technical focus

# Marker-assisted selection: new tools and strategies

The development of molecular genetics and associated technology has facilitated a quantum leap in our understanding of the underlying genetics of the traits sought through plant breeding. The usefulness of DNA markers for germplasm characterization, and of marker-assisted selection – the manipulation through DNA markers of genomic regions that are involved in the expression of traits of interest – for single-gene transfer, has been well demonstrated. However, when several genomic regions must be manipulated, marker-assisted selection has turned out to be less useful. The efficient and effective application of marker-assisted selection for polygenic trait improvement certainly needs new technology but, more importantly, it requires the development of innovative strategies that bypass the conceptual bottlenecks imposed by current approaches.

Since the beginning of agriculture, humans have sought to improve crops by selecting for desired traits. Although the process is now much more sophisticated, it is still mostly based on field observations. The challenge now is to integrate new molecular technology into breeding schemes and to develop efficient marker-assisted selection strategies aimed at plant improvement. With the development of molecular tools and the first molecular genetic maps for plants, marker-assisted selection has become possible, whether these traits are controlled by identified genes or quantitative trait loci (QTLs). During the past decade, the developing ability to transfer target genomic regions using DNA markers resulted in extensive mapping experiments aimed at the development of marker-assisted selection[1,2]. There is now a large amount of research that addresses marker-assisted selection in some form. However, there are few data on marker-assisted selection experiments: the majority of work is aimed at identifying genomic regions of interest, from which marker-assisted selection experiments are an attractive 'next step'.

## Success stories

The efficiency of marker-assisted selection experiments for the transfer of a single target region has been reported for several plant genomes, the integration of the *Bt* transgene into different genetic backgrounds being a good example[3]. When the expression of a target trait is regulated by a single gene, or by a gene responsible for a high percentage of the phenotypic variance of the trait, the transfer of a single genomic region from a donor to a recipient line can produce significant trait improvement. By making an allelic map of the genome with DNA markers, plants presenting a 'better' genome composition can be efficiently identified. The target genome must have the donor allele at the target segment and non-target regions should have the largest possible proportion of the recurrent genome (that of the line to be improved). Compared with conventional backcrossing, the use of DNA markers thus offers two advantages:

- Faster recovery of the recurrent genome.
- More efficient selection of genomes that have recombination events close to the target gene (Fig. 1).

Currently, marker-assisted selection of single alleles is perhaps the most powerful approach that uses DNA markers effectively. Its success is especially important in the field of genetic engineering, now that transformation methods are available for most plant species but transformation efficiency is still very germplasm dependent.

Another area in which the use of DNA markers has been successfully reported is in the movement of one or more alleles of interest from a wild relative into an elite cultivar[5]. These and similar results demonstrate the agronomic potential of favorable alleles present in wild relatives. Considering the allelic diversity present in nature and in germplasm banks, such results open new doors for investigations based on marker-assisted selection.

The use of DNA markers also makes the process of selecting parental lines more efficient. Based on the genetic diversity calculated from fingerprinting data, plant material can be classified into genetic pools. This information can be extremely helpful for identifying the most appropriate parental lines to be crossed. For crop species in which heterosis is exploited through the production of hybrid cultivars, measurements of genetic distance based on DNA markers can be useful for predicting the yields of crosses between lines from the same germplasm group[6]. When combined with phenotypic data, DNA marker-associated effects evaluated at an early generation testcross can be used efficiently to predict later-generation testcross performance[7]. Thus, DNA markers can have a strong impact on breeding programs by reducing the number of lines that have to be tested.

## The intriguing case of polygenic trait improvement

Most traits of agronomic importance are complex and regulated by several genes, with yield among the most polygenic and complex. Compared to a 'simpler' trait controlled by one or a few genes, the improvement of polygenic traits through marker-assisted selection raises more questions[2], as demonstrated by the abundance of studies based on computer simulations[8,9] and the paucity of published data on the topic[10]. In fact, no experiment has clearly demonstrated whether using DNA markers for quantitative trait improvement is superior to conventional breeding selection, although some encouraging results have been published[11,12].

