New Phytologist

Review

Tansley review

Markers and mapping revisited: finding your gene

Author for correspondence:
Neil Jones
Tel: +44 (0)1970 622230
Email: rnj@aber.ac.uk

Received: *13 March 2009*
Accepted: *10 May 2009*

Neil Jones[1], Helen Ougham[2], Howard Thomas[1] and Izolda Pašakinskienė[3,4]

[1]IBERS, Aberystwyth University, Edward Llwyd Building, Penglais Campus, Aberystwyth, Ceredigion SY23 3DA, UK; [2]IBERS, Aberystwyth University, Gogerddan Campus, Aberystwyth, Ceredigion SY23 3EB, UK; [3]Botanical Garden of Vilnius University, Kairenu 43, LT-10239 Vilnius, Lithuania; [4]Faculty of Natural Sciences, Department of Botany and Genetics, MK Čiurlionio g. 21, LT-03101 Vilnius, Lithuania

Contents

Summary

New Phytologist (2009) **183**: 935–966
doi: 10.1111/j.1469-8137.2009.02933.x

Key words: bioinformatics, genetic variation, mapping tools, marker mapping, marker types, *stay-green* case study.

This paper is an update of our earlier review (Jones *et al*., 1997, Markers and mapping: we are all geneticists now. *New Phytologist* 137: 165–177), which dealt with the genetics of mapping, in terms of recombination as the basis of the procedure, and covered some of the first generation of markers, including restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPDs), simple sequence repeats (SSRs) and quantitative trait loci (QTLs). In the intervening decade there have been numerous developments in marker science with many new systems becoming available, which are herein described: cleavage amplification polymorphism (CAP), sequence-specific amplification polymorphism (S-SAP), inter-simple sequence repeat (ISSR), sequence tagged site (STS), sequence characterized amplification region (SCAR), selective amplification of microsatellite polymorphic loci (SAMPL), single nucleotide polymorphism (SNP), expressed sequence tag (EST), sequence-related amplified polymorphism (SRAP), target region amplification polymorphism (TRAP), microarrays, diversity arrays technology (DArT), single-strand conformation polymorphism (SSCP), denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE) and methylation-sensitive PCR. In addition there has been an explosion of knowledge and databases in the area of genomics and bioinformatics. The number of flowering plant ESTs is *c*. 19 million and counting, with all the opportunity that this provides for gene-hunting, while the survey of bioinformatics and computer resources points to a rapid growth

New Phytologist

point for future activities in unravelling and applying the burst of new information on plant genomes. A case study is presented on tracking down a specific gene (*stay-green* (*SGR*), a post-transcriptional senescence regulator) using the full suite of mapping tools and comparative mapping resources. We end with a brief speculation on how genome analysis may progress into the future of this highly dynamic arena of plant science.

I. Introduction

1. Background

This paper discusses advances that have been made in the field of practical gene mapping in plants since the present authors published their guide for nonspecialists, 'Markers and mapping: we are all geneticists now' (referred to here as MM1), in connection with the 2nd *New Phytologist* Symposium over a decade ago (Jones *et al.*, 1997). During this period there has been nothing short of a revolution in understanding, in technologies and in the scale of application of molecular genetics approaches to physiological and ecological problems in the plant sciences. Substantially complete and annotated sequences of the nuclear genomes of five green plant species (*Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, *Vitis vinifera* and *Physcomitrella patens*) are publicly available and at least as many more are well on the way to completion, including those of representative legumes (*Medicago truncatula* and *Glycine max*), major crop species (*Zea mays*, *Solanum lycopersicum*, *Sorghum bicolor* and *Triticum* spp.), trees (*Eucalyptus*) and cryptogams (*Selaginella*). The sequenced gene spaces of several other species (for example *Lolium*) are also entering the databases, and the dawn of the era of the '$1000 genome', based on cheap and efficient sequencing technologies, means that these trends will continue at an ever-increasing pace. At the same time, new molecular marker systems, and improved versions of existing techniques, together with novel high-throughput sample handling and separation approaches, have greatly increased the information output and accuracy of mapping procedures. All this is against a background of increasing globalization of genetics research so that what at one time was a 'cottage industry', in which individual laboratories would create and analyse their own mapping families with home-made markers, has become an international and interdisciplinary effort comprising groups across the world that collaborate to contribute to and exploit common resources. This means that it is easier than ever to map your locus of interest and, in the best-characterized species and their syntenic relatives, to get from a map location to an isolated DNA sequence.

2. What are we trying to mark, and how?

The reasons for wishing to map a gene are as diverse as the fields of biology that have been touched by molecular genetics. For many users of mapping and related methods it may not be necessary to go as far as isolating the gene of interest – being able to track it with a reasonably close marker through a natural or breeding population will be good enough. But such are the benefits of getting the gene out and examining its allelic structure, regulatory features, physiological function and phylogeny that mapping as a means of gene isolation is becoming a desirable and feasible strategy across practically the entire range of disciplines within biology. The present account describes the tools available to the researcher who may not be a specialist in molecular genetics but for whom markers and mapping may be considered as a practical approach.

It is clearly impossible to capture the whole of this burgeoning field in a single review aimed at the nonspecialist seeking to understand enough of the topic to become an informed user of these powerful techniques, so we have been selective. In particular, we have confined ourselves to dealing mainly with principles rather than instrumentation and platform technologies, which quickly become superseded by new technical developments. The present discussion covers the issue of genetic variation and the choice of the right kind of mapping family; theoretical and practical features of the major molecular marker systems currently in use; and the bioinformatics resources available to turn the tsunami of molecular information into biological understanding. These elements are brought together in a case study in which an elusive gene was finally run to earth by application of the full range of mapping tools and resources. Finally, we try briefly to imagine what the world of genome analysis might look like in the future, when Markers and Mapping part 3 comes to be written.

II. Survey of genetic variation – sources and resources

1. Genetic variation

Mapping is possible only if there is genetic variation for the subject trait. A successful mapping study is critically dependent on the choice of contrasting parental lines from which the mapping population will be generated. Here we consider the nature of genetic variation and how it may be found or generated. Genetic variation is ultimately a matter of variation in DNA structure. Alteration to the base sequence of DNA may be in-frame or frame-shifted. It may take the form of point mutation, insertion or deletion and may have sense,

mis-sense or nonsense consequences for transcription, translation, protein function and phenotype expression. All types of DNA variation are potentially useful for the mapping of molecular markers, and with the exception of simple sense base substitutions with no phenotypic outcome, may also result in detectable variation in trait expression.

Phenotypic variation may be continuous or discontinuous. Discontinuous variation is characteristic of traits controlled by one or a small number of genes that behave in a relatively simple Mendelian fashion. A population displaying continuous variation for a given trait, in contrast, comprises a range of forms grouped uni- or multi-modally around a mean. Continuous variation is quantitative and discontinuous variation is qualitative. K. J. Mather introduced the term 'polygenic inheritance' to describe the basis of quantitative variation (Mather, 1943). Jones *et al.* (1997) referred to the difference between continuous/quantitative and discontinuous/qualitative variation in terms of their analogue and digital natures, respectively. In a sense, mapping quantitative traits is a digitizing process, allowing the contributing polygenes to be resolved into discrete loci.

Useful variation for genetic mapping may be generated experimentally. To target and map a particular trait, crossing contrasting lines will result in populations segregating for the character, either qualitatively or quantitatively, in first or subsequent generations depending on dominance relations. With the trend towards the development of generic genomics resources aimed ultimately at providing the means of characterizing any and every gene in a given species, populations segregating for the greatest possible range of traits are being actively identified, curated and documented. The source of variation, its scale and how it is maintained and annotated in an exploitable form depend on the biology and experimental properties of the species. Examples of mapping resources that have been developed in fully sequenced model plants follow.

2. *Arabidopsis thaliana*

The complete sequence of the *A. thaliana* nuclear genome was announced in 2000 (Arabidopsis Genome Initiative, 2000). Accessions of *A. thaliana* collected from across its ecological range represent a rich source of phenotypic and genetic variation from which many trait-targeted and generic mapping populations have been developed (Koornneef *et al.*, 2004). Amongst the most commonly used parental lines are Columbia (Col), Landsberg erecta (Ler) and Wassilewskija (Ws). Genetic variation among these accessions includes not just qualitative allelic differences but partial or complete gene deletions. For instance, Ler probably lacks more than 100 genes or gene-like sequences found in Col, including some encoding enzymes, mostly of secondary metabolism (Borevitz *et al.*, 2003). An example of mapping families produced by crossing Col, Ler and Ws parents and analysing segregating F_2 populations is the study by Gómez-Gómez *et al.* (1999) on the genetics of the elicitation of pathogen defence-related responses by bacterial flagellin. With the exception of Ws, all ecotypes studied (including Col and Ler) were shown to be sensitive to an elicitor flagellin peptide. In Ws × Col and Ws × Ler crosses, all F_1 progeny were sensitive and sensitive:insensitive segregated 3 : 1 in F_2, indicating that sensitivity is regulated by a single dominant Mendelian locus. Mapping was carried out in the F_2 populations from Ws × Col (74 plants) and Ws × Ler (117 plants), using cleavage amplification polymorphism (CAP) and microsatellite markers (see Section III 3). The *flagellin sensing 1* (*FLS-1*) locus was mapped to chromosome 5, within 2.9 cM of a marker sequence encoding the seed storage protein CRA1.

In a monocarpic annual species such as *A. thaliana*, F_2 populations are of limited value as mapping families because they are ephemeral, giving rise to F_3 progeny that do not breed true. 'Immortalizing' the mapping population requires development of a heterogeneous family comprising true-breeding homozygous individuals. This may be achieved by selfing/single-seed descent, or by sibling mating, from F_1 or F_2 individuals (Fig. 1). In principle six or more generations should ensure substantial homozygosity (Burr & Burr, 1991). The short generation time and self-pollinating habit of *A. thaliana* are helpful for quick and efficient development of such recombinant inbred lines (RILs). The occurrence of multiple meiotic events during repeated selfing promotes high rates of recombination (Jansen, 2003). Each individual in the RIL population will have a unique combination of linkage blocks derived from the original parental lines. This approach was taken by Lister & Dean (1993), who created an *A. thaliana* RIL resource by crossing Col and Ler and taking 300 F_2 individuals through to F_8. The genotype of a particular RIL will normally be a mosaic of genomic regions derived from the two parental lines. If an RIL is repeatedly backcrossed with the background parent, lines with single introgressions will be the eventual result. Such introgression lines (ILs) are sometimes referred to as near-isogenic lines (NILs) or backcross inbred lines (BILs). Keurentjes *et al.* (2007) developed an *A. thaliana* NIL population and compared its usefulness for mapping with that of an RIL family. Ler was used as one parental line and Cvi, an accession from Cape Verde Islands, as the other. A set of 161 RILs was taken through to F_{13} and, based on an assessment with PCR markers, shown to be essentially 100% homozygous. Appropriate Ler/Cvi RILs were repeatedly backcrossed to Ler as recurrent female parent to produce a population of 92 NILs, of which 40 contained a single introgression, 32 carried two, 19 had three and one carried four. Comparative mapping analyses for a number of traits, including flowering time, stature and branching pattern, did not reveal much difference between the two populations in total numbers of quantitative trait loci (QTLs) detected, but different loci were identified, depending on the nature and genetic architecture of the corresponding character. It is generally true that a given mapping family may be more suitable for mapping some traits than others. Recently Balasubramanian *et al.* (2009) described

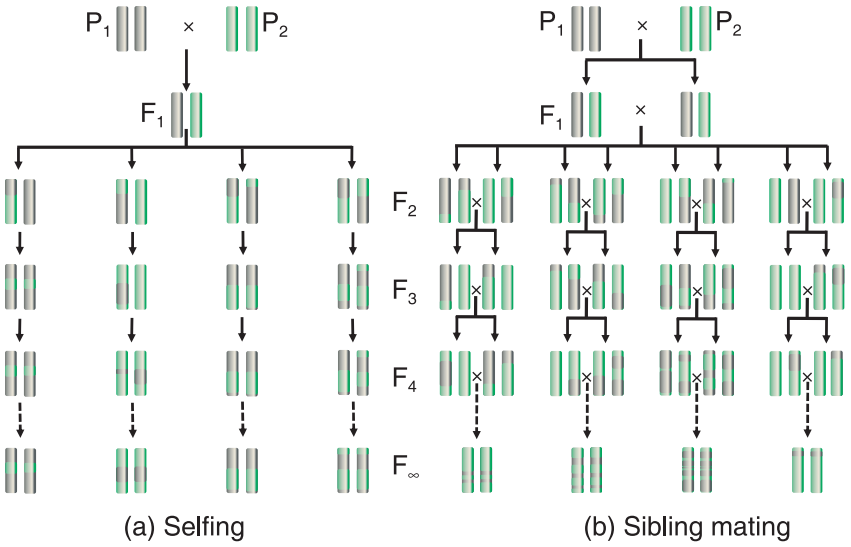


Fig. 1 The production of recombinant inbred lines by selfing (a) and by sibling mating (b). From Broman (2005).

Table 1 Selected characteristics of the three rice (*Oryza sativa*) subspecies (from Barker *et al*., 1985)

	Subspecies		
Characteristics	*indica**	*japonica*	*javanica*
Tillering	High	Low	Low
Height	Tall	Medium	Tall
Lodging	Easily	Not easily	Not easily
Photoperiod	Sensitive	Nonsensitive	Nonsensitive
Cool temperature	Sensitive	Tolerant	Tolerant
Shattering	Easily	Not easily	Not easily
Grain type	Long to medium	Short and round	Large and bold
Rice texture	Nonsticky	Sticky	Intermediate

*Traditional type before development of modern semidwarf varieties.

a further refinement of the RIL approach to developing *A. thaliana* mapping families. These authors crossed accessions Est-1 and Kend-L with Col as female and, from the F_2 progeny, intercrossed 75 nonoverlapping pairs for three generations to create advanced intercross (AI) lines. The resulting AI-RILs were genotyped after six rounds of selfing without any intentional selection. The advantage of the AI approach is that the opportunity for recombination is increased before genotypes are fixed upon selfing.

3. Rice

Detailed linkage maps have been produced for all of the world's staple cereal species and the practical needs of crop improvement and food security continue to drive the commercial and academic development of genetic resources (Lörz & Wenzel, 2005). Rice (*O. sativa*) is unique in being simultaneously the principal food for almost half of humanity and a model species for monocot genetics. Drafts of the entire genomic sequences of *indica* and *japonica* rice were announced in 2002 (Goff *et al*., 2002; Yu *et al*., 2002). The first molecular linkage maps for rice were published in 1988 (McCouch *et al*., 1988) and used restriction fragment length polymorphism (RFLP) markers (see Section III 2) applied to 50 individuals in an F_2 population derived from an *indica* × *javanica* cross. Trait and molecular variation among interfertile *indica*, *japonica* and *javanica* subspecies (Table 1) has been the basis of all the major mapping family resources developed subsequently. The use of ephemeral mapping families based on segregating F_2 populations has been largely supplanted by the development of RIL and doubled-haploid (DH) resources. A comparative study of F_2, DH and RIL populations derived from initial *indica* × *japonica* crosses was carried out by Xu *et al*. (1997). Two DH families (comprising 133 and 135 individuals, respectively) were produced by F_1 anther culture followed by regeneration of haploid plants and either spontaneous or colchicine-mediated doubling of chromosome number. This research highlighted the need to be aware of artefacts that may be associated with different ways of creating mapping families. In particular, skewed segregation patterns were observed in

the DH populations, probably associated with preferential regeneration from anther culture of *japonica*-derived genotypes. Since the first mapping populations were established, the densest linkage maps of any crop plant have been developed in rice and some of the most impressive examples of positional cloning based on QTLs have been accomplished in this species (see, for example, Heuer *et al*., 2009).

4. *Populus trichocarpa*

The genome sequence of *Populus trichocarpa* was announced in 2006 (Tuskan *et al*., 2006). Mapping in a woody perennial such as *P. trichocarpa* presents very different challenges and opportunities compared with herbaceous annuals such as rice and *A. thaliana*. Trees take many years to reveal their full phenotypic expression and large areas of land are required to maintain experimental populations. However, segregating F_2 families of perennials can be readily immortalized by vegetative propagation. Bradshaw *et al*. (1994) and Bradshaw & Stettler (1995) carried out QTL mapping in an F_2 population derived from an interspecific cross between *P. trichocarpa* and *Populus deltoides*. Two F_1 siblings were crossed to produce the F_2 mapping family. Genetic analysis of trees requires patience: it is significant that the parental cross was made in 1981, the F_1 cross in 1988 and the first maps published 6 yr later. However, this family has been maintained to the present day and continues to be a resource for analysis of traits unique to woody trees, for example bioenergy characters (Rae *et al*., 2008).

Populus is a dioecious, wind-pollinated, obligate outbreeder and, as a consequence, natural populations have high levels of phenotypic variation and heterozygosity. Across the ecotypic range, particular alleles at particular loci conferring fitness will tend to accumulate, along with molecular markers close enough to them to escape recombination and randomization within the population (a phenomenon referred to as linkage disequilibrium (LD), discussed further in Section IV 3). Such populations can be exploited for LD (sometimes called association) mapping (Ingvarsson, 2005). In this connection, the Swedish Aspen (SwAsp) Collection has been established as a resource for the analysis of natural phenotypic and genetic variation in *Populus tremula* (Luquez *et al*., 2008). One hundred and sixteen individual trees were collected from 12 locations in Sweden covering the latitude range 56.3° to 66.2°N. Randomized blocks of replicated clones were planted in two common gardens, located at Simlång (56.7°N) and Umeå (63.9°N), sites experiencing, respectively, 220 and 160 d per annum with a mean temperature over 5°C. Luquez *et al*. (2008) comment that 'association mapping in a natural tree population will likely require mapping populations with sizes greater than the 116 clones that make up the SwAsp collection', but the availability of this resource will promote the development of techniques for mapping and trait evaluation applicable to perennial species that would otherwise be difficult subjects for genetic analysis.

5. *Physcomitrella patens*

The genome of the bryophyte *Physcomitrella patens* is the first genome of a nonangiosperm land plant to have been sequenced and annotated (Rensing *et al*., 2008). Amongst a number of features that distinguish mosses from the flowering plants is the haploid status of the long-lived gametophytic vegetative phase during alternation of generations. This has important implications for genetic mapping. In particular, it means that segregating populations can be used as early as the F_1 generation. The first *P. patens* linkage map has been described by Kamisugi *et al*. (2008). The parental strains were the widely used accession Gransden2004, originally isolated as a single spore in Cambridge, UK, and the French genotype Villersexel (used as the male parent). The Gransden strain was marked with either an antibiotic resistance transgene or an auxotrophic mutation, allowing the products of successful crossing to be screened. Two F_1 populations, consisting of 188 and 94 individuals, respectively, were maintained by vegetative propagation and used to generate linkage maps. The integrated map from the two populations consisted of 31 linkage groups comprising 1420 markers. The chromosome number is thought to be around $n = x = 27$ (Reski *et al*., 1994; Rensing *et al*., 2007).

