Plant Biotechnology Journal (2003) **1**, pp. 23–31

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

Madan Thangavelu[1], Allan B. James[2], Alan Bankier[1], Glenn J. Bryan[2], Paul H. Dear[1] and Robbie Waugh[2]

[1]*MRC Laboratory of Molecular Biology, Protein and Nucleic Acid Chemistry Division, Cambridge CB2 2QH, UK.* [2]*Unit of Genomics, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK*

Received 20 April 2002;
revised 1 June 2002;
accepted 11 June 2002.
Correspondence (fax +44 1223 412178;
e-mail phd@mrc-lmb.cam.ac.uk)

Keywords: plant genome mapping, HAPPY mapping, PEP amplification, DNA markers

Summary

HAPPY mapping is an *in vitro* approach for defining the order and spacing of DNA markers directly on native genomic DNA. This cloning-free technique is based on analysing the segregation of markers amplified from high molecular weight genomic DNA which has been broken randomly and 'segregated' by limiting dilution into subhaploid samples. It is a uniquely versatile tool, allowing for the construction of genome maps with flexible ranges and resolutions. Moreover, it is applicable to plant genomes, for which many of the techniques pioneered in animal genomes are inapplicable or inappropriate. We report here its demonstration in a plant genome by reconstructing the physical map of a 1.9 Mbp region around the FCA locus of *Arabidopsis thaliana.* The resulting map, spanning around 10% of chromosome 4, is in excellent agreement with the DNA sequence and has a mean marker spacing of 16 kbp. We argue that HAPPY maps of any required resolution can be made immediately and with relatively little effort for most plant species and, furthermore, that such maps can greatly aid the construction of regional or genome-wide physical maps.

Introduction

Genetic linkage maps have been constructed for many crop and plant species, using a combination of polymorphic phenotypes and phenotypically silent sequence polymorphisms (Ramsay *et al*., 2000; Tanksley, 1993). Such maps have served as a framework within which genes controlling monogenic and polygenic traits can be located and subsequently isolated (Buschges *et al*., 1997; Cai *et al*., 1997), as a basis for comparative studies (Devos and Gale, 2000; Eckardt, 2001), a detailed analysis of plant genetic diversity (Russell *et al*., 2000) and for crop improvement (Swanston *et al*., 1999).

Genetic mapping, however, has two inherent limitations. First, its dependence on polymorphic loci prevents the mapping of chromosomal regions that are identical by descent (i.e. not polymorphic). Second, its resolution is severely limited, since the probability of recombination being observed over ever-shorter distances becomes vanishingly small. It is further confounded by recombination 'hot-spots' and segments of suppressed recombination, yielding maps with poor correlation between genetic and physical distances (Künzel *et al*., 2000).

The advent of physical mapping, based on the analysis of large insert clones such as YACs or BACs, has circumvented many of these drawbacks. 'Contigs' of overlapping cloned fragments are assembled to span a region or genome of interest, frequently as a prelude to genome sequencing. Thus, a physical map of the model plant *Arabidopsis thaliana* was constructed (Mozo *et al*., 1999) to enhance the genome sequencing effort, and physical maps are currently being made of many crop species (Budiman *et al*., 2000; Klein *et al*., 2000). Such maps serve both as archives of genomic information and as substrates for sequencing. However, physical mapping is limited by the cloning process on which it relies: regions recalcitrant to cloning lead to uncloseable gaps, and rearranged or co-ligated fragments, or large repeated regions, can lead to distortions (Konfortov *et al*., 2000). Moreover, high levels of redundancy are required to ensure complete coverage.

For these reasons, it has been proposed (Dear *et al*., 1998) that physical maps are most effective if they are built over an

independently constructed 'scaffold' of sequence-tagged sites (STSs), both to guide their completion and to safeguard against errors. Radiation Hybrid (RH) mapping has been successful, particularly in human and animal genomes, as a means of producing such scaffolds (Cox *et al*., 1990). RH mapping relies upon observing the segregation of STS markers amongst a panel of hybrid cells which contain a random subset of fragments from the genome of an irradiated 'donor' and the complete genome of a host cell line. RH panels require considerable effort to produce (even from mammalian genomes), and complications arise because the biological activity of donor fragments in the host cell lines biases their segregation. Furthermore, the presence of a host genome in the hybrids precludes the ability to map economical 'multiplex' generic markers such as AFLPs and RAPDs. Many other approaches which have been central to the analysis of animal genomes remain impractical or unattractive in the context of plants.

It is therefore apparent that there is still a need for an economical and easy means of accurately mapping plant genomes. HAPPY mapping is an *in vitro* technique that circumvents the problems associated with *in vivo* methods such as genetic, physical and RH mapping. We argue here that it can be applied to almost any plant genome and that, depending upon the density of markers mapped, it can serve as an aid to marker-based selection, comparative genomics, gene isolation and physical mapping.

