IAEA-TECDOC-1253

Radioactively labelled DNA probes for crop improvement

Proceedings of a final Research Co-ordination Meeting
organized by the
Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture
and held in Vienna, 6–8 September 1999

INTERNATIONAL ATOMIC ENERGY AGENCY IAEA

November 2001

The originating Section of this publication in the IAEA was:

Plant Breeding and Genetics Section
International Atomic Energy Agency
Wagramer Strasse 5
P.O. Box 100
A-1400 Vienna, Austria

RADIOACTIVELY LABELLED DNA PROBES FOR CROP IMPROVEMENT
IAEA, VIENNA, 2001
IAEA-TECDOC-1253
ISSN 1011–4289

Printed by the IAEA in Austria
November 2001

FOREWORD

With the advent of DNA molecular marker technology in the 1980s plant breeding had a new and powerful tool with which to increase its efficacy. Such markers are abundant and directly reveal information about the genotype and therefore are more useful than simple phenotypic markers. In plant breeding applications, molecular markers reveal information about variability and genetic relationships, and enable genetic mapping, which greatly assists the breeder in selection of parents and progeny, as well as in management of breeding strategies. Furthermore, molecular markers linked to phenotypic traits permit very early selection of superior progenies from breeding populations, therefore significantly reducing the need for field testing and greatly increasing efficiency of plant breeding programmes. For this to occur the oligonucleotide probes for labelling genetic markers and/or the primers for polymerase chain reactions to amplify genetic markers needed to be also accessible to scientists in developing Member States. In addition, technical information, training and troubleshooting were needed to support the utilization of DNA markers. In the early 1990s there was a dramatic increase in requests for access to this technology. This co-ordinated research project (CRP) facilitated the transfer of molecular marker technology, in terms of both material and information, from advanced laboratories to assist breeding programmes in developing countries.

Two other CRPs were conducted concurrently in order to assist developing Member States to utilise molecular markers — Application of DNA Based Marker Mutations for Improvement of Cereals and other Sexually Reproduced Crop Plants, and Use of Novel DNA Fingerprinting Techniques for the Detection and Characterisation of Genetic Variation in Vegetatively Propagated Crops (IAEA-TECDOC-1010 and IAEA-TECDOC-1047, respectively). The present CRP built upon the success of the former projects by ensuring the availability of probes/primers, many of them radioactively labelled, and the distribution of associated technical information for use in projects utilising molecular markers to improve local crop plant varieties in their resistance to biotic and abiotic stress, better yields, improved agronomic traits and enhanced harvested product quality.

The CRP successfully facilitated the distribution of DNA probes and primers, the establishment of an ordering and enquiry system for participants to source probes and primers including Web-based procurement services, and the distribution of protocols and related information in response to requests from participants. The project also encouraged the development and dissemination of background knowledge and data relating to the markers and their application. Networks between scientists in developed countries and those in developing countries were fostered and interactive forums for learning and troubleshooting were promoted.

The present publication summarizes the achievements of this CRP obtained through joint effort of all participants to facilitate application of molecular marker technology in plant breeding programmes in developing countries. The IAEA officers responsible for this publication were M. Maluszynski and L.S. Lee of the Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture. Special thanks are due to G. Caetano-Anollés for his assistance in the development of this report.

EDITORIAL NOTE

This publication has been prepared from the original material as submitted by the authors. The views expressed do not necessarily reflect those of the IAEA, the governments of the nominating Member States or the nominating organizations.

The use of particular designations of countries or territories does not imply any judgement by the publisher, the IAEA, as to the legal status of such countries or territories, of their authorities and institutions or of the delimitation of their boundaries.

The mention of names of specific companies or products (whether or not indicated as registered) does not imply any intention to infringe proprietary rights, nor should it be construed as an endorsement or recommendation on the part of the IAEA.

The authors are responsible for having obtained the necessary permission for the IAEA to reproduce, translate or use material from sources already protected by copyrights.