III. Survey of molecular marker types

1. Introductory remarks

Much has happened since MM1. New generations of markers have been developed which reflect advances in technical procedures and which enable more efficient and effective systems to be used in wide ranges of plant types and situations. Virtually all molecular markers (DNA markers) reveal neutral sites of variation at the DNA sequence level. 'Neutral' in this context means that, unlike morphological markers, these variations do not show themselves in the phenotype, and each might be nothing more than a single nucleotide difference in a gene or a piece of repetitive DNA. They have the big advantage that they are much more numerous than morphological markers, and they do not disturb the physiology of the organism. Restriction enzymes, electrophoretic separation of DNA fragments, Southern hybridization, the polymerase chain reaction (PCR), and labelled probes are the tools that allow us to access and to use these markers; although it should be noted that the best possible marker is the allele of the gene of interest, so that for example the use of the *stay-green* gene sequence (Section VI) as a marker in its own right means that one allelic form is phenotypically distinct from the other.

In this section we present a survey of marker types (summarized in Table 2), including some of those dealt with in MM1, for the convenience of the reader, and explain the underlying science as well as some of the potential uses of mapping and finding genes in plants.

Table 2 Classification of marker systems

Marker system	Advantages	Disadvantages
First-generation markers based on restriction fragment detection		
Restriction fragment length polymorphism (RFLP)	Co-dominant; highly reproducible.	Low multiplex ratio*; high on time/labour.
Second-generation markers based on PCR		
Cleavage amplification polymorphism (CAP)	Insensitive to DNA methylation; no requirement for radioactivity.	Produces informative PCR products.
Random amplified polymorphic DNA (RAPD)	Low on time/labour; medium multiplex ratio*.	Dominant; low reproducibility.
Amplified fragment length polymorphism (AFLP)	High reproducibility; high muliplex ratio*.	Dominant; moderate time/labour.
Sequence-specific amplification polymorphism (S-SAP)	Applicable for targeting any gene, transposon or sequence of interest.	Sequence must be known to enable design of element-specific PCR primers.
Simple sequence repeat (microsatellite) (SSR)	Co-dominant; highly reproducible; low on time and labour.	High cost of development; low multiplex ratio*.
Inter-simple sequence repeat (ISSR)	Technically simple; no prior genomic information needed to reveal both inter- and intraspecific variation.	Dominant markers; band staining can be weak.
Variable number tandem repeat (minisatellite) (VNTR)	Numerous multiallelic loci.	Low-resolution fingerprints in plants.
Sequence tagged sites (STS)	Co-dominant; useful for mapping.	Reproducibility; based on some degree of sequence knowledge.
Sequence characterised amplification region (SCAR)	May be dominant or co-dominant; better reproducibility than RAPDs.	More difficult to reproduce than RAPDs.
Sequence amplification of microsatellite polymorphic loci (SAMPL)	High multiplexing*; co-dominant markers; extensive polymorphism.	Some blurred banding; stutter bands.
Third-generation markers based on DNA sequencing		
Single nucleotide polymorphism (SNP)	Common; evenly distributed; detection easily automated; high throughput; low assay cost; useful for association studies; potentially high multiplex ratio*.	Usually only two alleles present.
Genome scanning for expressed genes		
Expressed sequence tag (EST)	Easy to collect and sequence; reveals novel transcripts; good representation of transcripts.	Error-prone; isolation of mRNA may be difficult.
Sequence-related amplified polymorphism (SRAP)	Simplicity; high throughput; numerous co-dominant markers; high reproducibility; targets coding sequences; detects multiple loci without previous knowledge of sequence information; PCR products directly sequenced.	Detects co-dominant and dominant markers, which can lead to complexity; null alleles detected directly.
Target recognition amplification protocol (TRAP)	Simple to use; highly informative; produces numerous markers by using existing public EST databases; uses markers targeted to a specific gene.	Requires cDNA or EST sequence information for primer development.
Markers using array technology		
Microarrays (arrangements of small spots of DNA fixed to glass slides)	Whole-genome scanning; high-throughput technology; genotype–phenotype relationship; expression analysis of large numbers of genes.	Expensive; needs gene sequence data; technically demanding.
Diversity array technology (DArT)	No sequence data required; high throughput; detects single base changes and indels; rapid germplasm characterization.	Dominant markers; technically demanding.
Other marker systems		
Single-strand conformational polymorphism (SSCP)	Detects DNA polymorphisms and mutations at multiple sites in DNA fragments.	Temperature-dependent; sensitivity affected by pH.
Denaturing gradient gel electrophoresis (DGGE)	Separates individual sequences from a complex mixture of microbes based on sequence differences.	PCR fragment size limited to about 500 bp; difficult to resolve fragments that differ by only one or two bases.
Temperature gradient gel electrophoresis (TGGE)	Almost identical to DGGE; more reliable; uses temperature gradient.	Technically demanding; little used in plants.
Methylation-sensitive PCR	Detects sites of methylated DNA.	

*The multiplex ratio is the number of independent loci detected in the assay.

2. First-generation markers based on restriction fragment detection

Restriction enzymes cut DNA at restriction sites. Each different restriction enzyme recognizes a specific and characteristic nucleotide sequence. A list of all the known restriction enzymes, together with their recognition sequences, methylation sensitivity and other useful information, is given in the REBASE database (Roberts *et al.*, 2003), available at http://rebase.neb.com/rebase/rebase.html. Because even a single nucleotide alteration can create or destroy a restriction site, mutations cause variation in the number of sites. Thus there is variation, or polymorphism, between individuals in the positions of cutting sites and the lengths of DNA between them, resulting in restriction fragments of different sizes. RFLPs can derive from the nuclear, chloroplast, and mitochondrial genomes. The main advantages of RFLP markers are their co-dominance and high reproducibility. Drawbacks as compared with PCR-based techniques are the tedious experimental procedures, and the requirement for microgram amounts of pure DNA.

Two different-sized fragments are alleles of one locus. The locus itself is identified by the probe used to detect it, and takes the name or number of that probe. The RFLP is a marker, and it can be used in genetic analysis like any other marker that has alleles identifying a locus; although we note also that the RFLP is co-dominant as we can distinguish all three morphs. This makes the RFLP more informative than the morphological marker with full dominance, where we can only identify two phenotypes: (*AA* or *Aa*) and *aa*. RFLPs arise as mutations that alter restriction sites, but the events giving rise to them, over evolutionary time, are as stable as the mutations giving any other form of allelic variation; that is, they are constant for all practical purposes. It follows that we might find large numbers of such markers, depending only on the level of polymorphism in a population and the availability of probes. In the numbers game this puts us orders of magnitude ahead of classical markers (such as isoenzymes and morphological features) in our capacity to detect selectively neutral allelic variation, and therefore far ahead also in the resolving power of our genetics. The method has now been largely superseded by protocols based on PCR or sequencing. PCR-RFLP, also called CAP, is considered below.

3. Second-generation markers based on PCR

Cleavage amplification polymorphism (CAP) CAP (also called PCR-RFLP) markers are generated by digesting PCR products with a restriction enzyme in a process involving two steps. The DNA sequence is first amplified using a sequence-specific primer pair, and this process itself can lead to a number of different-sized and thus informative PCR fragments. The inclusion of an *in vitro* amplification step makes CAP analysis insensitive to DNA methylation. The PCR product is then digested with a 4-cutter restriction enzyme. In contrast to conventional RFLPs, CAPs have no requirement for radioactivity or blotting. They are co-dominant, and can distinguish between homozygous and heterozygous individuals. An example of useful application of CAPs derived from rDNA internal transcribed spacer (ITS) has been demonstrated in the confirmation of the hybridity of interspecific hybrids of *Allium* spp., *Allium fistulosum* × *Allium schoenoprasum* (Umehara *et al.*, 2006).

Random amplified polymorphic DNA (RAPD) Williams *et al.* (1990) discovered that a single PCR primer of *c.* 10 arbitrary nucleotides in length will find homologous sequences in DNA, by chance, and will amplify several different regions of a genome. The primer amplifies a piece of DNA of between 200 and 2000 kb long, which lies between two inverted copies of itself, one copy binding to each strand of the DNA. Statistically, priming occurs once in every million base pairs. During the PCR reaction, a set of fragments of differing sizes will be generated, and because the fragments have been amplified there is enough DNA to be visualized by staining with ethidium bromide. In general, for the average-sized genome, between five and 10 fragments will be amplified to produce discrete DNA-banding patterns. Polymorphisms arise because sequence variation in the genome alters the primer binding sites. RAPDs are therefore dominant markers as a consequence of their presence/absence at particular loci, and they will segregate from a heterozygous diploid as Mendelian alleles. RAPDs are much simpler and less expensive to work with than RFLPs because no prior knowledge of sequences is required and there is no need for radioactive probes. Many different primers can be designed, and there is virtually no limit to the numbers of RAPDs in a genome. RAPDs can be used for mapping, but because of the random nature of their generation, and short primer length, they cannot easily be transferred between species. They are most often used as species-specific markers for diversity and phylogenetic studies; for example, genome relationships in Triticeae (Wei *et al.*, 1995). Their main disadvantages are poor reliability and reproducibility, and their sensitivity to experimental conditions (Karp *et al.*, 1996).

Amplified fragment length polymorphism (AFLP) The AFLP method combines the use of restriction enzymes with PCR amplification of fragments, and detects fragment length polymorphisms (Vos *et al.*, 1995; Meudt & Clarke, 2007; Fig. 2). The first step in the generation of AFLPs is to double-digest genomic DNA with two restriction enzymes. A rare cutter such a *Pst*I cuts in nonmethylated DNA and is used to create a bias towards low-copy fragments, and a frequent cutter such as *Mse*I then produces the smaller fragments with an average length of *c.* 256 bp. The use of frequent cutter enzymes only would generate too many fragments for gel electrophoresis. Next, a specific short DNA sequence is linked to one end of the fragment, and a different sequence added to

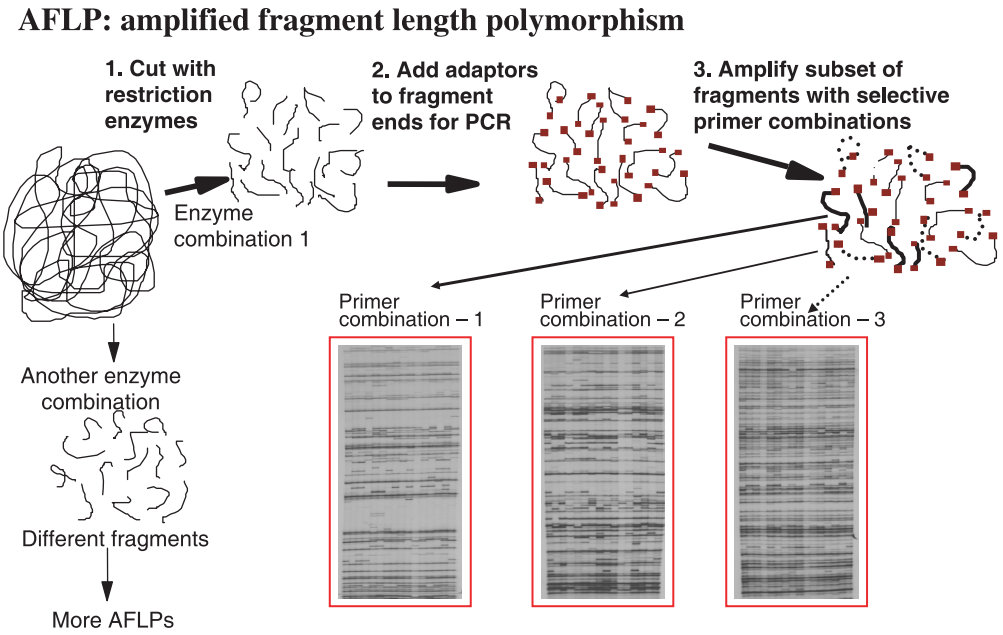


Fig. 2 Diagram showing the way in which restriction enzymes and adaptors for PCR are used to generate subsets of fragments with selective primer combinations, and the resulting amplified fragment length polymorphism (AFLP) gels. Different restriction enzymes will result in a different set of fragments. Courtesy of Angela Karp.

the other. These sequences, called adaptors, together with the adjacent restriction sites, serve as binding sites for PCR primers. The primers are designed to match the two different added sequences, and they also carry short extensions of 1–3 nucleotides to bring about selective amplification of those fragments with complementary 1–3 nucleotide sequence. Three kinds of fragment result: Type I are fragments with rare cutter ends only, and these are negligible; Type II have one rare cutter and one frequent cutter end; and Type III have two frequent cutter ends. Only Type II fragments are used in the PCR amplification. The current method of choice for selecting a suitable subset of Type II fragments uses fluorescent tagging and capillary electrophoresis, as described by Meudt & Clarke (2007). The AFLP system is technically difficult and expensive to set up, but it detects a large number of loci, reveals a great deal of polymorphism and produces high-complexity DNA fingerprints (Fig. 2) which can be used for identification and for high-resolution mapping and marker-assisted cloning.

Sequence-specific amplification polymorphism (S-SAP) S-SAP is a modified form of AFLP that uses only one restriction site-specific AFLP primer in the final amplification step, and a second primer complementary to a defined DNA sequence, which may be any gene, retrotransposon or sequence of interest. By using three selected bases at the 3′ end of the AFLP primer, banding complexity resulting from the abundance of BARE-1 retrotransposons in barley (*Hordeum vulgare*) was reduced to a useful level (Waugh *et al.*, 1997). High levels of polymorphism were revealed, and all fragments behaved as dominant markers and were evenly distributed across the barley genome. Several other kinds of mapping based on transposons have also been described, including amplification of insertion mutagenized sites (AIMS; Frey *et al.*, 1998), Mu AFLP (Edwards *et al.*, 2002), MITE-AFLP (Park *et al.*, 2003), inter-retrotransposon polymorphism and retrotransposon-microsatellite polymorphism (IRAP and REMAP, respectively; Kalendar *et al.*, 1999).

Simple sequence repeat (SSR or microsatellite) Plant genomes contain large numbers of simple sequence repeats, or microsatellites, of < 6 bp which are tandemly repeated and widely scattered at many hundreds of loci throughout the chromosome complement. Typically they may be dinucleotides (AC)n, (AG)n, (AT)n; trinucleotides (TCT)n, (TTG)n; tetranucleotides (TATG)n and so on, where n is the number of repeating units within the microsatellite locus. In addition to occurring at many different loci, they can also be polyallelic. (AT)n dinucleotides are the most abundant type of SSR in plants (Ma *et al.*, 1966). The methodology used to isolate an SSR at a particular locus starts with the construction of a small-insert genomic library. The library is then screened with a number of microsatellite probes to identify inserts carrying SSRs. The inserts are then sequenced and primers are chosen that match unique flanking sequences for particular loci (Fig. 3).

PCR amplification is used to generate DNA banding patterns on a gel and to reveal polymorphism based on different numbers of repeats at the two alleles of a locus. The marker thus has the advantage of being co-dominant. In addition, SSRs are simple, PCR-based and extremely polymorphic, and highly informative as a result of the number and frequency of

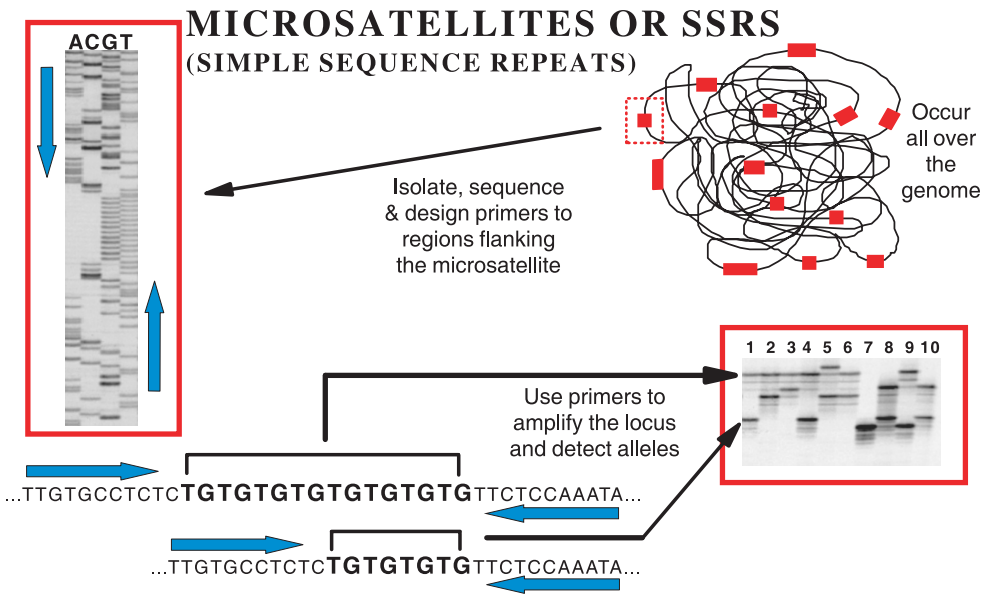


Fig. 3 Diagram to illustrate how primers are designed and used to generate simple sequence repeats (SSRs). Courtesy of Angela Karp.

alleles detected and their ability to distinguish between closely related individuals. They find application as markers for mapping, cultivar identification, protecting germplasm, determination of hybridity and analysis of gene pool variation and as diagnostic markers for traits of economic value (Powell *et al.*, 1996; King *et al.*, 2008). Microsatellites are, however, expensive to establish, have a long development time and need specific primers.

Inter-simple sequence repeat (ISSR) ISSRs are regions between microsatellite loci (Zietkiewicz *et al.*, 1994). Sequences complementary to those of two microsatellites are used as PCR primers. Provided that the variable region between the microsatellites is of an appropriate length for PCR to take place, then this variable region will be amplified. Limited PCR amplification cycles during PCR avoids excessive replication of long contiguous DNA sequences, and the result is a mix of a variety of amplified DNA strands which vary in length but which are generally quite short. About 10–60 fragments from multiple loci are generated simultaneously. ISSR markers can reveal genetic pholymorphisms and identify individual genotypes, and this method has been widely used (Reddy *et al.*, 2002; Ai-hua & Jan-bo, 2006; Meloni *et al.*, 2006). For example, SSRs are widely dispersed in all perennial ryegrass linkage groups, both in intergenic regions and in gene loci; and the identification and characterization of these sequences were useful for the enrichment of the *Lolium perenne* genetic map, and for genotyping cultivars and breeding germplasm (Pivorienė & Pašakinskienė, 2007; Pivorienė *et al.*, 2008).

Variable number tandem repeat (VNTR or minisatellite) RFLPs are probe-based markers, as we need a labelled probe in order to detect the polymorphism. The other well-known probe-based marker is the variable number tandem repeat, or minisatellite, first discovered in humans by Alec Jeffreys in the 1980s, and used extensively and sensationally in forensic DNA profiling (Jeffreys *et al.*, 1985). Minisatellites consist of tandem arrays of short repeated sequences highly dispersed throughout the genome at numerous loci. They are embedded in unique flanking sequences and the loci are hypervariable in terms of their number of repeat units. Not only are there many different loci, but multi-allelic forms of single loci exist as well at the population level as a result of unequal crossing over and replication slippage. A consensus sequence for different loci means that a 'polycore' probe can be constructed which can detect up to 30 loci simultaneously, to give a detailed 'DNA fingerprint'. VNTRs are particularly useful in vertebrates, but probes also exist that can be used to produce low-resolution fingerprints in plants, and these find their application in cultivar identification.