The difficulty of manipulating quantitative traits is related to their genetic complexity – principally the number of genes involved in their expression and the interactions between genes (epistasis). Because several genes are involved in the expression of polygenic traits, they generally have smaller individual effects on the plant phenotype. This implies that several regions (QTLs) must be manipulated at the same time in order to have a significant impact, and that the effect of individual regions is not easily identified. For this reason, repetitions of field tests are required to characterize accurately the effects of QTLs and to evaluate their stability across environments. Although significant QTL effects should be detected across several environments, the evaluation of the QTL by environmental interactions (Q×E) remains a major constraint on the efficiency of marker-assisted selection[13]. Regarding gene interactions, there remain clear limitations on evaluating epistatic effects between different regions of a genome. Epistatic interactions can induce a skewed evaluation of QTL effects, and if all the genomic regions involved in the interactions are not incorporated in the selection scheme, they can bias the selection process.

A major issue that needs to be addressed is how to increase the efficiency of marker-assisted selection for quantitative traits through better characterization of target genes. Fortunately, field design[14] and statistical approaches for QTL mapping have steadily progressed during the past decade. With the latest mathematical methods, such as composite interval mapping, field data from different environments can be integrated into a joint analysis to evaluate the Q×E and thus identify 'stable' QTLs across environments[15]. Moreover, when combined with a detailed linkage map, composite interval mapping allows a precise identification of the QTL in the genome and a better identification of coupled QTLs (linked QTLs at which the favorable allele comes from the same parental line). The manipulation of QTLs in repulsion (linked QTLs at which the favorable allele comes from a different parental line) is a typical example of how DNA markers can be used very efficiently to select those genotypes that have broken the linkage between two QTL regions at an early stage of


