

# Molecular Markers in Plants

COPYRIGHTED MATERIAL



# 1 Evolution of DNA Marker Technology in Plants

Robert J. Henry

## Contents

Introduction	4
Early Marker Technologies	4
Evolving Range of Applications of DNA Markers in Plants	12
Applications	13
Future Developments	15
References	15

## Introduction

Genetic markers are key tools for plant identification and plant improvement (Henry, 2001). Genetic marker technology has evolved rapidly with early methods based on phenotyping or isozymes being replaced by DNA-based methods of increasing sophistication. Early markers were few in number and difficult to assay. The ultimate development of the technology will provide simple methods to assess all genetic variation in the genome. This chapter provides a brief account of the development of genetic marker technology and its application to plants over the last 30 years. This perspective is provided as background and context for the accounts of the latest technologies (Henry *et al.*, 2012) to follow in subsequent chapters.

Molecular marker technology has evolved through several phases. Early methods employing non-DNA-based methods were replaced by DNA-based methods as the technologies for DNA analysis improved. Early hybridization-based methods were displaced rapidly following the development of polymerase chain reaction (PCR). PCR-based methods greatly increased the feasibility of high-throughput marker screening. Early PCR-based methods relied upon arbitrary primers because of a lack of sequence information for many species. These in turn were overtaken by the widespread adoption of more robust microsatellite or simple sequence repeat (SSR) markers. Single nucleotide polymorphisms (SNPs) have more recently replaced SSR markers as larger volumes of sequence data became available (Henry, 2008). Second-generation sequencing technologies have greatly accelerated the move to sequence-based markers. Ongoing improvements in DNA sequencing promise a continued convergence of sequencing and genotyping technologies. Third-generation sequencing promises delivery of technology for routine sequencing of even complex plant genomes enabling ready marker discovery and analysis.

Molecular markers have a wide range of applications in biological systems including plants. Molecular markers are very useful in identification of plants and in determining the relationships between plants. Plant identification may be important in plant breeding, plant production and processing, policing of intellectual property rights, and forensic applications. Determination of genetic relationships is required in evolutionary and conservation genetic analyses and in selection of germplasm in plant breeding. Plant breeding is often directly supported by marker-assisted selection.

This chapter will outline the evolution of molecular marker techniques and their applications to plants.

## Early Marker Technologies

Biochemical markers (e.g., isozymes) have been widely applied to the analysis of the genetics of plant populations. This approach has provided

cost-effective options for laboratories with little equipment studying poorly known biological systems. The most common of these techniques was the assay of isozymes.

Improving DNA analysis methods with greater discrimination, simplicity of analysis, and suitability for automation have largely replaced these non-DNA-based methods. However, early biochemical approaches have still been in limited use despite recent dramatic advances in DNA analysis technologies. Isozyme analysis has persisted for species for which little or no DNA sequence data was available and in laboratories not equipped with facilities for DNA analysis or where labor costs are low compared with reagent costs. Recent advances in DNA analysis tools should provide cost-effective and preferable DNA-based methods for almost all applications.

### ***DNA-Based Methods***

The development of DNA analysis methods provided an opportunity to directly analyze difference in the genome of the organism rather than rely on inference from analysis of expressed genes (as in isozyme analysis). Early DNA-based methods used hybridization of DNA to detect variation in the DNA samples. The development of restriction fragment length polymorphism (RFLP) analysis provided an approach that was widely adopted in the 1980s and became the standard approach until replaced by PCR-based methods in the 1990s. The evolving DNA-based methods are listed in Table 1.1. DNA extraction from plants is generally much more difficult than from typical animal sources because of the rigid cell wall and the high levels of secondary metabolites in plant cells. This requires tissue disruption techniques to break the cell wall that are likely to shear the DNA if too vigorous. Improved DNA extraction methods made a significant contribution to the advances in application of DNA analysis methods in plants (Thompson and Henry, 1993; Graham, Mayer, and Henry, 1994).

