Appl.* **88,** 283–288.

Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B. J., Mendes, P., Roessner-Tunali, U., Beale, M. H., Trethewey, R. N., Lange, B. M., Wurtele, E. S., and Sumner, L. W. (2004). Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* **9,** 418–425.

Birnbaum, K., Shasha, D. E., Wang, J. Y., Jung, J. W., Lambert, G. M., Galbraith, D. W., and Benfey, P. N. (2003). A gene expression map of the *Arabidopsis* root. *Science* **302,** 1956–1960.

Boivin, K., Acarkan, A., Mbulu, R. S., Clarenz, O., and Schmidt, R. (2004). The *Arabidopsis* genome sequence as a tool for genome analysis in Brassicaceae: A comparison of the *Arabidopsis* and *Capsella rubella* genomes. *Plant Physiol.* **135,** 735–744.

Bonierbale, M. W., Plaisted, R. L., and Tanksley, S. D. (1988). RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* **120,** 1095–1104.

Borevitz, J. O., Liang, D., Plouffe, D., Chang, H.-S., Zhu, T., Weigel, D., Berry, C. C., Winzeler, E., and Chory, J. (2003). Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res.* **13,** 513–523.

Börner, A., Schumann, E., Fürste, A., Cöster, H., Leithold, B., Röder, M. S., and Weber, W. E. (2002). Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **105,** 921–936.

Borner, G. H., Sherrier, D. J., Weimar, T., Michaelson, L. V., Hawkins, N. D., Macaskill, A., Napier, J. A., Beale, M. H., Lilley, K. S., and Dupree, P. (2005). Analysis of detergent-resistant membranes in *Arabidopsis*: Evidence for plasma membrane lipid rafts. *Plant Physiol.* **137,** 104–116.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32,** 314–331.

Bowers, J. E., Chapman, B. A., Rong, J., and Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422,** 433–438.

Bradshaw, H. D., Jr., Ceulemans, R., Davis, J., and Stettler, R. (2000). Emerging model systems in plant biology: Poplar (*Populus*) as a model forest tree. *J. Plant Growth Regul.* **19,** 306–313.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M. (2001). Minimum information about a microarray experiment (MIAME): Toward standards for microarray data. *Nat. Genet.* **29,** 365–371.

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., Eletr, S., Albrecht, G., Vermaas, E., Williams, S. R., Moon, K., Burcham, T., Pallas, M., Du Bridge, R. B., Kirchner, J., Fearon, K., Mao, J., and Corcoran, K. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18,** 630–634.

Breyne, P., Dreesen, R., Cannoot, B., Rombaut, D., Vandepoele, K., Rombauts, S., Vanderhaeghen, R., Inze, D., and Zabeau, M. (2003). Quantitative cDNA-AFLP analysis for genome-wide expression studies. *Mol. Genet. Genomics* **269,** 173–179.

Brown, G. G., Formanova, N., Jin, H., Wargachuk, R., Dondy, C., Patil, P., Laforest, M., Zhang, J., Cheung, W. Y., and Landry, B. S. (2003). The radish *Rfo* restorer gene of *Ogura* cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J.* **35,** 262–272.

Brown, P. O., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21,** 33–37.

Brunner, A. M., Busov, V. B., and Strauss, S. H. (2004). Poplar genome sequence: Functional genomics in an ecologically dominant plant species. *Trends Plant Sci.* **9,** 49–56.

Brunner, S., Keller, B., and Feuillet, C. (2003). A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the *Rph7* locus. *Genetics* **164,** 673–683.

Buckler, E. S., Thornsberry, J. M., and Kresovich, S. (2001). Molecular diversity, structure and domestication of grasses. *Genet. Res.* **77,** 213–218.

Burke, D. T., Carle, G. F., and Olson, M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236,** 806–812.

Calenge, F., Faure, A., Goerre, M., Gebhardt, C., Van de Weg, Parisi, L., and Durel, C.-E. (2004). Quantitative trait loci (QTL) analysis reveals both broad-spectrum and isolate-specific QTL for scab resistance in an apple progeny challenged with eight isolates of *Venturia inaequalis*. *Phytopathology* **94,** 370–379.

Canovas, F. M., Dumas-Gaudot, E., Recorbet, G., Jorrin, J., Mock, H. P., and Rossignol, M. (2004). Plant proteome analysis. *Proteomics* **4,** 285–298.

Carle, G. F., and Olson, M. V. (1984). Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. *Nucleic Acids Res.* **12,** 5647–5664.

Castelli, V., Aury, J.-M., Jaillon, O., Wincker, P., Clepet, C., Menard, M., Cruaud, C., Quetier, F., Scarpelli, C., Schächter, V., Temple, G., Caboche, M., Weissenbach, J., and Salanoubat, M. (2004). Whole genome sequence comparisons and "full-length" cDNA sequences: A combined approach to evaluate and improve *Arabidopsis* genome annotation. *Genome Res.* **14,** 406–413.

Castiglioni, P., Pozzi, C., Heun, M., Terzi, V., Müller, K. J., Rohde, W., and Salamini, F. (1998). An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. *Genetics* **149,** 2039–2056.

Causse, M., Duffe, P., Gomez, M. C., Buret, M., Damidaux, R., Zamir, D., Gur, A., Chevalier, C., Lemaire-Chamley, M., and Rothan, C. (2004). A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. *J. Exp. Bot.* **55,** 1671–1685.

Causse, M. A., Fulton, T. M., Cho, Y. G., Ahn, S. N., Chunwongse, J., Wu, K., Xiao, J., Yu, Z., Ronald, P. C., Harrington, S. E., Second, G., McCouch, S. R., and Tanksley, S. D. (1994). Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* **138,** 1251–1274.

Chang, C., Bowman, J. L., De John, A. W., Lander, E. S., and Meyerowitz, E. M. (1988). Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **85,** 6856–6860.

Chang, Y. L., Tao, Q., Scheuring, C., Ding, K., Meksem, K., and Zhang, H. B. (2001). An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence. *Genetics* **159,** 1231–1242.

Chao, S., Sharp, P. J., Worland, A. J., Warham, E. J., Koebner, R. M. D., and Gale, M. D. (1989). RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. *Theor. Appl. Genet.* **78,** 495–504.

Chen, M., SanMiguel, P., De Oliveira, A. C., Woo, S. S., Zhang, H., Wing, R. A., and Bennetzen, J. L. (1997). Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. *Proc. Natl. Acad. Sci. USA* **94,** 3431–3435.

Chen, M., Presting, G., Barbazuk, W. B., Goicoechea, J. L., Blackmon, B., Fang, G., Kim, H., Frisch, D., Yu, Y., Sun, S., Higingbottom, S., Phimphilai, J., Phimphilai, D., Thurmond, S., Gaudette, B., Li, P., Liu, J., Hatfield, J., Main, D., Farrar, K., Henderson, C., Barnett, L., Costa, R., Williams, B., Walser, S., Atkins, M., Hall, C., Budiman, M. A., Tomkins, J. P., Luo, M., Bancroft, I., Salse, J., Regad, F., Mohapatra, T., Singh, N. K., Tyagi, A. K., Soderlund, C., Dean, R. A., and Wing, R. A. (2002). An integrated physical and genetic map of the rice genome. *Plant Cell* **14,** 537–545.

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, Z., Presting, G. G., Buell, C. R., Wing, R. A., and Jiang, J. (2001). High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. *Genetics* **157,** 1749–1757.

Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C. R., Gu, M., Blattner, F. R., and Jiang, J. (2002). Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* **14,** 1691–1704.

Chivasa, S., Ndimba, B. K., Simon, W. J., Robertson, D., Yu, X. L., Knox, J. P., Bolwell, P., and Slabas, A. R. (2002). Proteomic analysis of the *Arabidopsis thaliana* cell wall. *Electrophoresis* **23,** 1754–1765.

Cho, R. J., Mindrinos, M., Richards, D. R., Sapolsky, R. J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T. L., Stammers, M., Federspiel, N., Theologis, A., Yang, W. H., Hubbell, E., Au, M., Chung, E. Y., Lashkari, D., Lemieux, B., Dean, C., Lipshutz, R. J., Ausubel, F. M., Davis, R. W., and Oefner, P. J. (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.* **23,** 203–207.

Christou, P. (1997). Rice transformation: Bombardment. *Plant Mol. Biol.* **35,** 197–203.

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.

Coe, E., Cone, K., McMullen, M., Chen, S. S., Davis, G., Gardiner, J., Liscum, E., Polacco, M., Paterson, A., Sanchez-Villeda, H., Soderlund, C., and Wing, R. (2002). Access to the maize genome: An integrated physical and genetic map. *Plant Physiol.* **128,** 9–12.

Cold Spring, Harbor Laboratory, Washington University Genome Sequencing Center, and PE Biosystems *Arabidopsis* Sequencing Consortium (2000). The complete sequence of a heterochromatic island from a higher eukaryote. *Cell* **100,** 377–386.

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.

Cone, K. C., McMullen, M. D., Bi, I. V., Davis, G. L., Yim, Y. S., Gardiner, J. M., Polacco, M. L., Sanchez-Villeda, H., Fang, Z., Schroeder, S. G., Havermann, S. A., Bowers, J. E., Paterson, A. H., Soderlund, C. A., Engler, F. W., Wing, R. A., and Coe, E. H., Jr. (2002). Genetic, physical, and informatics resources for maize: On the road to an integrated map. *Plant Physiol.* **130,** 1598–1605.

Cook, D., Fowler, S., Fiehn, O., and Thomashow, M. F. (2004). A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **101,** 15243–15248.

Copenhaver, G. P., Nickel, K., Kuromori, T., Benito, M.-I., Kaul, S., Lin, X., Bevan, M., Murphy, G., Harris, B., Parnell, L. D., McCombie, W. R., Martienssen, R. A., Marra, M., and Preuss, D. (1999). Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* **286,** 2468–2474.