 1360 - 1385/98/$19.00 PII: S1360-1385(98)01240-0

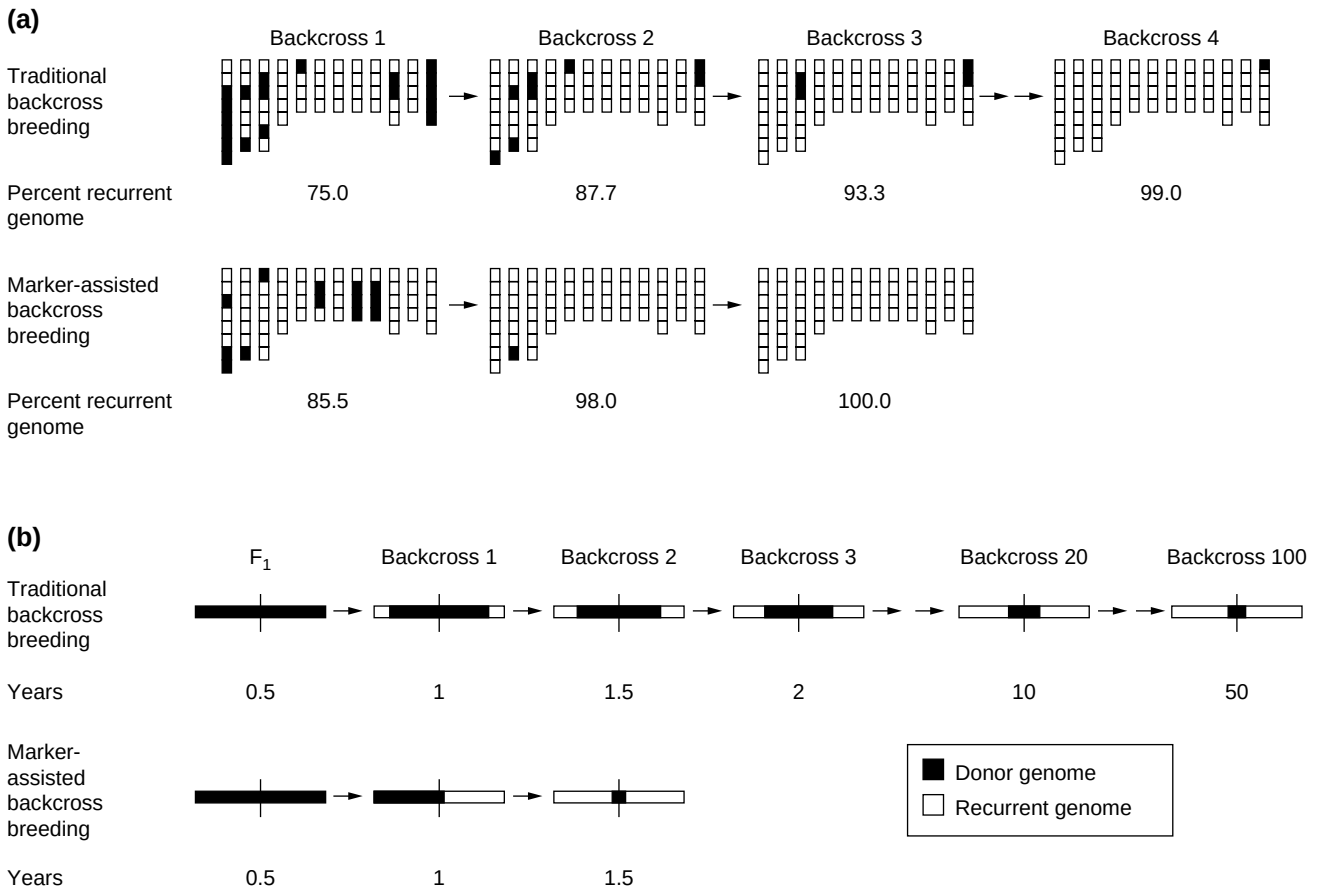



**Fig. 1.** Comparison through simulation of backcross breeding using either a traditional approach or marker-assisted selection. For traditional backcross breeding, 'graphical' genotypes were generated for randomly selected individuals from various backcross generations derived from a single backcross 1 individual. For marker-assisted selection, graphical genotypes were generated from a simulated RFLP-assisted backcross breeding program, 30 progeny were generated at each backcross generation and the best genotype was used as the parent for the next generation. (a) Rate of return to the recurrent parent genome in regions unlinked to genes being introgressed, indicating that marker-assisted selection achieves complete conversion in only three backcrosses as compared to a minimum of six for conventional selection. Chromosomes are indicated by vertical bars. (b) Rate of return to the recurrent parent genome in regions of the chromosome flanking the target gene, indicating that marker-assisted selection achieves the same level of conversion in only two backcrosses as would be achieved in 100 backcrosses through conventional methods. *Modified from Ref. 4.*

recombination. However, although the process of QTL identification has improved significantly, marker-assisted selection is still limited by three main factors:

- The number of samples that can be analyzed.
- The number of lines that can be improved within a given time.
- The belief that QTL identification is required whenever additional germplasm is used.

We believe that solutions exist to get past these constraints.

### PCR-based markers: a significant step forward

New and reliable PCR-based markers, such as sequence-tagged sites and microsatellites[16], are now available for several plant genomes. The use of microsatellites allows the detection of a high level of polymorphism, independent of the species considered, because they target highly variable regions of the genome. These markers represent important tools for dissecting the genomes of species with a low level of polymorphism. PCR-based markers have also opened new doors for genome manipulation, since their use allows:

- Earlier sampling, because of the small amount of tissue required.
- Faster DNA preparation, because of the small amount of template DNA required.
- More efficient handling of large sample sizes, because of the efficiency of PCR technology.