CONTENTS

SUMMARY

1. NEEDS AND CURRENT AND FUTURE APPLICATIONS

Molecular markers have played a major role in the genetic characterization, breeding, and improvement of many crop species. They also have contributed and greatly expanded our abilities to assess biodiversity, reconstruct accurate phylogenetic relationships, and understand the structure, evolution and interaction of plant and microbial populations. Molecular markers uncover sequence variability generally by using hydrogen-bonding interactions between nucleic acid strands ('hybridization') and/or enzymatic accumulation of defined nucleic acid sequences ('amplification'). A number of robust techniques that use these two properties have been developed in the past years, and many more are expected to be added to this arsenal of molecular tools. These techniques use cloned or synthetic nucleic acids as hybridization probes or synthetic oligonucleotides as amplification primers. However throughout the rest of this report, we will collectively referred to them as DNA probe-derived technologies.

This co-ordinated research project (CRP) focuses on transferring DNA probe technology for crops of economic importance to developing and developed countries. Technological transfer according to the UN Conference on Trade and Development (1990) is the "transfer of systematic knowledge for the manufacture of a product, for the application of a process, or for the rendering of a service". Two kinds of technologies can be transferred, 'hard' and 'soft' technologies. In this case, hard technologies involve tangible resources including DNA probes while soft technologies include know-how, skills and techniques. It should be noted that transfer of technology is considered complete only when a technology is successfully applied and adopted in a particular country and institution. This implies that technology transfer requires that the recipient has used both hard and soft technologies successfully.

The transfer of hard technologies involves the distribution of DNA probes and synthetic oligonucleotides from CRP participants (already in existence or under development), reagents related to the technology to be transferred (such as reference DNA, buffers and other components), and software. The transfer of soft technology involves written information including protocols, suggestions for troubleshooting, background references, access to information in databases, local and regional courses, expert visits to selected labs, and training of recipients of technology transfer in laboratories of either CRP participants or other experts of fields of relevance.

The first goal of this CRP has been to efficiently provide information on the nature of available probes and probe sets. A second goal has been to provide an avenue of communication on the best methods of application, and to provide for an interactive format for probe request and delivery to ensure that participating laboratories receive the most appropriate tools for their specific needs. The third and most important goal has been to provide the DNA probes to the requesting laboratories and institutions. In some cases, probes have been selected in groups ('anchor' sets) that serve as reference to link genetic maps. Whenever possible, these core or anchor sets were distributed to facilitate the integration of research results among projects of this CRP in order to allow the integration of the results with existing knowledge. As a fourth goal, considerable improvements have been made by the CRP participants in experimental procedures related to the development and application of amplification-based DNA markers. For example, a powerful cost-effective enrichment

procedure for microsatellite isolation promises the generation of a larger number of oligonucleotide primer pairs capable of amplifying highly variable regions in fungal and plant genomes. Similarly, methods of robust experimental design (the so-called Taguchi methods), widely and successfully used in the electronic and automotive industry, were applied to the optimization of experimental DNA marker protocols. This approach can save resources and effort devoted to the adaptation of these techniques to the analysis of plants and their pathogens.

The CRP has been highly successful in achieving the four goals mentioned above. Over 12 000 hybridization probes, 2800 microsatellite primer pairs and 2000 fingerprinting primers have been distributed throughout the world. The CRP has also been successful in distributing soft technology resources such as experimental protocols, troubleshooting information and background references. World Wide Web (WWW) sites have been made available and participants have been active in addressing questions about many aspects of the different technologies on an individual basis. What remains to be done is the continued monitoring of the successful transfer of technology to the recipient laboratories. Our recommendations are as follows:

(1) A mechanism must be put in place to ensure that transferred technology be correctly applied. Specifically, this can be achieved by expanding the existing web site of the IAEA to include useful links, an interactive question-and-answer forum to address specific problems for the implementation of techniques and to troubleshoot protocols. It should also provide a forum to publish results and improved protocols, share results, experiences, and materials between scientists. This site could also be enhanced by establishing a multimedia environment with visual aids (movies, images, etc.).

(2) The distribution of hard and soft technologies should be continued by establishing alternative mechanisms that will assure successful implementation of technology transfer. We believe that without an alternative mechanism to continue to distribute DNA probe technologies, this program will not achieve its full impact and potential.

(3) We also affirm the importance of matching appropriate DNA probe technologies with the specific applications of the recipient laboratories. Individual laboratories should be capable of selecting and performing the specific molecular biology technique in a precise and repeatable fashion. If these conditions are not met, then opportunities of training should be made available to them. Furthermore, the recipient laboratory must have or have access to the right equipment and reagents necessary to carry out such protocols.