Coulson, A., Sulston, H., Brenner, S., and Karn, J. (1986). Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **83,** 7821–7825.

Cove, D. (2000). The moss *Physcomitrella pateus*. *J. Plant Growth Regul.* **79,** 272–283.

Crick, F. (1970). Central dogma of molecular biology. *Nature* **227,** 561–563.

Cutler, S. R., Ehrhardt, D. W., Griffitts, J. S., and Somerville, C. R. (2000). Random GFP: cDNA fusions enable visualization of subcellular structures in cells of *Arabidopsis* at a high frequency. *Proc. Natl. Acad. Sci. USA* **97,** 3718–3723.

Damm, B., Schmidt, R., and Willmitzer, L. (1989). Efficient transformation of *Arabidopsis thaliana* using direct gene transfer to protoplasts. *Mol. Gen. Genet.* **217,** 6–12.

Dear, P. H., and Cook, P. R. (1993). Happy mapping: Linkage mapping using a physical analogue of meiosis. *Nucleic Acids Res.* **21,** 13–20.

Dehmer, K. J., and Hammer, K. (2004). Taxonomic status and geographic provenance of germplasm accessions in the *Solanum nigrum* L. complex: AFLP data. *Genet. Resources Crop Evol.* **51,** 551–558.

Delseny, M. (2004). Re-evaluating the relevance of ancestral shared synteny as a tool for crop improvement. *Curr. Opin. Plant Biol.* **7,** 126–131.

Desloire, S., Gherbi, H., Laloui, W., Marhadour, S., Clouet, V., Catholico, L., Falentin, C., Giancola, S., Renard, M., Budar, F., Small, I., Caboche, M., Delourme, R., and Bendahmane, A. (2003). Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Rep.* **4,** 588–594.

Devos, K. M., and Gale, M. D. (2000). Genome relationships: The grass model in current research. *Plant Cell* **12,** 637–646.

Diatchenko, L., Lau, Y. F. C., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D., and Siebert, P. D. (1996). Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* **93,** 6025–6030.

Dominguez, I., Graziano, E., Gebhardt, C., Barakat, A., Berry, S., Arús, P., Delseny, M., and Barnes, S. (2003). Plant genome archaeology: Evidence for conserved ancestral chromosome segments in dicotyledonous plant species. *Plant Biotechnol. J.* **1,** 91–99.

Dong, F., and Jiang, J. (1998). Non-Rabl patterns of centromere and telomere distribution in the interphase nuclei of plant cells. *Chromosome Res.* **6,** 551–558.

Dong, F., McGrath, J. M., Helgeson, J. P., and Jiang, J. (2001). The genetic identity of alien chromosomes in potato breeding lines revealed by sequential GISH and FISH analyses using chromosome-specific cytogenetic DNA markers. *Genome* **44,** 729–734.

Doudrick, R. L., Heslop-Harrison, J. S., Nelson, C. D., Schmidt, T., Nance, W. L., and Schwarzacher, T. (1995). Karyotyping slash pine *Pinus elliottii* var. *elliottii* using patterns of fluorescence *in situ* hybridization and fluorochrome banding. *J. Hered.* **86,** 289–296.

Dunford, R. P., Yano, M., Kurata, N., Sasaki, T., Huestis, G., Rocheford, T., and Laurie, D. A. (2002). Comparative mapping of the barley *Ppd-H1* photoperiod response gene region, which lies close to a junction between two rice linkage segments. *Genetics* **161,** 825–834.

Dunkley, T. P., Dupree, P., Watson, R. B., and Lilley, K. S. (2004). The use of isotope-coded affinity tags (ICAT) to study organelle proteomes in *Arabidopsis thaliana*. *Biochem. Soc. Trans.* **32,** 520–523.

Durrant, W. E., Rowland, O., Piedras, P., Hammond-Kosack, K. E., and Jones, J. D. (2000). cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles. *Plant Cell* **12,** 963–977.

Edwards, K. J., Thompson, H., Edwards, D., de Saizieu, A., Sparks, C., Thompson, J. A., Greenland, A. J., Eyers, M., and Schuch, W. (1992). Construction and characterisation of a yeast artificial chromosome library containing three haploid maize genome equivalents. *Plant Mol. Biol.* **19,** 299–308.

Egener, T., Granado, J., Guitton, M.-C., Hohe, A., Holtorf, H., Lucht, J. M., Rensing, S. A., Schlink, K., Schulte, J., Schween, G., Zimmermann, S., Duwenig, E., Rak, B., and Reski, R. (2002). High frequency of phenotypic deviations in *Physcomitrella patens* plants transformed with a gene-disruption library. *BMC Plant Biol.* **2,** 6.

Emerson, R. A., Beadle, G. W., and Fraser, A. C. (1935). A summary of linkage studies in maize. *Cornell Univ. Agric. Exp. Stn. Memoir* **180,** 1.

Enoki, H., Izawa, T., Kawahara, M., Komatsu, M., Koh, S., Kyozuka, J., and Shimamoto, K. (1999). *Ac* as a tool for the functional genomics of rice. *Plant J.* **19,** 605–613.

Eshed, Y., and Zamir, D. (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141,** 1147–1162.

European Union *Arabidopsis* Sequencing Consortium Cold Spring Harbor Laboratory, Washington University, St. Louis, and PE Biosystems (1999). Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.. *Nature* **402,** 769–777.

European Union Chromosome 3 *Arabidopsis* Sequencing Consortium, The Institute for Genomic Research, and Kazusa DNA Research Institute (2000). Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* **408,** 820–822.

Fedorova, M., van de Mortel, J., Matsumoto, P. A., Cho, J., Town, C. D., Van den Bosch, K. A., Gantt, J. S., and Vance, C. P. (2002). Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiol.* **130,** 519–537.

Fei, Z., Tang, X., Alba, R. M., White, J. A., Ronning, C. M., Martin, G. B., Tanksley, S. D., and Giovannoni, J. J. (2004). Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. *Plant J.* **40,** 47–59.

Feltus, F. A., Wan, J., Schulze, S. R., Estill, J. C., Jiang, N., and Paterson, A. H. (2004). An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome Res.* **14,** 1812–1819.

Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., Li, Y., Zhu, J., Liu, Y., Hu, X., Jia, P., Zhang, Y., Zhao, Q., Ying, K., Yu, S., Tang, Y., Weng, Q., Zhang, L., Lu, Y., Mu, J., Lu, Y., Zhang, L. S., Yu, Z., Fan, D., Liu, X., Lu, T., Li, C., Wu, Y., Sun, T., Lei, H., Li, T., Hu, H., Guan, J., Wu, M., Zhang, R., Zhou, B., Chen, Z., Chen, L., Jin, Z., Wang, R., Yin, H., Cai, Z., Ren, S., Lv, G., Gu, W., Zhu, G., Tu, Y., Jia, J., Zhang, Y., Chen, J., Kang, H., Chen, X., Shao, C., Sun, Y., Hu, Q., Zhang, X., Zhang, W., Wang, L., Ding, C., Sheng, H., Gu, J., Chen, S., Ni, L., Zhu, F., Chen, W., Lan, L., Lai, Y., Cheng, Z., Gu, M., Jiang, J., Li, J., Hong, G., Xue, Y., and Han, B. (2002). Sequence and analysis of rice chromosome 4. *Nature* **420,** 316–320.

Feuillet, C., and Keller, B. (1999). High gene density is conserved at syntenic loci of small and large grass genomes. *Proc. Natl. Acad. Sci. USA* **96,** 8265–8270.

Feuillet, C., and Keller, B. (2000). Comparative genomics in the grass family: Molecular characterization of grass genome structure and evolution. *Ann. Bot.* **89,** 3–10.

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

Fiehn, O. (2003). Metabolic networks of *Cucurbita maxima* phloem. *Phytochemistry* **62,** 875–886.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N., and Willmitzer, L. (2000a). Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18,** 1157–1161.

Fiehn, O., Kopka, J., Trethewey, R. N., and Willmitzer, L. (2000b). Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* **72,** 3573–3580.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein–protein interactions. *Nature* **340,** 245–246.

Fizames, C., Munos, S., Cazettes, C., Nacry, P., Boucherez, J., Gaymard, F., Piquemal, D., Delorme, V., Commes, T., Doumas, P., Cooke, R., Marti, J., Sentenac, H., and Gojon, A. (2004). The *Arabidopsis* root transcriptome by serial analysis of gene expression: Gene identification using the genome sequence. *Plant Physiol.* **134,** 67–80.

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S., IV. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* **54,** 357–374.

Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., and Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* **251,** 767–773.

Forsström, P. O., Merker, A., and Schwarzacher, T. (2002). Characterisation of mildew resistant wheat-rye substitution lines and identification of an inverted chromosome by fluorescent *in situ* hybridisation. *Heredity* **88,** 349–355.

Fransz, P., De Jong, J. H., Lysak, M., Castiglione, I., and Schubert, I. (2002). Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. USA* **99,** 14584–14589.

Fransz, P. F., Alonso-Blanco, C., Liharska, T. B., Peeters, A. J. M., Zabel, P., and de Jong, J. H. (1996). High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibres. *Plant J.* **9,** 421–430.

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., van der Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K. B., and Tanksley, S. D. (2000). *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science* **289,** 85–88.

Fridman, E., Pleban, T., and Zamir, D. (2000). A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proc. Natl. Acad. Sci. USA* **97,** 4718–4723.

Fridman, E., Carrari, F., Liu, Y.-S., Fernie, A., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* **305,** 1786–1789.

Fu, H., and Dooner, H. K. (2002). Intraspecific violation of genetic colinearity and its implications in maize. *Proc. Natl. Acad. Sci. USA* **99,** 9573–9578.