Principles of HAPPY mapping

The theory and principles of HAPPY mapping, initially proposed by Dear and Cook in 1989 (Dear and Cook, 1989), have been discussed more recently by Dear and co-workers (Dear, 1997; Konfortov *et al*., 2000; Piper *et al*., 1998). The rationale behind the method is outlined in Figure 1a. Briefly, genomic DNA is broken randomly by shearing or irradiation

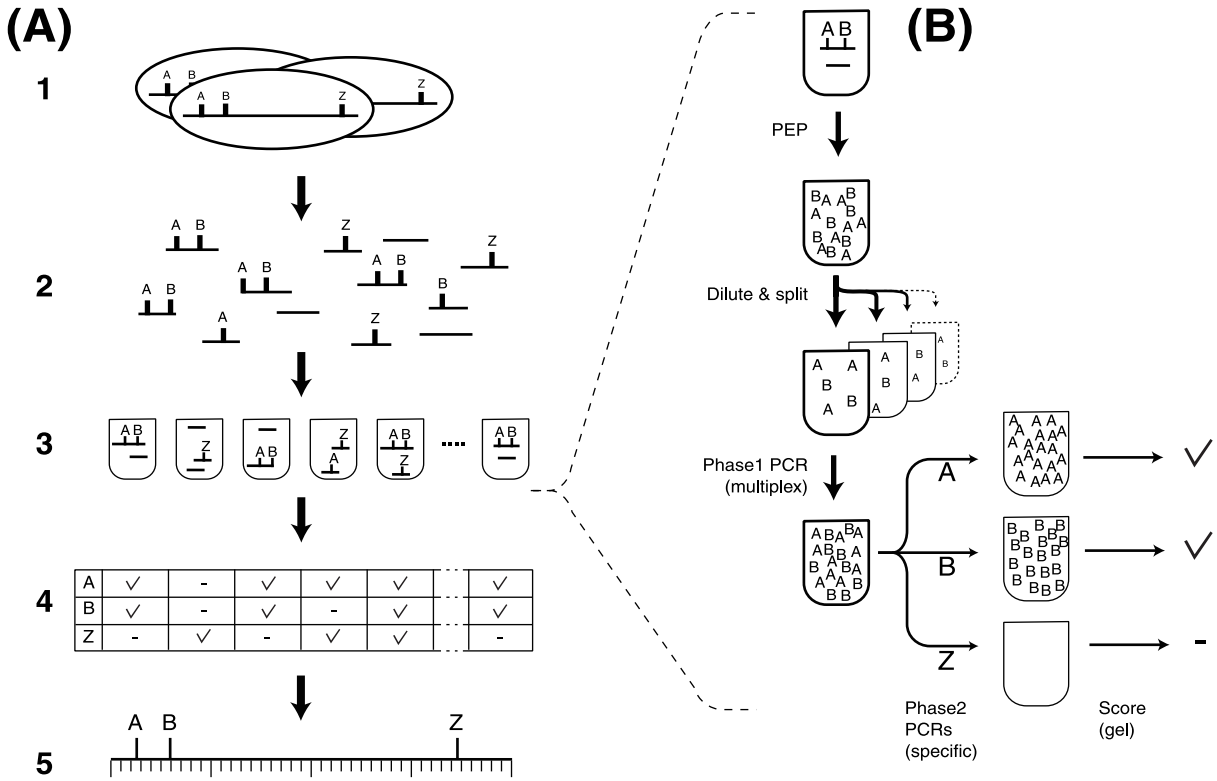

Figure 1 Principle of HAPPY mapping and marker typing. (A) Overview. DNA carrying STS markers (A, B, Z) is extracted from cells (1) and broken randomly to give a pool of fragments (2). These are dispensed at limiting dilution into a series of aliquots – the mapping panel (3). The panel is screened by PCR to produce a table (4) showing the marker content of each aliquot. Linked markers (A,B) are found to co-segregate; remote markers (B,Z) do not. Co-segregation frequencies reflect marker-to-marker distances, allowing a map (5) to be computed. (B) Expanded view of marker screening using a three-step PCR. The protocol is illustrated for one aliquot of the mapping panel. All DNA in the sample is first preamplified > 100-fold using PEP. This material is diluted and split into subfractions for multiple rounds of screening. One subfraction is amplified in a multiplex PCR for many markers (Phase 1). The products of this reaction are then diluted and split again, and screened for individual markers (A, B, Z) in turn, using hemi-nested primers (Phase 2). Results are scored on gels, determining the marker content of the aliquot.

(and may also be size-fractionated) and then sampled at a limiting dilution into multiple aliquots to generate a mapping panel. Each aliquot is apportioned such that it contains less than one genome's worth of DNA fragments and can therefore be considered an *in vitro* analogue of a radiation hybrid cell, containing a random subset of genomic fragments.

The mapping panel is screened by PCR (Figure 1b) to determine which markers are present in each member. If two markers lie physically very close together in the genome, they will seldom suffer an intervening break and will therefore tend to co-segregate. Conversely, remote markers will invariably be broken apart, and hence segregate independently. Therefore, co-segregation frequencies reflect the physical proximity between any pair of markers; analysis of all pairwise distance estimates allows the map to be calculated.

It will be clear from the above discussion that the useful resolution and range of a HAPPY mapping panel depends upon the mean size of the DNA fragments, which is under experimental control. Crudely speaking, a panel will be able to resolve the order of markers over distances as small as ≈ 0.1 times the mean fragment size; more closely spaced markers will co-segregate almost completely and hence cannot be resolved. Conversely, it will detect the linkage between markers which are separated by up to ≈ 0.8 times the mean fragment size; more widely spaced markers do not significantly co-segregate. Therefore, by controlling the degree of fragmentation of the DNA, panels may be created of any desired range or resolution.

HAPPY mapping is analogous both to genetic linkage analysis (which measures the frequency of meiotic recombination between markers) and to RH mapping (which measures the frequency of radiation-induced breaks in chromosomes). HAPPY mapping, however, possesses all the advantages of RH mapping (ability to map monomorphic markers; flexible resolution), but none of the drawbacks – it is immune to the biases caused by the biological activity of the DNA fragments, to cloning artefacts, or to the effects of chromosome structure. Constructing and screening a HAPPY mapping panel is relatively simple, and only small quantities of DNA are needed (Dear, 1997). In contrast to genetic maps, HAPPY maps are truly metric, showing a strong correlation between map distance and physical distance.