Sequence tagged site (STS) An STS is any site in a genome that is identified by a known and unique DNA sequence of *c.* 200–500 bp in length. It is characterized by the sequence of the primer that borders the fragment concerned, and is used to tag the larger region of the genome from which it originates, so that region can be identified wherever it occurs. STSs represent physical landmarks in the genome, and can be used with different kinds of maps and allow them to be aligned with one another. STSs derived from cDNA can be used to tag sequences that are expressed, namely expressed sequence tags (ESTs; see Section III 5). An STS is made from randomly cloned genomic DNA by a shotgun approach, and sequences of the clones generated are then used to design PCR primers which will amplify them anonymously. STSs that overlap form the basis of mapping these sequences.

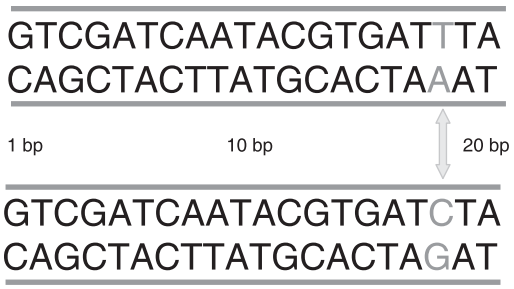


Fig. 4 Diagram showing a pair of homologous chromosomes each with a single chromatid, to illustrate the molecular basis of a single nucleotide polymorphism (SNP).

Sequence characterized amplification region (SCAR) SCARs are markers generated from cloned and sequenced RAPD fragments. They are produced by PCR with long (24-mer) oligonucleotide primers complementary to the ends of the original RAPD fragments. SCARs may have the dominant segregation behaviour of the original the RAPDs, or may be converted into co-dominant markers. They have been used by Liz *et al*. (2007) to map the *M/m* sex determining gene in cucumber (*Cucumis sativus*). Generating SCAR markers is not as convenient as working with RAPDs, but they give better reproducibility as a result of reduced competition between primer binding sites. In principle, creating locus-specific markers from anonymous PCR fragments is applicable to other multilocus marker systems such as AFLPs.

Selective amplification of microsatellite polymorphic loci (SAMPL) The SAMPL technique is similar to that of S-SAP described above. A single AFLP primer anneals with one of the sites at the ends of the restriction fragments in the second selective amplification step, and is then combined with a microsatellite-specific primer which targets a compound binucleotide repeat at some site internal to the restriction fragment. Compound microsatellites are comprised of two or more different simple sequence motifs directly adjacent to each other, and are common elements in plant genomes. The combination of a number of different restriction enzymes with the variability of the compound microsatellite motifs makes for a virtually unlimited level of detectable polymorphism. SAMPL brings together the desirable features of AFLPs and microsatellites in allowing for a high degree of multiplexing, and yields co-dominant markers with extensive polymorphism. An example of the application of SAMPL is the study of genetic diversity in bread wheat (*Triticum aestivum*) by Roy *et al*. (2002).

4. Third-generation markers based on DNA sequencing

Single nucleotide polymorphism (SNP) There are numerous sites within a genome where a short stretch of DNA in a pair of homologous chromosomes differs by a single nucleotide (Fig. 4). An SNP must be present in at least 1% of individuals in a population to qualify as polymorphic, that is, having at least two alternative allelic forms at the locus concerned. SNPs arise as mutations involving base pair substitutions, from errors in DNA replication or from mutagenic agents, and are formally distinguished from base pair insertions or deletions (indels). They were discovered during sequencing of the human genome, when regions of the genome from different individuals were aligned and found to show some sequence variation.

SNPs are the commonest variants in the genomes of eukaryotes. Mutations are rare events in eukaryotes, and nearly all SNPs are biallelic (i.e. with just two alleles); and as they have arisen by mutation over long periods of evolutionary time they represent stable sites in the genome. It follows, therefore, that their pattern of distribution will be different from one individual to another. Two individuals that are closely related, however, are more likely to share identical SNPs than those that are not, because they will have a common ancestor. In terms of their location in the genome, SNPs may be found in coding or regulatory sequences of a gene, or in noncoding regions. They comprise *c*. 90% of the genetic variation in any organism: for example, 1 SNP per 31 bp in the noncoding regions and 1 per 124 bp in the coding regions of maize (*Z. mays*), and 1 per 540 bp in the coding regions of wheat (Tyrka *et al*., 2004). When they occur in the coding sequences they may or may not affect the phenotype, and in some cases may be the causative agent of the mutation, depending on the base pair change concerned (same-sense, mis-sense or nonsense mutation). In any event, in such a situation the SNP will show 100% association with the gene or trait concerned, and will have obvious utility and economic value in 'finding your gene', or in association mapping and marker-assisted selection (MAS). It follows from what is written above that SNPs can have dense representation in genomic maps, and there is a high probability therefore that many of them will have very close linkage (be in LD – see Section IV 3) with genes near or adjacent to them, or within which they are actually located.

SNPs are often described as the 'new generation of molecular markers', as they are of recent discovery and are coming into use more than a decade after other established markers such as RFLPs, SSRs and AFLPs. The reason for this late arrival is the need to have sequence information in order to deal with them, and this has only become available in sufficient quantities outside of model plant species (*A. thaliana*, rice and maize) relatively recently.

There are numerous ways of detecting and genotyping SNPs (Lörz & Wenzel, 2005; comprehensively reviewed by Gupta *et al*., 2008), depending on the needs of particular situations, and what resources and time constraints are available. Some examples of well-known methods are given below. *Genotyping* means to characterize individuals for their SNP profile, whereas *detecting* SNPs means to find them in the first place. Genotyping to characterize individuals may be carried

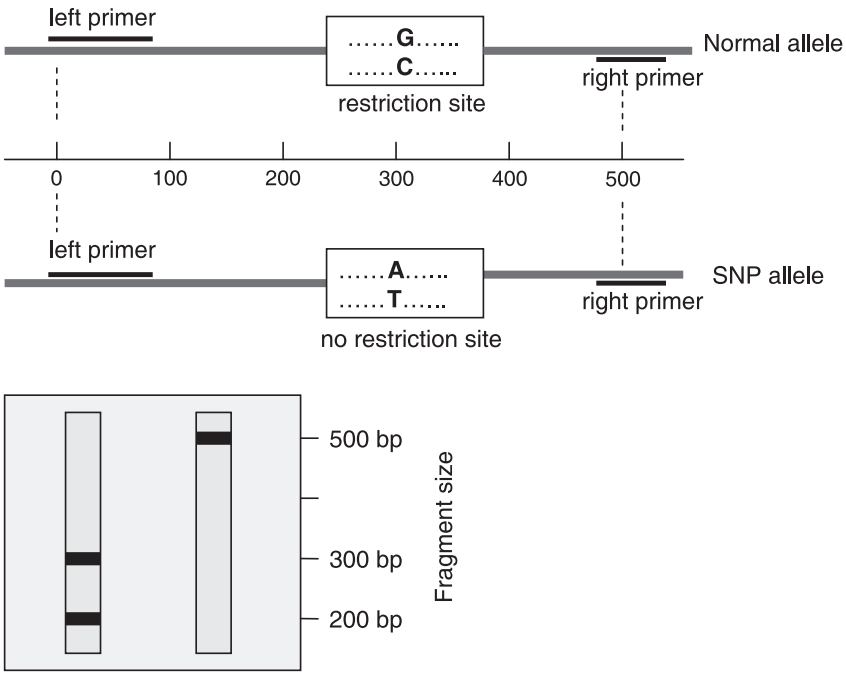


Fig. 5 Example of a genotype heterozygous for a single nucleotide polymorphism (SNP) which falls within a restriction site and deletes the site. Polymerase chain reaction (PCR) with primers spanning the locus reveals two fragment sizes: the long fragment represents the allele containing the SNP, and the two shorter fragments come from the normal allele. PCR with the same primers on a sample of individuals will associate the SNP with the locus, and thus provide a test for the presence of the allele concerned.

out on a large scale using high-throughput methods based on multiplexing (Gupta *et al.*, 2008; Table 2).

SNP detection by alignment of sequences The simplest way to detect SNPs is by the alignment of sequences from a particular region of the genome, from several individuals, and then to compare the region concerned *in silico* (i.e. from computer-read sequences) for any differences in both the coding and noncoding regions. In some cases there may be several SNPs within a short segment of the genome among which there is no, or virtually no, recombination; such a region constitutes an SNP haplotype.

Locus-specific PCR amplification Primers are used to amplify the DNA from a particular locus in several plants, followed by sequencing to search for SNPs. Any new SNP that is found will serve as a marker for the locus concerned. When new primers are made that include the sequence variation resulting from the SNP these primers can then be used to genotype plants, and act as a marker to find the gene for the SNP concerned. The system can be multiplexed to handle several loci simultaneously.

PCR for SNPs that alter restriction sites Where a SNP creates or alters a restriction site, PCR can be used to detect the polymorphism, provided that the sequence is known for several hundred bp encompassing the SNP. The procedure is illustrated in Fig. 5. Following PCR amplification, restriction enzymes are used to display fragment lengths by gel electrophoresis and ethidium bromide staining. Although RFLPs are usually based on SNPs, most SNPs do not occur in restriction sites; there are many more SNPs than there are RFLPs.

Allele-specific oligonucleotide (ASO) hybridization The ASO procedure is based on the principle that a short nucleotide sequence of *c*. 20 bp is long enough to detect a change resulting from an SNP as a consequence of the stringency of hybridization (a longer sequence has too many H-bonds and the hybrid DNA is unstable). The 20-bp normal sequence will only hybridize to one of the two alleles at the SNP locus, and will make a hybrid DNA that remains stable when the temperature is raised, whereas the SNP allele DNA strands will fall apart. The method has the advantage that it is not based on gels, and that large numbers of genomic fragments from many individuals on a DNA chip (up to *c*. 10 000) can be genotyped simultaneously. The microarray is 'interrogated' (hybridized) with the target DNA (allele-specific short sequence) to see which probe DNA fails to give stable hybridization when the probe is washed off, and thus represents an SNP.

Molecular beacons A molecular beacon is a single-stranded oligonucleotide probe with complementary regions at each end, allowing the probe to form a hairpin-like structure. A fluorophore is attached at one end of the beacon, and a fluorescence quencher at the other end, with the fluorophore adjacent to the quencher. This arrangement prevents the beacon from fluorescing (Fig. 6). The molecule is also engineered such that only the probe sequence is complementary to the genomic DNA that will be used in the assay. When the probe finds its target DNA it will anneal and hybridize with it, and the hairpin region of the probe will then form a longer and more stable linear hybrid with its target sequence. The fluorophore will be separated from the quencher and will fluoresce. If the probe encounters a target to which it cannot

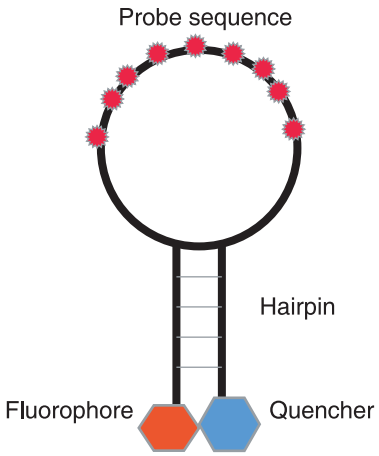


Fig. 6 The structure of a molecular beacon designed to detect a particular sequence in target DNA. When the probe finds its homologous sequence the hairpin takes on a linear form and fluorophore emits its specific wavelength of light; otherwise, it remains as a hairpin and the quencher prevents the fluorophore emitting.

anneal, the beacon will remain in its original hairpin state and will not fluoresce (Tyagi & Kramer, 1996). Two beacons can be used to identify SNPs at any given location, and to genotype an individual. The first beacon matches the wild-type allele and the other the mutant SNP allele. When the wavelength of the fluorophore of the first beacon is detected the individual is homozygous at the wild-type locus; when the wavelength of the second probe is found the locus is homozygous for the SNP allele; and if both beacons fluoresce the locus is heterozygous with both alleles. A single reaction mixture using several different fluorphores can be designed for several different targets, but within limits.

5. Genome scanning for expressed genes

Expressed sequence tag (EST) ESTs are short subsequences of *c*. 200–800 bp of cDNA sequence, prepared from mRNA, that can be used be used to 'tag' genes and to generate markers. They are useful for finding your gene, followed by determination of its gene sequence, and for gene mapping. ESTs are generated by sequencing either one or both ends of an expressed sequence. An EST can then be used to find the corresponding gene by hybridizing to homologous sequences in a genomic DNA library. The challenge in identifying genes within the genomes of different organisms is dependent upon genome size as well as the presence or absence of introns which interrupt the coding sequence of a gene and therefore reduce the similarity between an EST (a product of mRNA, from which intron regions have been removed) and the original gene. There has been rapid progress in EST identification because of its high utility, and at the time of writing 61 357 256 entries are listed in the National Center for Biotechnology Information (NCBI) public database, covering almost 20 million entries for plant species (http://www.ncbi.nlm.nih.gov/dbEST/). Over 19 million of these are from flowering plants, and a high proportion are from important crop plants, including wheat and other cereals, for which no complete genome sequence is yet available; ESTs are especially useful for gene discovery in such species. ESTs can also provide information about gene function, as an EST collection reports on which genes are being expressed in a particular tissue or developmental stage, or in response to a given environment or stress.

Sequence-related amplified polymorphism (SRAP) SRAP is a marker method developed by Li & Quiros (2001) and Li *et al*. (2003) using genomic DNA from *Brassica oleracea*, and is based on a two-primer system, using primers of arbitrary sequence, 17–21 nucleotides in length, that preferentially amplify open reading frames (ORFs). In principle, any primer pair, if short enough, will amplify many sequences in a genome, but the basis of SRAP, as with other methods described here, is to select the specific primer combinations to achieve the intended outcome. SRAP is considered to be a more powerful technique than RAPD and ISSR for revealing genetic diversity among closely related cultivars because a core sequence CCGG in the forward primer targets exonic gene sequences, while the core of the second primer (AATT) binds to the AT-rich sequences of noncoding regions. PCR with these two primers amplifies the junction fragments between genes and their noncoding flanker regions, and the SRAPs should therefore often be tightly linked to actual genes, and generate a fingerprint of the coding sequences. The amplified DNA fragments can be separated using denaturing acrylamide gels and detected by autoradiography to produce the cultivar fingerprints. The basis of the polymorphisms lies in allelic differences, which correspond to the 3′ ends of the primers. The early PCR cycles are run at 35°C, which tolerates a high level of mismatch, but in the later cycles at 50°C mismatches are not tolerated because the primers now pair with only those products that were amplified in the early cycles, and not with the genomic sequences. SRAPs generate a large number of polymorphic fragments in each reaction, are simple to use, are applicable to any species and are highly reproducible, and their PCR products can be directly sequenced without the need for cloning. On the downside, both dominant and co-dominant markers may be found, which is a complication, and null alleles are not directly detectable. SRAPs have been amplified in several crops (Li & Quiros, 2001) such as potato (*Solanum tuberosum*), rice, lettuce (*Lactuca sativa*), Chinese cabbage (*Brassica rapa*), rapeseed (*Brassica napus*), garlic (*Allium sativum*), apple (*Malus* × *domestica*), *Citrus*, and celery (*Apium graveolens*).

Target region amplification polymorphism (TRAP) TRAP has similarities with SRAP, in that it uses (1) two primers *c*. 18 nucleotides long and (2) nonstringent PCR conditions during the first five cycles. The difference between the two methods is in the configuration of the primers. TRAP requires cDNA

or EST sequence information for its fixed primer development, and the second primer has an arbitrary AT- or CG-rich core which anneals respectively to either intron or exon sequences. The amplification products are separated using polyacrylamide or agarose gels, and detected by autoradiography and silver or ethidium bromide staining. The major advantage of TRAP over SRAP is that it exploits the vast amount of EST sequence held in the public databases. SRAP and TRAP have utility in germplasm diversity analysis, genetic mapping, including transcriptome map construction, and trait gene tagging and gene cloning in many crops, for example in analysis of genetic diversity in sugar cane (*Saccharum*) collections (Alwala *et al.*, 2006), and varietal mapping in wheat (Liu *et al.*, 2005).

6. Markers using array technology

Diversity arrays technology (DArT) is a high-throughput method, independent of sequence information, that can be used to discover hundreds of markers in a single experiment (Jaccoud *et al.*, 2001; Wittenberg *et al.*, 2005). It is useful in biodiversity studies to detect differences between cultivars of a crop species. DNA is extracted from the plants of interest, and a part of the DNA from each cultivar is mixed, as the starting material. The mixture is then cut into smaller fragments using restriction enzymes that target specific regions of the genome. A pool of DNA fragments of varying sizes is then cloned to produce multiple copies of these small fragments. The result is known as a representation, and it is of reduced complexity compared with the original starting material. A random sample of cloned fragments are then arranged as a microarray. A representation of each cultivar is then prepared, in the same way as for the staring material, and the DNA is labelled with a fluorescent dye and bound onto individual microarrays for each cultivar. The labelled DNA hybridizes with its matching spots on the microarrays. The hybridization patterns will be different for each cultivar, and will reveal the degree of diversity between the two cultivars concerned. DArT was initially used to detect a large number of genetic differences between varieties of rice and wheat, but also has wide application for mapping across almost all plant species, for example barley (Wenzl *et al.*, 2006).

7. Other marker systems

Single-strand conformation polymorphism (SSCP) SSCP is based on the differences in the conformation of single-stranded DNA fragments, that is, the way in which single-stranded DNAs will fold into secondary structures (conformations) based on their nucleotide sequences and their physicochemical environments (Sunnucks *et al.*, 2000). It is considered to be the most suitable method for detecting mutations in short stretches of DNA (75–250 bp), or bacterial 16S rRNA (Schwieger & Tebbe, 1998); rice sequence information has also been used to amplify introns with primers targeting conserved exons, followed by SSCP detection (Bertin *et al.*, 2005). In the SSCP technique, PCR is carried out on the DNA sample with a fluorescently labelled primer, and the double-stranded DNA is then heat-denatured and cooled on ice to prevent renaturation. The single-stranded fragments are separated by polyacrylamide gel electrophoresis, on the basis of their three-dimensional conformation, and detected with UV light. The fragments, which may differ by as little as a single base pair out of several hundred bases, can then be cut out and sequenced to profile an individual from with a population sample of DNA sequences.