2. CURRENT AND FUTURE DNA MARKER TECHNOLOGIES IN CROP IMPROVEMENT

Restriction fragment length polymorphisms (RFLPs) were the first generation of hybridization-based markers with substantial impact in agricultural biotechnology. These were subsequently followed by amplification-based technologies derived from the polymerase chain reaction (PCR). Notably, the use of arbitrary oligonucleotide primers in the amplification reaction facilitated the study of previously uncharacterized genomes. The most common of these arbitrarily amplified DNA (AAD) techniques is randomly amplified polymorphic DNA (RAPD). Other techniques have also been developed and are widely used in genetic analysis. Some of these include amplified fragment length polymorphism (AFLP)

analysis, arbitrarily primed-PCR (AP-PCR) and DNA amplification fingerprinting (DAF). AAD techniques are versatile and can be coupled with hybridization-based approaches effectively, to increase their power [e.g., randomly amplified microsatellite polymorphism (RAMP)] and to address a wider range of applications. See Appendix I for a list of these and other related techniques.

Molecular markers are being used extensively to investigate the genetic basis of agronomic traits and to facilitate the transfer of desirable traits between breeding lines. A number of techniques have been particularly useful for genetic analysis. For example, collections of RFLP anchor probes have been very versatile and important for the generation of genetic maps, construction of physical maps and the establishment of syntenic relationships between genomes, and marker assisted breeding. Numerous examples of specific genes that have been identified as tightly linked to RFLP markers are available for the improvement of specific agronomic traits in almost all major crops. Specific examples include viral, fungal and bacterial resistance genes in maize, wheat, barley, rice, tomatoes and potatoes. Additional examples include insect resistance genes in maize, wheat and rice as well as drought and salt tolerance in sorghum. These markers used in conjunction with bulked segregant analysis provide a very efficient method of characterizing and mapping natural and induced mutants that reflect interesting agricultural traits. Markers have also been used to determine the genes underlying quantitative variation for height, maturity, disease resistance and yield in essentially all major crops. In particular, the AAD techniques have been useful in the assessment of biodiversity, the study of plant and microbial populations sometimes in interaction, and identification of plant varieties and cultivars. AAD techniques have produced sequence-tagged sites that serve as landmarks for genetic and physical mapping. Besides AAD-based genome scanning techniques (e.g., RAPD, DAF, AP-PCR, SAMPL, RAMP, AFLP, tecMAAP, ASAP) (see Appendix I), there are other important oligonucleotide-based marker techniques that use hybridization (e.g. microsatellite profiling) or amplification of DNA (e.g. STS, SCAR, CAPS, STMS analysis) or RNA (e.g. DD-PCR, PCR-based diagnostics of RNA viruses). It is envisioned that emerging oligonucleotide-based technologies derived from the use of hybridization arrays, the so-called DNA chips and oligonucleotide arrays, will be the choice for future genomic studies. However, many of them are still under development, are proprietary, or require the use of expensive equipment, and are therefore not suitable or cost-effective for adequate transfer to developing countries. Clearly, the initial transfer of technology has only involved a selected group of techniques that are well established and are of broad application (e.g. AAD, RFLP and STMS analysis). However, techniques are continuously changing and evolving, so technology transfer needs to keep pace with current developments in genomics.

New developments in genomic research have given access to an enormous amount of sequence information as well as new insights on the function and interaction of genes and the evolution of functional domains, chromosomes and genomes. In this context, functional and comparative genomics can help in comparative genetic mapping and linkage analysis of useful agricultural traits. Future DNA marker techniques, such as the use of oligonucleotide arrays will incorporate genomic information. Comparative analyses of sequence information in the growing databases now publicly available in the World Wide Web will be an invaluable resource for genetic characterisation of exotic crop germplasm. This will have an increasingly important role in prospection and conservation endeavors.