HAPPY mapping's utility has been demonstrated in human (Dear *et al*., 1998; Walter *et al*., 1993) and in a range of animal genomes (Konfortov *et al*., 2000; Piper *et al*., 1998). In this paper, we report its evaluation for high throughput, high resolution mapping in plants. We describe the mapping of a set of markers from a 1.9 Mbp region surrounding the *Arabidopsis thaliana* FCA locus (Bevan *et al*., 1998). We conclude that HAPPY mapping is a uniquely valuable addition to currently available plant genome mapping tools.

Results and discussion

The FCA region of *A. thaliana* was chosen as a test-bed for demonstrating HAPPY mapping for three reasons: it represented a small but significant proportion of the genome (1.9 Mbp, or ≈ 1.6% of the genome), providing a stringent test for the correct regional assignment of markers; it had been physically mapped with YACs and BACs; and it had been one of the earliest regions to be sequenced and annotated, allowing us to test the accuracy of the HAPPY map down to the base-pair level. The region is also of biological significance (the FCA locus being a gene which controls flowering time), and hence has some of the characteristics of other regions in which high resolution mapping would be important.

Chromosomal distribution of markers used in this study

The locations of the 120 STS markers which were used in this study were selected essentially at random across the 1.9 Mbp genome sequence encompassing FCA, with distances between adjacent markers ranging from < 0.5 kbp to > 45 kbp.

High molecular weight *Arabidopsis* DNA and preparation of mapping panel

DNA prepared in strings as described was routinely in excess of 2000–3000 kbp as determined by PFGE (results not shown). The mapping panel was prepared by irradiation-induced breakage of embedded DNA, followed by the size-selection of fragments from ≈ 450 to 550 kbp. A panel of such fragments would be expected to detect linkage over distances of up to ≈ 400 kbp, and to correctly resolve markers spaced more than ≈ 50 kbp apart.

It should be noted that the pulsed-field gel from which this panel was taken could also have furnished much larger fragments, suitable for detecting linkage over longer distances. In previous studies (Dear *et al*., 1998), we have used such gels to prepare panels with ranges of > 1 Mbp.

Pre-amplification of mapping panel

Primer extension PCR (PEP) was used as the first stage in the marker scoring process, pre-amplifying all sequences in each aliquot of the mapping panel by ≈ 100-fold (Figure 1b). We

had previously found (Konfortov *et al*., 2000; Piper *et al*., 1998) that the PEP of very small amounts of DNA (on the order of 0.1 pg) is more efficient if the reactions are supplemented with ≈ 1–2 pg of 'carrier' DNA. Accordingly, our PEP reactions (each containing ≈ 0.08 pg of *Arabidopsis* DNA) were supplemented with ≈ 1.5 pg of human DNA.

If mapping panels prepared in this manner were to be typed for 'non-specific' markers such as RAPDs or AFLPs, there would be no way to distinguish markers arising from the target genome (*Arabidopsis*) from those arising from the carrier DNA (human). It should be possible to overcome this problem by replacing the human carrier with simple-sequence DNA, although we have not tested this. However, most genomes for which HAPPY mapping would be considered are much larger than that of *Arabidopsis*, and the use of carrier DNA does not appear to be necessary when PEP is performed on correspondingly larger DNA Samples.

Marker-typing

Markers in the mapping panel were detected by a two-phase hemi-nested PCR (Figure 1b). Phase-1 PCRs were multiplexed, using up to 40 pairs of external primers simultaneously. Sub-samples of the phase-1 products were then screened for one marker at a time, using the appropriate forward-internal and reverse primer pair. Since the initial PEP products can be split into ≈ 40 subfractions, and since up to 40 markers were multiplexed in each phase-1 PCR, this procedure allows us to type ≈ 1600 markers on any one mapping panel. However, more recent studies on a variety of genomes have shown that ≥ 400 markers may be multiplexed in the phase-1 PCR without loss of efficiency. Accordingly, at least 16 000 (40 × 400) markers could be typed using a single PEP-amplified panel. There is no indication that 400 markers is the upper limit for multiplexing, and hence it is likely that even more could be mapped.

Of the 120 markers typed, 107 yielded clear (first-rate) results. Ten gave second-rate results (weak bands or anomalously low numbers of positives) and were discarded from further analysis. Only three failed to yield any results. Second-rate markers and failures were not re-examined to identify reasons for failure; however, the majority of these were markers for which primers had been designed by Primer3 rather than HOSP, suggesting that most failures were due to avoidable primer design weaknesses. All negative controls were clean, suggesting little or no contamination during the preparation or screening of the panel.

Analysis of the results for the first few markers indicated that the mean DNA content of the panel was ≈ 0.7 genomes per aliquot (GPA); analysis of all results gave a figure of 0.69 GPA.

LOD and linkage analysis

Results were scored and analysed as described in Experimental procedures. The entire set of 107 first-rate markers formed a single linkage group at a significance threshold of LOD = 5.0. Hence, HAPPY mapping was able to correctly co-assign all markers to a tightly defined (< 2%) region of the genome, meeting the first requirement for correct 'global positioning' of markers. The lod scores and θ-values were used to compute the HAPPY map using the distance geometry algorithm described by Newell *et al*. (1995). The HAPPY map and its correspondence with the DNA sequence is presented in Figure 2.

