Forward genetics and map-based cloning approaches

Janny L. Peters, Filip Cnudde and Tom Gerats

Department of Experimental Botany, Plant Genetics, University of Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands

Whereas reverse genetics strategies seek to identify and select mutations in a known sequence, forward genetics requires the cloning of sequences underlying a particular mutant phenotype. Map-based cloning is tedious, hampering the quick identification of candidate genes. With the unprecedented progress in the sequencing of whole genomes, and perhaps even more with the development of saturating marker technologies, map-based cloning can now be performed so efficiently that, at least for some plant model systems, it has become feasible to identify some candidate genes within a few months. This, in turn, will boost the use of forward genetics approaches, as applied (for example) to isolating genes involved in natural variation and genes causing phenotypic mutations as derived from (second-site) mutagenesis screens.

There are basically two ways to link the sequence and function of a specific gene: forward and reverse genetics. Reverse approaches rely upon sequence information as retrieved from genome and expressed sequence tag (EST) (see Glossary) sequencing or transcript profiling projects. The scientist starts with the selection of a specific (set of) sequence(s) and tries to gain insight into the underlying function(s) by selecting for mutations that disrupt the sequence and thereby, it is hoped, its function. Tagging by either endogenous or heterologous transposable elements or T-DNA constructs provides a range of opportunities, for a restricted number of model species [1,2]. These tagging approaches can identify the function of a specific gene by uncovering a specific phenotype. Moreover, when a particular function is encoded by more than one gene, reverse approaches are the only way to perform a step-by-step analysis of such redundant functions [3].

A less-focused set of reverse approaches includes antisense, cosuppression and RNA interference strategies [4]. Although they indicate particular overall functions, these strategies generally run the risk of putting a set of related genes out of order, thereby obscuring the contribution of each individual gene to the obtained phenotype. Nonetheless, constructs targeting a single specific gene can be made and can be of value when working with species for which tagging approaches are not feasible. A much more widely applicable and promising

Glossary

Allelism test crosses: (complementation test for functional allelism) a test to determine whether two mutations are caused by the same gene. By crossing the two mutants, one obtains an individual in which one homologue of a particular chromosome carries one of the recessive mutations and the other homologue carries the other recessive mutation. If the phenotype of the resulting heterozygote is wild type, we say that the two mutations complement each other, and they are usually in different genes. If the heterozygote shows the mutant phenotype, we say they do not complement each other and conclude that the mutations are in the same gene. A complementation group is a collection of mutations that affect the same gene.

Bacterial artificial chromosome (BAC): a vector used to clone DNA fragments of 100–300 kb in Escherichia coli cells.

Bulked segregant analysis (BSA): a rapid procedure for identifying markers in specific regions of the genome [34]. The method involves comparing two pooled DNA samples of individuals from a segregating population originating from a single cross. Within each pool, or bulk, the individuals are identical for a trait or gene of interest but are arbitrary for all other genes. Two pools with a different trait of interest are analysed to identify markers that distinguish them. Markers that are polymorphic between the pools will be genetically linked to the locus determining the trait used to construct the pools.

CentiMorgans (cM): a measure of recombination frequency. This unit of linkage refers to the distance between two loci based on the number of recombination events occurring between them. Two loci are said to be 1 cM apart if recombination is observed between them in 1% of meioses. In Arabidopsis, 1 cM is equivalent, on average, to 205 kb.

Genetic marker: a readily identifiable genetic element that exists in two or more allelic forms and is inherited in association with particular genes and genetic traits of interest.

Epistasis: a condition in which one gene suppresses the expression of another gene (and therefore whatever phenotype it is responsible for) when the two genes are not alternate alleles of the same phenotype. The gene that does the suppressing is called the epistatic gene. The suppressed gene is called hypostatic with regard to the epistatic gene.

Expressed sequence tag (EST): a partial sequence obtained by performing a single raw sequence read from a random cDNA clone. ESTs have uses in the discovery of new genes, mapping and the identification of coding regions in genomic sequences.

F2: the second generation of a cross, obtained by self-fertilization or intercrossing of F1 individuals.

Insertion/deletion (InDel) polymorphisms: DNA sequence variations that involve the insertion (or deletion) of one or more nucleotides.

Linkage: a measure of the degree to which alleles at two loci fail to assort independently during meiosis and are inherited together more often than not.