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

Galibert, F., Finan, T. M., Long, S. R., Pühler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M. J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R. W., Dréano, S., Federspiel, N. A., Fisher, R. F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R. W., Jones, T., Kahn, D., Kahn, M. L., Kalman, S., Keating, D. H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M. C., Pohl, T. M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thébault, P., Vandenbol, M., Vorhölter, F.-J., Weidner, S., Wells, D. H., Wong, K., Yeh, K.-C., and Batut, J. (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* **293,** 668–672.

Gall, J. G., and Pardue, M. L. (1969). Formation and detection of RNA–DNA hybrid molecules in cytological preparations. *Proc. Natl. Acad. Sci. USA* **63,** 378–383.

Gardiner, J., Schroeder, S., Polacco, M. L., Sanchez-Villeda, H., Fang, Z., Morgante, M., Landewe, T., Fengler, K., Useche, F., Hanafey, M., Tingey, S., Chou, H., Wing, R., Soderlund, C., and Coe, E. H., Jr. (2004). Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. *Plant Physiol.* **134,** 1317–1326.

Gebhardt, C., and Salamini, F. (1992). Restriction-fragment-length-polymorphism analysis of plant genomes and its application to plant-breeding. *Int. Rev. Cytol.* **135,** 201–237.

Gebhardt, C., and Valkonen, J. P. T. (2001). Organization of genes controlling disease resistance in the potato genome. *Annu. Rev. Phytopathol.* **39,** 79–102.

Gebhardt, C., Walkemeier, B., Henselewski, H., Barakat, A., Delseny, M., and Stüber, K. (2003). Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* reveals structurally conserved domains and ancient duplications in the potato genome. *Plant J.* **34,** 529–541.

Gebhardt, C., Ballvora, A., Walkemeier, B., Oberhagemann, P., and Schüler, K. (2004). Assessing genetic potential in germplasm collections of crop plants by marker–trait association: A case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Mol. Breeding* **13,** 93–102.

Giraudat, J., Hauge, B. M., Valon, C., Smalle, J., Parcy, F., and Goodman, H. M. (1992). Isolation of the *Arabidopsis ABI3* gene by positional cloning. *Plant Cell* **4,** 1231–1261.

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

Grandillo, S., Ku, H. M., and Tanksley, S. D. (1999). Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theor. Appl. Genet.* **99,** 978–987.

Grattapaglia, D., and Sederoff, R. (1994). Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudotestcoss mapping strategy and RAPD markers. *Genetics* **137,** 1121–1137.

Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L., and Aebersold, R. (2002). Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol. Cell Proteomics* **1,** 323–333.

Grill, E., and Somerville, C. (1991). Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. *Mol. Gen. Genet.* **226,** 484–490.

Grube, R. C., Radwanski, E. R., and Jahn, M. M. (2000). Comparative genetics of disease resistance within the Solanaceae. *Genetics* **155,** 873–887.

Haas, B. J., Volfovsky, N., Town, C. D., Troukhan, M., Alexandrov, N., Feldmann, K. A., Flavell, R. B., White, O., and Salzberg, S. L. (2002). Full-length messenger RNA sequences greatly improve genome annotation. *Genome Biol.* **3**RESEARCH0029.

Handberg, K., and Stougaard, J. (1992). *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *Plant J.* **2,** 487–496.

Hansen, M., Kraft, T., Ganestam, S., Säll, T., and Nilsson, N.-O. (2001). Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. *Genet. Res.* **77,** 61–66.

Harper, L. C., and Cande, W. Z. (2000). Mapping a new frontier: Development of integrated cytogenetic maps in plants. *Funct. Integr. Genomics* **1,** 89–98.

Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S. Y., Antonio, B. A., Parco, A., Kajiya, H., Huang, N., Yamamoto, K., Nagamura, Y., Kurata, N., Khush, G. S., and Sasaki, T. (1998). A high-density rice genetic linkage map with 2275 markers using a single F_2 population. *Genetics* **148,** 479–494.

Helentjaris, T., Slocum, M., Wright, S., Schaefer, A., and Nienhuis, J. (1986). Construction of genetic-linkage maps in maize and tomato using restricion-fragment-length-polymorphism. *Theor. Appl. Genet.* **72,** 761–769.

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

Heslop-Harrison, J. S. (2000). Comparative genome organization in plants: From sequence and markers to chromatin and chromosomes. *Plant Cell* **12,** 617–636.

Heun, M., Schäfer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., and Salamini, F. (1997). Site of Einkorn wheat domestication identified by DNA fingerprinting. *Science* **278,** 1312–1314.

Hiei, Y., Komari, T., and Kubo, T. (1997). Transformation of rice mediated by *Agrobacterium tumefaciens*. *Plant Mol. Biol.* **35,** 205–218.

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.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K., and Sugiura, M. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* **217,** 185–194.

Hirochika, H., Guiderdoni, E., An, G., Hsing, Y. I., Eun, M. Y., Han, C. D., Upadhyaya, 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.

Hirschhorn, J. N., and Daly, K. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6,** 95–108.

Höfte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M., Moisan, A., Jourjon, M. F., Charpenteau, J. L., Berthomieu, P., Guerrier, D., Giraudat, J., Quigley, F., Thomas, F., Yu, D. Y., Mache, R., Raynal, M., Cooke, R., Grellet, F., Delseny, M., Parmentier, Y., Marcillac, G., Gigot, C., Fleck, J., Phillipps, G., Axelos, M., Bardet, C., Tremousaygue, D., and Lescure, B. (1993). An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J.* **4,** 1051–1061.

Holub, E. B. (2001). The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* **2,** 516–527.

Hulbert, S. H., Richter, T. E., Axtell, J. D., and Bennetzen, J. L. (1990). Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proc. Natl. Acad. Sci. USA* **87,** 4251–4255.

Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A., and de Jong, P. J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* **6,** 84–89.

Isidore, E., van Os, H., Andrzejewski, S., Bakker, J., Barrena, I., Bryan, G. J., Caromel, B., van Eck, H., Ghareeb, B., de Jong, W., van Koert, P., Lefebvre, V., Milbourne, D., Ritter, E., Rouppe van der Voort, J., Rousselle-Bourgeois, F., van Vliet, J., and Waugh, R. (2003). Toward a marker-dense meiotic map of the potato genome: Lessons from linkage group I. *Genetics* **165,** 2107–2116.

Ivandic, V., Thomas, W. T. B., Nevo, E., Zhang, Z., and Forster, B. P. (2003). Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in. *Hordeum spontaneum. Plant Breeding* **122,** 300–304.

Jackson, M. T. (1997). Conservation of rice genetic resources: The role of the International Rice Genebank at IRRI. *Plant Mol. Biol.* **35,** 61–67.

Jackson, S. A., Cheng, Z., Wang, M. L., Goodman, H. M., and Jiang, J. (2000). Comparative fluorescence *in situ* hybridization mapping of a 431-kb *Arabidopsis thaliana* bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome. *Genetics* **156,** 833–838.

Jander, G., Norris, S. R., Rounsley, S. D., Bush, D. F., Levin, I. M., and Last, R. L. (2002). *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol.* **129,** 440–450.

Jannoo, N., Grivet, L., Dookun, A., D'Hont, A., and Glaszmann, J. C. (1999). Linkage disequilibrium among modern sugarcane cultivats. *Theor. Appl. Genet.* **99,** 1053–1060.

Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A. R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R. J., Hall, R., Kopka, J., Lane, G. A., Lange, B. M., Liu, J. R., Mendes, P., Nikolau, B. J., Oliver, S. G., Paton, N. W., Rhee, S., Roessner-Tunali, U., Saito,

K., Smedsgaard, J., Sumner, L. W., Wang, T., Walsh, S., Wurtele, E. S., and Kell, D. B. (2004). A proposed framework for the description of plant metabolomics experiments and their results. *Nat. Biotechnol.* **22,** 1601–1606.

Jiang, J. M., and Gill, B. S. (1994). Nonisotopic *in situ* hybridization and plant genome mapping: The first 10 years. *Genome* **37,** 717–725.

John, H. A., Birnstiel, M. L., and Jones, K. W. (1969). RNA–DNA hybrids at the cytological level. *Nature* **223,** 582–587.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., and Tabata, S. (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res.* **7,** 331–338.

Kazusa DNA Research Institute, The Cold Spring Harbor and Washington University in St. Louis Sequencing Consortium, and European Union *Arabidopsis* Genome Sequencing Consortium (2000). Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature* **408,** 823–826.

Khorasani, M. Z., Hennig, S., Imre, G., Asakawa, S., Palczewski, S., Berger, A., Hori, H., Naruse, K., Mitani, H., Shima, A., Lehrach, H., Wittbrodt, J., Kondoh, H., Shimizu, N., and Himmelbauer, H. (2004). A first generation physical map of the medaka genome in BACs essential for positional cloning and clone-by-clone based genomic sequencing. *Mech. Dev.* **121,** 903–913.

Kilian, A., Stiff, C., and Kleinhofs, A. (1995). Barley telomeres shorten during differentiation but grow in callus culture. *Proc. Natl. Acad. Sci. USA* **92,** 9555–9559.

Kilian, A., Chen, J., Han, F., Steffenson, B., and Kleinhofs, A. (1997). Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. *Plant Mol. Biol.* **35,** 187–195.

Kirch, H. H., Nair, A., and Bartels, D. (2001). Novel ABA- and dehydration-inducible aldehyde dehydrogenase genes isolated from the resurrection plant *Craterostigma plantagineum* and *Arabidopsis thaliana*. *Plant J.* **28,** 555–567.

Kistner, C., and Parniske, M. (2002). Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* **7,** 511–518.

Klein, P. E., Klein, R. R., Cartinhour, S. W., Ulanch, P. E., Dong, J., Obert, J. A., Morishige, D. T., Schlueter, S. D., Childs, K. L., Ale, M., and Mullet, J. E. (2000). A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. *Genome Res.* **10,** 789–807.