Denaturing gradient gel electrophoresis (DGGE) DGGE is a fingerprinting method for molecular ecology that separates PCR-generated DNA products of differing sequence from within a sample of DNA from, for example, individual bacteria out of a population. It is a form of electrophoresis where there is a chemical gradient across the denaturing gel, and where the melted strands fragment completely into single strands (Fischer & Lerman, 1980). The PCR products from a given reaction will be of a similar size (bp), but sequence differences in otherwise identical fragments will cause them to partially melt at different positions in the gradient and therefore to 'stop' at different positions in the gel. In other words, the DNA fragments will be separated according to their motilities under increasingly denaturing conditions. Differing sequences of DNA (from different bacteria) will generate different patterns of bands. Each type of band represents a different bacterial population within a community of microbes. The pattern of bands thus provides a fingerprint for populations within a community, or intervals within a population, and explains how the composition of the community may change according to environmental variables. DGGE is more complicated than SSCP because it requires gradient gels. See http://www.jove.com/index/details.stp?ID=164 for a video of DGGE.

Temperature gradient gel electrophoresis (TGGE) TGGE is almost identical to DGGE, but was developed as a more reliable technique. Like DGGE it utilizes the melting behaviour of DNA duplexes as a primary method for separating fragments on the gel; but the difference is that TGGE uses a temperature gradient instead of a chemical one (Lessa & Applebaum, 1993). DGGE and TGGE are considered to be technically demanding and are little used for plant genetic markers.

Methylation-sensitive PCR The methylation of CpG dinucleotides in genomic DNA is one of the main factors involved in gene silencing, and one of the most significant causes of epigenetic control over gene expression. There is much interest in mapping sites of methylation across genomes, and this can now be achieved by sequencing genomic DNA that has been treated with sodium bisulphite, a method that avoids the use

of restriction enzymes. Fronmer *et al.* (1992) first described the bisulphite reaction to distinguish between cytosine and 5mCytosine (5mC). Treatment of DNA with sodium bisulphite converts cytosine residues to uracil in single-stranded DNA, but 5mC does not react in the same way. This modification creates sequence differences between methylated and unmethylated samples. DNA sequences of interest are then amplified by primers specific for bisulphite-treated DNA, and there are a variety of protocols now used for this purpose, amongst which bisulphite genomic sequencing PCR and methylation-specific PCR (Herman *et al.*, 1996) are the most common. Using these methodologies it is possible to map methylation sites, as for any other marker. An example of a methylation map is shown in Fig. 7.

IV. Mapping approaches and tools

1. Quantitative trait locus (QTL) mapping

Because the loci of individual polygenes cannot be identified, QTLs must be mapped by a variation of the standard procedure used for molecular markers or major genes (Paterson *et al.*, 1988). The approach is to establish the statistical relationship between the inheritance of the trait and that of molecular markers whose map positions are known. The principles of QTL mapping were outlined by Jones *et al.* (1997; Fig. 8). Consider a quantitative trait, such as the greenness or yellowness of a given leaf at a given time in development of the whole plant. The mapping population is established by crossing two parent lines that are divergent in molecular markers (for example SSRs or RFLPs) and in leaf colour at time *t*. RFLP or SSR alleles and QTLs for colour will then segregate in the progeny. Fig. 8 presents the simplest case of a single QTL comprising a cluster of more or less adjacent polygenes interacting to give quantitative control of colour. The possibilities are shown for this QTL in relation to four molecular markers, one nearby, one more distant, one that is remote but still linked, and an unlinked marker. All the plants with the specific allele of the molecular marker from parent 1 will display a frequency distribution for colour phenotype, as will plants with the parent 2 allele. The degree to which these individual frequency distribution curves coincide is related to the genetic distance between marker and QTL. This is quantified by statistical procedures such as maximum likelihood, and allows a QTL to be located on the molecular marker map as a score above a threshold that represents the likelihood that the effect occurs by chance.

Recent examples of the application of QTLs can be found for the genetic components of fertility in perennial ryegrass (*Lolium perenne*) (Armstead *et al.*, 2004), reduced culm internode length in barley (Samuri *et al.*, 2009), the identification and verification of QTLs for agronomic traits in wild barley introgression lines (Schmalenbach *et al.*, 2009), and numerous other current references which appear regularly in the literature.

2. Expression quantitative trait locus (eQTL) mapping

Expression quantitative trait locus mapping, as with QTL mapping, seeks to find the underlying genomic variation to explain traits. It is a novel system, and it differs from traditional QTL mapping in that thousands of traits are analysed and thousands of QTLs are revealed, rather than just a few. eQTL mapping combines microarray expression profiling with genetic mapping, and presents logistical issues in terms of the volume of data requiring analysis that it generates.

Essentially, as in mapping in general, two parental inbred lines are crossed to give an F_1, and two F_1s are then crossed to produce the F_2, comprising many individuals unique for their traits and expression profiles through recombination in the heterozygous F_1. F_2 traits can be traced to particular chromosomal locations by statistically correlating the phenotypes and genotypes across the F_2 population. The expression of genes determining traits is determined by microarray analysis, and the expression levels of the genes (phenotypes) are treated as quantitative traits. Statistical analysis correlates DNA marker variation (e.g. SSRs and ESTs) with the levels of gene expression. Significant correlations relate the genes to their respective chromosomal regions, as has been determined for cell wall digestibility in maize (Shi *et al.*, 2007), for example.

3. Linkage disequilibrium and association mapping

Even if recombination across an entire genome, or chromosome, occurred randomly (which the almost universal existence of genomic hotspots and regions of little or no chiasma formation shows is practically unknown; Lichten & Goldman, 1995), the markers near to a given locus would have a tendency to remain linked to it as a consequence of the Poisson distribution of random crossing-over events (Hanson, 1959). The locus will carry some of its nearby markers along as it moves from generation to generation, a phenomenon known as linkage drag (Brown *et al.*, 1989). The statistical relationship between the inheritance pattern of these markers and the trait locus constitutes the haplotype and will deviate significantly from the random behaviour of markers in general. Such associated markers are in LD, and the degree of LD can be used as an estimate of map distance provided that allele frequencies are also taken into account. Note that LD as a mapping parameter is a population concept, whereas linkage is a chromosome-level phenomenon (Flint-Garcia *et al.*, 2003). Linkage drag can be exploited for rapidly tagging a locus in a mapping family by bulked segregant analysis (BSA; Michelmore *et al.*, 1991). This works well for a trait that is sufficiently qualitative in expression (for example disease resistance versus susceptibility) that the segregating family can be separated into two pools on the basis of the two phenotypic extremes. Screening DNA from the respective pools will distinguish unlinked (nonpolymorphic) markers from marker polymorphisms segregating with the trait in question (Fig. 9). Tagging does not give a map distance

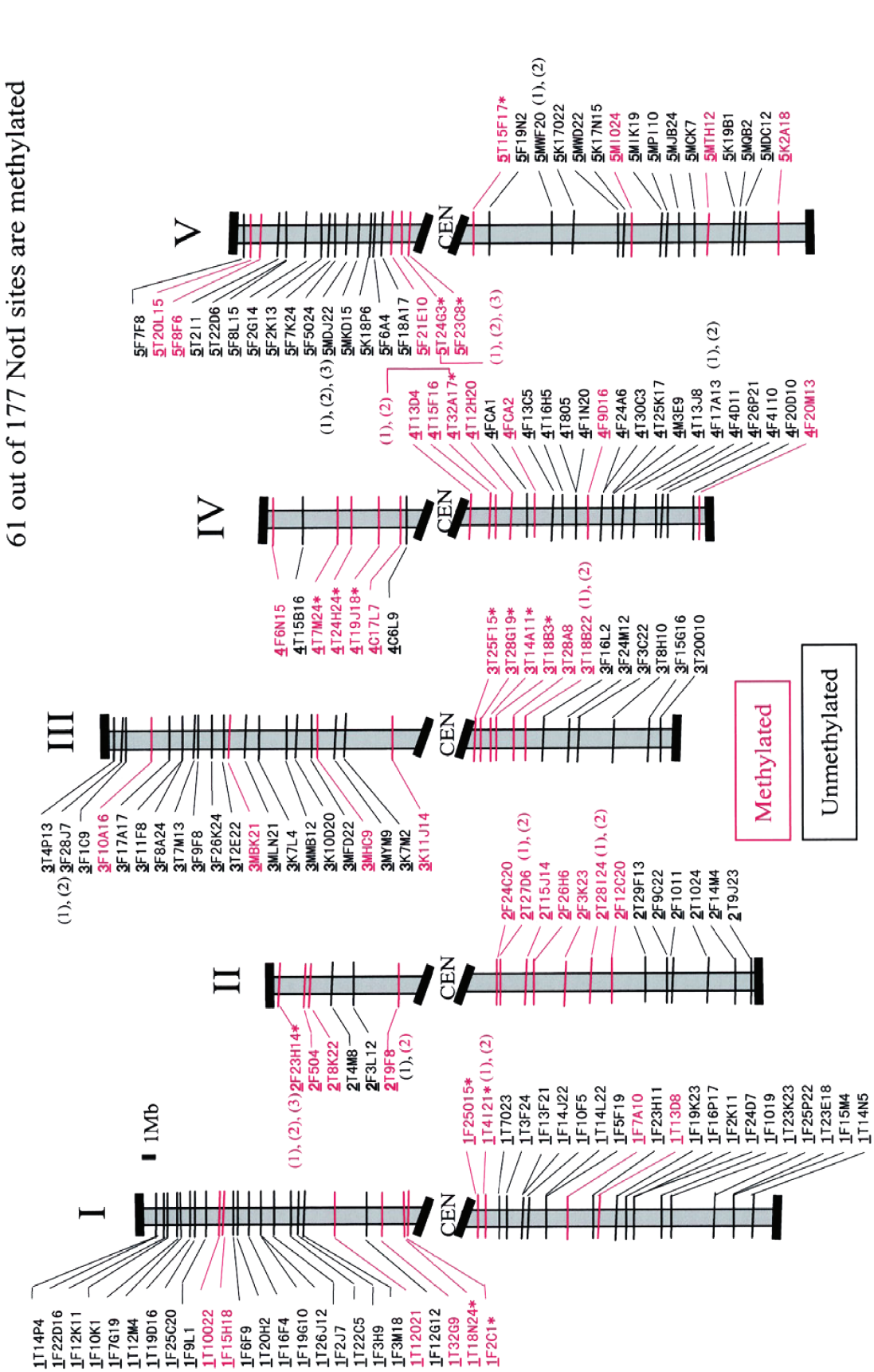


Fig. 7 A physical map of the methylation status of the *Arabidopsis thaliana* genome, constructed using the methylation-sensitive restriction enzyme *Not*I (from Matsuyama *et al.*, 2003). Loci with asterisks were not yet confirmed by polymerase chain reaction (PCR) in this study.

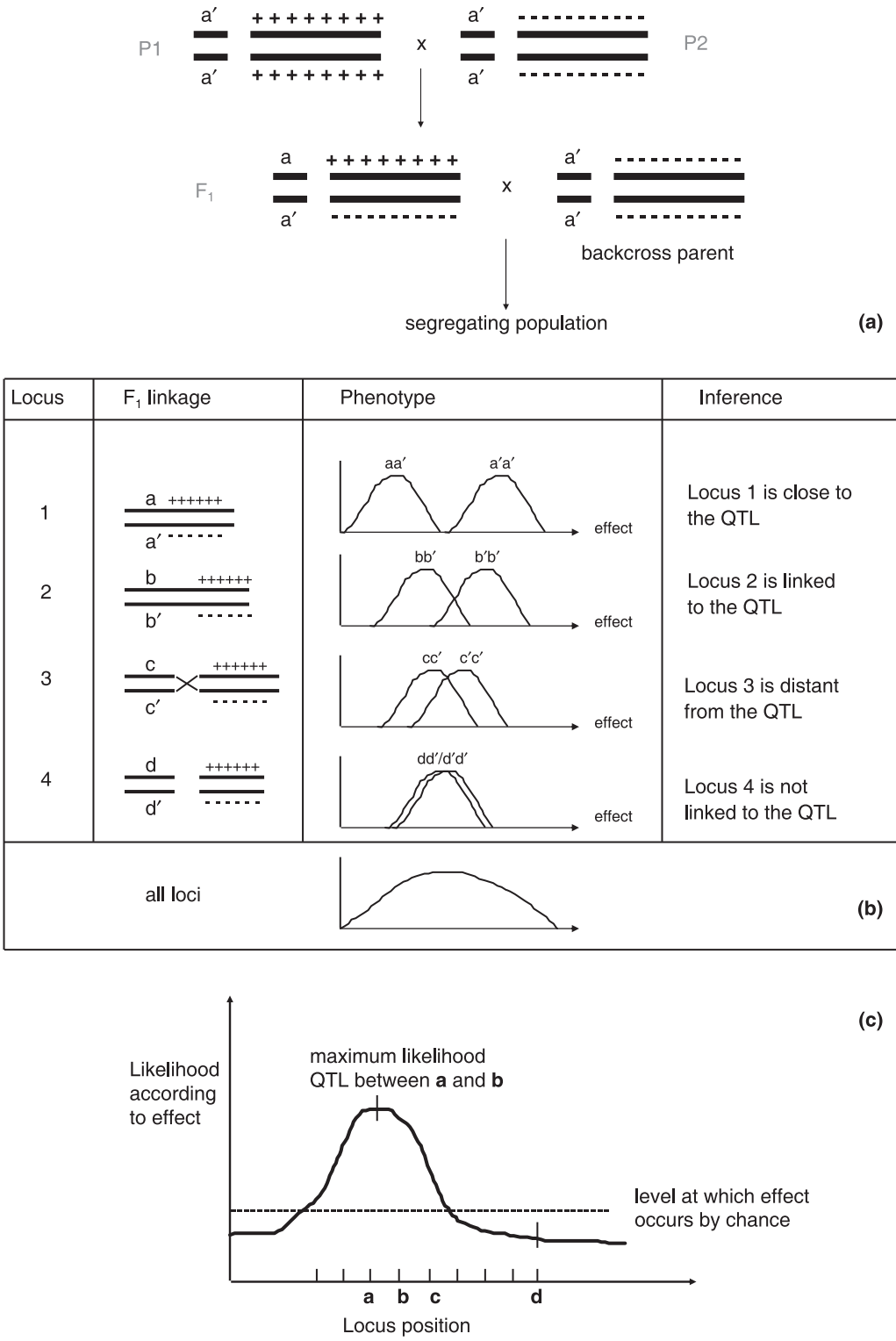


Fig. 8 Composite diagram of the procedure for mapping a quantitative trait locus (QTL). (a) A mapping population is established by crossing parents which are divergent for their restriction fragment length polymorphism (RFLP) markers and for the quantitative character concerned (plant height). The heterozygous F_1 is then backcrossed to one of the parents to give the segregating population. (b) The linkage between the QTL and various marker loci can then be ascertained by the way in which height distribution patterns are associated with the segregation of the two alleles at each locus. (c) The map position of the QTL is determined as the maximum likelihood from the distribution of likelihood values (ratio of likelihood that the effect occurs by linkage: likelihood that the effect occurs by chance) calculated for each locus. (Figure based on an idea by Glynis Giddings.)

Table 3 Comparison of mapping families derived from parental crosses and natural populations

Parental cross-derived mapping family	Natural population
• Parental lines must be genotypically and phenotypically contrasting and crossable	• No parents
• At least 2–3 generations must be grown before analysis can begin	• No generation time
• High-resolution mapping requires production and maintenance of large populations	• Natural populations often very large
• Alleles can be screened only in pairs in diploids	• Large numbers of alleles may be present and detected at any given locus
• Number of assessable phenotypic traits usually limited	• Large numbers of phenotypes may be screened in the same population
• Markers are often family-specific	• Markers are generic
• Less prone to false positives	• Spurious correlations related to population structure can be a problem

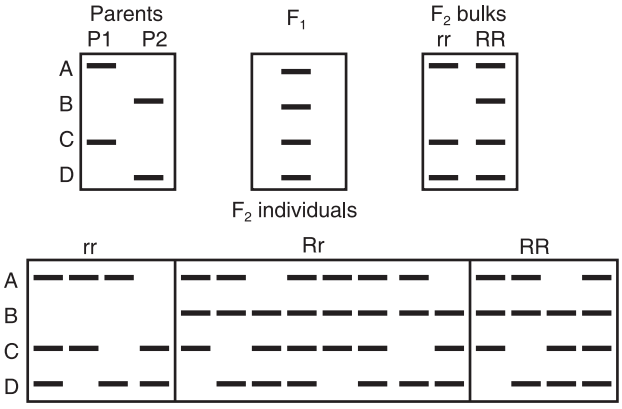


Fig. 9 The scheme of bulked segregant analysis (from Michelmore *et al*., 1991).

but, where comparative or syntenic relationships have been established, it may be sufficient to target a map location. It is also a practically useful tool for marker-assisted selection.

It is clearly not feasible to carry out linkage analysis in humans by creating synthetic inbred mapping families. This has given great impetus to the exploitation of LD in natural human populations for association mapping (Jorde, 1995). Because of the availability of rich sources of genetic variation in genebanks for conservation of wild species and breeders' germplasm, LD is a particularly appropriate approach for mapping in plants too. Table 3 lists a number of advantages of natural compared with synthetic populations.

The case of heading date in *L. perenne* is a typical example of the application of LD analysis to mapping a trait in a species in which good genetic resources are available (Skøt *et al*., 2005). The study employed 26 wild, semi-natural and cultivar accessions (each of which, because *L. perenne* is predominantly outbreeding, has a high degree of variability and heterozygosity in its own right). The ecotypic provenances of these populations covered the latitude range 37° to 53°N and longitude 8°W to 9°E. Observations of heading date in field trials allowed populations to be classified into five groups: very early, early, intermediate, late and very late. The total number of genotypes screened was 832, and 589 polymorphic AFLP markers were identified. Of the variance in allele frequencies, 61% was attributable to within-population variation and 39% to variation among populations. Mapping in natural populations is prone to errors arising from nonrandom complexities of population structure (Table 3). This may be tested and corrected for using various statistical procedures, including principal co-ordinate analysis and hierarchical clustering based on genetic distance (Nei, 1972; Pritchard *et al*., 2000). In this way Skøt *et al*. (2005) identified four populations of hybrid *Lolium* species as the major contributors to marker distortions, probably because of the presence of *Lollium multiflorum* alleles in the *L. perenne* background. Two AFLP markers were significantly linked with heading date across all populations and a further marker was linked when population structure artefacts were taken into account. Two of these three markers mapped on *Lolium* chromosome 7 to within 2 cM of a major QTL controlling 70% of the variation for heading date (Armstead *et al*., 2004).

The degree of LD reflects the balance between the frequency of new mutations, which create novel haplotypes, and recombination rate and location, which tend to destroy haplotypes by reassortment (Rafalski & Morgante, 2004; Kim *et al*., 2007). In highly homozygous inbreeders such as *A. thaliana*, recombination takes place between largely identical haplotypes, thus preserving LD across blocks of tens to many hundreds of kb. The LD patterns of outbreeders are much more heterogeneous because of the nonrandom nature of crossing-over, but in general LD extends over significantly smaller lengths of genome: 5–60 kb in humans, 0.1–0.2 kb in Norway spruce (*Picea abies*) and 0.2–1.5 kb in maize (Rafalski & Morgante, 2004). The existence of contrasting populations of a given plant species differing in the extent and diversity of LD opens the possibility to cross them and create powerful resources for whole-genome association analysis, predictive diagnostics for plant breeding and rapid fine-mapping and isolation of genes, complementary to (and perhaps in time supplanting) traditional linkage mapping family approaches.

