High density, high throughput physical mapping in plants

A.B. James, G. Bryan, P.H. Dear[1], M. Thangavelu[1], L. Ramsay & R. Waugh

Genetic linkage maps have been constructed for many plant and most crop species. By correlating the pattern of inheritance of trait information in a meiotic mapping population with that of individually mapped genetic markers, many monogenic and polygenic traits have been located to specific regions of the plant genome. The major impetus for meiotic mapping was the discovery of extensive, yet easily visualised, variability at the DNA level, which could be used as markers in most natural populations. Over the last 20 years, this approach has been extremely valuable and productive. However it suffers from a number of limitations. Firstly, it relies on recombination occurring between polymorphic loci during meiosis, thus excluding the ability to map chromosomal regions that are identical by descent. Secondly, the ability to order the entire complement of genes on a plant chromosome is unlikely since the probability of recombination events occurring over ever shorter distances becomes vanishingly small. Meiotic mapping is further compounded by the presence of recombination 'hot-spots', and segments of suppressed recombination. The result is a poor correlation between genetic and physical distances.

The advent of physical mapping, based on sequence tagged site (STS) or expressed sequence tag (EST) content analysis of large insert clones such as YACs or

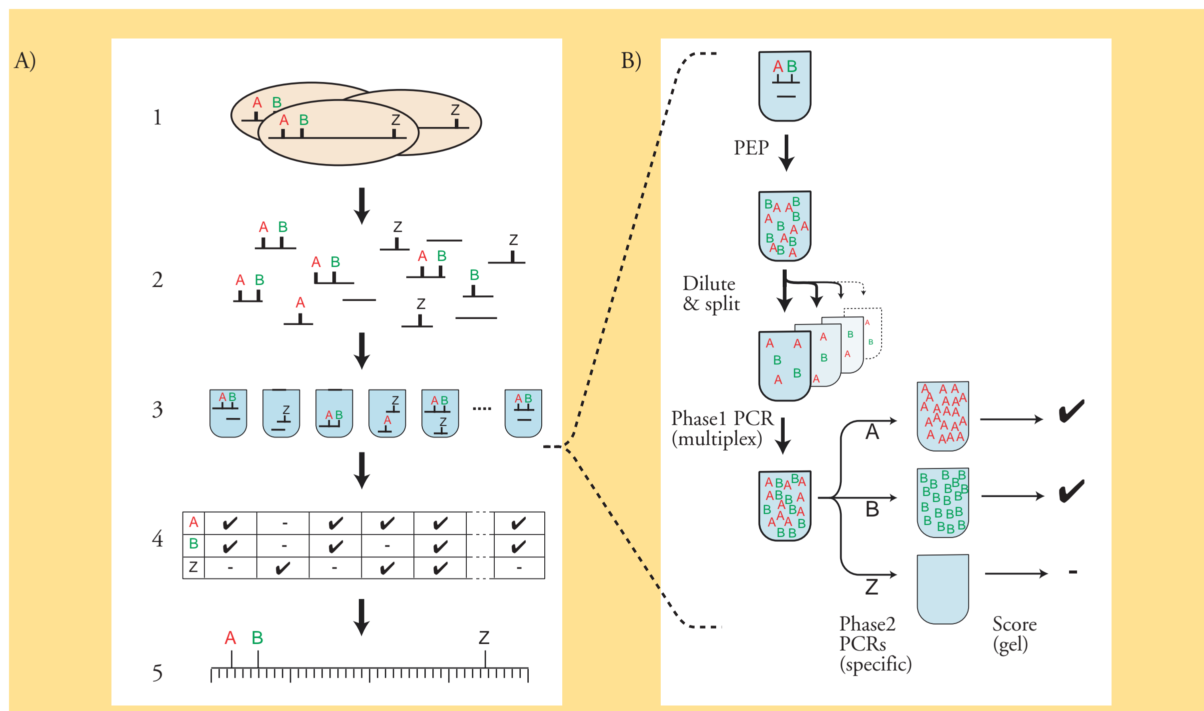

Figure 1 Principle of HAPPY mapping and marker typing. (A) Overview. DNA carrying STS markers (A, B, Z) is extracted from cells (1) and broken randomly to give a pool of fragments (2). These are dispensed at limiting dilution into a series of aliquots - the mapping panel (3). The panel is screened by 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 3-step PCR. The protocol is illustrated for one aliquot of the mapping panel. All DNA in the sample is first pre-amplified >100-fold using PEP. This material is diluted and split into sub-fractions for multiple rounds of screening. One sub-fraction is amplified in a multiplex PCR for many markers (Phase1). 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 (Phase2). Results are scored on gels, determining the marker content of the aliquot.