**Table 1.1 Evolving genotyping methods.**

<b>Method</b>	<b>Acronym</b>
Restriction fragment length polymorphism	RFLP
Random amplified polymorphic DNA	RAPD
Amplified fragment length polymorphism	AFLP
Diversity arrays	DArT
Sequence characterized amplified region	SCAR
Simple sequence repeat	SSR
Single nucleotide polymorphism	SNP

### ***Restriction Fragment Length Polymorphism***

RFLP analysis used one or more endonucleases (restriction enzymes) to digest genomic DNA from the sample. Differences in DNA sequence in the sample at the restriction site or between adjacent restriction sites resulted in differences in the length of the fragment in the digest. The digested DNA was separated by electrophoresis. The DNA was then transferred to a membrane for analysis of fragments. Specific probes derived from genomic DNA or cDNA were labeled and used to detect fragments in the digest containing related sequences. Labeling was with radioisotopes or later specific proteins that could be detected using antibodies. Membranes could be hybridized with a series of different probes to explore polymorphisms in different parts of the genome. The limitations of these early methods included the need for large amounts of DNA especially for species with large genome sizes.

### ***PCR-Based Methods***

The development of the PCR in the late 1980s provided a new tool that rapidly changed approaches to DNA analysis including DNA-based markers. PCR based techniques replaced the earlier non-PCR methods because of their greater sensitivity, discrimination power and ease of automation. The use of PCR has greatly accelerated the development and application of DNA markers in plants. PCR-based marker analysis has facilitated much greater automation of analysis improving throughput and reliability of analysis. PCR also allowed the use of much smaller quantities of DNA for plant DNA analysis. This greater sensitivity together with the high specificity of PCR resulted in rapid replacement of earlier methods. Simpler DNA extraction methods could be developed for use with PCR because of the need for less DNA that did not need to be of high molecular weight. The small amount of sequence data available at the time limited the application of PCR to amplification of specific or known genetic sequences and required the development of techniques based upon arbitrarily primed PCR to generate markers.

### ***Arbitrary Methods***

A lack of DNA sequence data for most species was overcome by use of PCR primers of arbitrary sequence. These primers were used to generate a fingerprint for the genotypes. Random amplified DNA polymorphism (RAPD) was followed by amplified fragment length polymorphism (AFLP) and DNA amplification fingerprinting (DAF). The diversity array technique method also generated large numbers of markers without sequence data.

### ***Random Amplified DNA Polymorphism***

The RAPD method used short (10 mer) primers to amplify fragments for analysis by gel electrophoresis (Williams *et al.*, 1990). This method required careful standardization of conditions for the low stringency amplifications involved and as a result was difficult to repeat in different laboratories.

Other variations on this approach included the DAF method.

### ***Amplified Fragment Length Polymorphism***

The AFLP method used restriction enzymes (Zabeau and Vos, 1993) to produce fragments. Oligonucleotides were ligated to the ends to produce priming sites for amplification of a subset of fragments. This produced complex mixtures of fragments. AFLP was widely used for DNA fingerprinting. The method persisted in use for longer than the RAPD method because of the greater reproducibility of the technique.

### ***Diversity Array Technique***

The most recent of these methods, not based on specific sequence targets, is the diversity array technique (Jaccoud *et al.*, 2001). In this method, a genomic representation of a species is arrayed for analysis by hybridization. This technique provided very large numbers of markers and has been widely applied especially in construction of genetic maps (Xia *et al.*, 2005). This method generates very large numbers of useful markers but requires significant development effort for each species.

### ***Specific Sequence-Based PCR***

As increasing amounts of sequence data became available from Sanger sequencing, PCR amplification has been widely applied to the analysis of known sequence polymorphisms (Garland *et al.*, 2000; McIntosh, Pacey-Miller, and Henry, 2005). Differences in the length of the amplified fragment or differences in the sequence of the amplified fragment can be detected by a range of methods: restriction digestion, melting temperature analysis (Shepherd *et al.*, 1998), and hybridization with a labeled probe. Differences in the primer site can be used to develop assays that only allow amplification from specific target sequences.

### ***Sequence Characterized Amplified Region***

Sequence characterized amplified region makers were often derived from RAPD, AFLP, or other markers. Sequencing of the amplified fragment was used to design a PCR with highly specific primers. This allowed

conventional robust PCR to be used to assay the marker rather than continuing to rely on the more complex or less reproducible arbitrary primed PCR used in the early methods. These early sequence-based markers indicated the way that markers would develop as sequence data became more readily available.

### ***Reverse Transcription–Polymerase Chain Reaction***

Further development of PCR marker methods included the monitoring of the progress of PCR by direct detection approaches as an alternative to electrophoretic separation of the products for analysis. A range of methods employing these real-time PCR methods have been adapted for the detection of DNA polymorphisms as markers in plants. These often employ specific probes to detect the PCR product (Kennedy *et al.*, 2006b; Bundock *et al.*, 2006).