2.1. Hybridization-based DNA marker technologies

2.1.1. Restriction fragment length (RFLP) analysis

RFLPs were the first DNA markers to be used in crop improvement. Compared to PCR-based DNA markers, RFLPs are particularly useful for comparative genome studies, and to link the newly developed genetic marker information with the existing RFLP maps. Genetic mapping of wheat, maize, and rice and other grass species with common DNA probes has revealed remarkable conservation of gene content and gene order. The consensus grass map aligning ten different genomes — diploid oats, Triticeae, the two genomes of maize, pearl millet, sorghum, sugarcane, foxtail millet, barley and rice — was described by 29 rice linkage blocks. This consensus map can be used to rapidly construct maps of other grass species by using a set of evenly spaced anchor probes, and to predict the locations of key genes for adaptation from one crop species to another. Consensus maps have also been constructed in *Solanum*, *Cucurbita*, crucifers and legumes. In the last 10 years, more than 12 000 RFLP clones or inserts have been distributed by the CRP participants. RFLP clones or inserts will be continually required by researchers since its inter-species locus specificity cannot be currently substituted by other simple DNA marker technologies.

2.1.2. DNA arrays and expressed sequence tags (ESTs)

One emerging and powerful strategy examines any nucleic acid sequence by direct hybridization with sets of probes laid on a grid. For example, oligonucleotide arrays can be confined to defined physical addresses in solid supports (nylon, glass, silicon, etc.) by solid-phase oligonucleotide synthesis, light-directed chemical synthesis using photolithographic masks, or accurate fluid micro-dispensing by pin transfer technologies. The simultaneous survey of hundreds or thousands of genes or general sequences has extended the efficiency of gene expression and genotyping studies. Some applications are listed in Appendix II. However, there is limited availability of suitable arrays for agricultural applications and the costs of such arrays are high.

Efforts in many laboratories have been initiated to produce large collections of partially sequenced DNA clones derived from mRNA (expressed sequenced tags, ESTs). Projects have been initiated in rice, maize, soybean, barley, tomato, *Arabidopsis*, pine and wheat. The conclusion of each EST project will result in the production of 50 000–100 000 ESTs for each species. Information of the status of the EST projects may be found on the WWW sites listed in Appendix III. The ESTs are a valuable resource for several types of DNA marker methods and investigations of gene expression with micoarrays or other techniques.

2.2. Amplification-based DNA marker technologies

2.2.1. Simple sequence repeat (SSR) markers

SSRs are abundant in most eukaryotic genomes, and their ease of use and high information content (see Appendix II) has ensured that they have largely replaced RFLP as a mapping tool. In addition to genetic mapping, SSRs are being used increasingly for genotyping, measuring the genetic diversity of breeding materials and cultivars, surveying genetic resources and gene bank collections. Recently, SSRs have been developed in most major crops, such as rice, wheat, maize, barley, and soybean. In the past three years, more than 2800 microsatellite primer pairs have been distributed by the CRP participants.

Development and use of SSRs have been extended to other crops as well, such as pearl millet, sorghum, cassava, banana and plantain. Moreover, microsatellite markers have been developed for two important phytopathogenic fungi. It is expected that SSRs will be widely used for crop improvement. We anticipate that SSR primers will be increasingly required by researchers and scientists in the developing countries in the next five years.

2.2.2. Arbitrarily amplified DNA

The use of arbitrary primers for mapping and fingerprinting has many advantages, including their relatively low cost, the optional use of radioactivity, the requirement of small amounts of template DNA, and the ability to obtain rapid results. Moreover, methods can be easily automated. However, these techniques are not an end-all solution for mapping because of the low information content of the individual markers.

Arbitrary primers can be used in combination with anchored poly-T primers to tag differentially expressed transcripts in complementary DNA (cDNA). Differential display reverse transcribed PCR (DDRT-PCR) is generally used to display differences in gene expression. Amplified fragments can be cloned and sequenced, but must be verified by Northern analysis. There are non-radioactive versions of the method available.

2.2.3. Amplified fragment length polymorphism (AFLP) analysis

The AFLP technique offers the potential for rapidly generating genotype data at multiple loci. It is an efficient way for genetic map construction, tagging important traits, cultivar fingerprinting, and genetic variation and diversity studies. Locus-specificity of AFLP markers within species can be used to anchor other markers to chromosomes or existing linkage maps when the standard AFLP profiles are publicly available (see an example of barley AFLP profiles in the GrainGenes database; Appendix III). Since AFLP can produce abundant and repeatable results by different laboratories, and the cost per data point is relatively low (see Appendix II), it is the most attractive PCR-based marker technology for researches in developing countries, where it has already been applied in crop improvement. However, making non-radioactive AFLP technique alternatives affordable for developing countries constitutes a great challenge.