4. Genetical and physical mapping

To isolate the DNA sequence responsible for the phenotype associated with a map locus it is necessary to understand the

relationship between linkage map distances and physical distances along the genome. It is also important, especially when working with a nonmodel species, to use comparative map information to zero in on sequences of interest and circumvent the need for time- and resource-consuming fine-scale mapping, which often runs up against limitations of ever-decreasing recombination frequency and polymorphism as the target region gets smaller.

Genetic maps do not tell us which linkage groups correspond to which chromosomes, nor how the markers within a linkage group relate to the physical structure of the chromosome. It is commonly the case in grasses, cereals and many other plants that recombination does not occur with equal frequency across the whole genome. Under these circumstances, a marker might appear tightly linked to a gene of economic value, but in reality be many kilobases away in the actual chromosome. For marker-assisted selection this tight linkage will serve the purpose anyway; but for map-based cloning the desired gene might be too distant from the marker to be reached using the marker probe. It is thus desirable to saturate the map with as many markers as possible and, for cloning purposes, to integrate genetic (linkage) and physical (chromosomal) maps. To assign linkage groups to specific chromosomes, use can be made of various chromosomal stocks, such as trisomics, monosomics, addition lines, translocations and deletions, which give modified segregation patterns and expose chromosome-specific markers. A particularly useful account of physical mapping in barley, using *in situ* hybridization (ISH) to chromosomes, has been presented by Pedersen *et al.* (1995). These authors show how mapping of single and low-copy genes by ISH can provide 'anchor sites' for integrating the physical and genetic maps.

The full power of moving between the linkage and physical maps is realized through the availability of large-insert clones (up to 250 kb) in the form of bacterial artificial chromosomes (BACs). Assembling BACs into overlapping regions spanning part or all of the genome ('contigs') makes it feasible to get directly from a mapped locus to candidate gene sequences, but there are challenges in applying BACs to large-genome species which have dispersed repeats common to different chromosomes. This problem has been addressed in wheat where a physical map of chromosome 3B has been constructed based on flow-sorted chromosomes and chromosome-specific BAC libraries (Paux *et al.*, 2008).

The entire *A. thaliana* genome has been cloned into BACs (Mozo *et al.*, 1999). The inserts have been sequenced and arranged into contigs by matching overlapping sequences at their ends. BACs specific for individual chromosomes are identified on the basis of the karyotype (chromosome morphology); and to visualize them they are nick translated to incorporate fluorescent nucleotides and then used as probes for ISH. Chromosome parts can be tracked at interphase to show their disposition in the nucleus and, in *A. thaliana*, for true chromosome painting. An integrated linkage-physical map of rice, one of several, was first published by Chen *et al.* (2002).

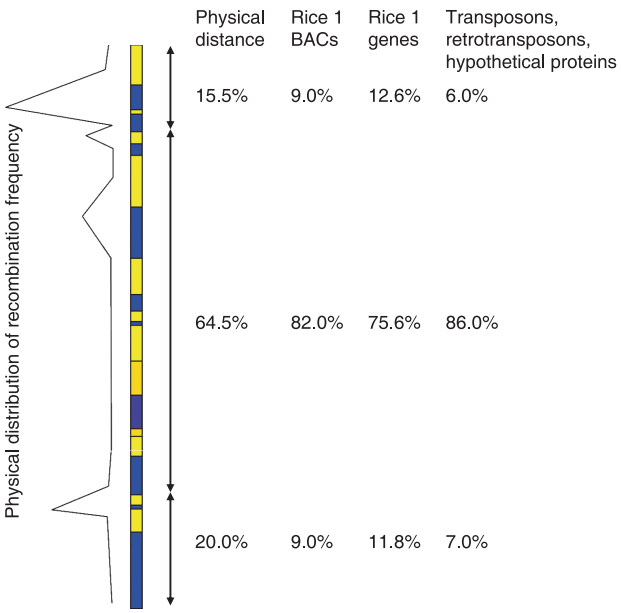


Fig. 10 Physical map of chromosome 3 of *Lolium perenne*, showing how recombination values, based on chiasma distribution, vary along the length of the chromosome, and how this relates to the distribution of a number of physical markers. Courtesy of Julie King.

The relationship between genetic distance (in Morgans) and physical distance (in base-pairs) is not consistent from one species, or part of the genome within a species, to another (Chen & Gustafson, 1995; Erayman *et al.*, 2004). Differences in recombination frequency along the chromosome may identify physically distant markers as genetically close to each other, or vice versa. In high-recombination regions of the wheat genome, for example, 1 cM is equal to 118 kb, but where recombination frequency is low the ratio is as high as 22 000 kb per cM (Gill *et al.*, 1996a,b). Structural features of chromosomes such as centromeres, telomeres, nucleolus organiser regions and pericentric regions are often associated with reduced recombination (Tanksley *et al.*, 1992; Erayman *et al.*, 2004; King *et al.*, 2007a). Conversely, recombination hotspots also occur (Künzel *et al.*, 2000; Erayman *et al.*, 2004). Figure 10 shows the mismatch between recombination frequency and the distribution of genes and markers along a single chromosome of *Lolium-Festuca* (King *et al.*, 2007a).

5. Introgression, radiation hybrid and gametocidal chromosome-based mapping

Introgression mapping Substituting an entire *Festuca pratensis* chromosome for each of the seven *L. perenne* homeologues generates a monosomic series. In turn each substitution line is the basis for a recombination series comprising progressively smaller *F. pratensis* segments in the *L. perenne* background (Fig. 11). The resource thus created permits rapid mapping, both linkage and physical, of *Lolium*- or *Festuca*-specific traits and markers, as well as supporting fundamental studies on the

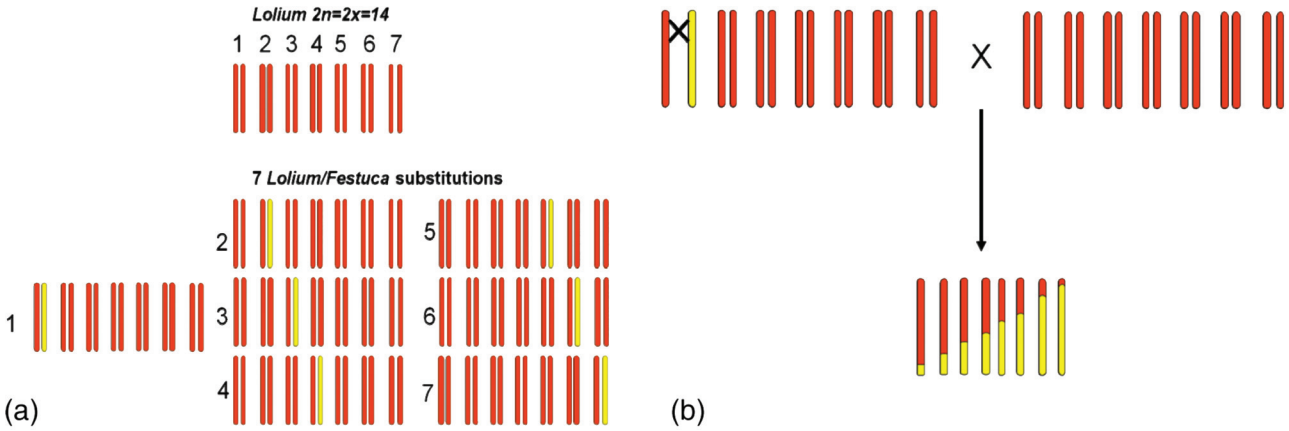


Fig. 11 Diagram showing (a) a chromosome substitution series for each of the seven chromosomes of *Festuca pratensis* into the background of *Lolium perenne* and (b) a recombination series of differently sized segments of a single *F. pratensis* chromosome introgressed into one of the chromosomes of *L. perenne*. Courtesy of Julie King.

distribution of genes and recombination sites in large-genome plants (King *et al.*, 2007b).

Radiation hybrids (RHs) Another approach to relating physical sizes and genetic distances in chromosome arms is the use of RHs (Riera-Lizarazu *et al.*, 2008). This is best exemplified by experiments with monosomic or disomic addition lines of each of the 10 chromosomes of maize into the background of hexaploid oat, *Avena sativa* (Kynast *et al.*, 2004). In these oat × maize crosses proembyos were rescued by embryo culture and found to retain one or more of the maize chromosomes in addition to the haploid oat genome (buffered by its polyploid status). To fragment the maize chromosomes, F_1 monosomic lines were irradiated with γ-rays to produce RH lines as a tool for low-resolution mapping of chromosome segments onto which genes and markers (SSRs) could be allocated. Physical mapping of the barley genome has also been undertaken in a similar way (Wardrop *et al.*, 2002, 2004), but the application of RH mapping in plants has so far been limited.

Gametocidal chromosome (GC)-based mapping. As an alternative to radiation treatment Masoudi-Nejad *et al.* (2005) used the GC from *Aegilops cylindrica* to fragment chromosome 7 of barley as an addition line in the background of hexaploid wheat, and to physically map a number of SSRs and AFLPs onto the barley chromosome, at high resolution. In both RHs and GC-based mapping, no polymorphic markers are needed and these two systems, their complexity notwithstanding, have potential for ordering the mass of genetic markers now available in relation to the physical dimensions of the chromosomes.

6. Comparative mapping

Comparative mapping aligns genetic maps among a number of related species, as in the Gramineae. When this is done we find that parts of the genome for a number of crop species show similarity for their map profiles. Rice, as the model for the grass family, presents us with full sequence information. The syntenic relationships are such that genes are collinear and conserved: the whole of the Gramineae can now be looked upon as a single genetic system built up from *c.* 30 rice linkage blocks (Devos & Gale, 2000). This stroke of evolutionary good fortune enables us to use the rice map to transfer and to fill in detail into the relatively skeletal maps of the large genomes of cereals and other grasses.

Molecular marker mapping has strengthened our realization that, in several taxonomic groups of crop plants, notably within the Gramineae, which share the same common basic chromosome number, the linkage groups and the individual chromosome maps look very similar. When we take out the repetitive DNA and compare the maps for single copy sequences (essentially RFLPs) we find that they are syntenic. This means that even between crops as diverse as wheat and rice the genes we are interested in are basically the same in both species, and they line up into maps that are broadly very similar (Moore *et al.*, 1995), although it should be noted that synteny frequently breaks down at the micro-level (Bennetzen, 2000). The added value to mapping is that not only can we use the same set of RFLP probes across wide species gaps, but we can transfer map information, even entire maps, from one species to another. It becomes possible to know the location of genes of interest in, for example, wheat by reading the marker map of the rice model. Increasingly too we are able to relate genetic maps to the physical dimensions and organization of chromosomes, and this greatly enhances prospects for gene isolation and manipulation.

7. HAPPY mapping – linkage without sex

In a sense, the business of choosing parents, making crosses, establishing families and manipulating populations to render them suitable for genetic analysis is the price that has to be paid for using the biological tools of meiosis, sex and recombination to

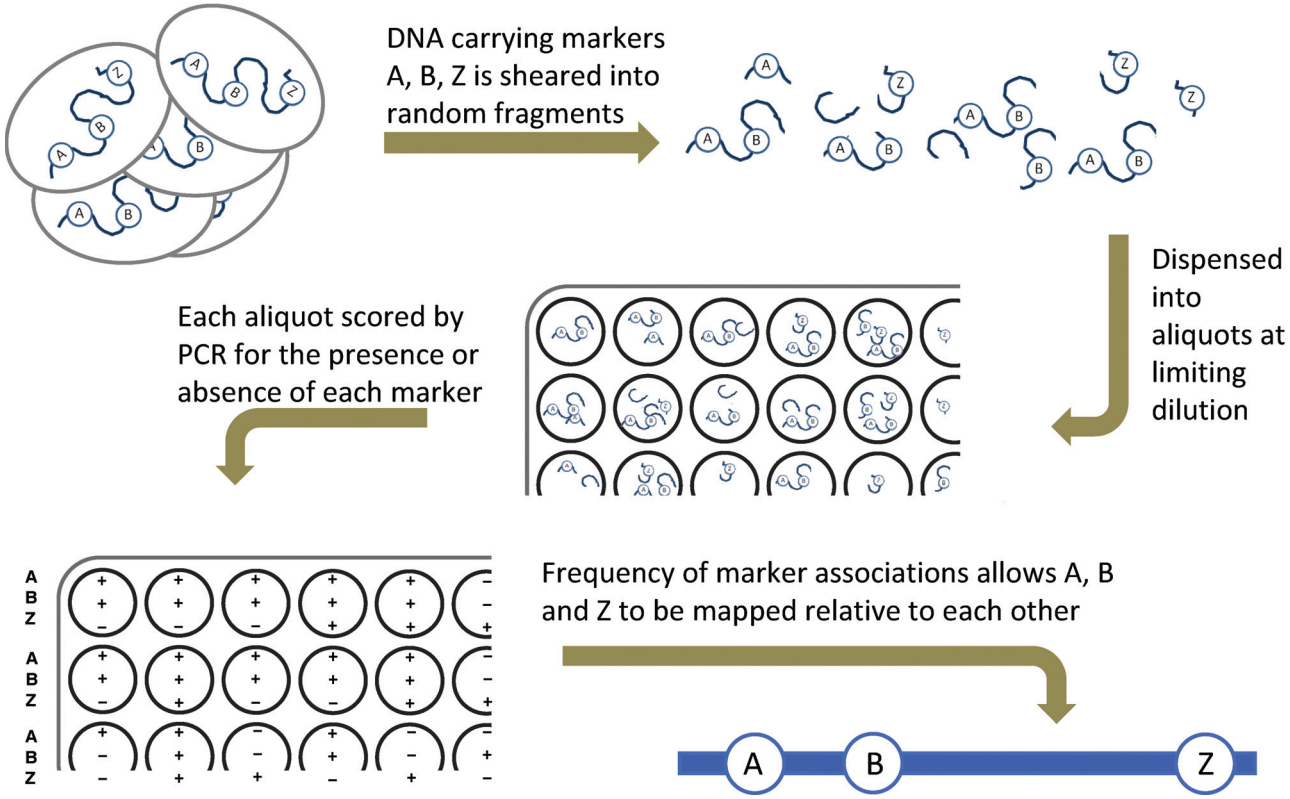


Fig. 12 Principle of HAPPY mapping and marker typing (after Thangavelu *et al*., 2003).

reveal marker–trait associations. But in principle, determining the statistical relationships between sequences could be achieved by using a 'physical analogue of meiosis' (Dear & Cook, 1993) – cutting the genome into pieces *in vitro* and visualizing which markers remain associated on the same DNA fragment and which are separated. This is the premise of HAPPY mapping (the name comes from haploid and polymerase chain reaction). The procedure was first proposed as a mapping approach for the human genome by Dear & Cook (1989), and has subsequently been applied to a number of animal genomes, and to *A. thaliana* (Thangavelu *et al*., 2003; Fig. 12). DNA carrying molecular markers A, B and Z is extracted from cells and broken randomly. The pool of fragments is diluted and dispensed into a series of aliquots (known as the mapping panel). The panel is screened by PCR and the marker content of each aliquot scored. Linked markers (A and B) co-segregate; remote markers (B and Z) do not. Co-segregation frequencies allow marker-to-marker distances to be computed, resulting in a linkage map of A, B and Z.

V. Survey of bioinformatics and computer resources

1. Mapping needs bioinformatics

Just as in the case of DNA sequence and gene expression data, the advent of higher-throughput, more economical practical methodologies for markers and mapping has led to the need for bioinformatics tools to manage and make accessible the flood of data. For many crop species, web-accessible databases provide interactive genetic maps, including QTL maps for traits of agronomic interest. This availability greatly facilitates the selection of appropriate markers for marker-assisted selection programmes, not only for the species in question but for related species where collinearity of gene (and marker) order is sufficient. Bioinformatics is also valuable for comparative genomics research, leading to new tools for gene function discovery. The *stay-green* gene story related in Section VI could not have been elucidated without the online availability of key bioinformatics resources, including the Gramene website (see Table 4), the rice genome annotation project (http://rice.plantbiology.msu.edu/), and the Genevestigator tool for visualizing patterns of gene expression (see Table 4).

2. Bioinformatics resources for model and major crop species

For those plant species where a complete or near-complete genome sequence is available, genome browsers using the ENSEMBL software package (http://www.ensembl.org/index.html; Hubbard *et al*., 2007) have been developed. ENSEMBL browsers collate all available data on features of the genome sequence, including the positions of known or predicted genes, ESTs, homologous sequences from other species, and available

Table 4 Web-accessible plant bioinformatics databases

Website	URL	Species covered	Major data types
The Arabidopsis Information Resource (TAIR)	http://www.arabidopsis.org/	*Arabidopsis thaliana*	Genes, gene families, proteins, sequences, markers, metabolic pathways
Gramene	http://www.gramene.org/	Rice (*Oryza sativa*) as focus; wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), sorghum and other cereals. Grape (*Vitis vinifera*) and poplar (ENSEMBL browsers)	Genetic and comparative maps, markers, traits, QTLs, ENSEMBL genome browsers, proteins, biochemical pathways
GrainGenes	http://wheat.pw.usda.gov/GG2/index.shtml	Small-grain cereals: wheat, barley, rye, and wild relatives	Genetic maps, markers, polymorphisms, QTLs, sequences
The Legume Information System	http://www.comparative-legumes.org/	Legumes, including *Medicago truncatula*, soybean (*Glycine max*), chickpea, *Lotus japonicus*, and *Phaseolus* and *Vigna* species	Sequence, genetic and comparative maps, transcript data
SoyBase	http://soybase.org/index.php	Soybean	Genetic, physical and sequence maps, markers, traits, QTLs, pathologies
Dendrome/ TreeGenes	http://dendrome.ucdavis.edu/treegenes/	Forest trees (conifers)	Genetic and comparative maps, markers, polymorphisms including SNPs, ESTs, phenotypes, QTLs
PLEXdb	http://www.plexdb.org/	Higher plants and fungal pathogens of plants	Gene expression
NASCArrays	http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl	*A. thaliana*	Microarray gene expression data and analysis tools
Genevestigator	https://www.genevestigator.ethz.ch/gv/index.jsp	*A. thaliana*, barley and rice	Gene expression, including visualization and clustering tools

EST, expressed sequence tag; QTL, quantitative trait locus; SNP, single nucleotide polymorphism.

insertional knockouts, against a background of the primary DNA sequence itself. The ENSEMBL browser is highly customizable and browsers for different plants tend to have different features depending on data availability and the requirements of their user communities. Currently the *A. thaliana* and rice ENSEMBL genome browsers, available from URLs listed in Table 4, are the most comprehensive; others, including those for maize and sorghum (*Sorghum bicolor*), are in development.

A caveat concerning the interpretation of data in ENSEMBL browsers concerns the positioning of QTLs, which are presented in, for example, the rice ENSEMBL. Because the ENSEMBL views of the genome relate to sequence-derived physical maps, whereas QTL locations are determined by statistical genetic mapping techniques, QTL positions in ENSEMBL should be regarded as approximate. Nevertheless, where a QTL has been very closely linked to a genetic marker whose physical location within the genome sequence is known, it is possible to use the ENSEMBL view to examine the possible candidate genes which may underlie a QTL.