[1] MRC Laboratory of Molecular Biology, Protein and Nucleic Acid Chemistry Division, Cambridge CB2 2QH, United Kingdom.

BACs, has circumvented many of these drawbacks. In particular, it precludes the requirement for polymorphic genetic markers. Using this approach, 'contigs' of overlapping cloned fragments are assembled to span the entire genome in question, or the particular region of interest. A genome-wide STS content physical map of *Arabidopsis thaliana* has been constructed and maps of several crops species, including rice and sorghum, are being prepared. Physical mapping in this manner is generally regarded as enhancing the molecular genetics of the particular organism, since it serves as an archive of genomic information. Furthermore, integration of physical maps with genetic maps is extremely valuable for map-based gene isolation, comparative genome analysis, and as sources of sequence-ready clones for genome sequencing projects. However, mapping chromosomes or genomes in this way is limited by the cloning process on which it relies; regions recalcitrant to cloning lead to uncloseable gaps, while rearranged or co-ligated fragments, or repeated regions larger than the size of the clones, can lead to distortions. Hence, it has been proposed that physical maps are most effective if built over an independently constructed STS 'scaffold'.

HAPPY mapping[1] has been developed as an *in vitro* physical mapping technique that addresses the problems associated with other *in vivo* methods, such as Radiation Hybrid (RH) mapping. HAPPY mapping's utility has been demonstrated by mapping human chromosome 14 at high resolution[2], the complete genome of the parasite *Cryptosporidium parvum*[3], and chromosome 6 of *Dictyostelium discoideum*[4]. The technique involves breaking intact genomic DNA at random, segregating the fragments into aliquots (the 'mapping panel') each of which contains less than one genome's worth of DNA

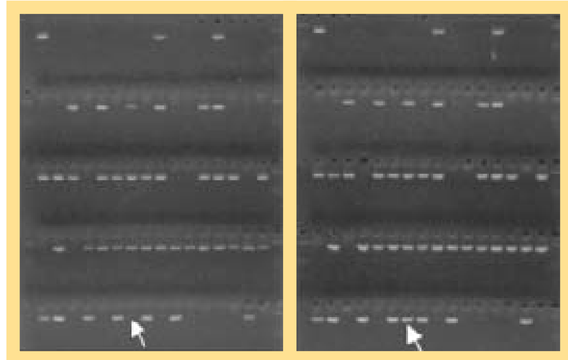

Figure 2 Typing two physically linked PCR-STS markers on a High resolution HAPPY mapping panel of barley. Only a single 'recombinant' is observable between the two markers which are separated by 10kb on the barley genome.

fragments (the *in vitro* analogue of a radiation hybrid cell), and measuring the frequency of co-segregation of markers among the aliquots (Fig. 1). Closely linked markers are rarely separated by an intervening break and therefore tend to co-segregate. In this way, it is analogous both to classical genetic linkage analysis, which measures the frequency of recombination between markers during meiosis, and to RH mapping. HAPPY mapping, however, possesses all the advantages of RH mapping (no requirements for polymorphisms; flexible resolution of markers depending on the size of DNA fragments utilised), but none of the drawbacks - it is immune to artefacts caused by the biological activity of the DNA fragments, to cloning artefacts, or to the effects of chromosome structure. In addition, constructing and screening a HAPPY mapping panel is relatively simple, and only small quantities of DNA are needed.