3. DISTRIBUTION OF DNA CLONES, PRIMERS AND ASSOCIATED INFORMATION

3.1. Resources for wheat and barley (John Innes Centre)

A core set of 73 probes was selected as an ‘anchor set’ for distribution, and a further 31 probes were selected to complement the original set. The probes were available as a set or individually and were chosen based on genome coverage, copy number, signal strength, polymorphism levels and cross hybridisation to wheat and barley. Data sheets with all relevant details, including coloured maps, were also sent out with the probes. In the past three years, more than 150 microsatellite markers in wheat were developed and characterised. A set of 42 mapped wheat microsatellite primer pairs has been selected for distribution. A set of 31 barley RFLP probes were also developed and distributed. Overall, more than 8000 wheat probes, 2000 wheat microsatellite primer pairs, and 200 barley probes were distributed to more than 250 research groups in 40 countries.

3.2. Resources for pearl and foxtail millets

From the results of pearl millet mapping projects at the John Innes Centre, a core set of probes has been selected for distribution either individually or as a complete set. Moreover, many of the millet DNA probes have been end-sequenced and primers prepared for use in sequence-tagged-site (STS) analysis. As a result, 52 primer pairs were available for distribution. A detailed foxtail millet map has been developed and a core set, containing 48 probes, was available upon request. A millet genes database in ACEDB format has been developed and is available on the WWW (see Appendix III). It includes information on probes, end-sequences, RFLP/STS polymorphisms and genetic maps. During the CRP, more than 700 pearl millet probes have been distributed to China, Colombia, France, India, Japan, South Africa, the USA and the UK. Some probes were requested by developed countries, but their target areas were the developing countries.

3.3. Resources for maize, rye, sorghum and rice

The genetic maps of maize and sorghum were improved through comparative mapping of RFLP loci detected by 124 maize cDNA clones and through the development of a new mapping population of maize at Iowa State University. Comparative mapping between maize and sorghum and maize and rice, using the set of 124 maize cDNA clones (and other clones) in each study, substantiated previous observations of extensive conservation of locus order but it also provided strong evidence of numerous large-scale chromosomal rearrangements.

The new mapping population for maize (intermated B73 × Mo17, 'IBM') was created by random intermating during the first segregating generation. Intermating for four generations prior to the derivation of recombinant inbred lines (RILs) increased the frequency of recombinants at many regions of the maize genome and provided better genetic resolution of locus order. Expansion of the maize genetic map was not uniform along the length of a linkage group and was less than the theoretical expectation. The 350 IBM RILs were genotyped at 512 loci detected by DNA clones, including 76 of the 124 supported by this contract. The production of the sorghum mapping population of RILs from the cross CK60 × PI229828 has been delayed by weather conditions that were not conducive to plant growth and seed development.

Seed of the IBM RILs (4900) have been distributed to 16 research organizations in the public and private sector. The DNA clones (1206) have been distributed to nine research laboratories. Further distribution of the seed and clones will be managed by curators at stock centers.

There are two major categories of probes in the collection of the University of Missouri at Columbia, cDNAs and genomic RFLP clones (generally derived from *Pst*I inserts). For general mapping, distribution involved maize core probe sets consisting of 90 probes selected on the basis of even coverage of the maize genome, simple hybridization patterns and high levels of polymorphism among maize lines. Many of the core set of markers have also been mapped in rice and sorghum. There are a total of 6400 clones in the collection and more than 4700 are available for distribution. In addition, sets of 203 probes for rye and 10 minisatellites probes for rice are also available.

3.4. Resources for mungbean, cowpea, common bean and soybean

Vigna is a genus that includes the cultivated species mungbean and cowpea. It is closely related to common bean (*Phaseolus*) and soybean (*Glycine*). Thus, *Vigna* DNA clones can be easily used in a variety of legume species. The University of Minnesota has sent, upon request,

packages of *Vigna* clones including the following items: 1) a set of 73 single-copy DNA clones from mungbean and cowpea; 2) two moderately repetitive and highly polymorphic mungbean clones; 3) genomic DNA from mungbean, cowpea, soybean, common bean, and pigeon pea; 4) a spreadsheet printout with locus name, cloning vector, insert size, map position, mapping enzyme, and chromosomal locations in common bean and soybean, if known; 5) the current RFLP linkage maps for mungbean and cowpea; and 6) DNA transformation, plasmid mini-preparation, and PCR amplification protocols.