ENSEMBL browsers present pictorial representations of genomes; alternative and useful compilations of sequence-based data are the genome annotation projects for rice (http://rice.plantbiology.msu.edu/) and *A. thaliana* (http://www.tigr.org/tdb/e2k1/ath1/). These databases provide, among much other information, listings of the known or predicted protein-coding regions, in order, on each of the BAC clones used in the corresponding genome sequencing project (Fig. 13), facilitating identification of candidate gene targets.

Genome databases for some individual plant species, or groups of related species, were in existence well before the sequencing of any plant genomes. Some of the most important and widely used are listed in Table 4. The Arabidopsis Information Resource (TAIR) is the central resource for *A. thaliana*; it provides comprehensive information about *A. thaliana* genes, their products, and the consequences of knockouts, together with genetic maps, relevant literature and a plethora of other information useful not only to *A. thaliana* scientists but to the plant community as a whole.

Database resources are particularly well developed for grasses and cereals. Gramene is designed as a resource for comparative grass genomics, capitalizing on the availability of the complete rice genome sequence as the reference. Where genetic maps from two or more species have been constructed using some common markers, Gramene uses the free software package CMap, developed by the Generic Model Organism Database project (http://gmod.org/wiki/Cmap) to display alignments of these maps, highlighting the common markers. This makes it possible to examine the conservation or otherwise of marker order within linkage groups, and to visualize potential inversion or translocation events (Fig. 14). Gramene hosts the ENSEMBL browsers for sorghum, grape (*Vitis vinifera*) and poplar (*Populus trichocarpa*) as well as rice. GrainGenes includes map data for barley, rye (*Secale cereale*), oats, and diploid, tetraploid and hexaploid wheat (*Triticum*), together with wild relatives of these species.

The Legume Information System (LIS; Gonzales *et al.*, 2005) brings together genetic maps from a number of leguminous

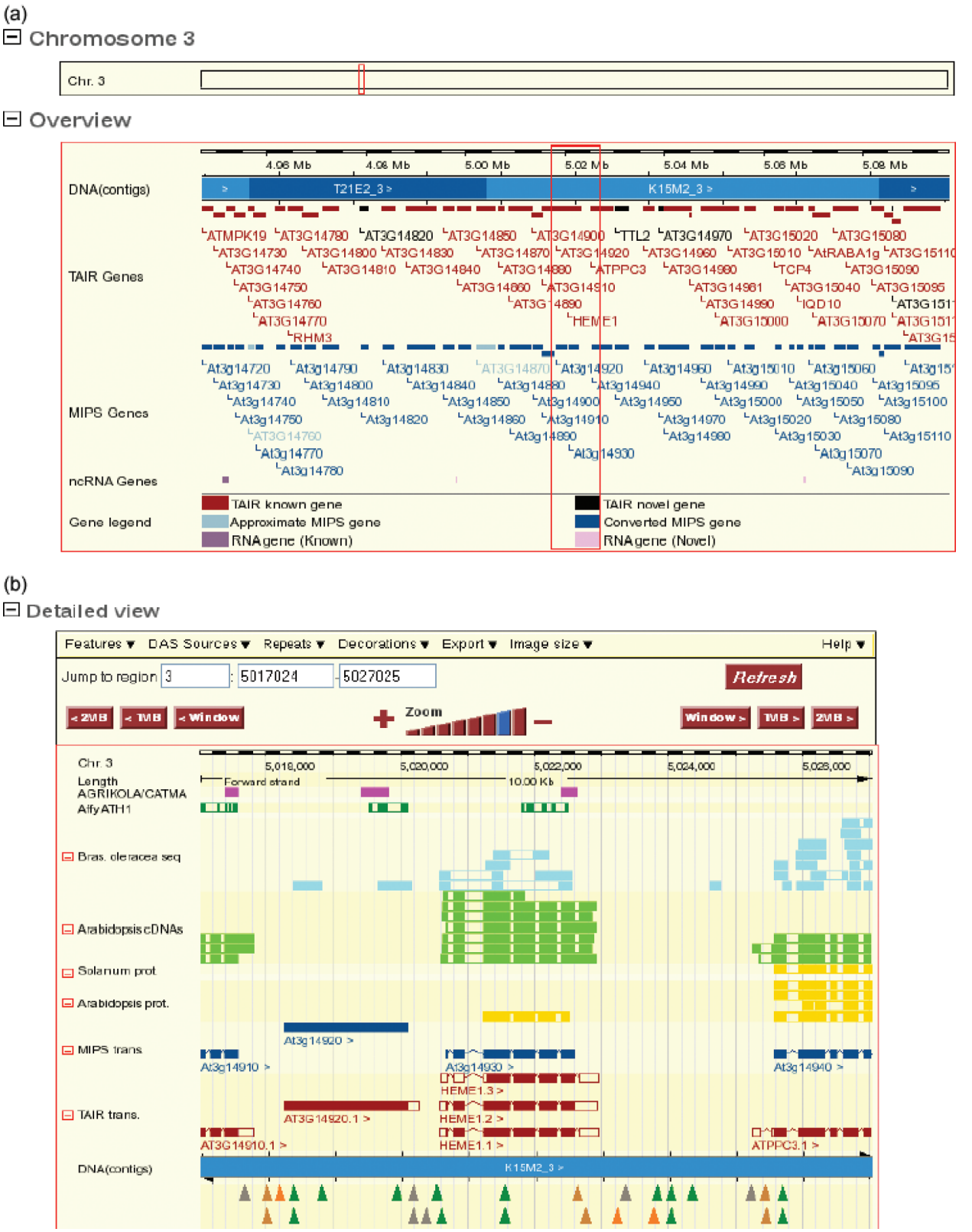


Fig. 13 AtEnsembl view of a part of *Arabidopsis thaliana* chromosome 3. (a) Overview, showing the genes identified within two adjacent contigs from the *A. thaliana* sequencing project. (b) Detailed view, showing a range of features aligned with a 10-kb region of the chromosome. In addition to genes, features include cDNAs and proteins from *A. thaliana*, *Brassica* (*Brassica oleracea*) and other species, and the positions of T-DNA insertional knockouts in genes and other regions. MIPS: *Arabidopsis thaliana* sequences annotated by the Munich Information Centre for Protein Sequences, http://mips.helmholtz-muenchen.de/index.html. ncRNA: non-coding RNA sequences. AtEnsembl is accessible at http://atensembl.arabidopsis.info/index.html (James *et al*., 2007).

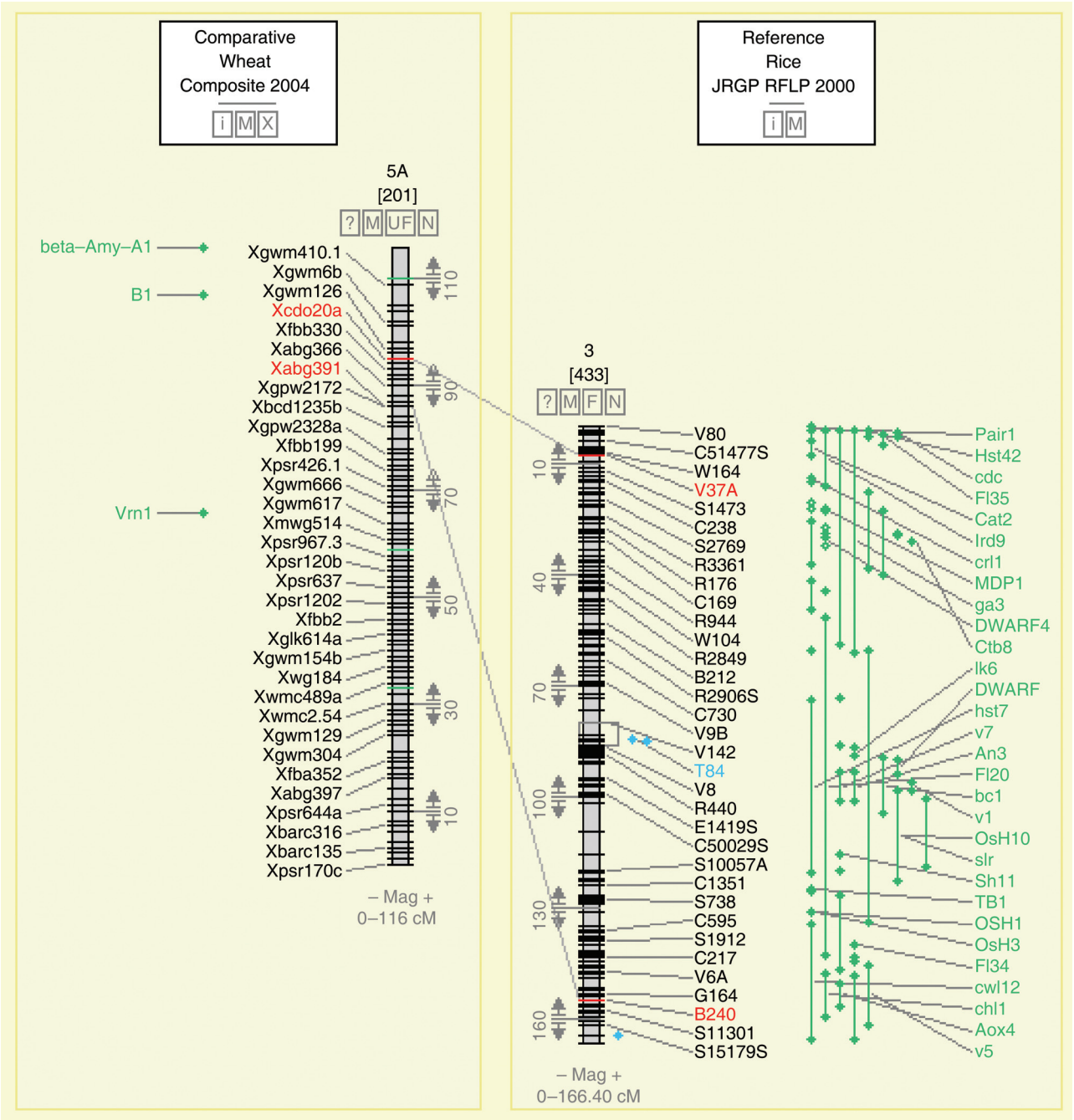


Fig. 14 Alignment between regions of a wheat (*Triticum aestivum*) genetic map (chromosome 5a) and a restriction fragment length polymorphism (RFLP)-based map of rice (*Oryza sativa*) chromosome 3, on which a number of quantitative trait loci (QTLs) have also been located. This figure was produced using the comparative mapping software CMap at the Gramene website (http://www.gramene.org/). Grey lines linking maps indicate where the same marker has been used in both mapping studies, enabling conservation of gene order to be analysed. Note that in this instance the wheat and rice mapping projects have used different names for the same RFLP markers. Gramene resolves this problem by maintaining a list of all possible aliases for each marker.

crops with the model legume *Medicago truncatula*. Like Gramene, LIS uses CMap to enable comparative map displays, although there has been less use of anchor probes in the legumes than in the grasses and cereals, resulting in a paucity of common features to link maps across the species divide. Soybean (*Glycine max*) is one of the species covered by LIS, but it also has its own web portal, SoyBase and the Soybean Breeder's Toolkit, which provides extensive information about traits, diseases and QTLs as well as maps. Details of the resources described above, and other useful plant bioinformatics sites, are given in Table 4.

3. Gene expression databases

In addition to the general gene expression databases ArrayExpress (maintained by the European Bioinformatics Institute (EBI)) and Gene Expression Omnibus (maintained by NCBI), several plant-specific databases for storing and displaying gene expression data are available via the Internet. The data contained in them are mainly derived from microarray experiments, but they also contain results from more targeted studies, including northern blot and RT-PCR analyses. As high-throughput sequencing approaches replace microarray experiments for analysing transcriptomic data, data sets of this type will also be included. In the context of markers and mapping, gene expression data are particularly useful in assigning functions to candidate genes potentially underlying traits of interest, as the tissue specificity of gene expression, and its response to environmental factors, frequently provides supporting evidence for a postulated role. The most comprehensive plant gene expression databases are PLEXdb (several higher plants and plant pathogens) and NASCArrays (*A. thaliana*); these and others are summarized in Table 4.

A useful suite of tools for examining expression profiles of genes across many experiments is Genevestigator (Zimmermann *et al.*, 2005; https://www.genevestigator.com/gv/index.jsp), which currently includes *A. thaliana*, barley and rice amongst the species covered; others are human, mouse and rat. Registration is required for access to all but the most basic functionality, and there is a fee for use of some tools, including clustering. Genevestigator allows the user to examine patterns of gene expression at different developmental stages, in different tissues, and in response to biotic and abiotic stress treatments (Figs 15, 16c).

4. Other online resources useful for mapping

While not always regarded as bioinformatics resources *per se*, the databases that document germplasm collections are as valuable to plant researchers and breeders as the others discussed above. Most larger genebanks now provide web-based access to catalogues of their stocks, often with descriptions of plant phenotype and performance under different environmental conditions as well as information about the origin of the plant material. A comprehensive review of these resources is beyond the scope of this paper, but some particularly useful sources are the sixty European Central Crop Databases (http://www.bioversityinternational.org/networks/ecpgr/links/ecpgr_search.asp), which cover all major crop species and their relatives conserved in genebanks throughout Europe; EURISCO (http://eurisco.ecpgr.org), which brings together the National Inventories of Plant Genetic Resources from 38 European countries and, by December 2008, included over 1 100 000 accessions; and the National Plant Germplasm System of the US Department of Agriculture (http://www.ars-grin.gov/npgs/index.html).

Many of the plant genome databases contain information about mutant phenotypes, with links making it possible to order mutant stocks. TAIR provides links to the Arabidopsis Biological Resource Centre (http://www.biosci.ohio-state.edu/~plantbio/Facilities/abrc/abrchome.htm) in the USA and the UK's Nottingham Arabidopsis Stock Centre (http://arabidopsis.info/), from where it is possible to order natural or chemically induced mutants, insertional knockouts, and other transgenic lines in which individual genes have been over-expressed or down-regulated. Such stocks are of great value in determining gene function.

5. Mapping software

Two of the software packages most widely used for constructing genetic linkage maps, particularly for plants, are JoinMap and mapmaker. JoinMap (van Ooijen, 2006; available from http://www.kyazma.nl/index.php/mc.JoinMap/), which determines linkage maps for diploid species, was originally developed specifically for plant data. It handles many different population types, and provides a wide range of visualization options. In addition to the construction of individual linkage maps, it has also been exploited for integrating data from a number of separate genetic maps where a subset of markers has been used on two or more of the maps. mapmaker/exp (http://www.broad.mit.edu/tools/software.html) is a another popular linkage analysis package designed to help construct linkage maps of markers segregating in experimental crosses. The original version of mapmaker (Lander *et al.*, 1987) could also be used with natural populations, but this is no longer the case for mapmaker/exp. It performs full linkage analysis for dominant (e.g. RAPDs), recessive, and co-dominant (e.g. RFLP or AFLP) markers.

Neither JoinMap nor mapmaker/exp includes the capacity to add QTL data to a genetic linkage map, but each has a companion software package with this functionality. MapQTL, the partner to JoinMap, will add QTL positions to a linkage map produced using JoinMap or other mapping software. It handles population types including BC1, F_2, RILs (any generation), cross-pollinators and doubled haploids. The mapmaker/qtl companion to mapmaker allows genes controlling polygenic quantitative traits in F_2 intercrosses and BC1 backcrosses to be placed on a genetic linkage map. The JoinMap and mapmaker/exp packages are both deployed as local installations on the user's own computer; both are available for Windows operating systems on PCs, and mapmaker/exp is also available for some other operating systems.

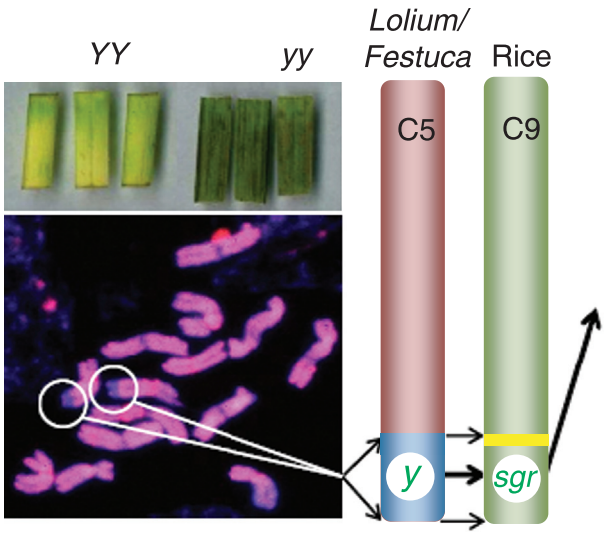
Many other packages are available for the analysis of mapping and marker data. For a comprehensive listing of genetic mapping software, the reader is referred to the website http://www.nslij-genetics.org/soft/, which is maintained by the North Shore LIJ Research Institute. The listing has been under active development since the mid-1990s, and

Results:

Graph in log(2) scale

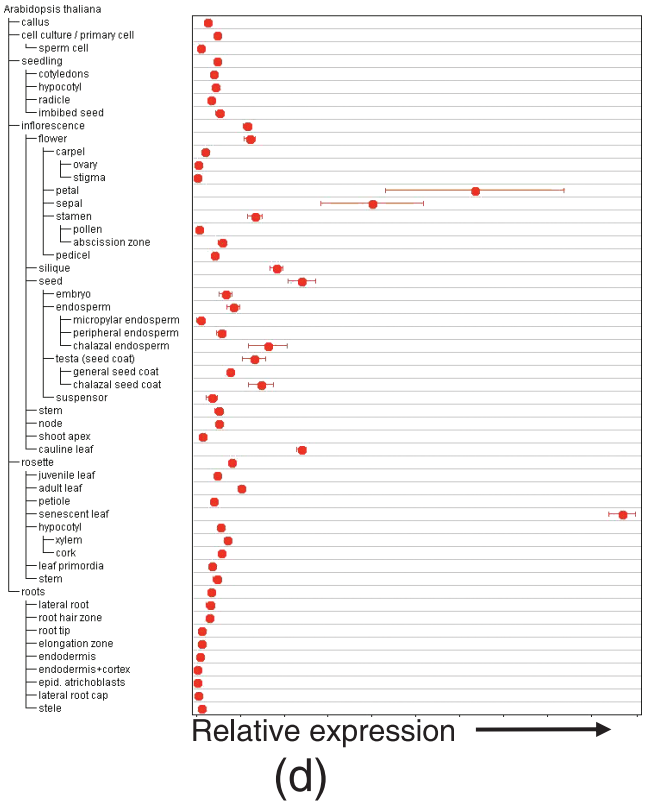
Treatment	# of Chips	Mean	Std-Err	250207_at IAT5G13930 Linear 51000 25500 0	Ratio	250207_at IAT5G13930 Linear 0 25500 51000	Std-Err	Mean	# of Chips	Control
Biotic: A. brassiciola (+)	0			n/a	n/a	n/a			0	Biotic: A. brassiciola (-)
Biotic: A. tumefaciens (+)	2	2999	2688		0.44		3879	6858	2	Biotic: A. tumefaciens (-)
Biotic: B. cinerea (+)	6	790	376		0.94		384	840	6	Biotic: B. cinerea (-)
Biotic: E. cichoracearum (+)	4	5842	790		0.89		396	6545	4	Biotic: E. cichoracearum (-)
Biotic: E. orontii (+)	24	5030	308		0.37		1547	13642	24	Biotic: E. orontii (-)
Biotic: F. occidentalis (+)	0			n/a	n/a	n/a			0	Biotic: F. occidentalis (-)
Biotic: M. persicae (+)	0			n/a	n/a	n/a			0	Biotic: M. persicae (-)
Biotic: M. persicae (+)	3	375	42		0.91		65	412	3	Biotic: M. persicae (-)
Biotic: mycorrhiza (+)	2	89	33		1.62		2	55	2	Biotic: mycorrhiza (-)
Biotic: nematode (+)	1	6371	0		1.91		0	3334	1	Biotic: nematode (-)
Biotic: P. infestans (+)	6	4139	1577		0.33		4476	12479	6	Biotic: P. infestans (-)

Fig. 15 Example of output from the Genevestigator Response Viewer Tool, which displays the relative levels of expression of a gene in response to a range of biotic and abiotic stresses. The number of chips (microarrays) contributing to each result is shown. Data derived from a large number of chips are likely to be the most reliable. Results are displayed as horizontal bars, with standard errors, and as numerical ratios in colour-coded rectangles. A ratio of greater than 1, shown on a red background, indicates that the gene is more strongly expressed in the treatment (e.g. pathogen) than in the control. A ratio of less than 1, on a green background, indicates less expression in the treatment than the control. n/a means no data available for this gene in response to this treatment. Genevestigator URL: https://www.genevestigator.com/gv/index.jsp.