Kolesnik, T., Szeverenyi, I., Bachmann, D., Kumar, C. S., Jiang, S., Ramamoorthy, R., Cai, M., Ma, Z. G., Sundaresan, V., and Ramachandran, S. (2004). Establishing an efficient *Ac/Ds* tagging system in rice: Large-scale analysis of *Ds* flanking sequences. *Plant J.* **37,** 301–314.

Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L., Deciu, C., Haynes, P. A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J., Wolters, D., and Yates, J. R., III. (2002). Proteomic survey of metabolic pathways in rice. *Proc. Natl. Acad. Sci. USA* **99,** 11969–11974.

Komatsu, S., and Tanaka, N. (2005). Rice proteome analysis: A step toward functional analysis of the rice genome. *Proteomics* **5,** 938–949.

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

Koornneef, M., van Eden, J., Hanhart, C. J., Stam, P., Braaksma, F. J., and Feenstra, F. J. (1983). Linkage map of *Arabidopsis thaliana*. *J. Hered.* **74,** 265–272.

Kotani, H., Sato, S., Fukami, M., Hosouchi, T., Nakazaki, N., Okumura, S., Wada, T., Liu, Y. G., Shibata, D., and Tabata, S. (1997). A fine physical map of *Arabidopsis thaliana* chromosome 5: Construction of a sequence-ready contig map. *DNA Res.* **4,** 371–378.

Kraakman, A. T. W., Niks, R. E., van den Berg, P. M. M. M., Stam, P., and van Eeuwijk, F. A. (2004). Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* **168,** 435–446.

Kraft, T., Hansen, M., and Nilsson, M.-O. (2000). Linkage disequilibrium and fingerprinting in sugar beet. *Theor. Appl. Genet.* **101,** 323–326.

Kramer, A., Feilner, T., Possling, A., Radchuk, V., Weschke, W., Burkle, L., and Kersten, B. (2004). Identification of barley CK2α targets by using the protein microarray technology. *Phytochemistry* **65,** 1777–1784.

Ku, H. M., Vision, T., Liu, J., and Tanksley, S. D. (2000). Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci. USA* **97,** 9121–9126.

Kubis, S., Baldwin, A., Pate, R., Razzaq, A., Dupree, P., Lilley, K., Kurth, J., Leister, D., and Jarvis, P. (2003). The *Arabidopsis ppi1* mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins. *Plant Cell* **15,** 1859–1871.

Kynast, R. G., Okagaki, R. J., Galatowitsch, M. W., Granath, S. R., Jacobs, M. S., Stec, A. O., Rines, H. W., and Phillips, R. L. (2004). Dissecting the maize genome by using chromosome addition and radiation hybrid lines. *Proc. Natl. Acad. Sci. USA* **101,** 9921–9926.

Landry, B. S., Kesseli, R. V., Farrara, B., and Michelmore, R. W. (1987). A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance, and morphological markers. *Genetics* **116,** 331–337.

Lapitan, N. L. V. (1992). Organization and evolution of higher-plant nuclear genomes. *Genome* **35,** 171–181.

Lee, J. M., Grant, D., Vallejos, C. E., and Shoemaker, R. C. (2001). Genome organization in dicots. II. *Arabidopsis* as a ''bridging species'' to resolve genome evolution events among legumes. *Theor. Appl. Genet.* **103,** 765–773.

Lee, S., Kim, J., Son, J.-S., Nam, J., Jeong, D.-H., Lee, K., Jang, S., Yoo, J., Lee, J., Lee, D.-Y., Kang, H.-G., and An, G. (2003). Systematic reverse genetic screening of T-DNA tagged genes in rice for functional genomic analyses: MADS-box genes as a test case. *Plant Cell Physiol.* **44,** 1403–1411.

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

Leitch, I. J., Chase, M. W., and Bennett, M. D. (1998). Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. *Ann. Bot.* **82,** 85–94.

Leonards-Schippers, C., Gieffers, W., Schäfer-Pregl, R., Ritter, E., Knapp, S. J., Salamini, F., and Gebhardt, C. (1994). Quantitative resistance to *Phytophthora infestans* in potato: A case study for QTL mapping in an allogamous plant species. *Genetics* **137,** 67–77.

Li, L., Strahwald, J., Hofferbert, H.-R., Lübeck, J., Tacke, E., Junghans, H., Wunder, J., and Gebhardt, C. (2005). DNA variation at the invertase locus *invGE/GF* is associated with tuber quality traits in populations of potato breeding clones. *Genetics* **170,** 813–822.

Li, X., Song, Y., Century, K., Straight, S., Ronald, P., Dong, X., Lassner, M., and Zhang, Y. (2001). A fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant J.* **27,** 235–342.

Liang, P., and Pardee, A. B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257,** 967–971.

Lilly, J. W., Havey, M. J., Jackson, S. A., and Jiang, J. (2001). Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants. *Plant Cell* **13,** 245–254.

Lin, X., Kaul, S., Rounsley, S., Shea, T. P., Benito, M.-I., Town, C. D., Fujii, C. Y., Mason, T., Bowman, C. L., Barnstead, M., Feldblyum, T. V., Buell, C. R., Ketchum, K. A., Lee, J., Ronning, C. M., Koo, H. L., Moffat, K. S., Cronin, L. A., Shen, M., Pai, G., Van Aken, S., Umayam, L., Tallon, L. J., Gill, J. E., Adams, M. D., Carrera, A. J., Creasy, T. H.,

Goodman, H. M., Somerville, C. R., Copenhaver, G. P., Preuss, D., Nierman, W. C., White, O., Eisen, J. A., Salzberg, S. L., Fraser, C. M., and Venter, J. C. (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402,** 761–768.

Lister, C., and Dean, C. (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J.* **4,** 745–750.

Liu, J., Blaylock, L. A., Endre, G., Cho, J., Town, C. D., Van den Bosch, K. A., and Harrison, M. J. (2003). Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis. *Plant Cell* **15,** 2106–2123.

Livingstone, K. D., Lackney, V. K., Blauth, J. R., van Wijk, R., and Kyle-Jahn, M. (1999). Genome mapping in capsicum and the evolution of genome structure in the Solanaceae. *Genetics* **152,** 1183–1202.

Lloyd, A. M., Barnason, A. R., Rogers, S. G., Byrne, M. C., Fraley, R. T., and Horsch, R. B. (1986). Transformation of *Arabidopsis thaliana* with *Agrobacterium tumefaciens*. *Science* **234,** 464–466.

Lukens, L., Zou, F., Lydiate, D., Parkin, I., and Osborn, T. (2003). Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. *Genetics* **164,** 359–372.

Lysak, M. A., Pecinka, A., and Schubert, I. (2003). Recent progress in chromosome painting of *Arabidopsis* and related species. *Chromosome Res.* **11,** 195–204.

Mace, E. S., Gebhardt, C., and Lester, R. N. (1999). AFLP analysis of genetic relationships in the tribe Datureae (Solanaceae). *Theor. Appl. Genet.* **99,** 634–641.

Marra, M., Kucaba, T., Sekhon, M., Hillier, L., Martienssen, R., Chinwalla, A., Crockett, J., Fedele, J., Grover, H., Gund, C., McCombie, W. R., McDonald, K., McPherson, J., Mudd, N., Parnell, L., Schein, J., Seim, R., Shelby, P., Waterston, R., and Wilson, R. (1999). A map for sequence analysis of the *Arabidopsis thaliana* genome. *Nat. Genet.* **22,** 265–270.

Martin, G. B., Brommonschenkel, S. H., Chunwongse, J., Frary, A., Ganal, M. W., Spivey, R., Wu, T., Earle, E. D., and Tanksley, S. D. (1993). Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262,** 1432–1436.

Martin, G. B., Bogdanove, A. J., and Sessa, G. (2003). Understanding the functions of plant disease resistance proteins. *Annu. Rev. Plant Biol.* **54,** 23–61.

McClintock, B. (1929). Chromosome morphology in *Zea mays*. *Science* **69,** 629.

McCouch, S. R., and Doerge, R. W. (1995). QTL mapping in rice. *Trends Genet.* **11,** 482–487.

McCouch, S. R., Kochert, G., Yu, Z. H., Wang, Z. Y., Kush, G. S., Coffman, W. R., and Tanksley, S. D. (1988). Molecular mapping of rice chromosomes. *Theor. Appl. Genet.* **76,** 815–829.

McGinnis, K., Chandler, V., Cone, K., Kaeppler, H., Kaeppler, S., Kerschen, A., Pikaard, C., Richards, E., Sidorenko, L., Smith, T., Springer, N., and Wulan, T. (2005). Transgene-induced RNA interference as a tool for plant functional genomics. *Methods Enzymol.* **392,** 1–24.

McMullen, M. D., and Simcox, K. D. (1995). Genomic organization of disease and insect resistance genes in maize. *Mol. Plant Microbe Interact.* **8,** 811–815.

Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., and Koornneef, M. (1998). *Arabidopsis thaliana*: A model plant for genome analysis. *Science* **282,** 662, 679–682.

Meinke, D. W., Meinke, L. K., Showalter, T. C., Schissel, A. M., Mueller, L. A., and Tzafrir, I. (2003). A sequence-based map of *Arabidopsis* genes with mutant phenotypes. *Plant Physiol.* **131,** 409–418.

Meissner, R. C., Jin, H., Cominelli, E., Denekamp, M., Fuertes, A., Greco, R., Kranz, H. D., Penfield, S., Petroni, K., Urzainqui, A., Martin, C., Paz-Ares, J., Smeekens, S., Tonelli, C., Weisshaar, B., Baumann, E., Klimyuk, V., Marillonnet, S., Patel, K., Speulman, E., Tissier, A. F., Bouchez, D., Jones, J. J. D., Pereira, A., Wisman, E., and Bevan, M. (1999). Function search in a large transcription factor gene family in *Arabidopsis*: Assessing the potential of

reverse genetics to identify insertional mutations in R2R3 *MYB* genes. *Plant Cell* **11,** 1827–1840.