Marker-assisted selection (MAS): the use of linked genetic markers to select a specific character or trait.

Quantitative trait loci (QTLs): the genes that affect a trait that is measured on a quantitative scale. These traits are typically affected by more than one gene and by environmental factors.

Second-site mutation: a secondary mutation that suppresses another mutation, so that the wild-type phenotype is restored in the double mutant.

Single nucleotide polymorphisms (SNPs): a DNA sequence variation that involves a change in a single nucleotide.

Targeting-induced local lesions in genomes (TILLING): a technique designed to identify ethyl methanesulfonate (EMS)-mutagenized plants that have a mutation in a gene of interest whose sequence is known. TILLING combines pooling of DNA prepared from EMS-treated plants with the detection of a mutation by heteroduplex formation.

Transcript profiling: measuring the mRNA levels of all expressed genes relative to one another.
species-independent reverse approach is targeting-induced local lesions in genomes (TILLING) [5].

For specific model systems, such as Arabidopsis, tagging strategies now require only the screening of several databases to identify relevant insertion lines (http://www.arabidopsis.org/links/insertion.html). Unfortunately, this is not the general situation. The availability of such outstanding services is limited to a handful of species, such as rice (http://www.tigr.org/tdb/e2k1/osa1/), maize (http://www.tigr.org/tdb/tgi/maize/), wheat (http://wheat.pw.usda.gov/index.shtml), tomato (http://www.sgn.cornell.edu/), and to a lesser degree Antirrhinum (http://www.antirrhinum.net/) and Petunia, with Lotus (http://www.kazusa.or.jp/lotus/) and Medicago (http://www.noble.org/medicago/) as strong runners-up. But even for these, the situation is in general hardly comparable to what is available for Arabidopsis.

The main subject of this article relates to the opposite set of approaches to gene function analysis: forward genetics, which aims to identify the sequence change that underlies a specific mutant phenotype. The starting point is an already available or a specifically searched for and predicted phenotypic mutant of interest. Such mutants can be the result of deliberate mutagenesis or defined on the basis of existing variation. When the causative mutation is the result of a T-DNA or transposon insertion, rapid identification of the gene of interest is at least theoretically possible by locating the sequence tag and analysing its neighbouring sequences. However, the identification of a gene affected by a chemically or radiation-induced mutation requires a more-laborious map-based cloning (MBC) approach in which markers linked to the mutated gene are used to delimit the region containing the gene of interest. There are several good reasons to screen for mutants in chemical or radiation mutagenized populations: mutagenesis can be applied to essentially any species of interest, the spectrum of induced mutants is broader than with tagging approaches, mutagenesis is usually more efficient and second-site mutations are thus easier to obtain.

Until recently, MBC was a rather tedious, time-consuming and often unsuccessful approach, comparable to trying to reach the top of Mount Everest: although progress is initially easy, the final 100 kb or meters, respectively, are increasingly difficult to conquer. Combined with a fashionable view of functional gene definition by sequencing and reverse genetics approaches, this has diminished interest in MBC approaches. Lack of phenotypes in reverse screenings, mainly because of redundancy, and improvements in MBC approaches, especially the availability of saturating marker systems, have put both strategies back on a level playing field. Finally, quantitative trait loci (QTLs) underlying natural variation simply require map-based cloning for their sequence identification. Cloning QTLs is, in principle, no different from cloning any other gene but can at times be problematic because of phenotype definition. The specifics of QTL mapping have been described in several recent reviews [6–9] (http://naturalvariation.org/).

For reasons stated above, an efficient MBC procedure appears to be an important research tool. Today, the advances in sequencing projects, the wealth of available marker systems and the progress made in methods to detect DNA polymorphisms make fast MBC of a gene in a model species such as Arabidopsis feasible [10]. The definition of polymorphisms and thus of markers is no longer a limiting factor in MBC.