As proof of principle, we have constructed a high resolution HAPPY

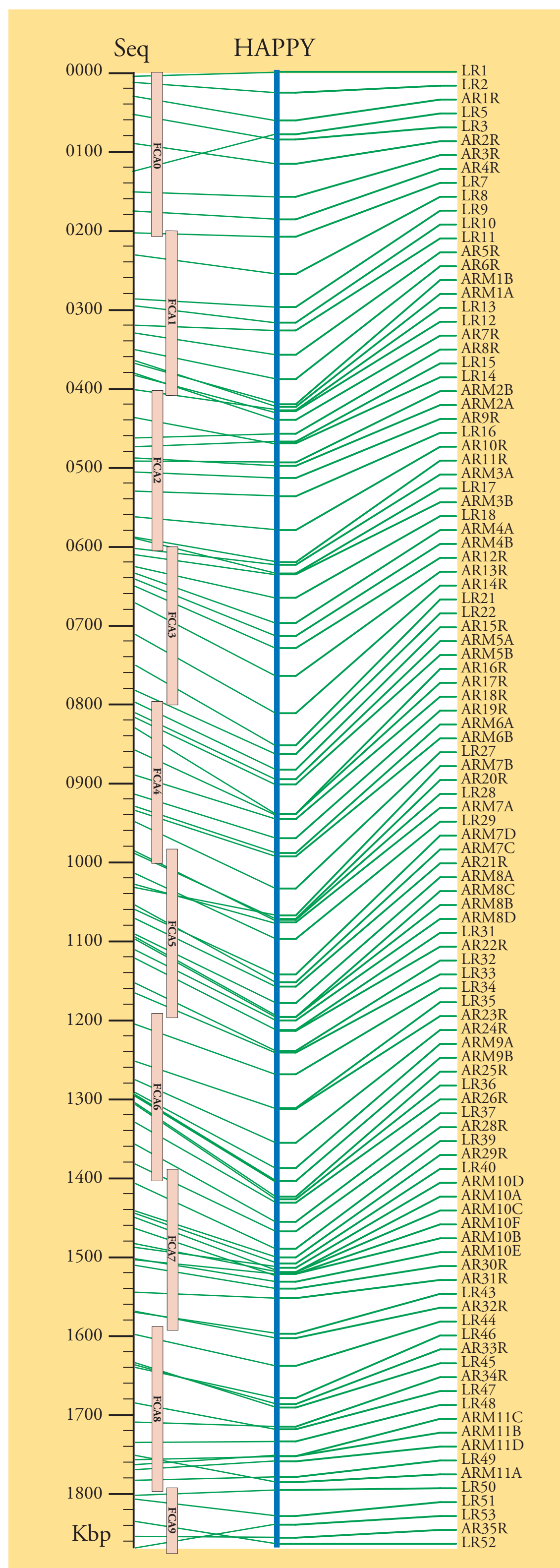

Figure 3 Comparison of HAPPY map, physical map and nucleotide sequence. The mapped positions of 107 markers (named at right) are shown on the HAPPY map (central heavy line) and nucleotide sequence (left).

map of a 2.0 Mbp section of *A. thaliana* chromosome 4. This map, offering an average marker to marker distance of 16 kbp, spans approximately 10% of the chromosome and approximately 1.4% of the entire *A. thaliana* genome. In order to construct the map, a HAPPY mapping panel was prepared comprising 80 aliquots of genomic DNA broken at random, with each aliquot containing less than one genome's worth of DNA. The panel was analysed for the presence of PCR amplified STS marker sequences spanning the sequenced FCA locus of *A. thaliana* chromosome 4 (~2.0 Mbp) (Fig. 1). Analysis of the frequency of co-segregation of 107 markers enabled the order and physical distance between markers to be estimated and compared with the true physical distance (Fig. 2). The map revealed a good overall correspondence with the true physical map; the order of all but 16 markers (< 15%) was as given in the sequenced FCA locus and the errors in the positions of all but one of these markers were well within the margins of error expected from the mapping panel. We have also shown that these local inversions could be corrected by re-typing the markers on a shorter range mapping panel comprising smaller DNA fragments. Having demonstrated the potential of the method in *Arabidopsis,* we have also started to look at crop plant genomes which, physically, can be significantly larger. Preliminary experiments carried out to compare physical maps of the *mla* locus in barley cultivars, with different introgressed *mla* disease resistance specificities (genome size c. 40x that of *Arabidopsis*), indicate that HAPPY mapping will be a useful 'comparative physical mapping' tool. This obviates the need to construct large insert DNA libraries (e.g. BACs) for many separate accessions, followed by comparative structural analysis. In conclusion, we believe that HAPPY mapping panels could be developed for any plant species and utilised as a high resolution physical mapping resource by the plant research community.

References

[1] Dear, P. H. & Cook, P. R. (1993). *Nucleic Acids Research* **21,** 13-20.

[2] Dear, P. H., Bankier, A. T. & Piper, M. B. (1998). *Genomics* **48,** 232-241.

[3] Piper, M. B., Bankier, A. T. & Dear, P. H. (1998). *Genome Research* **8,** 1299-1307.

[4] Konfortov, B. A., Cohen, H. M., Bankier, A. T. & Dear, P. H. (2000). *Genome Research* **10,** 1737-1742.