Comparison of the HAPPY map with genomic sequence

The sequence of the contiguous 1.9 Mbp region enabled the comparison of local marker order and spacing to be made over distances which were much smaller than the anticipated ≈ 50 kbp resolution of the mapping panel. Of the 107 markers, the order of 84 was in perfect agreement with the sequence data, even where the intermarker spacing was only a few kilobases. Of the remainder, all but one were placed well within the ≈ 50 kbp resolution predicted for the panel. The only exception (LR5, misplaced by ≈ 70 kbp) lies very close to one end of the map; it is a general observation that any segregation-based map (HAPPY, RH or genetic) is more error-prone near the ends of the linkage groups.

In summary, we have shown that the mapping panel provided results which met or exceeded the expected ≈ 50 kbp resolution. It is important to note that, in this and in earlier studies, we find no errors which can be attributed to peculiarities of sequence, chromosome structure or genome. Such maps can be used to gain information directly from genomic DNA at a level which is comparable to that achieved by a detailed restriction analysis of cloned DNA fragments or by sequencing, and are more than adequate to serve as robust frameworks for physical mapping.

In this study (data not shown) and other contexts (Konfortov *et al*., 2000; Walter *et al*., 1993), we have shown that HAPPY mapping panels produced with smaller DNA fragments can obtain resolution at the kilobase level. Such maps unambiguously order dense clusters of markers

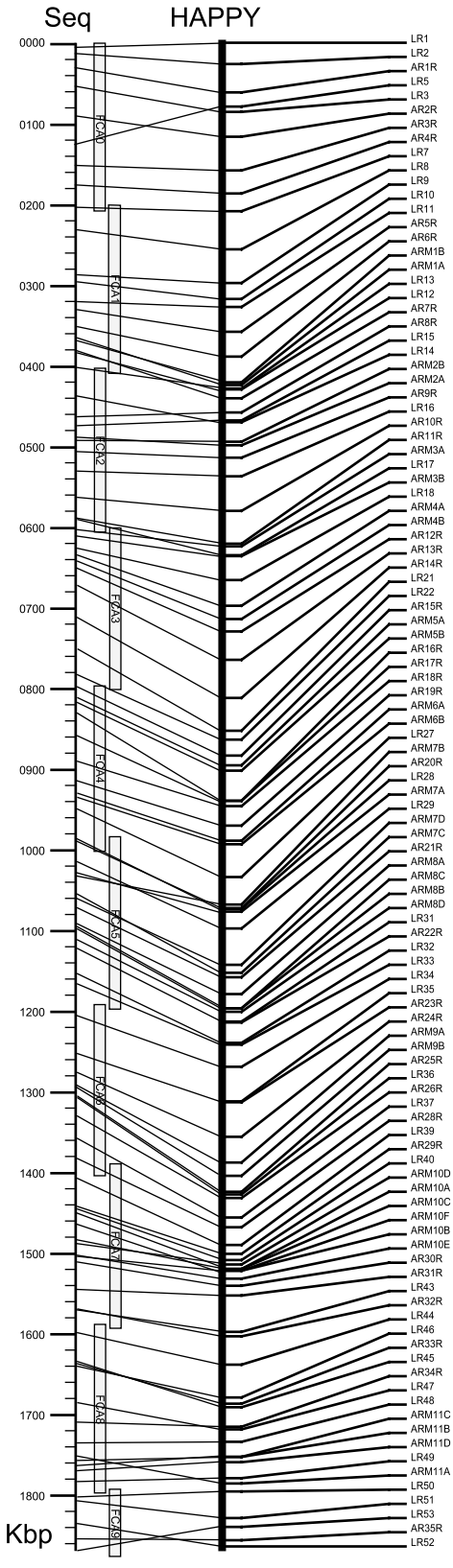

Figure 2 Comparison of HAPPY map and nucleotide sequence. The mapped positions of 107 markers (named at right) are shown on the HAPPY map (central heavy line) and nucleotide sequence (left). The regions covered by the sequence contigs FCA0-9 are indicated (boxes).

and can guide the assembly of data produced by whole-genome shotgun sequencing or by the subcloning of large-insert clones.

HAPPY mapping of plant genomes: some general considerations

The size and complexity of many plant genomes continues to be regarded as an obstacle to their analysis. Current models of genome evolution involve repeated rounds of genome duplication followed by progressive gene loss, and expansion resulting from the activity of retrotransposable elements (Gaut *et al*., 2000; Kumar and Bennetzen, 1999). Such models explain the very high proportion of gene families and non-coding repetitive DNA found in plant genomes. Indeed, the general picture that is emerging of many large plant genomes is that of a vast landscape of essentially featureless (non-coding) repetitive DNA interspersed with densely populated gene islands (Moore, 2000).

This architecture poses two major challenges for genome mapping. The first is that of spanning (if not analysing) the large tracts of featureless, repeated and possibly uncloneable DNA. The second is that many of the most interesting genes (for example, many disease resistance genes) will themselves be present as low copy, but repetitive, sequences (i.e. gene families). Clearly, any mapping technique will experience difficulties in defining a unique location for a marker which is duplicated in the genome!

With regard to the first problem – that of mapping clusters of genes embedded in featureless DNA tracts – there is an apparent conflict arising from the way in which the range and resolution of a HAPPY mapping panel are interrelated. Panels made from small DNA fragments will easily resolve the order of markers within gene-rich islands, but will lack the range necessary to map the positions of the islands relative to one another. Conversely, panels made with larger fragments will enable the positions of the islands to be mapped, but may lack the resolution to reveal their internal structure. This trade-off between range and resolution is characteristic of many mapping techniques.