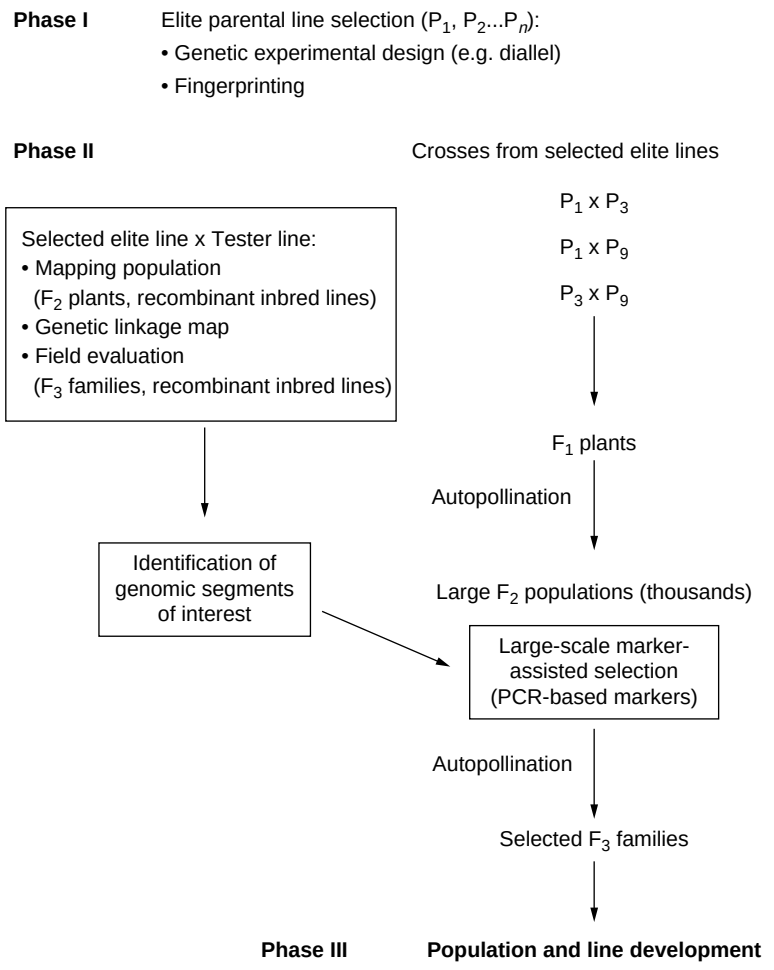
Thousands of plants at each cycle of selection can be screened without the use of sophisticated equipment[17]. An increase in the size of the screened population improves the efficiency of the selection at each cycle, thus reducing the total number of selection cycles required. However, the jump from small- to large-scale selection, although always increasing efficiency, is not always required, considering the additional cost and time requirements. The efficiency of screening a large population increases with the pressure of selection imposed by the model, as defined by the number of genomic regions to be transferred and, to a lesser extent, the chromosomic composition of the genome.

If a jump in the population size under study makes marker-assisted selection more efficient when several QTLs have to be transferred from a donor to a recipient line, the limitations related to the number of crosses that can be manipulated at the same time and the QTL identification per cross remain present. To overcome these constraints, new marker-assisted selection strategies must be developed.

### New strategies

To date, several simulation studies have looked at the integration of molecular marker

**Box 1. New theoretical selection scheme for marker-assisted selection**




In this new marker-assisted selection approach, the selection of suitable parental lines (Phase I) and the development of new lines (Phase III) is overseen by crop breeders, while DNA markers are used at an early stage of recombination to fix alleles at selected genomic regions (Phase II). Selection of parental lines is conducted among outstanding elite material for the trait to be improved to identify elite lines with the best allelic complementarity. By crossing each selected parental line with a tester (elite material lacking the target trait), segregating populations are developed. Genomic regions of interest for each parental line can be identified by combining the allelic segregation (e.g. $F_2$ plants and recombinant inbred lines) and field performance (e.g. $F_3$ families and recombinant inbred lines) of those segregating populations. The marker-assisted selection step, based on the use of reliable PCR-based markers to fix favorable alleles at target genomic regions, is conducted only once on large segregating populations derived from crosses between elite lines. This new strategy offers two major advantages. First, favorable alleles selected to improve a specific trait are derived from two or more sources of elite parental materials in a complementary scheme, disregarding the 'recipient/donor' line concept. Second, plants with fixed favorable alleles at specific genomic regions are selected at an early generation of recombination; no pressure of selection is applied outside the targeted regions. This assures good allelic variability in the rest of the genome for future line development under various conditions and environments.

technology into conventional breeding strategies. In practice, the backcross approach has been widely adopted, but presents limitations in terms of outputs that fit breeder requirements when several QTLs have to be manipulated. Now is the time to consider the development of new breeding strategies that take into account genetic characteristics, such as: the complexity of the genome; the nature and number of molecular markers available; the complexity of the traits to be improved; the number of plants that can be screened at each selection step; and the number of populations that can be concurrently manipulated. It is also crucial to explore the complementarity between marker-assisted selection and conventional breeding, and to develop overall strategies that tightly and interactively integrate the two approaches. In facing the challenge of improving several lines for quantitative traits in the same scheme or in parallel, marker-assisted selection strategies should probably concentrate on using DNA markers in one key selection step to maximize their impact. They could be used at the very beginning of the scheme as predictive tools to reduce the number of crosses; at an early stage of recombination to fix target genomic regions (Box 1); or at an advanced stage of germplasm development as a diagnostic tool (when the allelic value has already been identified).