3.5. Primers for yam, chickpea, banana, plantain and phytopathogenic fungi

About 1000 SSR oligonucleotides, locus-specific microsatellite primers and PCR primers for disease diagnosis, DNA fingerprinting and mapping have been distributed by the Plant Molecular Biology Group at Frankfurt University in the past 5 years. These include primers and oligonucleotides for RAPD, MP-PCR, and RAMPO cultivar identification in yam (*Dioscorea* spp.), locus-specific primer pairs for the amplification of sequence-tagged microsatellite sites (STMS) in chickpea, banana and plantain cultivars. Primers and oligonucleotides were also applied for SSR fingerprinting of the phytopathogenic fungi, *Mycosphaerella fijiensis* and *Ascochyta rabiei.* Recently, about 20 species-specific SSR primer pairs have been developed for *M. fijiensis* and were distributed to several laboratories in Latin America.

3.6. Arbitrary oligonucleotide kits

Operon Technologies through this CRP provided free of charge 10-mer primer kits each containing twenty 10-base arbitrary primers for use in genetic mapping. Operon Technologies currently has 1200 different 10-base primers in stock. The sequences were selected randomly, with the requirement that their (G+C) content be 60% to 70%, and that they have no self-complementary ends. The use of 10-mers for mapping or fingerprinting has many advantages, including a relatively low cost, no radioactivity, easy adaptation to automation, requirement for very small amounts of input DNA, rapid results, existing data bases for many organisms, and low cost equipment requirements. RAPDs can provide cost-effective and thorough methods for mapping and fingerprinting any genome.

3.7. Other arbitrary primers

A group of experimentally more demanding techniques (e.g. DAF, ASAP, MP-PCR) can produce more bands than using RAPDs. For example, DAF using high primer-to-template ratios of conventional or mini-hairpin primers can generate over 100 amplified products when analyzed in silver-stained polyacrylamide gels. A set of 10 conventional 8-mer primers was distributed by the University of Tennessee and the University of Oslo. Other more specialized primer sets (e.g. 11-mer and 12-mer mini-hairpin primers) were also made available. Detailed protocols on arbitrary primer techniques, electrophoresis, silver staining and recovery of DNA from polyacrylamide gels were placed in a newly developed WWW site at *http://biologi.uio.no/FellesAvdelinger/DNA_KAFFE/default.html*.

3.8. Primers for viral diagnosis

The CIBCM, University of Costa Rica (San Jose) provided primers (enough for 200 reactions) for the detection of plant RNA viruses that infect basic crops and ornamentals. RNA viruses can be reverse transcribed using the negative-strand primer and then a conventional PCR reaction is done to amplify the viral sequence. In principle, primers can be designed to detect groups of viruses or specific strains. For instance, in the case of the

Tenuiviridae a terminal palindromic sequence is present at both ends of all viral RNAs. More specific primers can be designed to detect each specific virus type.

Tenuiviruses	Primer specific to the whole family
	Primers specific for Rice Hoja Blanca virus (capsid protein)
Luteoviruses	Primers for Beet Western Yellows Virus and
	Barley Yellow Dwarf Virus
	Potato Leafroll Virus
Triste zaviruses	Primers for Citrus Tristeza Virus
Potyviruses	Potato Virus X
	Potato Virus Y

4. CONCLUSIONS AND RECOMMENDATIONS

(1) The distribution of DNA clones should be continued. However, support is not needed for all crop species. For example, many wheat and maize clones may be obtained from the Albany (CA) and Columbia (MO) USDA-ARS centres. Curators at germplasm banks and genetic stock centres may already have responsibility and the financial support for maintaining and distributing the DNA clones and related material and information. The DNA clones remain an important resource for research but their role has diminished with the advent of PCR-based marker systems. In developing countries, it is difficult for researchers to design and synthesize primers. Therefore, one recommendation from this CRP is to develop the means of enabling scientists in developing countries to obtain the necessary primers at reasonable cost and in a timely manner.

(2) The science and technology related to the utilization of DNA markers in crop improvement are advancing at an increasing rate. The rate and magnitude of the advancements make it difficult for researchers and research institution, anywhere, to review and assimilate the new information, knowledge and technology. Without special efforts in education, researchers and their facilities may become quickly out of date and disconnected from the opportunities afforded by the frontiers of science in this era. Thus, it is important to provide more frequent short-term (days or weeks instead of months or years in accordance with the goals of the session) training and educational opportunities in the principles and practice (theory and technical) of the utilization of DNA markers in crop improvement.