Menendez, C. M., Hall, A. E., and Gepts, P. (1997). A genetic linkage map of cowpea (*Vigna unguiculata*) developed from a cross between two inbred, domesticated lines. *Theor. Appl. Genet.* **95,** 1210–1217.

Menendez, C. M., Ritter, E., Schäfer-Pregl, R., Walkemeier, B., Kalde, A., Salamini, F., and Gebhardt, C. (2002). Cold sweetening in diploid potato: Mapping quantitative trait loci and candidate genes. *Genetics* **162,** 1423–1434.

Messing, J., Bharti, A. K., Karlowski, W. M., Gundlach, H., Kim, H. R., Yu, Y., Wei, F., Fuks, G., Soderlund, C. A., Mayer, K. F., and Wing, R. A. (2004). Sequence composition and genome organization of maize. *Proc. Natl. Acad. Sci. USA* **101,** 14349–14354.

Mewes, H. W., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Maierl, A., Oliver, S. G., Pfeiffer, F., and Zollner, A. (1997). Overview of the yeast genome. *Nature* **387**(Suppl.), 7–65.

Meyers, B. C., Scalabrin, S., and Morgante, M. (2004a). Mapping and sequencing complex genomes: Let's get physical!. *Nat. Rev. Genet.* **5,** 578–588.

Meyers, B. C., Tej, S. S., Vu, T. H., Haudenschild, C. D., Agrawal, V., Edberg, S. B., Ghazal, H., and Decola, S. (2004b). The use of MPSS for whole-genome transcriptional analysis in *Arabidopsis*. *Genome Res.* **14,** 1641–1653.

Milioni, D., Sado, P. E., Stacey, N. J., Roberts, K., and McCann, M. C. (2002). Early gene expression associated with the commitment and differentiation of a plant tracheary element is revealed by cDNA-amplified fragment length polymorphism analysis. *Plant Cell* **14,** 2813–2824.

Millar, A. H., Heazlewood, J. L., Kristensen, B. K., Braun, H. P., and Moller, I. M. (2005). The plant mitochondrial proteome. *Trends Plant Sci.* **10,** 36–43.

Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K., and Hirochika, H. (2003). Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* **15,** 1771–1780.

Moore, G., Gale, M. D., Kurata, N., and Flavell, R. B. (1993). Molecular analysis of small grain cereal genomes: Current status and prospects. *Biotechnology* **11,** 584–589.

Mozo, T., Fischer, S., Meier-Ewert, S., Lehrach, H., and Altmann, T. (1998). Use of the IGF BAC library for physical mapping of the *Arabidopsis thaliana* genome. *Plant J.* **16,** 377–384.

Mozo, T., Dewar, K., Dunn, P., Ecker, J. R., Fischer, S., Kloska, S., Lehrach, H., Marra, M., Martienssen, R., Meier-Ewert, S., and Altmann, T. (1999). A complete BAC-based physical map of the *Arabidopsis thaliana* genome. *Nat. Genet.* **22,** 271–275.

Murata, M. (2002). Telomeres and centromeres in plants. *Curr. Genomics* **3,** 527–538.

Nagaki, K., Tsujimoto, H., and Sasakuma, T. (1998). Dynamics of tandem repetitive Afa-family sequences in Triticeae, wheat-related species. *J. Mol. Evol.* **47,** 183–189.

Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P. B., Kim, M., Jones, K. M., Henikoff, S., Buell, C. R., and Jiang, J. (2004). Sequencing of a rice centromere uncovers active genes. *Nat. Genet.* **36,** 138–145.

Newman, T., Bruijn de, F. J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E., and Sommerville, C. (1994). Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Phys.* **106,** 1241–1255.

Nordborg, M., Borewitz, J. O., Bergelson, J., Berry, C. C., Chory, J., Hagenblad, J., Kreitman, M., Maloof, J. N., Noyes, T., Oefner, P. J., Stahl, E. A., and Weigel, D. (2002). The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* **30,** 190–193.

Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., and Kadowaki, K. (2002). The complete sequence of the rice (*Oryza sativa* L.) mitochondrial

genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genomics* **268,** 434–445.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., and Ohyama, K. (1992). Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: A primitive form of plant mitochondrial genome. *J. Mol. Biol.* **223,** 1–7.

O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250,** 4007–4021.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H., and Ozeki, H. (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* **322,** 572–574.

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

Olson, M. V., Dutchik, J. E., Graham, M. Y., Brodeur, G. M., Helms, C., Frank, M., MacCollin, M., Scheinman, R., and Frank, T. (1986). Random-clone strategy for genomic restriction mapping in yeast. *Proc. Natl. Acad. Sci. USA* **83,** 7826–7830.

O'Neill, C. M., and Bancroft, I. (2000). Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *Plant J.* **23,** 233–243.

Ouellet, F., Overvoorde, P. J., and Theologis, A. (2001). IAA17/AXR3: Biochemical insight into an auxin mutant phenotype. *Plant Cell* **13,** 829–841.

Paal, J., Henselewski, H., Muth, J., Meksem, K., Menéndez, C. M., Salamini, F., Ballvora, A., and Gebhardt, C. (2004). Molecular cloning of the potato *Gro1–4* gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. *Plant J.* **38,** 285–297.

Palmer, J. D. (1992). Cell Organelles. *In* "Plant Gene Research" (R. G. Hermann, Ed.), pp. 99–122. Springer-Verlag, Vienna.

Palmer, J. D., and Shields, C. R. (1984). Tripartite structure of the *Brassica campestris* mitochondrial genome. *Nature* **307,** 437–440.

Paterson, A. H., Bowers, J. E., Burow, M. D., Draye, X., Elsik, C. G., Jiang, C.-X., Katsar, C. S., Lan, T.-H., Lin, Y.-R., Ming, R., and Wright, R. J. (2000). Comparative genomics of plant chromosomes. *Plant Cell* **12,** 1523–1539.

Paterson, A. H., Bowers, J. E., and Chapman, B. A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. USA* **101,** 9903–9908.

Patterson, S. D., and Aebersold, R. H. (2003). Proteomics: The first decade and beyond. *Nat. Genet. Suppl.* **33,** 311–323.

Peltier, J. B., Emanuelsson, O., Kalume, D. E., Ytterberg, J., Friso, G., Rudella, A., Liberles, D. A., Soderberg, L., Roepstorff, P., von Heijne, G., and van Wijk, K. J. (2002). Central functions of the lumenal and peripheral thylakoid proteome of *Arabidopsis* determined by experimentation and genome-wide prediction. *Plant Cell* **14,** 211–236.

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D., and Harberd, N. P. (1999). "Green revolution" genes encode mutant gibberellin response modulators. *Nature* **400,** 256–261.

Peterson, D. G., Schulze, S. R., Sciara, E. B., Lee, S. A., Bowers, J. E., Nagel, A., Jiang, N., Tibbitts, D. C., Wessler, S. R., and Paterson, A. H. (2002). Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. *Genome Res.* **12,** 795–807.

Pflieger, S., Lefebvre, V., and Causse, M. (2001). The candidate gene approach in plant genetics: A review. *Mol. Breed.* **7,** 275–291.

Potokina, E., Sreenivasulu, N., Altschmied, L., Michalek, W., and Graner, A. (2002). Differential gene expression during seed germination in barley (*Hordeum vulgare* L.). *Funct. Integr. Genomics* **2,** 28–39.

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

Quackenbush, J., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Pertea, G., Sultana, R., and White, J. (2001). The TIGR Gene Indices: Analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Res.* **29,** 159–164.

Rabilloud, T. (1996). Solubilization of proteins for electrophoretic analyses. *Electrophoresis* **17,** 813–829.

Rabilloud, T. (2002). Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *Proteomics* **2,** 3–10.

Rabinowicz, P. D., Schutz, K., Dedhia, N., Yordan, C., Parnell, L. D., Stein, L., McCombie, W. R., and Martienssen, R. A. (1999). Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. *Nat. Genet.* **23,** 305–308.

Ramalingam, J., Vera Cruz, C. M., Kukreja, K., Chittoor, J. M., Wu, J.-L., Lee, S. W., Baraoidan, M., George, M. L., Cohen, M. B., Hulbert, S. H., Leach, J. E., and Leung, H. (2003). Candidate defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. *Mol. Plant Microbe Interact.* **16,** 14–24.

Rana, D., van den Boogaart, T., O' Neill, C. M., Hynes, L., Bent, E., Macpherson, L., Park, J. Y., Lim, Y. P., and Bancroft, I. (2004). Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. *Plant J.* **40,** 725–733.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovitch, S., Goodman, M. M., and Buckler, E. S., IV (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* **98,** 11479–11484.

Rensing, S. A., Fritzowsky, D., Lang, D., and Reski, R. (2005). Protein encoding genes in an ancient plant: Analysis of codon usage, retained genes and splice sites in a moss, *Physcomitrella patens*. *BMC Genomics* **6,** 43.

Rensink, W. A., and Buell, C. R. (2004). *Arabidopsis* to rice: Applying knowledge from a weed to enhance our understanding of a crop species. *Plant Physiol.* **135,** 622–629.

Reski, R., and Cove, D. J. (2004). *Physcomitrella patens*. *Curr. Biol.* **14,** R261–R262.

Rhee, S. Y., Weng, S., Flanders, D., Cherry, J. M., Dean, C., Lister, C., Anderson, M., Koornneef, M., Meinke, D. W., Nickle, T., Smith, K., and Rounsley, S. D. (1998). Genome maps 9: *Arabidopsis thaliana*. Wall chart. *Science* **282,** 663–667.