This conflict can be resolved in one of three ways. First, if a few unique-sequence markers can be found in the featureless inter-island tracts, then a shorter-range panel will suffice to link them (and the islands) together, while still resolving the dense cluster of markers within each gene-rich island. Second, if no suitable markers can be found between widely spaced islands, the genome can be tackled using two mapping panels: a long-range panel to link and order the islands, and a shorter-range panel to map markers within each island.

(Such an approach remains economical, because only a few markers from each island need be mapped on the long-range panel.) Third, it may be preferable to use HAPPY mapping only to map clusters of physically close markers, producing a number of 'patch maps', each covering one gene-rich island in detail. Genetic or cytogenetic analysis (e.g. FISH) can then provide the coarse mapping data needed to position the islands relative to one another.

With regard to the problem of widespread gene duplication and the consequent repetitive nature of many gene-derived markers, this is less of a difficulty in HAPPY mapping than in many other approaches. The hemi-nested PCR described here is highly sequence-specific and discriminates between related sequences more effectively than techniques which depend on hybridization, restriction-fingerprinting of clones or non-nested PCR (Hong, 1997; Lin *et al*., 2000). Nevertheless, preliminary studies with the genome of potato (results not shown) have revealed problems arising from the near-perfect duplication of markers, presumably due to genes undergoing paralogous amplification. While it may be possible to distinguish these 'post-amplification' by assaying the products for gene-specific polymorphisms (e.g. by restriction digestion), in general it will be desirable to derive single copy markers from the faster-evolving non-translated parts of genes (such as promoters or 3′ untranslated regions), which are more likely to have diverged since the duplication of the genes.

Clearly these issues of range, resolution and duplicated sequences will have to be investigated in greater detail in the context of specific genomes. Nevertheless, by using examples of agronomically important plants it is possible to outline the role that HAPPY mapping can have in improving our understanding of plant genomes.

HAPPY maps of smaller plant genomes (1C < 1000 Mbp), which are likely to have comparable gene densities to *Arabidopsis*, should be readily achievable at very high densities and high resolution. It is feasible in such genomes to map markers with an average spacing of a few tens of kilobases (comparable to the 16 kb marker-to-marker spacings achieved in this study), and very large tracts of featureless repeated sequence are unlikely to be found. Examples of plants whose genomes fall into this size category include rice (*Oryza sativa*), grapevine (*Vitis vinifera*), common bean (*Phaseolus vulgaris*) and potato (*Solanum tuberosum*).

As genome size increases beyond ≈ 1000 Mbp, one would expect to encounter gene densities perhaps an order of magnitude lower than that of *Arabidopsis*. Contiguous maps of such genomes should be within the reach of longer-range HAPPY mapping panels. Plants such as coffee (*Coffea arabica*), sugar beet (*Beta vulgaris*), sweet potato (*Ipomoea batatas*), finger millet (*Eleusine coracana*), groundnut (*Arachis hypogaea*), tomato (*Lycopersicon esculentum*), soybean (*Glycine max*), rape (*Brassica napus*), sorghum (*Sorghum bicolor*) and rubber (*Hevea brasiliensis*) probably fall within this group.

The larger genomes (1C > 2500 Mbp) are more likely to have large featureless repetitive DNA tracts divided by gene-rich islands. It may never be worthwhile attempting to analyse the tracts, but it is both feasible and attractive to generate 'patch maps' of individual gene islands, as outlined above, with genetic or cytogenetic approaches resolving the relative positions of the islands. Prominent within this group are maize (*Zea mays*), cotton (*Gossypium* spp.), pea (*Pisum sativum*), barley (*Hordeum vulgare*), oats (*Avena sativa*), onion (*Allium sativum*) and bread wheat (*Triticum aestivum*). In the Triticeae, in particular, it is clear from the detailed comparison of physical and genetic maps of barley, developed using microdissected chromosome translocation stocks and PCR-based genetic markers, that such a strategy is applicable. This comparison revealed that 4.9% of the physical map contained > 47% of the single-locus genetic markers or genes used in the construction of the genetic map (Künzel *et al*., 2000). In these regions the ratio of physical to genetic distance was estimated to be in the range of 0.1–0.9 Mbp/cM. Compared to *Arabidopsis*, with an average ≈ 0.2 Mbp/cM, this indicates that, even in the 5300 Mbp barley genome, HAPPY mapping can facilitate the construction of high resolution local physical gene maps, if not complete genome maps.

In all instances therefore, it becomes possible to produce accurate HAPPY maps of the complete genome or of the interesting gene-rich regions. Relative to other approaches, HAPPY mapping is technically non-demanding. It is not subject to the biological distortions that afflict *in vivo* techniques, is less severely impeded by the repetitive elements and duplications which characterize many plant genomes, and is highly flexible in the range and resolution which it affords.

We see two main roles for HAPPY mapping in plant genomics. In many genomes it will serve as a 'stand-alone' technique, defining the location of genes, arbitrary sequences and polymorphic markers. It will refine genetic maps, aiding the search for quantitative trait loci and genes of commercial significance and provide guidance for marker-assisted breeding programmes. In this context, it will also be a valuable comparative tool, as it is the only technique which does not require the creation and maintenance of an extensive set of resources (clones, radiation hybrids or multigeneration families) for each genome to be mapped. Markers

which have been successfully mapped on one genome can, at little additional cost, immediately be mapped on other cultivars or on closely related species.