Maps and markers

The first concept of a genetic map was presented by Alfred H. Sturtevant [11], who ordered five sex-linked characters in a linear fashion on the Y chromosome of Drosophila melanogaster. Today, whole genomes are being sequenced with ever-increasing speed. In addition to a range of microbial and animal genomes, the Arabidopsis [12] and rice [13,14] genomes have almost been sequenced. In total, there are around 40 smaller and larger genome-sequencing projects in progress for plants, including species such as Avena sativa (oat), Medicago sativa and Medicago trunculata, Lotus corniculata, different Brassica species, banana, barley, coffee, cotton, Eucalyptus, maize, Populus, soybean and tomato [15]. Between the conception of the first map and the sequencing of whole genomes lie about 100 years of incredible developments in theoretical understanding and technological breakthroughs, leading to the building of increasingly dense genetic maps, culminating in physical maps in which genes or markers are located at a precise sequence position. As a result, the definition of the sequence and function of a gene in model species such as Arabidopsis or rice can be carried out within a four-year PhD project, something completely unimaginable less than 20 years ago.

A genetic map is constructed on the basis of recombination events between two non-sister chromatids of each pair of homologous chromosomes during meiosis (Box 1). A genetic localization experiment determines the order of linked markers. The distance determination [in centi-MORGANS (cm) or percentage recombination] is relative (as in medieval times, when the distance between towns was measured in hours). Recombination frequencies vary between different chromosome parts, physical conditions and sexes. As a consequence, the ratio between genetic and physical distance is not constant over the length of the chromosome. Also, genetic distance depends on the parental combination used: closely related lines will exhibit an intrinsically higher recombination frequency than distantly related lines do.

In the early days, progress in mapping was hampered by the lack of sufficient markers that did not exhibit epistatic interactions. Mutants could only be analysed as phenotypic black boxes. The analysis of polymorphisms in isozymes was a first step towards denser maps that were easier to create [16]. Further significant advancements included the invention of environmentally insensitive DNA-based marker systems such as restriction fragment length polymorphism (RFLP) analysis [17] and, even more so, the development of PCR-based markers such as random amplified polymorphic DNA (RAPD), simple sequence repeats (SSR) and amplified fragment length polymorphisms (AFLPs) (Table 1). Thus, the availability of markers is no longer a bottleneck and, today, this is true for essentially any living organism. Moreover, sequencing
Map-based cloning relies on a biological process that takes place at the heart of meiosis: the high-frequency genetic exchange events of meiotic recombination. Meiotic recombination not only ensures proper chromosome disjunction but also contributes to genetic diversity among the gametes. Each pair of homologous chromosomes undergoes at least one crossover to allow correct segregation at the first meiotic division. The double-strand-break repair model [48] still serves as a guideline for the basic description of meiotic recombination in yeast and other eukaryotes.

Recombination events are not uniformly distributed throughout the genome. Instead, the frequency of crossovers per unit physical distance can vary substantially, even within a single chromosome [49]. In Saccharomyces cerevisiae, recombination hot spots correspond to the preferred sites for meiosis-specific double-strand breaks that initiate recombination. In yeast, there is also a positive correlation between the occurrence of hot spots and an open chromatin configuration [50], rather than with specific DNA sequences. Furthermore, in most organisms, recombination is concentrated in the gene-rich (proximal) regions of the chromosome and strongly suppressed over large intervals around the centromeres. Analysis using classical markers forms the basis for the present understanding of meiotic recombination, but genetic recombination values can also be estimated directly by counting chiasmata. In Arabidopsis, for example, meiotic cells exhibit an average of 9.24 chiasmata [51].

Environmental factors, such as temperature, can influence the frequency of meiotic recombination. However, most frequently, such differences have a genetic basis. In many organisms, including tomato [52] and Arabidopsis [53], recombination during male and female gametogenesis occurs at different rates. Sequence diversity in interspecific crosses [52] and heterozygosity [54] also influence meiotic recombination. Finally, several genes have been identified that are essential for normal meiotic levels of recombination. These mostly encode proteins required for the synopsis of chromosomes or for catalysing key steps in DNA breakage, repair and recombination [55]. Disrupting these genes usually results in depressed levels of recombination [56]. Mutant loci that increase meiotic recombination frequencies, such as Rm1 in Petunia [57] or xrd4 in Arabidopsis [58], remain exceptional. The variation in meiotic recombination frequencies among eight accessions of Arabidopsis [59] also indicates the existence of recombination regulatory elements, which are interesting to plant breeders.

projects enable us to assign markers a physical position on the map (units in base pairs). For example, a map harbouring >1250 AFLP markers localized directly on the Arabidopsis sequence was developed by combining experimental and in silico AFLP analysis [18]. The physical position of these and several hundreds of other markers can be found at the TAIR website (http://www.arabidopsis.org/). Consequently, in sequenced model species, we can now focus on a precisely defined region of the genome that contains the gene of interest.