## The future

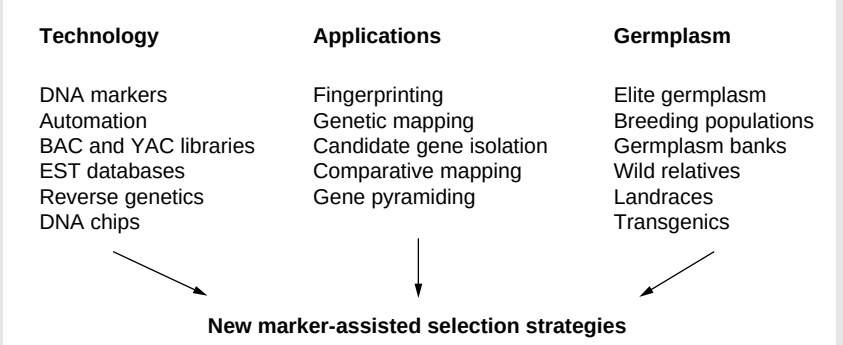
Marker-assisted selection for polygenic trait improvement is in an important transition phase, and the field is on the verge of producing convincing results. Given the plethora of ongoing experiments and the explosion of new molecular technology and applications, new or improved selection schemes should be developed and applied very soon (Box 2). Notable among these developments will be the multiplication of QTL studies and the expanded identification of gene functions using expressed sequence databases and reverse genetic analyses[18]. More data will make gene manipulation easier and more efficient[9], while also enabling the development of new genome concepts such as the clustering of genes with similar developmental functions[19]. Recent efforts in comparative genetic analysis allow the identification across different plant species of gene sequences involved in the expression of target traits. The superior alleles identified among genomes at those target genes can be used as DNA markers to develop efficient screening techniques[20]. Finally, technological developments, including automation, allele-specific diagnostics and DNA chips, will make marker-assisted selection approaches based on large-scale screening much more powerful and effective.

Today, the optimism of a decade ago has been tempered somewhat by constraints encountered by some current marker-assisted selection approaches. However, considering the potential for the development of new strategies, the future for polygenic trait improvement through DNA markers, and the contribution of this to plant breeding efforts worldwide, appears bright.

## Acknowledgements

Our thanks to Javier Betran, Daniel Grimanelli, Mireille Khairallah and Olivier Leblanc for their helpful comments and suggestions on the manuscript. The assistance of David Poland in the writing of the paper is also gratefully acknowledged.

**Box 2. Factors involved in the development of new marker-assisted selection strategies**




Development of any new marker-assisted selection strategy will need to consider progress in molecular technology, new applications of molecular genetics and the diversity of germplasm available. Abbreviations: BAC, bacterial artificial chromosome; YAC, yeast artificial chromosome; EST, expressed sequences tag.

**Jean-Marcel Ribaut*** **and David Hoisington**
International Maize and Wheat Improvement Center (CIMMYT), Lisboa 27,
Apdo. Postal 6-641, 06600 Mexico D. F., Mexico

*Author for correspondence (tel +52 595 54400; fax +52 595 54425; e-mail jribaut@cimmyt.mx)