(3) The development of WWW sites devoted to the implementation of various DNA marker techniques and their use in crop improvement is extremely important. Such sites could include detailed descriptions of methods, supplementary video or images, literature citations, sources of reagents, suggestions for appropriate data collection and interpretation, descriptions of applications, variations of the primary method and the many informal and often unwritten details that enhance the performance of a procedure or an analysis. The existing web sites should be updated. This can be achieved by including useful links, an interactive question-and-answer forum to address specific problems of the implementation of techniques and by troubleshooting protocols. It should also provide a forum to publish results and improved protocols, share results and experiences, and materials between scientists.

DEVELOPMENT, DISTRIBUTION AND APPLICATION OF DNA MARKERS FOR CEREAL RESEARCH

X. QI, P. STEPHENSON, K.M. DEVOS, M.D. GALE
Comparative Genetics Unit,
John Innes Centre, Norwich Research Park,
Colney, Norwich, United Kingdom

Abstract

DNA probes and primers are important resources for molecular genetic research and molecular breeding. Presently, more than 2500 wheat probes, 400 barley probes, 800 foxtail, pearl millet and finger millet probes, and approximately 150 wheat microsatellite (SSR) primer pairs have been developed and maintained in our DNA Resource Centre at the John Innes Centre (JIC). To accelerate probe and primer distribution, an 'anchor set' and a 'supplementary anchor set', containing 73 and 31 wheat RFLP probes, respectively, and a standard set of 42 primer pairs for wheat SSR markers were selected. Similarly, a set of 52 pearl millet probes has been selected for distribution. More than 8000 wheat RFLP probes, 2000 wheat SSR primer pairs, 700 millet probes and 200 barley probes have been distributed to more than 250 research groups in 40 countries. Our wheat and millet probes and other grass cDNA probes have been used for comparative genetic studies. The revealed conservation of gene content and gene order has been used to construct maps of many grass species and to predict the locations of key genes from one crop species to another. Developed SSR and AFLP markers in wheat, barley and millet are particularly suited for genetic diversity analyses and map construction.

1. INTRODUCTION

Several research groups at the John Innes Centre (JIC) are involved in cereal molecular research, and as a result, a large number of DNA probes and primers have been developed for a range of cereal crops. Since the first FAO/IAEA Research Co-ordination Meeting on Radioactively Labelled DNA Probes for Crop Improvement in Vienna, 1995, we have been actively developing, collecting and maintaining DNA probes and primers, and distributing materials world-wide. Presently, more than 2500 wheat probes, 400 barley probes, 800 foxtail, pearl and finger millet probes and approximately 150 wheat microsatellite (SSR) primer pairs developed at JIC are maintained in our DNA Resource Centre. Moreover, RFLP anchor probes for wheat, barley, pearl millet and foxtail millet and a set of 42 mapped wheat microsatellite primer pairs have been selected for distribution as sets. To date the JIC Cereals DNA Resource Centre has distributed more than 8000 wheat PSR RFLP probes, 2000 wheat PSP microsatellite primer pairs, 700 millet probes and 200 barley probes to more than 250 research groups in 40 countries. To accelerate the application of molecular markers in pearl millet breeding in India and Africa, we are now developing SSRs in this crop. Very soon, a reasonable number of SSR markers with known map locations will be ready for distribution.

Comparative genetic studies using our wheat and millet probes, and other grass cDNAs have revealed remarkable conservation of gene content and gene order [1,2]. This colinearity can be used to construct maps of any grass species using a set of anchor probes evenly spaced on existing maps, and to predict the locations of key genes for adaptation from one crop species to another. Genetic diversity analyses of wheat and barley have shown that the high polymorphism information content (PIC) and the ease of use of microsatellites is particularly suited for high throughput genotyping studies [3].

2. DNA PROBES AND PRIMERS: DEVELOPMENT, MAINTENANCE AND DISTRIBUTION

2.1. Wheat anchor probes and microsatellite primers

A core set of 73 PSR probes has been selected as an 'anchor set' for distribution, and recently a further 31 probes have been selected to complement these (the 'supplementary' set). The probes are available as a set or individually and are chosen based on genome coverage, copy number, signal strength, polymorphism levels and cross hybridization to wheat and barley. Map locations of both the anchor and supplementary anchor RFLP probes are shown in Figure 1 [4]. Data sheets with all relevant details, including colored maps, are sent out with the probes.