Map-based cloning strategies: balancing the available marker systems

Map-based cloning strategies use the fact that, as distances between the gene of interest and the analysed markers decrease, so does the frequency of recombination. To position any locus in the Arabidopsis genome in a 10-kb interval, one would need ~12 000 well-positioned markers. In itself, this is feasible with presently available technologies. In recent Arabidopsis MBC projects, simple sequence length polymorphism (SSLP) markers have most commonly been used to map the gene of interest [19–22].

SSLP markers are PCR based, abundant and co-dominant, and their detection is straightforward and inexpensive. However, specific primers need to be designed and tested for each putative SSLP marker. In sequenced organisms, this is a relatively undemanding task but, in species being examined for the first time, SSLP markers need to be developed de novo, which is a time-consuming and costly business: first, sequence segments harbouring SSLPs have to be identified; then, specific PCR primers have to be developed and tested. The use of random primers (as in RAPD analysis) avoids this up-front input, but the reproducibility of this approach is usually not up to the required standards, particularly between different laboratories. The detection of single nucleotide polymorphisms (SNPs) and insertion/deletion (InDel) polymorphisms has been simplified by recent developments in sequencing technology and is relatively straightforward even in crop species, particularly when many EST sequences are available. SNPs and InDels represent a virtually inexhaustible source of polymorphic markers. For Arabidopsis, 37 344 SNPs and 18 579 InDels have been identified (http://www.arabidopsis.org/Cereon/index.html). Our current knowledge of SNPs and their application to crop genetics has recently been reviewed [23].

The AFLP technique [24] is highly reproducible and does not require prior sequence information. It can thus be applied directly to any organism. Genomic restriction fragments are adaptor-ligated, and PCR primers are designed based on adaptor and restriction site sequences. Subsets of fragments can subsequently be specifically amplified by adding random selective nucleotides at the 3’ end of the primers (Figure 1). The number of selective nucleotides needed depends on the size of the genome studied. Thus, although it is based on restriction and random selection, the targeted subsets of PCR templates are highly specific and the procedure is robust. A great advantage is that one basic set of primers can be used for organisms with comparable genome sizes.

Map-based cloning in Arabidopsis

To initiate MBC of a gene, genome-wide mapping procedures have been developed based on SSLP markers [25] and SNPs [26], among other techniques. We have exploited the fact that multiple AFLP markers can be obtained per primer combination to develop an AFLP-based genome-wide mapping strategy [18,27]. Like all genome-wide mapping strategies, it is helpful at the first steps of the MBC process only, when no knowledge about the location of the gene of interest is available. Our strategy establishes linkage to a <6 Mb region within 3 days by analysing only 20–30 mutant individuals from a segregating population with eight primer combinations that provide a well-dispersed grid of 85 AFLP markers covering the genome. Further analysis of ~120 mutant individuals will typically exhaust the available Arabidopsis AFLP map [27] and identifies a 200–800 kb region within 1–2 weeks. In addition to linkage, the AFLP-based approach will show non-linkage to the rest of the genome, which is essential for recognizing situations involving gross chromosomal re-arrangements.
Before embarking on a large screen to select recombinants for fine mapping, it is crucial to confirm that there are no comparable (and cloned) mutants located in the identified region. The sequence-based map of Arabidopsis genes with mutant phenotypes [28] is a great help to this end. Potential candidates should be tested in allelism test crosses. Once it is established that an unidentified gene is being dealt with, fine mapping should be initiated. Because the advantage of AFLP (many physically dispersed markers per primer combination) is lost at this point, we recommend that selective amplification of microsatellite polymorphic loci (SAMPL) be performed. This method is based on oligonucleotide primers. Variation is based on size differences of the microsatellite.