**References**

**1** Dudley, J.W. (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits, *Crop Sci.* 33, 660–668
**2** Lee, M. (1995) DNA markers and plant breeding programs, *Adv. Agron.* 55, 265–344
**3** Ragot, M. *et al.* (1995) Marker-assisted backcrossing: a practical example, in *Techniques et Utilisations des Marqueurs Moléculaires* (Les Colloques, Vol. 72) (Berville, A. and Tersac, M., eds), pp. 45–56, Institut National de la Recherche Agronomique
**4** Tanksley, S.D. *et al.* (1989) RFLP mapping in plant breeding: new tools for an old science, *Biotechnology* 7, 257–264
**5** Tanksley, S.D. and McCouch, S.R. (1997) Seed banks and molecular maps: unlocking genetic potential from the wild, *Science* 277, 1063–1066
**6** Melchinger, A.E. *et al.* (1992) Genetic diversity for RFLPs in European maize inbreds: II. Relation to performance of hybrids within versus between heterotic groups for forage traits, *Theor. Appl. Genet.* 84, 672–681
**7** Eathington, S.R. *et al.* (1997) Usefulness of marker–QTL associations in early generation selection, *Crop Sci.* 37, 1686–1693
**8** Lande, R. and Thompson, R. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits, *Genetics* 124, 743–756
**9** Hospital, F. and Charcosset, A. (1997) Marker-assisted introgression of quantitative trait loci, *Genetics* 147, 1469–1485
**10** Mohan, M. *et al.* (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants, *Mol. Breed.* 3, 87–103
**11** Stuber, C.W. (1995) Mapping and manipulating quantitative traits in maize, *Trends Genet.* 11, 477–481
**12** Johnson, G.R. and Mumm, R.H. (1996) Marker assisted maize breeding, in *Proceedings of the Fifty-first Annual Corn and Sorghum Research Conference* (Wilkinson, D., ed.), pp. 75–84, American Seed Trade Association
**13** Beavis, W.D. and Keim, P. (1996) Identification of quantitative trait loci that are affected by environment, in *Genotype-by-environment Interaction* (Kang, M.S. and Gauch, H.G., eds), pp. 123–149, CRC Press
**14** Gleeson, A.C. (1997) Spatial analysis, in *Statistical Methods for Plant Variety Evaluation* (Kempton, R.A. and Fox, P.N., eds), pp. 68–85, Chapman & Hall
**15** Jiang, C. and Zeng, Z-B. (1995) Multiple trait analysis of genetic mapping for quantitative trait loci, *Genetics* 140, 1111–1127
**16** Powell, W. *et al.* (1996) Polymorphism revealed by simple sequence repeats, *Trends Plant Sci.* 1, 215–222
**17** Ribaut, J-M. *et al.* (1997) Use of STSs and SSRs as rapid and reliable preselection tools in a marker-assisted selection-backcross scheme, *Plant Mol. Biol. Rep.* 15, 154–162
**18** Bensen, R.J. *et al.* (1995) Cloning and characterization of the maize *An1* gene, *Plant Cell* 7, 75–84
**19** Khavkin, E. and Coe, E. (1997) Mapped genomic locations for developmental functions and QTLs reflect concerted groups in maize (*Zea mays L.*), *Theor. Appl. Genet.* 95, 343–352
**20** Sorrells, M.E. and Wilson, W.A. (1997) Direct classification and selection of superior alleles for crop improvement, *Crop Sci.* 37, 691–697

## book reviews

# Nematode parasitism

**Cellular and Molecular Aspects of Plant–Nematode Interactions**
*edited by C. Fenoll, S. Ohl and F. Grundler*
Kluwer, 1997. £76.00/$124.00 hbk
(vii + 286 pages) ISBN 0 7923 4637 8

Plant-parasitic nematodes are obligate feeders on plants, and interactions with their hosts can be very complex and dynamic. Particularly intriguing are the sedentary endoparasitic nematodes, which induce profound modifications in root-cell phenotype and function to form permanent feeding sites. These nematodes are among nature's most successful parasites and elicit some of the most elaborate responses in plant tissue of any parasite. Concerns for sustainable and environmentally safe nematode management tactics, combined with the tools of modern biology, have stimulated a research emphasis on the molecular biology of nematodes and their interactions with plants. This activity is rapidly expanding our knowledge of plant–nematode interactions, and the progress being made is reviewed in this book.

The molecular analysis of plant–nematode interactions is being approached in two ways: elucidation of the molecular responses elicited in the infected plant tissues; and characterization of the molecular signals from the nematode that trigger the plant responses. The driving force for understanding the molecular basis of these interactions is the development of novel target-specific resistance strategies that are economical and environmentally benign.

The intensely studied *Arabidopsis* is being used as a model host for dissecting the cellular and molecular responses of the plant to nematode parasitism. These studies have provided new information on nematode feeding behavior and form the beginning of a detailed analysis of host-gene activity during nematode infection and feeding-site development. Several chapters discuss the recent developments in understanding the molecular responses to parasitism by sedentary endoparasitic nematodes in *Arabidopsis* and other hosts. Transcriptional changes in nematode-parasitized roots include expression of cell cycle genes, genes critical for feeding-site function, wound-response genes and other genes of unknown function. Although our knowledge of plant genes whose expression is altered during nematode infection is rapidly increasing, little is yet known about the regulatory mechanisms involved.

Sedentary endoparasitic nematodes appear to have the ability to alter host gene expression locally. Here again, the mechanisms by which