Rhee, S. Y., Beavis, W., Berardini, T. Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M., Miller, N., Mueller, L. A., Mundodi, S., Reiser, L., Tacklind, J., Weems, D. C., Wu, Y., Xu, I., Yoo, D., Yoon, J., and Zhang, P. (2003). The *Arabidopsis* Information Resource (TAIR): A model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res.* **31,** 224–228.

Rice Chromosome 10 Sequencing Consortium (2003). In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* **300,** 1566–1569.

Rice Full-Length cDNA Consortium (2003). Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. *Science* **301,** 376–379.

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

Riera-Lizarazu, O., Vales, M. I., and Phillips, R. L. (2001). A compilation of molecular genetic maps of cultivated plants. *In* "Advances in Cellular and Molecular Biology of Plants" (R. L. Phillips and I. K. Vasil, Eds.), 2nd ed., Vol. 6, pp. 463–497. Kluwer Academic Publishers, Dordrecht.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature* **405,** 847–856.

Rishi, A. S., Munir, S., Kapur, V., Nelson, N. D., and Goyal, A. (2004). Identification and analysis of safener-inducible expressed sequence tags in *Populus* using a cDNA microarray. *Planta* **220,** 296–306.

Röder, M. S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.-H., Leroy, P., and Ganal, M. W. (1998). A microsatellite map of wheat. *Genetics* **149,** 2007–2023.

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

Rose, J. K., Bashir, S., Giovannoni, J. J., Jahn, M. M., and Saravanan, R. S. (2004). Tackling the plant proteome: Practical approaches, hurdles, and experimental tools. *Plant J.* **39,** 715–733.

Ross, J., Aviv, H., Scolnick, E., and Leder, P. (1972). *In vitro* synthesis of DNA complementary to purified rabbit globin mRNA: RNA-dependent DNA polymerase-reticulocyte-hemoglobin-density gradient centrifugation-oligo(dT) primer. *Proc. Natl. Acad. Sci. USA* **69,** 264–268.

Rossberg, M., Theres, K., Acarkan, A., Herrero, R., Schmitt, T., Schumacher, K., Schmitz, G., and Schmidt, R. (2001). Comparative sequence analysis reveals extensive microcolinearity in the lateral suppressor regions of the tomato, *Arabidopsis*, and *Capsella* genomes. *Plant Cell* **13,** 979–988.

Rounsley, S. D., Glodek, A., Sutton, G., Adams, M. D., Somerville, C. R., Venter, J. C., and Kerlavage, A. R. (1996). The construction of *Arabidopsis* expressed sequence tag assemblies: A new resource to facilitate gene identification. *Plant Physiol.* **112,** 1177–1183.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G. T., Mullis, K. B., and Ehrlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487–491.

Saji, S., Umehara, Y., Antonio, B. A., Yamane, H., Tanoue, H., Baba, T., Aoki, H., Ishige, N., Wu, J., Koike, K., Matsumoto, T., and Sasaki, T. (2001). A physical map with yeast artificial chromosome (YAC) clones covering 63% of the 12 rice chromosomes. *Genome* **44,** 32–37.

Sallaud, C., Gay, C., Larmande, P., Bès, M., Piffanelli, P., Piégu, B., Droc, G., Regad, F., Bourgeois, E., Meynard, D., Périn, C., Sabau, X., Ghesquière, A., Glaszmann, J. C., Delseny, M., and Guiderdoni, E. (2004). High throughput T-DNA insertion mutagenesis in rice: A first step towards *in silico* reverse genetics. *Plant J.* **39,** 450–464.

Salse, J., Piegu, B., Cooke, R., and Delseny, M. (2004). New *in silico* insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome. *Plant J.* **38,** 396–409. [Erratum in *Plant J.* 2004; **38**, 873.]

San Miguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y., and Bennetzen, J. L. (1998). The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20,** 43–45.

Saranga, Y., Menz, M., Jiang, C.-X., Wright, R. J., Yakir, D., and Paterson, A. H. (2001). Genomic dissection of genotype × environment interactions conferring adaptation of cotton to arid conditions. *Genome Res.* **11,** 1988–1995.

Saravanan, R. S., and Rose, J. K. (2004). A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics* **4,** 2522–2532.

Sasaki, T., Song, Y., Koga-Ban, Y., Matsui, E., Fang, F., Higo, H., Nagasaki, H., Hori, M., Miya, M., Murayama-Kayano, E., Takiguchi, T., Takasuga, A., Niki, T., Ishimaru, K., Ikeda, H., Yamamoto, Y., Mukai, Y., Ohta, I., Miyadera, N., Havukkala, I., and Minobe, Y. (1994). Toward cataloguing all rice genes: Large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. *Plant J.* **6,** 615–624.

Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K., Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K., Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y., Idonuma, A., Iijima, M., Ikeda, M., Ikeno, M., Ito, S., Ito, T., Ito, Y., Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S., Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T., Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M., Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M., Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T., Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K., Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K., Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H.-I., Eun, M. Y., Yano, M., Jiang, J., and Gojobori, T. (2002). The genome sequence and structure of rice chromosome 1. *Nature* **420,** 312–316.

Sato, S., Kotani, H., Hayashi, R., Liu, Y. G., Shibata, D., and Tabata, S. (1998). A physical map of *Arabidopsis thaliana* chromosome 3 represented by two contigs of CIC YAC, P1, TAC and BAC clones. *DNA Res.* **5,** 163–168.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999a). Complete structure of the chloroplast genome of *Arabidopsis thaliana. DNA Res.* **6,** 283–290.

Sato, Y., Sentoku, N., Miura, Y., Hirochika, H., Kitano, H., and Matsuoka, M. (1999b). Loss-of-function mutations in the rice homeobox gene *OSH15* affect the architecture of internodes resulting in dwarf plants. *EMBO J.* **18,** 992–1002.

Schaefer, D. G. (2001). Gene targeting in *Physcomitrella patens*. *Curr. Opin. Plant Biol.* **4,** 143–150.

Schafer, A. J., and Hawkins, J. R. (1998). DNA variation and the future of human genetics. *Nat. Biotechnol.* **16,** 33–39.

Scheible, W. R., Morcuende, R., Czechowski, T., Fritz, C., Osuna, D., Palacios-Rojas, N., Schindelasch, D., Thimm, O., Udvardi, M. K., and Stitt, M. (2004). Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of *Arabidopsis* in response to nitrogen. *Plant Physiol.* **136,** 2483–2499.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270,** 467–470.

Schmidt, R., Putterill, J., West, J., Cnops, G., Robson, F., Coupland, G., and Dean, C. (1994). Analysis of clones carrying repeated DNA sequences in two YAC libraries of *Arabidopsis thaliana* DNA. *Plant J.* **5,** 735–744.

Schmidt, R., West, J., Love, K., Lenehan, Z., Lister, C., Thompson, H., Bouchez, D., and Dean, C. (1995). Physical map and organization of *Arabidopsis thaliana* chromosome 4. *Science* **270,** 480–483.

Schnable, P. S., Hochholdinger, F., and Nakazono, M. (2004). Global expression profiling applied to plant development. *Curr. Opin. Plant Biol.* **7,** 50–56.

Schneider, K., Weisshaar, B., Borchardt, D. C., and Salamini, F. (2001). SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. *Mol. Breed.* **8,** 63–74.

Schön, C. C., Lee, M., Melchinger, A. E., Guthrie, W. D., and Woodman, W. L. (1993). Mapping and characterization of quantitative trait loci affecting resistance against second-generation European corn borer in maize with the aid of RFLPs. *Heredity* **70,** 648–659.

Schwarzacher, T. (2003). DNA, chromosomes, and *in situ* hybridization. *Genome* **46,** 953–962.

Schwarzacher, T., Anamthawat-Jonsson, K., Harrison, G. E., Islam, A. K. M. R., Jia, J. Z., King, I. P., Leitch, A. R., Miller, T. E., Reader, S. M., Rogers, W. J., Shi, M., and Heslop-Harrison, J. S. (1992). Genomic *in situ* hybridization to identify alien chromosomes and chromosome segments in wheat. *Theor. Appl. Genet.* **84,** 778–786.

Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M., Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., and Shinozaki, K. (2002). Functional annotation of a full-length *Arabidopsis* cDNA collection. *Science* **296,** 141–145.

Service, R. F. (2001). Proteomics: High-speed biologists search for gold in proteins. *Science* **294,** 2074–2077.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B. Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., and Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J.* **5,** 2043–2049.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89,** 8794–8797.

Simillion, C., Vandepoele, K., Van Montagu, M. C. E., Zabeau, M., and Van de Peer, Y. (2002). The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **99,** 13627–13632.

Simko, I., Costanzo, S., Haynes, K. G., Christ, B. J., and Jones, R. W. (2004). Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. *Theor. Appl. Genet.* **108,** 217–224.

Soderlund, C., Longden, I., and Mott, R. (1997). FPC: A system for building contigs from restriction fingerprinted clones. *Comput. Appl. Biosci.* **13,** 523–535.

Soellick, T. R., and Uhrig, J. F. (2001). Development of an optimized interaction-mating protocol for large-scale yeast two-hybrid analyses. *Genome Biol.* **2,** RESEARCH0052.

Somerville, C., and Dangl, J. (2000). Genomics: Plant biology in 2010. *Science* **290,** 2077–2078.