### ***Simple Sequence Repeat***

Short repeated sequences are found widely in the genome and show great variation in most populations making them very useful markers. PCR also accelerated the analysis of these SSRs or “microsatellites.” Most SSR analysis targeted dinucleotide or trinucleotide repeats.

SSR markers have been widely applied in plants (Cordeiro, Taylor, and Henry, 2000). Analysis required that the sequences of the DNA on either side of the repeat sequence be determined to allow design of specific primers for PCR amplification (Ablett, Hill, and Henry, 2006). Analysis of variations in the length of the amplified fragment was originally achieved by gel electrophoresis but later was widely conducted using capillary electrophoresis systems including those developed for DNA fragment length separation in DNA sequencing by the Sanger method (Rossetto, McNally, and Henry, 2002). Capillary electrophoresis allowed automated data capture and facilitated analysis of much larger numbers of samples. Analysis of many loci for each sample was achieved efficiently by combining loci with very different lengths and with primers labeled with different colored fluorophores in the same capillary. Instruments with up to 96 capillaries were used to achieve high throughput.

Innovative approaches were required to discover SSR loci and obtain the flanking sequence data needed for their analysis. Early markers were based on sequencing libraries of genomic clones isolated by hybridization with repeat sequences. As more DNA sequence data have become available, these have been identified efficiently in gene sequences (e.g., EST data; Ablett, Hill, and Henry, 2006) and more recently whole-genome sequence data.



### **Single Nucleotide Polymorphism**

The ultimate minimum unit of genetic variation is the single nucleotide. Analysis of SNP markers has increased as more sequence data has become available (Henry and Edwards, 2009). Often insertions and deletions (indels) are assayed as SNP.

### **Discovery**

The discovery of SNP has been achieved in a variety of ways but is now probably best achieved by DNA sequencing using the latest technologies (Arai-Kichise, 2011; Barbazuk *et al.*, 2007; Deschamps and Campbell, 2010; Imelfort *et al.*, 2009; Novaes *et al.*, 2008; Trebbi *et al.*, 2011; Trick *et al.*, 2009; Yamamoto *et al.*, 2010). Early techniques relied upon a range of mutation detection methods (Cross *et al.*, 2008).

### **Analysis**

Protocols for the analysis of SNP have evolved (Pacey-Miller and Henry, 2003) and have until recently been limited by the availability of known SNP in most organisms. This situation has changed with recent advances in DNA sequencing and we may now need to develop new SNP analysis tools to adequately explore variation in the very large numbers of SNP that can be easily discovered by sequencing.

Platforms for automated SNP analysis allow large numbers of samples and SNP to be assayed (Pattimore *et al.*, 2010; Sexton *et al.*, 2010; Kharabian-Masouleh *et al.*, 2009). Some are most effective for very large numbers of samples and others for very large numbers of SNPs. The mass array platform (Sequenom) automates the processing of samples in 384 well format with analysis of up to 40 SNP per sample by mass spectrometry of primer extension reactions designed to distinguish the SNP. Illumina platforms are designed for automation of larger numbers of SNP. These platforms allow very large numbers of SNP to be screened with panels of SNP defined for each species or system.

### **Impact of Advancing DNA Sequencing Technology**

DNA sequencing has been subjected to continual improvement with method based on conventional Sanger sequencing predominating until the last few years. Since then, second-generation (or next generation) sequencing technologies have dramatically increased sequencing efficiencies (Varshney and May, 2012). Further advances in DNA sequencing (Thudi *et al.*, 2012) promise to continue the trend to delivery of larger volumes of DNA sequence data at lower costs.

**Table 1.2 Sequence-based approaches to marker discovery.**

Whole-genome sequencing	Krishnan <i>et al.</i> (2012)
Organelle sequencing	Nock <i>et al.</i> (2011)
Transcriptome sequencing	Gillies, Furtado, and Henry (2012)
Enriched genome sequencing	Bundock, Casu, and Henry (2012)

A range of applications of these new sequencing technologies to marker discovery in plants are described in Table 1.2.