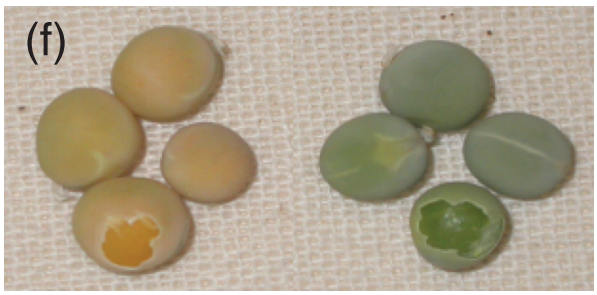


Os09g36110	subtilisin N-terminal region
Os09g36120	expressed protein
Os09g36130	expressed protein
Os09g36140	RNA recognition motif 2 family protein
Os09g36150	retrotransposon protein
Os09g36160	expressed protein
Os09g36170	hypothetical protein
Os09g36180	Glycosyltransferase family 8 protein
Os09g36190	Glycosyltransferase QUASIMODO1
Os09g36200	**senescence-inducible chloroplast stay-green protein**
Os09g36210	harpin-induced protein 1
Os09g36220	Two-component response regulator-like PRR95
Os09g36230	hypothetical protein
Os09g36240	deoyribopyrimidine photolyase
Os09g36250	ODORANT1 protein
Os09g36260	hypothetical protein
Os09g36270	pantothenate kinase 2
Os09g36280	glycosyl hydrolases family 17
Os09g36290	Ser/Thr protein phosphatase

(c)



Arabidopsis RNAi knockout of *At4g22920*



Mendel's *I/i* (yellow/green peas)

Fig. 16 Map-based isolation of a senescence gene. (a) The recessive *stay-green* mutation (*y*) introgressed from *Festuca* into *Lolium* is located by genomic *in situ* hybridization as a pair of terminal chromosomal segments. (b) Using molecular markers *Y* is mapped to a sector of *Lolium/Festuca* chromosome 5 which was shown to be syntenic with a region of rice (*Oryza sativa*) chromosome 9. (c) Several studies had located *sgr*, a major QTL for leaf senescence, in this region of rice 9. Fine mapping in *Lolium/Festuca* using common rice markers narrowed down the number of candidate genes to *c*. 30 on a single rice bacterial artificial chromosome (BAC). (d) The most likely candidate sequence, Os09g36200, is homologous with *Arabidopsis thaliana* At4g22920, which shows very clearly the expected senescence-associated expression pattern. (e) Knocking out At4g2290 in *A. thaliana* gives a *stay-green* phenotype with all the biochemical features of the *Festuca y* phenotype. (f) *SGR* is highly conserved across plant species, and further comparative mapping confirmed that allelic variation in it is responsible for the phenotypes of Mendel's green and yellow pea cotyledons. Figure based on Ougham *et al*. (2008).

information about new software is added on a frequent basis; approx. 480 programs were included as of November 2008. Although some of these programs are specific to (for example) human or farm animal pedigree data, many are relevant to plant scientists, including tools for QTL analysis, genetic association analysis, haplotype construction and population genetics. In each case a brief description of the package is provided, together with a URL for download, a summary of computer system requirements, and, where available, references.

VI. Bringing it all together: *stay-green* gene case study

1. Introductory remarks

By calling on the powerful armoury of genomics and bioinformatics weapons described in the present review, it is becoming feasible to dig out some of the most recalcitrant genes from some of the most unpromising species. We describe here how a gene, first identified by classical inheritance analysis almost 40 yr ago in a forage grass, was finally isolated by a combination of comparative mapping, informatics and functional genomics. The message is that moving from genome location out to the trait can be successful where the phenotype-to-genotype strategy has been unproductive.

2. Chasing *stay-green*, a leaf senescence gene

Thomas & Stoddart (1975) described a mutant of *Festuca pratensis* with a defect in senescence that prevents green tissues from turning yellow. Although senescing leaves remain green, Calvin cycle enzymes such as Rubisco (Roca *et al*., 2004) and other enzymes, for example those of chlorophyll biosynthesis (Thomas *et al*., 2002), are degraded normally, and there is no significant extension in photosynthetic capacity (Hauck *et al*., 1997). The metabolic lesion is located in the pathway of chlorophyll catabolism at the step that unpacks pigment-protein complexes and opens the macrocycle of phaeophorbide *a* (Vicentini *et al*., 1995; Roca *et al*., 2004; Armstead *et al*., 2006). Directly associated with retention of chlorophyll and catabolic intermediates upstream of the blockage is an extended stability of pigment-binding photosynthetic proteins (Davies *et al*., 1990). The cell biology of the protein pointed to a central role in dismantling thylakoid membrane structure during senescence (Thomas, 1997; Thomas *et al*., 2002). Several attempts were made to clone the gene by exploiting presumed differences in expression between mutant and wild type (Thomas *et al*., 1992, 1997, 2002) but the outcomes were inconclusive. A combination of the development of mapping resources in *Festuca* and *Lolium* with the advent of accessible genomics resources and comparative genetics information from the fully sequenced model species rice and *A. thaliana* enabled a successful positional cloning strategy to be adopted.

3. Introgression mapping in *Lolium–Festuca*

Conventional inheritance studies in *F. pratensis* established that *stay-green* (*y*) represents a single recessive Mendelian locus (Thomas, 1987). The mutant gene has been introgressed into a number of *Lolium* species, including *Lollium temulentum* and *L. perenne*, via an initial cross between *F. pratensis* and *L. multiflorum* (Thomas *et al*., 1994, 1999; Armstead *et al*., 2006). Interspecific and intergeneric introgression within the *Lolium–Festuca* complex represents a powerful tool for genetic analysis in large-genome grasses (see Section IV 5). Homeologous recombination between *F. pratensis* and *L. perenne* chromosomes occurs at practically homologous frequencies, and high levels of polymorphism and divergence in repetitive sequences allow introgressed alien segments to be tracked using genomic ISH and molecular markers (King *et al*., 2007b; Fig. 16a). Transferring the *Festuca stay-green* mutation into *L. perenne* made the *Lolium–Festuca* introgression series resource (Section IV 5) available for the mapping and, ultimately, isolation of the corresponding gene.

4. Mapping *stay-green* in *Lolium* and reading across into rice

A *Lolium/Festuca* mapping population was created, segregating for *y* and comprising 100 individuals (Moore *et al*., 2005). Test backcrosses with *Lolium* homozygous for the mutation were screened for the *stay-green* trait. Genomic *in situ* hybridisation analysis identified an individual in the mapping population carrying *y* and a single small *F. pratensis* introgression (Fig. 16a) and, by screening this individual and the parental genotypes, 22 of 266 AFLP primer pairs were found to generate 28 *F. pratensis*-specific polymorphisms. This study enabled an initial map of the introgressed segment carrying *y* to be constructed. Subsequently Armstead *et al*. (2006) applied anchored comparative mapping markers to this introgression population and showed that the *y* locus is associated with *L. perenne* chromosome 5 (C5). The C5 location carrying *y* is known to be syntenic with a region of rice C9 (Jones *et al*., 2002; Alm *et al*., 2003; Fig. 16b). Significantly, three independent mapping studies had identified a QTL for variation in greenness during leaf senescence associated with rice C9 (Ishimaru *et al*., 2001; Cha *et al*., 2002, Abdelkhalik *et al*., 2005). Armstead *et al*. (2006) confirmed that *F. pratensis*-derived and rice *stay-green* phenotypes were determined from syntenically equivalent genomic regions by using rice sequences flanking the position of the C9 greenness QTL. An extended *Lolium–Festuca* introgression mapping family consisting of 1627 individuals was screened and recombination within 10 cM of *y* was identified in 60 genotypes. Fine mapping in these 60 individuals followed by read-across into rice narrowed down the syntenic region to *c*. 200 kb of C9, carrying *c*. 30 gene models.

5. Functional and molecular characterization of *stay-green*

Getting from 30 candidates to a single sequence corresponding to *y* required a combination of bioinformatics and functional testing. Temporal and organ-specific expression patterns of *A. thaliana* homologues of the candidate rice sequences are available from microarray data via the Genevestigator Meta-Analyzer database (see Section V 3). The cell biology phenotype of *y* strongly suggests that the wild-type gene is senescence-inducible and its translation product is targeted to the chloroplast. Of the candidates one, rice sequence LOC_Os09g36200 and its *A. thaliana* homologue At4g22920, clearly had the expected properties (Fig. 16c,d). When At4g22920 was knocked out by RNAi, the resulting plants had the same *stay-green* phenotype as the mutant *F. pratensis* and the derived *Lolium* introgressions (Armstead *et al.*, 2007; Fig. 16e). The molecular basis of the lesion in *F. pratensis* was found to reside in a 4-bp insertion in the first exon, possibly representing the footprint of a departed mobile element. The *stay-green* gene (now called *SGR*) is highly conserved across plant species. Armstead *et al.* (2007) extended their mapping study from *Lolium–Festuca* and rice into pea (*Pisum sativum*) and showed that *SGR* co-locates with *I*, the locus regulating green-yellow cotyledon colour originally described by Gregor Mendel (Fig. 16f).

The availability of the *SGR* gene is opening up new opportunities for detailed insights into the mechanisms regulating chloroplast structure and function. We know that *SGR* encodes a post-transcriptional factor that controls the unpacking of thylakoid membrane, thereby facilitating catabolism of chlorophylls, pigment-binding proteins and lipids (Thomas *et al.*, 2002). The SGR protein of rice has been shown to interact physically with light-harvesting complex of photosystem II (Park *et al.*, 2007) and there is also evidence for direct modulation of the chlorophyll-degrading enzyme phaeophorbide a oxygenase by the SGR protein (Thomas *et al.*, 1996; Roca *et al.*, 2004; Ren *et al.*, 2007). It is beginning to look like the SGR factor is part of a multicomponent machine that systematically dismantles thylakoids in senescing plastids (Armstead *et al.*, 2006). Development of these new concepts in plant cell biology would not have been possible without the power of mapping, comparative genetics and bioinformatics.

VII. The future of mapping

The pace of research in the broad area of genetics is so great that anyone attempting to consult a crystal ball is at risk of being already overtaken by current and imminent developments that continue to make their way at astonishing speed from the cutting edge to the daily routine. Genetic mapping is changing beyond recognition as a result of technical advances such as barcoding, cheap sequencing (with the prospect of generating an entire *Escherichia coli* genome sequence in 15 s, and a human genome in 15 min), *in silico* methods, huge databases and efficient tools to exploit them, high-throughput trait analysis (phenomics) and systems biology approaches. Genetic mapping finds itself rather in the same circumstances as geographical mapping, where ever more sensitive remote sensing equipment on ever more cost-effective orbiting or airborne platforms is building minutely detailed pictures of the planet and its resources that outstrip our understanding of what they mean. In the remote sensing business, making the observations on the ground enabling satellite and aircraft images to be interpreted is known as establishing 'ground truth'. The future of mapping in the genetic sense is likely to be focused as much on establishing the ground truth of genomic structure–function predictions from sequence information, epigenetic features, regulatory networks and phenomics as on the physical act of making the molecular map itself. In turn these technologies and insights will have the increasingly urgent task of meeting humanity's need for food, feed, fuel and fibre in a world under pressure.

Acknowledgements

HT acknowledges the award of a Leverhulme Emeritus Fellowship. HO's research is partially supported by a UK Biotechnology and Biological Sciences Research Council Strategic Programme Grant. The authors also thank three anonymous reviewers for their critical reading of the manuscript and their valuable suggestions for its improvement.

References

Abdelkhalik AF, Shishido R, Nomura K, Ikehashi H. 2005. QTL-based analysis of leaf senescence in an indica/japonica hybrid in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **110**: 1226–1235.

Ai-hua L, Jan-bo W. 2006. Genomic evaluation of *Brassica* allopolyploids revealed by ISSR marker. *Genetic Resources and Crop Evaluation* **53**: 603–611.

Alm V, Fang C, Busso CS, Devos KM, Vollan K, Grieg Z, Rognli OA. 2003. A linkage map of meadow fescue (*Festuca pratensis* Huds) and comparative mapping with other Poaceae species. *Theoretical and Applied Genetics* **108**: 5–40.

Alwala S, Suman A, Arro JA, Veremis JC, Kimbeng CA. 2006. Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. *Crop Science* **46**: 448–455.

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

Armstead I, Donnison I, Aubry S, Harper J, Hörtensteiner S, James C, Mani J, Moffet M, Ougham H, Roberts L *et al.* 2006. From crop to model to crop: identifying the genetic basis of the staygreen mutation in the *Lolium/Festuca* forage and amenity grasses. *New Phytologist* **172**: 592–597.

Armstead I, Donnison I, Aubry S, Harper J, Hörtensteiner S, James C, Mani J, Moffet M, Ougham H, Roberts L *et al.* 2007. Cross-species identification of Mendel's I locus. *Science* **315**: 73.

Armstead IP, Turner LB, Farrell M, Skøt L, Cisneros P, Montoya T, Donnison IS, King IP, Humphreys MO. 2004. Synteny between a major heading-date QTL in perennial ryegrass (*Lolium perenne* L.) and the Hd3 heading-date locus in rice. *Theoretical and Applied Genetics* **108**: 822–828.

Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO *et al.* 2009. QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. *PLoS ONE* **4**: e4318.

Barker R, Herdt RW, Rose B. 1985. *The rice economy of Asia*. Manila, Philippines: International Rice Research Institute.

Bennetzen JL. 2000. Comparative sequence analysis of plant nuclear genes: microcolinearity and its many exceptions. *Plant Cell* **12**: 1021–1029.

Bertin I, Zhu J-H, Gale MD. 2005. SSCP-SNP in pearl millet: a new marker system for comparative genetics. *Theoretical and Applied Genetics* **110**: 1467–1472.

Borevitz JO, Liang D, Plouffe D, Chang H-S, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J. 2003. Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research* **13**: 513–523.

Bradshaw HD, Stettler RF. 1995. Molecular genetics of growth development in *Populus*. IV. Mapping QTLs with large effects on growth, form and phenology traits in a forest tree. *Genetics* **139**: 963–973.

Bradshaw HD, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF. 1994. Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS and RAPD markers. *Theoretical and Applied Genetics* **89**: 167–178.

Broman KW. 2005. The genomes of recombinant inbred lines. *Genetics* **169**: 1133–1146.

Brown AHD, Lawrence GJ, Jenkin M, Douglass J, Gregory E. 1989. Linkage drag in backcross breeding in barley. *Journal of Heredity* **80**: 234–239.

Burr B, Burr FA. 1991. Recombinant inbred lines for molecular mapping in maize. Theoretical and practical considerations. *Trends in Genetics* **7**: 55–60.

Cha KW, Lee YJ, Koh HJ, Lee BM, Nam YW, Paek NC. 2002. Isolation, characterization, and mapping of the stay green mutant in rice. *Theoretical and Applied Genetics* **104**: 526–532.

Chen JM, Gustafson JP. 1995. Physical mapping of restriction fragment length polymorphisms (RFLPs) in homoeologous group 7 chromosomes of wheat by *in situ* hybridization. *Journal of Heredity* **75**: 225–233.

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S *et al.* 2002. An integrated physical and genetic map of the rice genome. *Plant Cell* **14**: 537–545.

Davies TGE, Thomas H, Thomas BJ, Rogers LJ. 1990. Leaf senescence in a non yellowing mutant of *Festuca pratensis*: metabolism of cytochrome *f*. *Plant Physiology* **93**: 588–595.

Dear PH, Cook PR. 1989. Happy mapping – a proposal for linkage mapping the human genome. *Nucleic Acids Research* **17**: 6795–6807.

Dear PH, Cook PR. 1993. Happy mapping: linkage mapping using a physical analogue of meiosis. *Nucleic Acids Research* **21**: 13–20.

Devos KM, Gale MD. 2000. Genome relationships: the grass model in current research. *Plant Cell* **12**: 637–646.

Edwards D, Coghill J, Batley J, Holdsworth M, Edwards KJ. 2002. Amplification and detection of transposon insertion flanking sequences using fluorescent MuAFLP. *Biotechniques* **32**: 1090–1097.

Erayman M, Sandhu D, Sidhu D, Dilbirligi M, Baenziger PS, Gill KS. 2004. Demarcating the gene-rich regions of the wheat genome. *Nucleic Acids Res* **32**: 3546–3565.

Fischer SG, Lerman LS. 1980. Separation of random fragments of DNA according to properties of their sequences. *Proceedings of the National Academy of Science, USA* **77**: 4420–4424.

Flint-Garcia S, Thornsberry JM, Buckler ES. 2003. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* **54**: 357–74.

Frey M, Stettner C, Gierl A. 1998. A general method of gene isolation in tagging approaches: amplification of insertion mutagenised sites (AIMS). *Plant Journal* **213**: 717–721.

Fronmer M, McDonald LE, Miller DS, Collis M, Watt F, Grigg GW, Molloy PL, Paul CL. 1992. A genomic sequencing protocol that yields a positive display in 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences, USA* **89**: 1827–1831.

Gill KS, Gill BS, Endo TR, Boyko EV. 1996a. Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. *Genetics* **143**: 1001–1012.

Gill KS, Gill BS, Endo TR, Taylor T. 1996b. Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. *Genetics* **144**: 1883–1891.