Sorrells, M. E., La Rota, M., Bermudez-Kandianis, C. E., Greene, R. A., Kentety, R., Munkvold, J. D., Miftahudin, Mahmoud, A., Ma, X. F., Gustafson, P. J., Qi, L. L., Echalier, B., Gill, B. S., Matthews, D. E., Lazo, G. R., Chao, S., Anderson, O. D., Edwards, H., Linkiewicz, A. M., Dubcovsky, J., Akhunov, E. D., Dvorak, J., Zhang, D., Nguyen, H. T., Peng, J., Lapitan, N. L., Gonzalez-Hernandez, J. L., Anderson, J. A., Hossain, K., Kalavacharla, V., Kianian, S. F., Choi, D. W., Close, T. J., Dilbirligi, M., Gill, K. S., Steber, C., Walker-Simmons, M. K., McGuire, P. E., and Qualset, C. O. (2003). Comparative DNA sequence analysis of wheat and rice genomes. *Genome Res.* **13,** 1818–1827.

Sterky, F., Bhalerao, R. R., Unneberg, P., Segerman, B., Nilsson, P., Brunner, A. M., Charbonnel-Campaa, L., Lindvall, J. J., Tandre, K., Strauss, S. H., Sundberg, B., Gustafsson, P., Uhlén, M., Bhalerao, R. P., Nilsson, O., Sandberg, G., Karlsson, J., Lundeberg, J., and Jansson, S. (2004). A *Populus* EST resource for plant functional genomics. *Proc. Natl. Acad. Sci. USA* **101,** 13951–13956.

Stougaard, J. (2001). Genetics and genomics of root symbiosis. *Curr. Opin. Plant Biol.* **4,** 328–335.

Stuber, C. W., Lincoln, S. E., Wolff, D. W., Helentjaris, T., and Lander, E. S. (1992). Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* **132,** 823–839.

Suzuki, K., Hattori, A., Tanaka, S., Masumura, T., Abe, M., and Kitamura, S. (2005). High-coverage profiling analysis of genes expressed during rice seed development, using an improved amplified fragment length polymorphism technique. *Funct. Integr. Genomics* **5,** 117–127.

Takahashi, Y., Shomura, A., Sasaki, T., and Yano, M. (2001). *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the α subunit of protein kinase CK2. *Proc. Natl. Acad. Sci. USA* **98,** 7922–7927.

Taketa, S., Harrison, G., and Heslop-Harrison, J. S. (1999). Comparative physical mapping of the 5S and 18S–25S rDNA in nine wild *Hordeum* species and cytotypes. *Theor. Appl. Genet.* **98,** 1–9.

Tang, J., Xia, H., Cao, M., Zhang, X., Zeng, W., Hu, S., Tong, W., Wang, J., Wang, J., Yu, J., Yang, H., and Zhu, L. (2004). A comparison of rice chloroplast genomes. *Plant Physiol.* **135,** 412–420.

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., Miller, L., Paterson, A. H., Pineda, O., Röder, M. S., Wing, R. A., Wu, W., and Young, N. D. (1992). High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132,** 1141–1160.

Tao, Q., Chang, Y. L., Wang, J., Chen, H., Islam-Faridi, M. N., Scheuring, C., Wang, B., Stelly, D. M., and Zhang, H. B. (2001). Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. *Genetics* **158,** 1711–1724.

Tarchini, R., Biddle, P., Wineland, R., Tingey, S., and Rafalski, A. (2000). The complete sequence of 340 kb of DNA around the rice *Adh1–adh2* region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* **12,** 381–391.

Tautz, D., and Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* **12,** 4127–4138.

Tautz, D., Trick, M., and Dover, G. A. (1986). Cryptic simplicity in DNA is a major source of genetic variation. *Nature* **322,** 652–656.

Taylor, G. (2002). *Populus: Arabidopsis* for forestry. Do we need a model tree? *Ann. Bot.* **90,** 681–689.

Tenaillon, M. I., Sawkins, M. C., Long, A. D., Gaut, R. L., Doebley, J. F., and Gaut, B. S. (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl. Acad. Sci. USA* **98,** 9161–9166.

Theologis, A., Ecker, J. R., Palm, C. J., Federspiel, N. A., Kaul, S., White, O., Alonso, J., Altafi, H., Araujo, R., Bowman, C. L., Brooks, S. Y., Buehler, E., Chan, A., Chao, Q., Chen, H., Cheuk, R. F., Chin, C. W., Chung, M. K., Conn, L., Conway, A. B., Conway, A. R., Creasy, T. H., Dewar, K., Dunn, P., Etgu, P., Feldblyum, T. V., Feng, J., Fong, B., Fujii, C. Y., Gill, J. E., Goldsmith, A. D., Haas, B., Hansen, N. F., Hughes, B., Huizar, L., Hunter, J. L., Jenkins, J., Johnson-Hopson, C., Khan, S., Khaykin, E., Kim, C. J., Koo, H. L., Kremenetskaia, I., Kurtz, D. B., Kwan, A., Lam, B., Langin-Hooper, S., Lee, A., Lee, J. M., Lenz, C. A., Li, J. H., Li, Y., Lin, X., Liu, S. X., Liu, Z. A., Luros, J. S., Maiti, R., Marziali, A., Militscher, J., Miranda, M., Nguyen, M., Nierman, W. C., Osborne, B. I., Pai, G., Peterson, J., Pham, P. K., Rizzo, M., Rooney, T., Rowley, D., Sakano, H., Salzberg, S. L., Schwartz, J. R., Shinn, P., Southwick, A. M., Sun, H., Tallon, L. J., Tambunga, G., Toriumi, M. J., Town, C. D., Utterback, T., Van Aken, S., Vaysberg, M., Vysotskaia, V. S., Walker, M., Wu, D., Yu, G., Fraser, C. M., Venter, J. C., and Davis, R. W. (2000). Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature* **408,** 816–820.

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

Thomas, W. T. B., Powell, W., Swanston, J. S., Ellis, R. P., Chalmers, K. J., Barua, U. M., Jack, P., Lea, V., Forster, B. P., Waugh, R., and Smith, D. B. (1996). Quantitative trait loci for germination and malting quality characters in a spring barley cross. *Crop Sci.* **36,** 265–273.

Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D., and Buckler, E. S., IV (2001). *Dwarf8* polymorphisms associate with variation in flowering time. *Nat. Genet.* **28,** 286–289.

Tian, G. W., Mohanty, A., Chary, S. N., Li, S., Paap, B., Drakakaki, G., Kopec, C. D., Li, J., Ehrhardt, D., Jackson, D., Rhee, S. Y., Raikhel, N. V., and Citovsky, V. (2004). High-throughput fluorescent tagging of full-length *Arabidopsis* gene products in planta. *Plant Physiol.* **135,** 25–38.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y., and Martin, W. (2004). Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5,** 123–135.

Tsugita, A., and Kamo, M. (1999). 2-D electrophoresis of plant proteins. *Methods Mol. Biol.* **112,** 95–97.

Tulsieram, L. K., Glaubitz, J. C., Kiss, G., and Carlson, J. E. (1992). Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. *Biotechnology* **10,** 686–689.

Tuskan, G. A., DiFazio, S. P., and Teichmann, T. (2004). Poplar genomics is getting popular: The impact of the poplar genome project on tree research. *Plant Biol.* **6,** 2–4.

Tymowska-Lalanne, Z., and Kreis, M. (1998). The plant invertases: Physiology, biochemistry and molecular biology. *Adv. Bot. Res.* **28,** 71–117.

Uchimiya, H., Kidou, S., Shimazaki, T., Takamatsu, S., Hashimoto, H., Nishi, R., Aotsuka, S., Matsubayashi, Y., Kidou, N., Umeda, M., and Kato, A. (1992). Random sequencing of cDNA libraries reveals a versatility of expressed genes in cultured cells of rice (*Oryza sativa*). *Plant J.* **2,** 1005–1009.

Unlu, M., Morgan, M. E., and Minden, J. S. (1997). Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis* **18,** 2071–2077.

Unseld, M., Marienfeld, J. R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* **15,** 57–61.

Van den Berg, R. G., Bryan, G., del Rio, A., and Spooner, D. M. (2002). Reduction of species in the wild potato *Solanum* section *Petota* series *Longipedicellata*: AFLP, RAPD and chloroplast SSR data. *Theor. Appl. Genet.* **105,** 1109–1114.

Van Driel, R., and Fransz, P. (2004). Nuclear architecture and genome functioning in plants and animals: What can we learn from both? *Exp. Cell Res.* **296,** 86–90.

Van Eck, H. J., Rouppe van der Voort, J., Draaistra, J., van Zandvoort, P., van Enckevort, E., Segers, B., Peleman, J., Jacobsen, E., Helder, J., and Bakker, J. (1995). The inheritance and chromosomal localization of AFLP markers in a non-inbred potato offspring. *Mol. Breed.* **1,** 397–410.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* **270,** 484–487.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van der Lee, T., Fornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* **23,** 4407–4414.

Wang, R., Okamoto, M., Xing, X., and Crawford, N. M. (2003). Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol.* **132,** 556–567.

Ward, B. L., Anderson, R. S., and Bendich, A. J. (1981). The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell* **25,** 793–803.

Wardrop, J., Snape, J., Powell, W., and Machray, G. C. (2002). Constructing plant radiation hybrid panels. *Plant J.* **31,** 223–228.

Waugh, R., Dear, P. H., Powell, W., and Machray, G. C. (2002). Physical education: New technologies for mapping plant genomes. *Trends Plant Sci.* **7,** 521–523.

Weckwerth, W., Wenzel, K., and Fiehn, O. (2004). Process for the integrated extraction identification, and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* **4,** 78–83.

Welsh, J., and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18,** 7213–7218.

Wheeler, N. C., Jermstad, K. D., Krutovsky, K., Aitken, S. N., Howe, G. T., Krakowski, J., and Neale, D. B. (2005). Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. IV. Cold-hardiness QTL verification and candidate gene mapping. *Mol. Breed.* **15,** 145–156.