The second role for HAPPY mapping will be as an aid to physical mapping and genome sequencing, either regional or global. A relatively low-resolution HAPPY map will suffice to anchor large-insert clones in the genome, regardless of the presence of intervening uncloneable or repeated sequences which would confound conventional contig assembly. At higher resolution, HAPPY maps can directly aid the assembly of regional or genome-wide DNA sequence data. Indeed, it can be seen that the map produced in this study would be of adequate resolution to position, align and orientate large-insert clones such as BACs (which are typically similar in size to the FCA contigs spanning the region), as well as to reveal any major deletions, rearrangements or co-ligations within the clones.

In the results presented above we offer an illustration of the HAPPY mapping strategy for preparing high resolution maps of plant chromosomal DNA. We show that a PEP-amplified mapping panel can be easily prepared, and enables the mapping of thousands or tens of thousands of markers. We have demonstrated the accuracy of such mapping at two levels: perfect assignment of markers to a defined region within the genome; and exceptional agreement between the computed marker order and that known from the DNA sequence, surpassing in most cases the expected ≈ 50 kb resolution of the mapping panel.

Based on these results, and on our experience with other genomes, we propose that HAPPY mapping in its current form has great potential for the high resolution mapping of plant genomes in general and particularly for preparing maps of entire 'smaller' (≤ 1000 Mbp) plant genomes. The HAPPY mapping protocol as described offers immediate and immense benefits of cost and speed, and should accelerate the comparative analysis of plant genomes.

Experimental procedures

Plant material

Seeds of *A. thaliana* (Columbia ecotype) were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC), University of Nottingham, School of Biological Sciences, Plant Science Division, University Park, Nottingham NG7 2RD <http://nasc.nott.ac.uk> and grown under long day greenhouse conditions (16/8 h light/dark) to flowering stage, whereupon the leaves were harvested, frozen in liquid nitrogen, and stored at –70 °C.

Design of markers and primers spanning the FCA contig

Primers were designed using the DNA sequence of the 10 overlapping 200 kb FCA sequence contigs termed FCA0 to FCA9 (EMBL accession nos Z97335–Z97344). These contigs have been assembled from an assortment of BACs, cosmids and YAC subclones from the Columbia ecotype. A hemi-nested set of primers (forward-external, forward-internal, and reverse) was designed for each of 120 locations scattered randomly across the region, using the programs Primer 3 <http://www/genome.wi.mit.edu/genome_software/other/primer3.html> or HOSP (PhD, unpublished). The primers contained a two-base 3′ GC-clamp, a one-base 5′-GC clamp, and had an average melting temperature of 58–60 °C. Primer sequences and co-ordinates are detailed in the *Supplementary Material*.

Preparation of megabase genomic DNA in agarose 'strings'

Leaf nuclei were prepared from 1.5 g of frozen *Arabidopsis* leaf material using a protocol developed by Clemson University Genomics Institute <http://www.genome.clemson.edu/groups/bac/protocols/megabasedna.html#Nucl>, resuspended in 1 mL of 1 × homogenization buffer (10 mM Tris-HCl (pH 9.5), 80 mM KCl, 10 mM EDTA, 1 mM spermidine, 1 mM spermine, 0.5 M sucrose, 0.15% v/v β-mercaptoethanol), warmed briefly to 45 °C, and mixed gently with an equal volume of 1% (w/v) low-melting-point agarose (*ultra*PURE™, GibcoBRL) in 1 × HB, also at 45 °C. The mixture was taken up into glass capillaries (100 μL Supracaps, 1.2 mm internal diameter; Scientific Laboratory Supplies) and chilled at 4 °C for 10 min. When set, the strings of agarose-embedded nuclei were allowed to fall under gravity from the capillaries into ≈ 50 mL lysis buffer (0.5 M EDTA, pH 9.0, 1% lauryl sarcosine sodium salt, 0.1 mg/mL proteinase K) and incubated at 45 °C for 48 h with gentle mixing. The strings were washed in 0.5 M EDTA, pH 9.0 for 1 h at 45 °C; in 0.05 M EDTA, pH 8.0 for 1 h on ice; and stored in 0.05 M EDTA, pH 8.0 at 4 °C for up to several months.

Preparation of HAPPY mapping panel

The panel was prepared from DNA which had been broken by irradiation and size-fractionated by gel electrophoresis. Two agarose strings (see above) were irradiated (35 J/kg; Gravatom RX30/55M ^{137}Cs source) and placed in a 140 mm × 3 mm well in a 3 mm thick gel (LKB Gene Navigator) of 1% chromosomal grade agarose (Bio-Rad) in 0.5 × TBE (90 mM

Tris borate, 2 mM EDTA, pH 8.3), flanked by *Saccharomyces cerevisiae* chromosomal markers (Bio-Rad). Samples were sealed into the gel using 1% agarose in 0.5 × TBE prior to electrophoresis. After electrophoresis (180 V, 100 s pulse time, 12 h), the sides of the gel, containing the yeast standards, were excised, stained with ethidium bromide, and visualized with UV light. The unstained central portion was incubated for > 3 h in sterile TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA), then for 30 min in 0.1 × TE, both at 4 °C. The gel was removed from 0.1 × TE and a glass capillary (internal diameter 0.56 mm; Drummond Scientific Co.) was used to punch out the agarose plugs from across a region of the gel in the size range 450–550 kbp, as judged by alignment with the *S. cerevisiae* standards.