### Whole-Genome Sequencing

Assembly of complex plant genomes remains difficult (Souza *et al.*, 2011). Many genomes have now been assembled using next-generation sequencing data (TPGSC, 2011; Kim *et al.*, 2010; Chan *et al.*, 2010; Argout *et al.*, 2010; Lai *et al.*, 2010; Yan *et al.*, 2011). Shotgun resequencing of whole genomes of genotypes is now relatively simple for species with a high-quality reference genome sequence (Krishnan *et al.*, 2012). Short sequence reads can be aligned to the reference genome sequence for analysis of SNP and indels.

For example, resequencing of six lines of rice identified 2 819 086 SNP, 160 478 insertions, and 163 556 deletions relative to the Nipponbare reference genome (Krishnan *et al.*, 2012). These were parents used in hybrid production allowing all polymorphisms between the parents to be defined for possible crosses within this group.

### Organellar Sequencing

Sequencing of organelle genomes has widespread application in research on evolutionary relationships in plants (Zhang, Ma, and Li, 2011; Whittall *et al.*, 2010; Doorduyn *et al.*, 2011). Maternal genome markers may be found in non-nuclear genomes (Table 1.3). Sequencing of organellar genomes (mitochondria and chloroplast) has required the isolation of the organelle or the specific amplification of sequences from the organelle. Advances in DNA sequencing now allow a simpler approach of analysis of the

**Table 1.3 Sources of DNA for marker analysis in plants.**

	Genome size in rice (bp)
Nucleus	382 000 000
Mitochondria	491 000
Chloroplast	135 000

organellar sequence using a total shotgun sequence of DNA from the sample (Nock *et al.*, 2011).

For example, sequencing of total rice DNA allowed a complete chloroplast sequence to be extracted and revealed a single base error in the published sequence for cultivar Nipponbare. This error in the reference genome was confirmed by Sanger sequencing (Nock *et al.*, 2011).

### **Transcriptome Sequencing**

In many studies, the sequence of the genes expressed in the tissue defines the biological role of the tissue, and studies of differential expression between tissues, developmental stages, and genotypes support gene discovery (Dubey *et al.*, 2011). Sequencing of the transcriptome has been an important way to discover functional differences in plant genomes (Ablett *et al.*, 2000). EST sequencing with Sanger sequencing has been widely used to find SNP in plant genomes. More recent sequencing technologies have made this an even more useful approach to marker discovery (Hiremath *et al.*, 2011).

Sequencing of the transcriptome (Jain, 2012) or RNA-sequencing (RNA-seq) methods have become the preferred way to study differential expression of genes (Auer, Srivastava, and Doerge, 2012).

For example, Pattemore *et al.* (unpublished) sequenced the transcriptome of barley varieties to explore variation between and within cultivars. Gillies, Furtado, and Henry (2012) sequenced the transcriptome of the aleurone and starchy endosperm of developing wheat seeds to explore the differentiation of these two main tissues within the grain and potential for genetic improvement of wheat from a human nutrition and functional properties perspective.

### **Amplicon Sequencing**

The discovery of markers in specific genes or groups of genes can be targeted by amplicon sequencing (Kharabian-Masouleh *et al.*, 2011; Sexton *et al.*, 2012). Very large numbers of genotypes and genes can be screened simultaneously by sequencing the products of amplification from large numbers of genes and or genotypes (Bundock *et al.*, 2009; Kharabian-Masouleh *et al.*, 2011; Malory *et al.*, 2011).

For example, Kharabian-Masouleh *et al.* (2011) sequenced genes of starch biosynthesis in a diverse collection of rice genotypes to define the contribution of variation in gene sequence to starch properties and the resulting cooking and quality traits of rice.

Sexton *et al.* (2012) sequenced the genes of cell wall metabolism in eucalypts to discover the control of wood properties, identifying a specific gene that was associated with shrinkage of wood.

### ***Enriched Genome Sequencing***

The coverage of gene encoding regions of the genome can be improved by enrichment prior to sequencing (Hodges *et al.*, 2007). Temperature-controlled annealing can be used to separate the nonrepetitive sequences in the genomic DNA. Arrays and solution-based hybridization techniques can also be used.

For example, Bundock, Casu, and Henry (2012) used probes designed from the sorghum genome to enrich sugarcane genomic DNA. This allowed a 10-fold enrichment in sequences matching to genes and identified SNP in a large proportion of the genes in the whole genome.

### ***Genotyping by Sequencing***

The ultimate genotype data set is the genome sequence of the individual. As the technology to obtain a genome sequence improves, it will become an attractive alternative to the assay of a set of genetic markers as a routine tool for genotyping.