Whitelaw, C. A., Barbazuk, W. B., Pertea, G., Chan, A. P., Cheung, F., Lee, Y., Zheng, L., van Heeringen, S., Karamycheva, S., Bennetzen, J. L., SanMiguel, P., Lakey, N., Bedell, J., Yuan, Y., Budiman, M. A., Resnick, A., Van Aken, S., Utterback, T., Riedmuller, S., Williams, M., Feldblyum, T., Schubert, K., Beachy, R., Fraser, C. M., and Quackenbush, J. (2003). Enrichment of gene-coding sequences in maize by genome filtration. *Science* **302,** 2118–2120.

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.* **118,** 6531–6535.

Wong, G. K.-S., Wang, J., Tao, L., Tan, J., Zhang, J., Passey, D. A., and Yu, J. (2002). Compositional gradients in Gramineae genes. *Genome Res.* **12,** 851–856.

Wu, C. C., Nimmakayala, P., Santos, F. A., Springman, R., Scheuring, C., Meksem, K., Lightfoot, D. A., and Zhang, H. B. (2004a). Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. *Theor. Appl. Genet.* **109,** 1041–1050.

Wu, C. C., Sun, S., Nimmakayala, P., Santos, F. A., Meksem, K., Springman, R., Ding, K., Lightfoot, D. A., and Zhang, H. B. (2004b). A BAC- and BIBAC-based physical map of the soybean genome. *Genome Res.* **14,** 319–326.

Wu, J., Kurata, N., Tanoue, H., Shimokawa, T., Umehara, Y., Yano, M., and Sasaki, T. (1998). Physical mapping of duplicated genomic regions of two chromosome ends in rice. *Genetics* **150,** 1595–1603.

Wu, J., Maehara, T., Shimokawa, T., Yamamoto, S., Harada, C., Takazaki, Y., Ono, N., Mukai, Y., Koike, K., Yazaki, J., Fujii, F., Shomura, A., Ando, T., Kono, I., Waki, K., Yamamoto, K., Yano, M., Matsumoto, T., and Sasaki, T. (2002). A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell* **14,** 525–535.

Wu, J., Mizuno, H., Hayashi-Tsugane, M., Ito, Y., Chiden, Y., Fujisawa, M., Katagiri, S., Saji, S., Yoshiki, S., Karasawa, W., Yoshihara, R., Hayashi, A., Kobayashi, H., Ito, K., Hamada, M., Okamoto, M., Ikeno, M., Ichikawa, Y., Katayose, Y., Yano, M., Matsumoto, T., and Sasaki, T. (2003). Physical maps and recombination frequency of six rice chromosomes. *Plant J.* **36,** 720–730.

Wu, J., Yamagata, H., Hayashi-Tsugane, M., Hijishita, S., Fujisawa, M., Shibata, M., Ito, Y., Nakamura, M., Sakaguchi, M., Yoshihara, R., Kobayashi, H., Ito, K., Karasawa, W., Yamamoto, M., Saji, S., Katagiri, S., Kanamori, H., Namiki, N., Katayose, Y., Matsumoto, T., and Sasaki, T. (2004). Composition and structure of the centromeric region of rice chromosome 8. *Plant Cell* **16,** 967–976.

Xiao, J., Li, J., Yuan, L., and Tanksley, S. D. (1995). Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genetics* **140,** 745–754.

Xu, J., and Earle, E. D. (1996). High resolution physical mapping of 45S (5.8S, 18S and 25S) rDNA gene loci in the tomato genome using a combination of karyotyping and FISH of pachytene chromosomes. *Chromosoma* **104,** 545–550.

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., Karlin-Newmann, G., Liu, S. X., Lam, B., Sakano, H., Wu, T., Yu, G., Miranda, M., Quach, H. L., Tripp, M., Chang, C. H., Lee, J. M., Toriumi, M., Chan, M. M. H., Tang, C. C., Onodera, C. S., Deng, J. M., Akiyama, K., Ansari, Y., Arakawa, T., Banh, J., Banno, F., Bowser, L., Brooks, S., Carninci, P., Chao, Q., Choy, N., Enju, A., Goldsmith, A. D., Gurjal, M., Hansen, N. F., Hayashizaki, Y.,

Johnson-Hopson, C., Hsuan, V. W., Iida, K., Karnes, M., Khan, S., Koesema, E., Ishida, J., Jiang, P. X., Jones, T., Kawai, J., Kamiya, A., Meyers, C., Nakajima, M., Narusaka, M., Seki, M., Sakurai, T., Satou, M., Tamse, R., Vaysberg, M., Wallender, E. K., Wong, C., Yamamura, Y., Yuan, S., Shinozaki, K., Davis, R. W., Theologis, A., and Ecker, J. R. (2003). Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **302,** 842–846.

Yano, M., and Sasaki, T. (1997). Genetic and molecular dissection of quantitative traits in rice. *Plant Mol. Biol.* **35,** 145–153.

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y., and Sasaki, T. (2000). *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell* **12,** 2473–2483.

Yim, Y. S., Davis, G. L., Duru, N. A., Musket, T. A., Linton, E. W., Messing, J. W., McMullen, M. D., Soderlund, C. A., Polacco, M. L., Gardiner, J. M., and Coe, E. H., Jr. (2002). Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. *Plant Physiol.* **130,** 1686–1696.

Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., and Yang, H. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296,** 79–92.

Yu, J., Wang, J., Lin, W., Li, S., Li, H., Zhou, J., Ni, P., Dong, W., Hu, S., Zeng, C., Zhang, J., Zhang, Y., Li, R., Xu, Z., Li, S., Li, X., Zheng, H., Cong, L., Lin, L., Yin, J., Geng, J., Li, G., Shi, J., Liu, J., Lv, H., Li, J., Wang, J., Deng, Y., Ran, L., Shi, X., Wang, X., Wu, Q., Li, C., Ren, X., Wang, J., Wang, X., Li, D., Liu, D., Zhang, X., Ji, Z., Zhao, W., Sun, Y., Zhang, Z., Bao, J., Han, Y., Dong, L., Ji, J., Chen, P., Wu, S., Liu, J., Xiao, Y., Bu, D., Tan, J., Yang, L., Ye, C., Zhang, J., Xu, J., Zhou, Y., Yu, Y., Zhang, B., Zhuang, S., Wei, H., Liu, B., Lei, M., Yu, H., Li, Y., Xu, H., Wei, S., He, X., Fang, L., Zhang, Z., Zhang, Y., Huang, X., Su, Z., Tong, W., Li, J., Tong, Z., Li, S., Ye, J., Wang, L., Fang, L., Lei, T., Chen, C., Chen, H., Xu, Z., Li, H., Huang, H., Zhang, F., Xu, H., Li, N., Zhao, C., Li, S., Dong, L., Huang, Y., Li, L., Xi, Y., Qi, Q., Li, W., Zhang, B., Hu, W., Zhang, Y., Tian, X., Jiao, Y., Liang, X., Jin, J., Gao, L., Zheng, W., Hao, B., Liu, S., Wang, W., Yuan, L., Cao, M., McDermott, J., Samudrala, R., Wang, J., Wong, G. K.-S., and Yang, H. (2005). The genomes of *Oryza sativa*: A history of duplications. *PLoS Biol.* **3,** e38.

Yuan, Y., SanMiguel, P. J., and Bennetzen, J. L. (2003). High-Cot sequence analysis of the maize genome. *Plant J.* **34,** 249–255.

Zachgo, E. A., Wang, M. L., Dewdney, J., Bouchez, D., Camilleri, C., Belmonte, S., Huang, L., Dolan, M., and Goodman, H. M. (1996). A physical map of chromosome 2 of *Arabidopsis thaliana*. *Genome Res.* **6,** 19–25.

Zhang, H. B., and Wu, C. (2001). BAC as tools for genome sequencing. *Plant Physiol. Biochem.* **39,** 195–209.

Zhong, X. B., Bodeau, J., Fransz, P. F., Williamson, V. M., van Kammen, A., de Jong, J. H., and Zabel, P. (1999). FISH to meiotic pachytene chromosomes of tomato locates root-knot nematode resistance gene *Mi-1* and the acid phosphatase gene *Aps-1* near the junction of

euchromatin and pericentromeric heterochromatin of chromosome arms 6S and 6L, respectively. *Theor. Appl. Genet.* **98,** 365–370.

Zhou, Y., Tang, J., Walker, M. G., Zhang, X., Wang, J., Hu, S., Xu, H., Deng, Y., Dong, J., Ye, L., Lin, L., Li, J., Wang, X., Xu, H., Pan, Y., Lin, W., Tian, W., Liu, J., Wei, L., Liu, S., Yang, H., Yu, J., and Wang, J. (2003). Gene identification and expression analysis of 86,136 expressed sequence tags (EST) from the rice genome. *Genomics Proteomics Bioinformatics* **1,** 26–42.

Zhu, H., Kim, D. J., Back, J. M., Choi, H. K., Ellis, L. C., Küster, 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.

Zimmermann, I. M., Heim, M. A., Weisshaar, B., and Uhrig, J. F. (2004). Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J.* **40,** 22–34.

Zivy, M., el Madidi, S., and Thiellement, H. (1995). Distance indices in a comparison between the A, D, I and R genomes of the Triticeae tribe. *Electrophoresis* **16,** 1295–1300.

Further Reading

Chang, Y. L., Henriquez, X., Preuss, D., Copenhaver, G. P., and Zhang, H. B. (2003). A plant-transformation-competent BIBAC library from the *Arabidopsis thaliana* Landsberg ecotype for functional and comparative genomics. *Theor. Appl. Genet.* **106,** 269–276.

Seki, M., Satou, M., Sakurai, T., Akiyama, K., Iida, K., Ishida, J., Nakajima, M., Enju, A., Narusaka, M., Fujita, M., Oono, Y., Kamei, A., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2004). RIKEN *Arabidopsis* full-length (RAFL) cDNA and its applications for expression profiling under abiotic stress conditions. *J. Exp. Bot.* **55,** 213–223.