Individual agarose plugs (≈ 0.75 μL), were dispensed into 88 wells of a 96-well PCR plate (Thermowell, Costar); an equal volume of HPLC-grade water was dispensed into the remaining eight wells as negative controls. Each well was overlaid with a drop of light mineral oil (Sigma) and used immediately or stored at –70 °C.

Panel pre-amplification and marker typing

The scoring of markers on the mapping panel involved three steps: pre-amplification; multiplexed amplification (phase-1); and hemi-nested marker-specific amplification (phase-2) (Figure 1b).

Pre-amplification was carried out using PEP (primer extension preamplification) with a random 15-mer primer (Zhang *et al*., 1992). Reactions (5 μL) contained, in addition to the agarose plugs, 1 × Amplitaq Buffer II (PE Applied Biosystems), 2.5 mM $MgCl_2$, 200 μM each dNTP, 10 μM N15 primer (Operon Technologies, Inc., Alameda CA), 1 U *Taq* polymerase (AmpliTaq, PE Applied Biosystems), and ≈ 1.5 pg of human genomic DNA as 'carrier' (see Results and discussion). Cycling conditions were 93 °C for 5 min; then 50 cycles of 94 °C for 30 s, 37 °C for 2 min, 37–55 °C ramp over 3 min, and 55 °C for 4 min. Each pre-amplified aliquot was then diluted to 200 μL in HPLC-grade water and stored at –70 °C.

Phase-1 reactions (10 μL) were highly multiplexed (see Results and discussion), and contained 5 μL of the diluted PEP product, 0.25 μM of each external primer (forward-external and reverse for each of up to 40 markers), 1 × AmpliTaq Gold reaction buffer, 1 U AmpliTaq Gold polymerase (PE Applied Biosystems), 4 mM $MgCl_2$ and 200 μM each dNTP. Cycling conditions were 93 °C for 10 min; then 25 cycles of 94 °C for 20 s, 55 °C for 30 s and 72 °C for 1 min. Phase-1 products were diluted in HPLC grade water to a final volume of 200 μL and stored at –20 °C.

Phase-2 reactions for individual markers were performed using single primer pairs consisting of the forward-internal and the corresponding reverse primer. Reactions (10 μL) contained 5 μL of the diluted phase-1 product, 1 × AmpliTaq Gold buffer, 0.25 U AmpliTaq Gold polymerase, 1.5 mM $MgCl_2$, 200 μM each dNTP, and 1 μM each forward-internal and reverse primer for the marker in question. Cycling conditions were 93 °C for 10 min; then 33 cycles of 94 °C for 20 s, 55 °C for 30 s and 72 °C for 1 min

Phase-2 products were supplemented with an equal volume of loading dyes (15% w/v Ficoll 400 [Pharmacia], 0.15 mg/mL bromophenol blue, 4 × SyBr Green1 [FMC Bioproducts], 1 × TBE), resolved on 96-well 6% polyacrylamide gels in 0.5 × TBE (MIRAGE gels; Genetix, Dorset UK; 20 V/cm 11 min) and visualized by UV transillumination.

Rigorous precautions were taken at all steps prior to phase-2 to exclude contaminating DNA (Dear and Cook, 1989).

Checking the DNA content of the mapping panel

To check that the mapping panel had the correct DNA content (≈ 0.5–0.8 genomes per aliquot [GPA]), results for the first few markers were analysed before the remainder were typed. The proportion of aliquots found to contain a given marker is related to the mean DNA content of the aliquots according to the formula $GPA = -\log_e([T - N]/T)$, where T is the total number of aliquots analysed (in this case, 88) and N is the number of aliquots positive for the marker.

Data entry and analysis

The results of marker typing were scored manually (as positive, negative or ambiguous for each aliquot for each marker) using a graphical data entry program (GelEnter; unpublished PhD). Based on the quality of the results (the intensity of the bands and the proportion of aliquots scored as 'positive'), each marker was designated as 'first-' or 'second-rate'. Only first-rate markers were used in the subsequent analysis.

Pair-wise lod and θ values, reflecting, respectively, the likelihood of linkage between any two markers and the probability of a break between them (which reflects physical distance) were calculated as described in Dear *et al*. (1998), and the markers sorted into linkage groups (programs Lodulator and Lontig; unpublished PhD). Maps were computed from these pair-wise data using a distance geometry algorithm (Newell *et al*., 1995) as described in Dear *et al*. (1998). Finally, map distances were re-scaled linearly in accordance with the known total size of the mapped region.

Acknowledgements

We are indebted to the Leverhulme Trust for funding this project (Grant F/766/A), the Medical Research Council and the Scottish Crop Research Institute for additional funding of consumables. We would also like to thank Sir Aaron Klug for comments and suggested improvements to this document.

Supplementary material

Supplementary material containing primer sequences and co-ordinates see the Supplementary Table on the web at <http://wip.blackwellpublishing.com/products/journals/suppmat/PBI/PBI001/PBI001sm.htm>.