## **Evolving Range of Applications of DNA Markers in Plants**

The recent significant advances in sequencing technology are resulting in an expanded potential for DNA markers to be applied in plants (Table 1.4). Easier marker discovery allows application to systems and questions for which marker technology was previously considered too expensive.

### ***Plant Identification for IP Protection***

The protection of the intellectual property of plant breeders requires reliable and efficient identification of plant varieties. Molecular markers are now widely used to characterize and protect plant varieties. Policing of plant variety rights requires that evidence of a standard that can be present in a legal court is produced. DNA evidence has been successfully used to

**Table 1.4 Applications of DNA markers in plants.**

IP protection
Regulatory applications
Biosecurity
Conservation biology
Plant functional biology
Evolutionary biology
Plant improvement

prosecute cases of infringement of plant breeders' rights. New molecular technology may make these cases easier and allow more cost-effective protection of plant variety IP.

### ***Plant Variety Identification for Production and Quality Control***

Technology developments suggest that early applications of DNA technology to protection of valuable intellectual property will continue to expand to more routine general potential applications in process and quality control in food processing and marketing.

## **Applications**

The identification of plants using DNA markers maybe used to prosecute cases involving prohibited or restricted plants. Conventional phenotypic identification has been used to prosecute legal cases. DNA technology may allow investigations that link to the specific source of the plant material.

### ***Biosecurity Applications***

The identification of pest and diseases and the control of weeds can be aided by DNA marker methods. Speed of analysis maybe important in quarantine situations and rapid DNA analysis tools are required. DNA analysis may have special advantages when genotypes very similar in appearance pose very different biosecurity risks.

### ***Applications in Conservation Biology***

Conservation of biodiversity requires knowledge of the genetic relationships between individuals and species in wild populations. Molecular tools have been widely applied to conservation biology. However, recent advances in DNA sequencing promise to allow cost-effective extension of these molecular approaches to many more species and systems. DNA analysis has the potential to become a more routine tool in the management of biodiversity.

### ***Application in Evolutionary Biology***

Phylogenetic relationships have been increasingly based on DNA sequencing information. Much effort has been devoted to defining appropriate genetic loci for "barcoding" different groups of species. Advancing sequencing technologies (Henry, 2012) may allow complete analysis of the DNA in an organism to be used avoiding many of the difficult decision in applying more limited sampling of DNA sequence data. Variations in wild plant

populations can be used to study adaptation to different environmental conditions (Fitzgerald *et al.*, 2011; Shapter *et al.*, 2012).

### ***Applications in Understanding Biological Functions***

The ability to analyze the whole genome has dramatically reduced the effort required to define the molecular genetic basis of many important plant characteristics. This is rapidly advancing plant biology and plant production and also offers great potential in plant improvement. Association genetics with large-scale sequencing is allowing rapid identification of genes controlling important traits (Sexton *et al.*, 2012).

For example, using large-scale gene sequencing, Sexton *et al.* (2012) identified a pectin methylesterase associated with wood properties and Kharabian-Masouleh *et al.* (2011) define the starch biosynthesis genes that determined food properties in rice.

### ***Application in Plant Improvement***

#### **Plant Germplasm Characterization**

The characterization of genetic diversity in collections of plant germplasm (seed, cell cultures, or whole plant collections) can now be approached at the whole-genome level (Kilian and Graner, 2012). This will allow unprecedented improvements in the management and characterization of genetic resources. For the first time, it will be easier to define variation both within and between accessions in germplasm collections (seed banks).

#### **Plant Breeding Applications**

Selection of parents and screening of populations can now be based on whole-genome information. Rapid screening protocols can be used to apply the outcomes of whole-genome discovery (Kennedy *et al.*, 2006a and 2006b). Discovery of natural or induced mutations is simplified by recent advances (Cross *et al.*, 2008). Domestication of new species and introduction of genes from distant wild relatives will be greatly simplified by the power of genome sequencing in these systems (Malory *et al.*, 2011; Henry, 2012).

### ***Food Industry Applications***

Molecular tools are being applied more widely in the identification of food ingredients and in establishing the authenticity of food along the production chain. Important food safety applications may also emerge. The species and the genotypes or varieties of the species that make up the food (including some highly processed foods) can now be readily determined.

This is likely to be widely used by competing food companies to analyze their competitors' products.