Before embarking on a large screen to select recombinants for fine mapping, it is crucial to confirm that there are no comparable (and cloned) mutants located in the identified region. The sequence-based map of Arabidopsis genes with mutant phenotypes [28] is a great help to this end. Potential candidates should be tested in allelism test crosses. Once it is established that an unidentified gene is being dealt with, fine mapping should be initiated. Because the advantage of AFLP (many physically dispersed markers per primer combination) is lost at this point, we recommend that selective amplification of microsatellite polymorphic loci (SAMPL) be performed. This method is based on oligonucleotide primers. Variation is based on size differences of the microsatellite.
point, we advise designing other PCR-based markers as flanking markers. The InDel polymorphisms (http://www.arabidopsis.org/Cereon/index.html) are a practical source for designing such PCR markers but any other easy-to-use marker will do. The flanking PCR markers are tested on 1000–2000 individuals from a segregating population. The markers can be analysed simultaneously, which makes identifying recombinants in the region of interest easy [27]. All recombinant plants are selected and F3 seeds collected to determine whether wild-type recombinants are homozygous or heterozygous for the locus. The selected recombinants can subsequently be used to further delimit the area containing the gene of interest to a region of approximately 0.025–0.050 cM, corresponding to, on average, 5.5–11.0 kb. Markers necessary for this process can be developed from the available Cereon InDel and SNP collection [10]. During the process of MBC, it is important to keep checking the existing databases for candidate genes in the identified region because, for model organisms such as Arabidopsis, an increasing number of genes is being characterized, which implies that the chance that a gene has already been cloned is increasing. A diagram of the described MBC strategy is presented in Figure 2 and can be compared to similar schemes [10,19].

**Map-based cloning in other systems**

Although MBC is now relatively straightforward in Arabidopsis, these conclusions cannot be extrapolated directly to other systems. However, for the first steps of MBC projects, the AFLP-based genome-wide mapping procedure can confidently be used in any organism for which a dense AFLP map is available (e.g. barley [29], maize [30], tomato [31], lettuce [32], Petunia [33]) or can be created. When no AFLP map is available, the AFLP procedure can still be used because no prior sequence information is required. In case of a single-gene project, a useful modification of the strategy is first to identify linked markers using Bulked Segregant Analysis (BSA) [34], by performing a large set of AFLP primer combinations on DNA pools of mutants and wild types. Instead of AFLP markers, any other efficient marker system can be used in combination with BSA. As a result, one should be able easily to identify closely linked markers, enabling subsequent fine mapping, followed by joining of Bacterial Artificial Chromosome (BAC) sequences to develop a physical map. For fine mapping, microsynteny of the species under investigation with Arabidopsis (or rice) has been used successfully in some cases [35]. An interesting and, for plants, relatively new approach to

---

**Figure 1.** The amplified fragment length polymorphism (AFLP®) technique [24]. Genomic DNA is digested with a rare cutter (e.g. EcoRI) and a frequent cutter (e.g. MseI), which results in three types of fragments. Double-stranded adapters are ligated to the ends of the fragments to generate template DNA for amplification. Point mutations are incorporated into the adapter sequence to prevent restriction after successful ligation. Amplification of the fragments usually takes place in two steps called pre-amplification and selective amplification. Selective nucleotides are included at the 3' end of the PCR primers so that only a subset of the fragments will be amplified (* indicates a labelled primer). The number of selective nucleotides (indicated by +) used depends on the size of the genome under investigation. To avoid nonspecific amplification, the difference between the number of selective nucleotides in the pre- and selective amplification steps should be ≤ 2. The resulting fragments are visualized by polyacrylamide gel electrophoresis. Polymorphisms detectable on the gels are caused by mutation(s) in the restriction site(s), selective nucleotide(s) or insertion or deletions within the AFLP fragments. AFLP is a registered trademark of Keygene (http://www.keygene-products.com/).
identifying linked markers is linkage disequilibrium mapping in natural populations [23,36]. Although fine mapping in plants with large genomes remains difficult and labour intensive in comparison to Arabidopsis, a rather surprising development is taking place in MBC of genes from organisms with larger genomes. Although genome sizes can vary up to a 1000-fold, genetic maps (and number of genes) in general tend to be more or less equal in size. Where one thus would expect to get into trouble with a huge physical: genetic distance ratio for the larger genomes, there are now ample examples showing a remarkably favourable local ratio [8]. From the data obtained so far, it appears that in large genomes, genes are often positioned in gene-rich regions with a local Arabidopsis-like kb/cM ratio [37,38]. The mentioning of an average ratio then becomes irrelevant. In tomato, for instance, the average is 700 kb/cM [39] but, in specific cases, 1cM equalled ～100 kb [40], 160 kb [41], 280 kb [42] and even 5 kb [43]. By contrast, in low-recombination regions, the gene density might be equally low, thus restricting candidate gene analysis to one or a few genes per 100 kb. Overall, it might appear that cloning a gene from large genomes requires an effort similar to that of cloning from small genomes, once sequence information and marker availability improves for the species under investigation.