References

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., Van Staveren, M., Stiekema, W., Drost, L., Ridley, P., Hudson, S.A., Patel, K., Murphy, G., Piffanelli, P., Wedler, H., Wedler, E., Wambutt, R., Weitzenegger, T., Pohl, T.M., Terryn, N., Gielen, J., Villarroel, R., De Clerck, R., Van Montagu, M., Lecharny, A., Auborg, S., Gy, I., Kreis, M., Lao, N., Kavanagh, T., Hempel, S., Kotter, P., Entian, K.D., Rieger, M., Schaeffer, M., Funk, B., Mueller-Auer, S., Silvey, M., James, R., Montfort, A., Pons, A., Puigdomenech, P., Douka, A., Voukelatou, E., Milioni, D., Hatzopoulos, P., Piravandi, E., Obermaier, B., Hilbert, H., Dusterhoft, A., Moores, T., Jones, J.D.G., Eneva, T., Palme, K., Benes, V., Rechman, S., Ansorge, W., Cooke, R., Berger, C., Delseny, M., Voet, M., Volckaert, G., Mewes, H.W., Klosterman, S., Schueller, C. and Chalwatzis, N. (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature*, **391**, 485–488.

Budiman, M.A., Mao, L., Wood, T.C. and Wing, R.A. (2000) A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. *Genome Res*. **10**, 129–136.

Buschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., vanDaelen, R., vanderLee, T., Diergaarde, P., Groenendijk, J., Topsch, S., Vos, P., Salamini, F. and SchulzeLefert, P. (1997) The barley *mlo* gene: a novel control element of plant pathogen resistance. *Cell*, **88**, 695–705.

Cai, D.G., Kleine, M., Kifle, S., Harloff, H.J., Sandal, N.N., Marcker, K.A., KleinLankhorst, R.M., Salentijn, E.M.J., Lange, W., Stiekema, W.J., Wyss, U., Grundler, F.M.W. and Jung, C. (1997) Positional cloning of a gene for nematode resistance in sugar beet. *Science*, **275**, 832–834.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S. and Myers, R.M. (1990) Radiation hybrid mapping – a somatic-cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science*, **250**, 245–250.

Dear, P.H. (1997) *Genome Mapping – a Practical Approach* (Ed. Dear, P.H.), pp. 95–124. Oxford, UK: IRL Press.

Dear, P.H., Bankier, A.T. and Piper, M.B. (1998) A high-resolution metric HAPPY map of human chromosome 14. *Genomics*, **48**, 232–241.

Dear, P.H. and Cook, P.R. (1989) Happy mapping – a proposal for linkage mapping the human genome. *Nucl. Acids Res*. **17**, 6795–6807.

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

Eckardt, N.A. (2001) Everything in its place: Conservation of gene order among distantly related plant species. *Plant Cell*, **13**, 723–725.

Gaut, B.S., d'Ennequin, M.L., Peek, A.S. and Sawkins, M.C. (2000) Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl Acad. Sci. USA*, **97**, 7008–7015.

Hong, G.F. (1997) A rapid and accurate strategy for rice contig map construction by combination of fingerprinting and hybridization. *Plant Mol. Biol*. **35**, 129–133.

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

Konfortov, B.A., Cohen, H.M., Bankier, A.T. and Dear, P.H. (2000) A high-resolution HAPPY map of *Dictyostelium discoideum* chromosome 6. *Genome Res*. **10**, 1737–1742.

Kumar, A. and Bennetzen, J.L. (1999) Plant retrotransposons. *Annu. Rev. Genet*. **33**, 479–532.

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

Lin, Y.-R., Draye, X., Qian, X., Ren, S., Zhu, L.-H., Tomkins, J., Wing, R.A., Li, Z. and Paterson, A.H. (2000) Locus-specific contig assembly in highly-duplicated genomes, using the BAC-RF method. *Nucl. Acids Res.* **28**, e23.

Moore, G. (2000) Cereal chromosome structure, evolution, and pairing. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. **51**, 195–222.

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

Newell, W.R., Mott, R., Beck, S. and Lehrach, H. (1995) Construction of genetic maps using distance geometry. *Genomics*, **30**, 59–70.

Piper, M.B., Bankier, A.T. and Dear, P.H. (1998) A HAPPY map of *Cryptosporidium parvum*. *Genome Res*. **8**, 1299–1307.

Ramsay, L., Macaulay, M., Ivanissevich, S.D., MacLean, K., Cardle, L., Fuller, J., Edwards, K.J., Tuvesson, S., Morgante, M., Massari, A., Maestri, E., Marmiroli, N., Sjakste, T., Ganal, M., Powell, W. and Waugh, R. (2000) A simple sequence repeat-based linkage map of barley. *Genetics*, **156**, 1997–2005.

Russell, J.R., Ellis, R.P., Thomas, W.T.B., Waugh, R., Provan, J., Booth, A., Fuller, J., Lawrence, P., Young, G. and Powell, W. (2000) A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. *Mol. Breed*. **6**, 553–568.

Swanston, J.S., Thomas, W.T.B., Powell, W., Young, G.R., Lawrence, P.E., Ramsay, L. and Waugh, R. (1999) Using molecular markers to determine barleys most suitable for malt whisky distilling. *Mol. Breed*. **5**, 103–109.

Tanksley, S.D. (1993) Mapping polygenes. *Annu. Rev. Genet*. **27**, 205–233.

Walter, G., Tomlinson, I.M., Cook, G.P., Winter, G., Rabbitts, T.H. and Dear, P.H. (1993) Happy mapping of a YAC reveals alternative haplotypes in the human-immunoglobulin V(H) locus. *Nucl. Acids Res*. **21**, 4524–4529.

Zhang, L., Cui, X.F., Schmitt, K., Hubert, R., Navidi, W. and Arnheim, N. (1992) Whole genome amplification from a single cell: implications for genetic analysis. *Proc. Natl Acad. Sci. USA*, **89**, 5847–5851.