## Future Developments

The continued development of DNA analysis tools (Edwards and Henry, 2011) can be expected to advance molecular methods for plant analysis. This will continue to expand the range of practical applications of molecular technologies. More reliable data collected more easily and at lower cost will ensure plant identification in nature and cultivation becomes more reliant on DNA sequence information. Many advances will be required in the tools for analysis and interpretation of the large volumes of data that are likely to be available as this technology continues to develop. Improvements in sample and data handling will be critical in the expanding applications of DNA markers in plants. Availability of reference collections of seeds, DNA, and data will also be important.

Molecular markers have been based on measurement of sequence differences that are linked to the trait of interest. Advances in sequencing technology now make possible a move to analysis of the sequences that are causal for the trait of interest ("perfect markers") rather than just linked to them. This has many advantages. Recombination will not eliminate the association between the marker and the trait. The linkage will not depend upon the genetic background and hold for all genotypes. Knowledge of the causal polymorphism provides an explanation of the biological basis of the trait variation. This allows the identification of other genetic and nongenetic strategies to work with the trait.

Molecular markers will be of increasing value and will find more widespread application as they move to become "perfect markers."

## References

- Ablett, E., Seaton, G., Scott, K. *et al.* (2000) Analysis of grape ESTs: global gene expression patterns in leaf and berry. *Plant Science*, **159**, 87–95.
- Ablett, G., Hill, H. and Henry, R.J. (2006) Sequence polymorphism discovery in wheat microsatellite flanking regions using pyrophosphate sequencing. *Molecular Breeding*, **17**, 281–289.
- Argout, X., Salse, J., Aury, J.-M. *et al.* (2010) The genome of theobroma cacao. *Nature Genetics*, **43**, 101–108.
- Arai-Kichise, Y., Shiwa, Y., Nagasaki, H. *et al.* (2011) Discovery of genome-wide DNA polymorphisms in a landrace cultivar of japonica rice by whole-genome sequencing. *Plant Cell Physiology*, **52**, 274–282.
- Auer, P.L., Srivastava, S. and Doerge, R.W. (2012) Differential expression—the next generation and beyond. *Briefings in Functional Genomics*, **11**, 57–62.
- Barbazuk, W.B., Emrich, S.J., Chen, L.L. *et al.* (2007) SNP discovery via 454 transcriptome sequencing. *Plant Journal*, **51**, 910–918.

- Bundock, P.C., Cross, M.J., Shapter, F.M. and Henry, R.J. (2006) Robust allele-specific PCR markers developed for SNP's in expressed barley sequences. *Theoretical and Applied Genetics*, **112**, 358–365.
- Bundock, P.C., Elliott, F., Ablett, G. *et al.* (2009) Targeted SNP discovery in sugarcane using 454 sequencing. *Plant Biotechnology Journal*, **7**, 347–354.
- Bundock, P.C., Casu, R. and Henry, R.J. (2012) Enrichment of genomic DNA for polymorphism detection in a non-model highly polyploidy crop plant. *Plant Biotechnology Journal*, **10**, 657–667.
- Cordeiro, G.M., Taylor, G.O. and Henry, R.J. (2000) Characterisation of microsatellite markers from Sugarcane (*Saccharum* sp.) a highly polyploid species. *Plant Science*, **155**, 161–168.
- Chan, A.P., Crabtree, J., Zhao, Q. *et al.* (2010) Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnology*, **28**, 951–956.
- Cross, M., Waters, D., Lee, L.S. and Henry, R.J. (2008) Endonucleolytic mutation analysis by internal labeling (EMAIL). *Electrophoresis*, **29**, 1291–1301.
- Deschamps, S. and Campbell, M.A. (2010) Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. *Molecular Breeding*, **25**, 553–570.
- Doorduyn, L., Gravendeel, B., Lammers, Y. *et al.* (2011) The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Research*, **18**, 93–105.
- Dubey, A., Farmer, A., Schlueter, J. *et al.* (2011) Defining the transcriptome assembly and its use for genome dynamics and transcriptome profiling studies in pigeonpea (*Cajanus cajan* L.) *DNA Research*, **18**, 153–164.
- Edwards, M.A. and Henry, R.J. (2011) DNA sequencing methods contributing to new directions in cereal research. *Journal of Cereal Science*, **54**, 395–400.
- Fitzgerald, T.L., Shapter, F.M., McDonald, S. *et al.* (2011) Genome Diversity in Wild Grasses under Environmental Stress. *Proceeding of the National Academy of Sciences of the United States of America*, **108**, 21139–21144.
- Garland, S., Lewin, L., Blakeney, A. *et al.* (2000) PCR based molecular marker for the fragrance gene in rice (*Oryza sativa* L). *Theoretical & Applied Genetics*, **101**, 364–371.
- Gillies, S., Furtado, A. and Henry, R.J. (2012) Gene expression in the developing aleurone and starchy endosperm of wheat. *Plant Biotechnology Journal*, **10**, 668–679.
- Graham, G.C., Mayer, P. and Henry, R.J. (1994) A simplified method for the preparation of fungal genomic DNA for PCR and RAPD analysis. *Biotechniques*, **16**, 48–50.
- Henry, R.J. (ed.) (2001) *Plant Genotyping: The DNA Fingerprinting of Plants*, CABI Publishing, Oxon, pp. 325.
- Henry, R.J. (ed.) (2008) *Plant Genotyping II: SNP Technology*, CABI Publishing, Wallingford, pp. 280.
- Henry, R.J. (2012) Next generation sequencing for understanding and accelerating crop domestication. *Briefings in Functional Genomics*, **11**, 51–56.
- Henry, R.J. and Edwards, K. (2009) New tools for single nucleotide polymorphism (SNP) discovery and analysis accelerating plant biotechnology. *Plant Biotechnology Journal*, **7**, 311.