### Mapping applications

For applications such as MARKER-ASSISTED SELECTION (MAS) [44] and breeding by design [45], the definition of a limited interval is sufficient and one can stop once the segment carrying the gene of interest has been minimized to satisfaction. This application can also be used in plants with larger genomes. If one decides to proceed ultimately to isolate the gene of interest, some points should be taken into consideration. In principle, the mapping procedure can be continued until two markers flank the gene of interest. Even for Arabidopsis, this would require the screening of an extremely large population, which effectively disqualifies this approach because it is impractical. A more sensible strategy is to aim to identify a relatively small region (e.g. 50 kb in Arabidopsis) containing the

---

**Figure 2.** A map-based cloning procedure. The time (left) and material (right) needed to map a mutation in Arabidopsis thaliana is given for each step in the procedure. See Ref. [18] for the AFLP map referred to in steps 2 and 3. For the sequence-based map of Arabidopsis genes with mutant phenotypes referred to in step 4, see Ref. [38]. See http://www.arabidopsis.org for the TAIR database; see http://www.arabidopsis.org/Cereon/index.html for further details of InDel polymorphisms and SNPs. Abbreviations: AFLP, amplified fragment length polymorphism; TAIR, The Arabidopsis Information Resource; InDel, insertion/deletion; SNP, single-nucleotide polymorphism.
gene of interest. In **Arabidopsis**, a 50 kb region on average contains between four and ten genes. One way to identify the gene of interest is to order the appropriate TILLING [5] lines and/or lines in which the candidate genes are tagged by a T-DNA or transposon. A comprehensive list of sequence tagged lines can be found at the website of the Nottingham **Arabidopsis** Stock Centre (http://nasc.life.nott.ac.uk/); for obtaining TILLING lines, one should consult the **Arabidopsis** TILLING project website (http://tilling.fhcrc.org:9366/). Studying the phenotypes of such lines and eventually performing allelism tests might direct you to the desired gene. Another rational approach is to transform the mutant with subclones of the BAC(s) containing the wild-type version of the gene of interest and to test whether the clone used for transformation can complement the mutant of interest. Recently, BIBAC (binary-BAC) transformation libraries for **Arabidopsis** and rice have been presented [46,47]. When the gene is eventually pinpointed, sequencing the mutant allele(s) and comparing these to the wild-type sequence will definitely identify the desired gene.

**Conclusion**

Reverse mutagenesis screens are popular because of their genome-wide reach and theoretical promise to capture all genes involved in a particular process. However, lack of phenotypes is a general and disappointing feature of these approaches. Here, we have argued that well-designed forward mutagenesis screens together with a balanced approach. Here, we have argued that well-designed forward mutagenesis screens together with a balanced strategy for MBC are an excellent approach for gene function analysis. Although this is presently true only for model plant systems such as **Arabidopsis** and rice, once we have more sequence and marker availability it might also become the case for crop plants.

**References**

56 Grelon, M. et al. (2001) ATSP01-1 is necessary for efficient meiotic recombination in plants. EMBO J. 20, 589–600
62 Bell, C.J. and Ecker, J.R. (1994) Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. Genomics 19, 137–144
63 Zietkiewicz, E. et al. (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20, 176–183
71 Williams, J.G. et al. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18, 6531–6535
76 Witsenboer, H. et al. (1998) Identification, genetic localization and allelic diversity of selectively amplified microsatellite polymorphic loci (SAMPL) in lettuce and wild relatives (Lactuca spp.). Genome 40, 923–936

Free journals for developing countries

The WHO and six medical journal publishers have launched the Access to Research Initiative, which enables ~70 developing countries to gain free access to biomedical literature through the Internet. The science publishers, Blackwell, Elsevier Science, the Harcourt Worldwide STM group, Wolters Kluwer International Health and Science, Springer-Verlag and John Wiley, were approached by the WHO and the British Medical Journal in 2001. Initially, >1000 journals will be available for free or at significantly reduced prices to universities, medical schools, research and public institutions in developing countries.

The second stage involves extending this initiative to institutions in other countries.

Gro Harlem Brundtland, director-general for the WHO, said that this initiative was ‘perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries’.

See http://www.healthinternetwork.net for more information.