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

Gómez-Gómez L, Felix G, Boller T. 1999. A single locus determines sensitivity to bacterial flagellin in *Arabidopsis thaliana*. *Plant Journal* **18**: 277–284.

Gonzales MD, Archuleta E, Farmer A, Gajendran K, Grant D, Shoemaker R, Beavis WD, Waugh ME. 2005. The Legume Information System (LIS): an integrated information resource for comparative legume biology. *Nucleic Acids Research* **33** (Database issue): D660–D665.

Gupta PK, Rustgi S, Mir RR. 2008. Array-based high-throughput DNA markers for crop improvement. *Heredity* **101**: 5–18.

Hanson WD. 1959. Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing and selfing. *Genetics* **44**: 833–837.

Hauck B, Gay AP, Macduff J, Griffiths CM, Thomas H. 1997. Leaf senescence in a nonyellowing mutant of *Festuca pratensis*: implications of the stay-green mutation for photosynthesis, growth and nitrogen nutrition. *Plant, Cell & Environment* **20**: 1007–1018.

Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences, USA* **93**: 9821–9826.

Heuer S, Lu X, Chin JH, Tanaka JP, Kanamori H, Matsumoto T, De Leon T, Ulat VJ, Ismail AM, Yano M *et al.* 2009. Comparative sequence analyses of the major quantitative trait locus phosphorus uptake 1 (Pup1) reveal a complex genetic structure. *Plant Biotechnology Journal* **7**: doi: 10.1111/j.1467-7652.2009.00415.x.

Hubbard TJP, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T *et al.* 2007. Ensembl 2007. *Nucleic Acids Research* **35** (Database issue): D610–D617.

Ingvarsson PK. 2005. Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). *Genetics* **169**: 945–953.

Ishimaru K, Yano M, Aoki N, Ono K, Hirose T, Lin SY, Monna L, Sasaki T, Ohsugi R. 2001. Toward the mapping of physiological and agronomic characters on a rice function map: QTL analysis and comparison between QTLs and expressed sequence tags. *Theoretical and Applied Genetics* **102**: 793–800.

Jaccoud D, Peng K, Feinstein D, Kilian A. 2001. Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research* **29**: 25–31.

James N, Graham N, Clements D, Schildknecht B, May S. 2007. AtEnsEMBL. *Methods in Molecular Biology* **406**: 213–27.

Jansen RC. 2003. Quantitative trait loci in inbred lines. In: Balding DJ, Bishop M, Cannings C, eds. *Handbook of statistical genetics*. Chichester, UK: Wiley, 445–476.

Jeffreys AJ, Wison V, Thein SL. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**: 67–73.

Jones ES, Mahoney NL, Hayward MD, Armstead IP, Jones JG, Humphreys MO, King IP, Kishida T, Yamada T, Balfourier F *et al.* 2002. An enhanced molecular marker-based map of perennial ryegrass (*Lolium perenne* L.) reveals comparative relationships with other Poaceae species. *Genome* **45**: 282–295.

Jones RN, Ougham H, Thomas H. 1997. Markers and mapping: we are all geneticists now. *New Phytologist* **137**: 165–177.

Jorde LB. 1995. Linkage disequilibrium as a gene-mapping tool. *American Journal of Human Genetics* **56**: 11–14.

Kalendar R, Grob T, Regina M, Suoniemi A, Schulman A. 1999. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theoretical and Applied Genetics* **98**: 704–711.

Kamisugi Y, von Stackelberg M, Lang D, Care M, Reski R, Rensing SA, Cuming AC. 2008. A sequence-anchored genetic linkage map for the moss, *Physcomitrella patens. Plant Journal* **56**: 855–866.

Karp A, Seberg O, Buiatti M. 1996. Molecular techniques in the assessment of botanical diversity. *Annals of Botany* **78**: 143–149.

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M. 2007. Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* **175**: 891–905.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics* **39**: 1151–1155.

King J, Armstead IP, Donnison I, Roberts LA, Harper J, Skot K, Eldorough K, King IP. 2007a. Comparative analyses between *Lolium/Festuca* introgression lines and rice reveal the major fraction of functionally annotated gene models is located in recombination-poor/very recombination-poor regions of the genome. *Genetics* **177**: 597–606.

King J, Armstead IP, Donnison IS, Harper JA, Roberts LA, Thomas H, Ougham HJ, Thomas A, Huang L, King IP. 2007b. Introgression mapping in the grasses. *Chromosome Research* **15**: 105–113.

King J, Thorogood D, Edwards KJ, Armstead IP, Roberts L, Skøt K, Hanley Z, King IP. 2008. Development of a genomic microsatellite library in perennial ryegrass (*Lolium perenne*) and its use in trait mapping. *Annals of Botany* **101**: 845–853.

Koornneef M, Alonso-Blanco C, Vreugdenhil D. 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology* **55**: 141–172.

Künzel G, Korzun L, Meister A. 2000. Cytologically integrated physical restriction length polymorphism maps of the barley genome based on translocation breakpoints. *Genetics* **154**: 397–412.

Kynast RG, Okagaki RJ, Galatowitsch MW, Granath SR, Jacobs MS, Stec AO, Rines HW, Phillips R. 2004. Dissecting the maize genome by using chromosome addition and radiation hybrid lines. *Proceedings of the National Academy of Sciences, USA* **101**: 9921–9266.

Lander E, Green P, Abrahamson J, Barlow A, Daly M, Lincoln S, Newburg L. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.

Lessa EP, Applebaum G. 1993. Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology* **2**: 119–129.

Li G, Quiros CF. 2001. Sequence-related amplified polymorphism (SRAP), a new marker system base on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. *Theoretical and Applied Genetics* **103**: 455–461.

Li GM, Yang B, Quiros CF. 2003. Gene for gene alignment between *Brassica* and *Arabidopsis* genomes by direct transcriptome mapping. *Theoretical and Applied Genetics* **107**: 168–180.

Li Z, Pan J, Guan Y, Tao Q, He H, Si L, Cai R. 2007. Development and fine mapping of three co-dominant SCAR markers linked to the M/m gene in the cucumber plant (*Cucumis sativus* L.). *Theoretical and Applied Genetics* **117**: 1253–1260.

Lichten M, Goldman ASH. 1995. Meiotic recombination hotspots. *Annual Review of Genetics*. **29**: 423–444.

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

Liu ZH, Anderson JA, Hu J, Friesen TL, Rasmussen JB, Faris JD. 2005. A wheat intervarietal genetic linkage map based on microsatellite and target region amplified polymorphism markers and its utility for detecting quantitative trait loci. *Theoretical and Applied Genetics* **111**: 782–794.

Lörz H, Wenzel G. eds. 2005. *Molecular marker systems in plant breeding and crop improvement*. Biotechnology in agriculture and forestry, Vol. 55 (series eds T. Nagata, H. Lorz and J. M. Widholm). Berlin: Springer-Verlag.

Luquez V, Hall D, Albrectsen BR, Karlsson J, Ingvarsson P, Jansson S. 2008. Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. *Tree Genetics and Genomes* **4**: 279–292.

Ma ZQ, Roder M, Sörells ME. 1966. Frequencies and sequence characteristics of di-, tri-, and tetra nucleotide microsatellites in wheat. *Genome* **39**: 123–130.

Masoudi-Nejad A, Nasuda S, Bihoreau M-T, Waugh R, Endo TR. 2005. An alternative to radiation hybrid mapping for large-scale genome analysis in barley. *Molecular Genetics and Genomics* **274**: 589–594.

Mather KM. 1943. Polygenic inheritance and natural selection. *Biological Review* **18**: 32–64.

Matsuyama T, Kimura MT, Koike K, Abe T, Nakano T, Asami T, Ebisuzaki T, Held WA, Yoshida S, Nagase H. 2003. Global methylation screening in the *Arabidopsis thaliana* and *Mus musculus* genome: applications of virtual image restriction landmark genomic scanning (Vi-RLGS). *Nucleic Acids Research* **31**: 4490–4496.

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD. 1988. Molecular mapping of rice chromosomes. *Theoretical and Applied Genetics* **76**: 815–829.

Meloni M, Perini D, Filigheddu R, Binelli G. 2006. Genetic variation in five Mediterranean populations of *Juniperus phoenicea* as revealed by Inter-simple sequence repeat (ISSR) markers. *Annals of Botany* **97**: 299–304.

Meudt HM, Clarke AC. 2007. Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science* **12**: 106–117.

Michelmore RW, Paran I, Kesseli RV. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences, USA* **88**: 9828–9832.

Moore BJ, Donnison I, Harper JA, Armstead I, King J, Thomas H, Jones RN, Jones TH, Thomas HM, Morgan WG *et al.* 2005. Molecular tagging of a senescence gene by introgression mapping of a mutant stay-green locus from *Festuca pratensis*. *New Phytologist* **165**: 801–806.

Moore G, Devos KM, Wang Z, Gale MD. 1995. Grasses line up and form a circle. *Current Biology* **5**: 737–739.

Mozo T, Dewar K, Dunn P, Ecker JR, Fischer S, Kloska S, Lehrach H, Marra M, Martienssen R, Meier-Ewert S *et al.* 1999. A complete BAC-based physical map of the *Arabidopsis thaliana* genome. *Nature Genetics* **22**: 271–275.

Nei M. 1972. Genetic distance between populations. *American Naturalist* **106**: 283–292.

van Ooijen JW. 2006. *Joinmap®: Software for the calculation of genetic linkage maps in experimental populations, Version 4*. Wageningen, The Netherlands: Kyazma BV.

Ougham H, Hörtensteiner S, Armstead I, Donnison I, King I, Thomas H, Mur L. 2008. The control of chlorophyll catabolism and the status of yellowing as a biomarker of leaf senescence. *Plant Biology* **10** (Supplement 1): 4–14.

Park KC, Kim NH, Cho YS, Kang KH, Lee JK, Kim N-S. 2003. Genetic variations of AA genome *Oryza* species measured by MITE-AFLP. *Theoretical and Applied Genetics* **197**: 203–209.

Park S-Y, Yu J-W, Park J-S, Li J, Yoo S-C, Lee N-Y, Lee S-K, Jeong S-W, Seo HS, Koh H-J *et al.* 2007. The senescence-induced staygreen protein regulates chlorophyll degradation. *Plant Cell* **19**: 1649–1664.

Paterson AH, Lander ES, Had JD, Paterson S, Lincoln SE, Tanksley SD. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**: 721–726.

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W *et al.* 2008. A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* **322**: 101–104.

Pedersen C, Giese H, Linde-Laursen I. 1995. Towards an integration of the physical and genetic chromosome maps of barley by *in situ* hybridisation. *Hereditas* **123**: 77–88.

Pivorienė O, Pašakinskienė I. 2007. Genetic differences between annual and perennial ryegrass revealed by ISSR markers and their sequence characteristics. *Zemdirbyste-Agriculture* **94**: 97–204.

Pivorienė O, Pašakinskienė I, Brazauskas G, Lideikytė L, Jensen LB, Lübberstedt T. 2008. Inter-simple sequence repeat (ISSR) loci mapping in the genome of perennial ryegrass. *Biologija* **54**: 17–21.

Powell W, Machray GC, Provan J. 1996. Polymorphism as revealed by simple sequence repeats. *Trends in Plant Science* **1**: 215–222.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Rae AM, Pinel MPC, Bastien C, Sabatti M, Street NR, Tucker J, Dixon C, Marron N, Dillen SY, Taylor G. 2008. QTL for yield in bioenergy *Populus*: identifying G×E interactions from growth at three contrasting sites. *Tree Genetics and Genomes* **4**: 97–112.

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

Reddy MP, Sarla N, Siddiq EA. 2002. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. *Euphytica* **128**: 9–17.

Ren G, An K, Liao Y, Zhou X, Cao Y, Zhao H, Ge X, Kua B. 2007. Identification of a novel chloroplast protein AtNYE1 regulating chlorophyll degradation during leaf senescence in Arabidopsis. *Plant Physiology* **144**: 1429–1441.

Rensing SA, Ick J, Fawcett JA, Lang D, Zimmer A, Van de Peer Y, Reski R. 2007. An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens. BMC Evolutionary Biology* 7: 130.

Rensing SA, Lang D, Zimmer AD, Astrid Terry A, Salamov A. 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**: 64–69.

Reski R, Faust M, Wang X-H, Wehe M, Abel WO. 1994. Genome analysis of the moss *Physcomitrella patens* (Hedw.) B.S.G. *Molecular and General Genetics* **244**: 352–359.

Riera-Lizarazu O, Vales MI, Kianian SF. 2008. Radiation hybrid (RH) and happy mapping in plants. *Cytogenetic and Genome Research* **120**: 233–240.

Roberts RJ, Vincze T, Posfai J, Macelis D. 2003. REBASE: restriction enzymes and methyltransferases. *Nucleic Acid Research* **31**: 418–420.

Roca M, James C, Pružinská A, Hörtensteiner S, Thomas H, Ougham H. 2004. Analysis of the chlorophyll catabolism pathway in leaves of an introgression senescence mutant of *Lolium temulentum*. *Phytochemistry* **65**: 1231–1238.

Roy JK, Balyan HS, Prasad M, Gupta PK. 2002. Use of SAMPL for a study of DNA polymorphism, genetic diversity and possible gene tagging in bread wheat. *Theoretical and Applied Genetics* **104**: 465–472.

Samuri M, Nakamura S, Nair SK, Takeda K, Komatsuda T. 2009. A quantitative trait locus for reduced culm internode length in barley segregates as a Mendelian gene. *Theoretical and Applied Genetics* **118**: 643–652.

Schmalenbach I, Léon J, Pillen K. 2009. Identification and verification of QTLs for agronomic traits using wild barley introgression lines. *Theoretical and Applied Genetics* **118**: 483–497.

Schwieger F, Tebbe C. 1998. A new approach to utilize PCR-Single-Strand-Conformation Polymorphism for 16S rRNA gene-based microbial community analysis. *Applied and Environmental Microbiology* **64**: 4870–4876.

Skøt L, Humphreys MO, Armstead I, Heywood S, Skøt KP, Sanderson R, Thomas ID, Chorlton KH, Sackville Hamilton NR. 2005. An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). *Molecular Breeding* **15**: 233–245.

Shi C, Uzarowska A, Ouzunova M, Landbeck M, Wenzel G, Lübbersted T. 2007. Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint × Flint maize recombinant inbred line population. *BMC Genomics* **8**: 1–16.

Sunnucks P, Rispe C, Simon JC. 2000. SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Molecular Ecology* **9**: 1699–1710.

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB *et al.* 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**: 1141–1160.

Thangavelu M, James AB, Bankier A, Bryan GJ, Dear PH, Waugh R. 2003. HAPPY mapping in a plant genome: reconstruction and analysis of a high-resolution physical map of a 1.9 Mbp region of *Arabidopsis thaliana* chromosome 4. *Plant Biotechnology Journal* **1**: 23–31.

Thomas H. 1987. Sid: a Mendelian locus controlling thylakoid membrane disassembly in senescing leaves of *Festuca pratensis*. *Theoretical and Applied Genetics* **73**: 551–555.

Thomas H. 1997. Chlorophyll: a symptom and a regulator of plastid development. *New Phytologist* **136**: 163–181.

Thomas H, Evans C, Thomas HM, Humphreys MW, Morgan G, Hauck B, Donnison I. 1997. Introgression, tagging and expression of a leaf senescence gene in *Festulolium*. *New Phytologist* **137**: 29–34.

Thomas HM, Morgan WG, Meredith MR, Humphreys MW, Thomas H, Leggett JM 1994. Identification of parental and recombined chromosomes of *Lolium multiflorum* × *Festuca pratensis* by genomic in situ hybridization. *Theoretical and Applied Genetics* **88**: 909–913.

Thomas H, Morgan WG, Thomas AM, Ougham HJ. 1999. Expression of the stay-green character introgressed into *Lolium temulentum* Ceres from a senescence mutant of *Festuca pratensis*. *Theoretical and Applied Genetics* **99**: 92–99.

Thomas H, Ougham H, Canter P, Donnison I. 2002. What stay-green mutants tell us about nitrogen remobilisation in leaf senescence. *Journal of Experimental Botany* **53**: 801–808.

Thomas H, Ougham HJ, Davies TGE. 1992. Leaf senescence in a nonyellowing mutant of *Festuca pratensis*: transcripts and translation products. *Journal of Plant Physiology* **139**: 403–412.

Thomas H, Schellenberg M, Vicentini F, Matile P. 1996. Gregor Mendel's green and yellow pea seeds. *Botanica Acta* **109**: 3–4.

Thomas H, Stoddart JL. 1975. Separation of chlorophyll degradation from other senescence processes in leaves of a mutant genotype of meadow fescue (*Festuca pratensis*). *Plant Physiology* **56**: 438–441.

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

Tyagi S, Kramer FR. 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnology* **14**: 303–308.

Tyrka M, Błaszcyk, L, Chełkowski J, Lind V, Kramer I, Weilepp M, Wiśniewśka H, Ordon F. 2004. Development of the single nucleotide polymorphism marker of the wheat *Lr1* leaf rust resistance gene. *Cellular & Molecular Biology Letters* **9**: 879–889.

Umehara M, Sueyoshi T, Shimomura K, Iwai M, Shigyo M, Hirashima K, Nakahara T. 2006. Interspecific hybrids between *Allium fistulosum* × *Allium schoenoprasum* reveal carotene-rich phenotype. *Euphytica* **148**: 295–301.

Vicentini F, Hörtensteiner S, Schellenberg M, Thomas H, Matile P. 1995. Chlorophyll breakdown in senescent leaves: identification of the biochemical lesion in a stay-green genotype of *Festuca pratensis* Huds. *New Phytologist* **129**: 247–252.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M *et al.* 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407–4414.

Wardrop J, Fuller J, Powell W, Machray G. 2004. Exploiting plant somatic

radiation hybrids for physical mapping of expressed sequence tags. *Theoretical and Applied Genetics* **108**: 343–348.

Wardrop J, Snape J, Powell W, Machray G. 2002. Constructing plant radiation hybrid panels. *The Plant Journal* **31**: 223–228.

Waugh R, MacLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W. 1997. Genetic distribution of the Bare-1-like retro transposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Molecular and General Genetics* **253**: 687–694.

Wei J-Z, Wang RR-C. 1995. Genome- and species-specific markers and genome relationships of diploid perennial species in Triticeae based on RAPD analysis. *Genome* **38**: 1230–1236.

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V *et al.* 2006. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits, *BMC Genomics* 7: 206.

Williams JGK, Kubelic AR, Livak KJ, Rafalsky JA, Tingey SV. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**: 6531–6335.

Wittenberg AHJ, van der Lee T, Cayla C, Kilian A, Visser RGF, Schouten HJ. 2005. Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. *Molecular Genetics and Genomics* **274**: 30–39.

Xu Y, Zhu L, Xiao J, Huang N, McCouch SR. 1997. Chromosomal regions associated with segregation distortion of molecular markers in F_2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). *Molecular and General Genetics* **253**: 535–545.

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

Zietkiewicz E, Rafalski A, Labuda D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20**: 176–183.

Zimmermann P, Hennig L, Gruissem W. 2005. Gene-expression analysis and network discovery using Genevestigator. *Trends in Plant Science* **10**: 407–409.