- Henry, R.J., Edwards, M., Waters, D.L.E. *et al.* (2012) Application of large scale sequencing to plants. *Journal of Bioscience*, in press.
- Hiremath, P.J., Farmer, A., Cannon, S.B. *et al.* (2011) Large-scale transcriptome analysis of chickpea (*Cicer arietinum* L.) an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnology Journal*, **9**, 922–931.
- Hodges, E., Xuan, Z., Balija, V. *et al.* (2007) Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, **39**, 1522–1527.
- Imelfort, M., Duran, C., Batley, J. and Edwards, D. (2009) Discovering genetic polymorphisms in next-generation sequencing data. *Plant Biotechnology Journal*, **7**, 312–317.
- Jaccoud, D., Peng, K., Feinstein, D. and Kilian, A. (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research*, **28**, e25.
- Jain, M. (2012) Next-generation sequencing technologies for gene expression profiling in plants. *Briefings in Functional Genomics*, **11**, 63–70.
- Kennedy, B., Arar, K., Reja, V. *et al.* (2006a) Locked nucleic acids for optimizing displacement probes for quantitative real-time PCR. *Analytical Biochemistry*, **348**, 294–299.
- Kennedy, B., Waters, D.L.E. and Henry, R.J. (2006b) Screening for the rice blast resistance gene Pi-ta using LNA displacement probes and real-time PCR. *Molecular Breeding*, **18**, 185–193.
- Kharabian-Masouleh, A., Waters, D., Reinke, R. *et al.* (2009) A high-throughput assay for rapid and simultaneous analysis of perfect markers for important quality and agronomic traits in rice using multiplexed MALDI-TOF Mass Spectrometry. *Plant Biotechnology Journal*, **7**, 355–363.
- Kharabian-Masouleh, A., Waters, D.L.E., Reinke, R.F. *et al.* (2011) Discovery of polymorphisms in starch related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing. *Plant Biotechnology Journal*, **9**, 1074–1085.
- Kilian, B. and Graner, A. (2012) NGS technologies for analysing germplasm diversity in genebanks. *Briefings in Functional Genomics*, **11**, 38–50.
- Kim, M.Y., Lee, S., Van, K., *et al.* (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proceedings of National Academy of Sciences of the United States of America*, **107**, 22032–22037.
- Krishnan, S.G., Waters, D.L.E., Katiyar, S.K. *et al.* (2012) Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnology Journal*, **10**, 623–634.
- Lai, J., Li, R., Xu, X. *et al.* (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics*, **42**, 1027–1030.
- McIntosh, S.R., Pacey-Miller, T. and Henry, R.J. (2005) A universal protocol for identification of cereals. *Journal of Cereal Science*, **41**, 37–46.
- Malory, S., Shapter, F.M., Elphinstone, M.S. *et al.* (2011) Characterizing homologues of crop domestication genes in poorly described wild relatives by high-throughput sequencing of whole genomes. *Plant Biotechnology Journal*, **9**, 1131–1140.
- Nock, C., Waters, D.L.E., Edwards, M.A. *et al.* (2011) Chloroplast genome sequence from total DNA for plant identification. *Plant Biotechnology Journal*, **9**, 328–333.

- Novaes, E., Drost, D., Farmerie, W. *et al.* (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics*, **9**, 312.
- Pacey-Miller, T. and Henry, R.J. (2003) Single-nucleotide polymorphism detection in plants using a single-stranded pyrosequencing protocol with a universal biotinylated primer. *Analytical Biochemistry*, **317**, 165–170.
- Pattemore, J.A., Rice, N., Marshall, D.F. *et al.* (2010) Cereal Variety Identification using MALDI-TOF mass spectrometry SNP Genotyping. *Journal of Cereal Science*, **52**, 351–362.
- Rossetto, M., McNally, J. and Henry, R.J. (2002) Evaluating the potential of SSR flanking regions for examining taxonomic relationships in Vitaceae. *Theoretical and Applied Genetics*, **104**, 61–66.
- Sexton, T., Henry, R., McManus, L.J. *et al.* (2010) Capture of Assay Template by Multiplex PCR of Long Amplicons for Genotyping SNPs and InDels with MALDI-TOF Mass Spectrometry. *Molecular Breeding*, **25** 471–480.
- Sexton, T.R., Henry, R.J., Harwood, C.E. *et al.* (2012) Pectin methylesterase genes influence solid wood properties of *Eucalyptus pilularis*. *Plant Physiology*, **158**, 531–541.
- Shapter, F.M., Fitzgerald, T.L., Waters, D.L.E. *et al.* (2012) Analysis of adaptive ribosomal gene diversity in wild plant populations from contrasting climatic environments. *Plant Signaling & Behavior*, **7**, 1–3.
- Shepherd, M. and Henry, R.J. (1998) Monitoring of fluorescence during DNA melting as a method for discrimination and detection of PCR products in variety identification. *Molecular Breeding*, **4**, 509–517.
- Souza, G.M., Berges, H., Bocs, S. *et al.* (2011) The sugarcane genome challenges: Strategies for sequencing a highly complex genome. *Tropical Plant Biology*, **4**, 145–156.
- Thomson, D. and Henry, R.J. (1993) Use of DNA from dry leaves for PCR and RAPD analysis. *Plant Molecular Biology Reporter*, **11**, 202–206.
- Thudi, M., Li, Y., Jackson, S.A. *et al.* (2012) Current state-of-art of sequencing technologies for plant genomics. *Briefings in Functional Genomics*, **11**, 3–11.
- TPGSC (The Potato Genome Sequencing Consortium) (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, **475**, 189–195
- Trick, M., Long, Y., Meng, J. and Bancroft, I. (2009) Single nucleotide polymorphism (SNP) discovery in the polyploidy Brassica napus using Solexa transcriptome sequencing. *Plant Biotechnology Journal*, **7**, 334–346.
- Trebbi, D., Maccaferri, M., de Heer, P. *et al.* (2011) High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.) *Theoretical and Applied Genetics*, **123**, 555–569.
- Varshney, R.K. and May, G.D. (2012) Next-generation sequencing technologies: opportunities and obligations in plant genomics. *Briefings in Functional Genomics*, **11**, 1–2.
- Whittall, J.B., Syring, J., Parks, M. *et al.* (2010) Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology*, **19**(Suppl 1), 100–114.
- Williams, J.G.K., Kubelik, K.J., Livak, K.J. *et al.* (1990) DNA polymorphisms amplified by arbitrary primers are useful genetic markers. *Nucleic Acids Research*, **18**, 6531–6535.

- Xia, L., Peng, K.M., Yang, S.Y. *et al.* (2005) DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. *Theoretical and Applied Genetics*, **110**, 1092–1098.
- Yamamoto, T., Nagasaki, H., Yonemaru, J. *et al.* (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genomics*, **11**, 267.
- Yan, J., Kandianis, C.B., Harjes, C.E. *et al.* (2011) Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. *BMC Genomics*, **12**, 59.
- Zabeau, M. and Vos, P. (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. EPO Patent 0534858A1.
- Zhang, Y., Ma, P. and Li, D. (2011) High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (*Poaceae: Bambusoideae*). *PLoS ONE*, **6**(5), e20596.