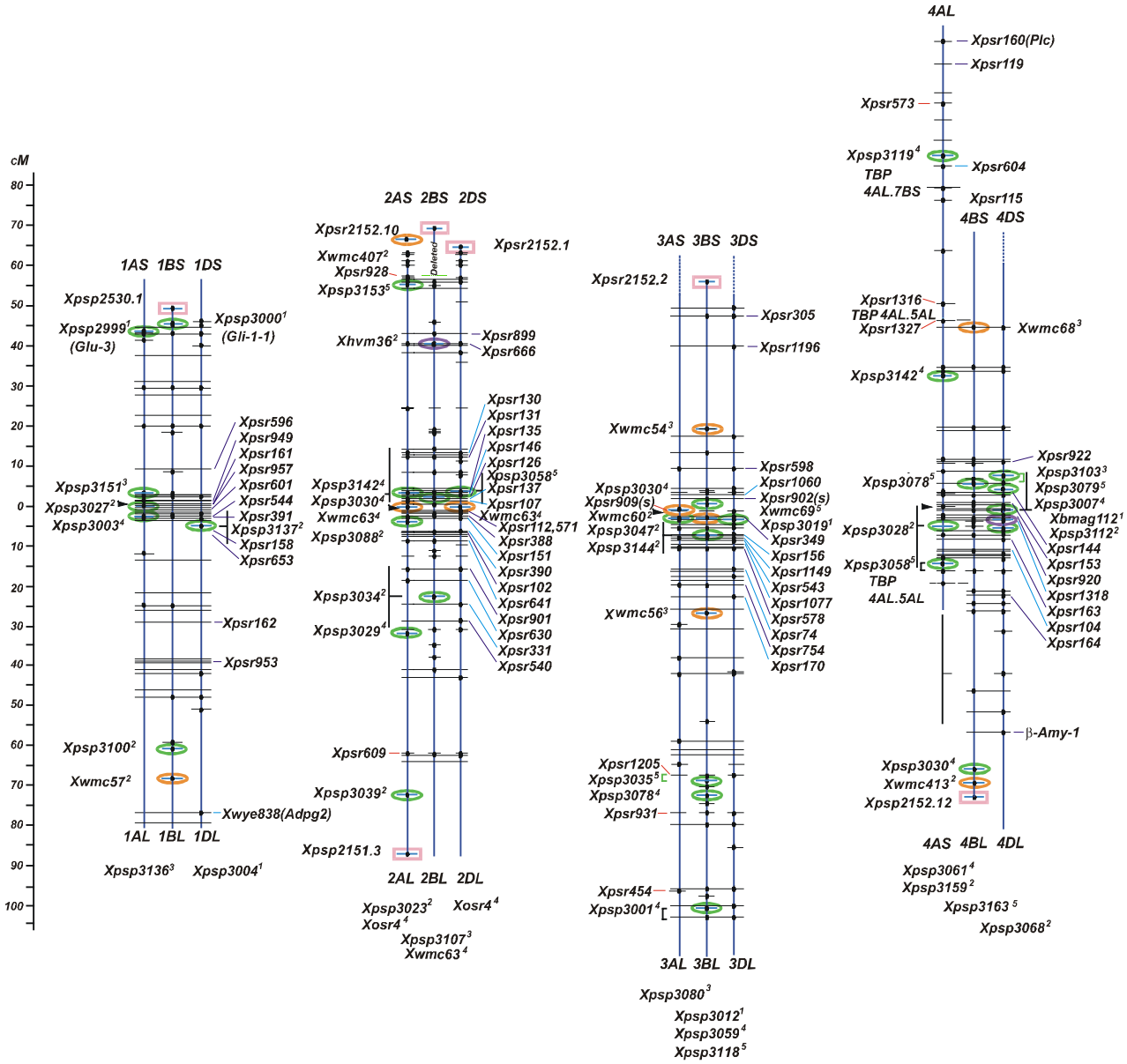

Fig. 1. John Innes Centre hexaploid wheat map, September 11999. RFLP markers generated by using anchor and supplementary anchor probes and SSRs are presented on the map.

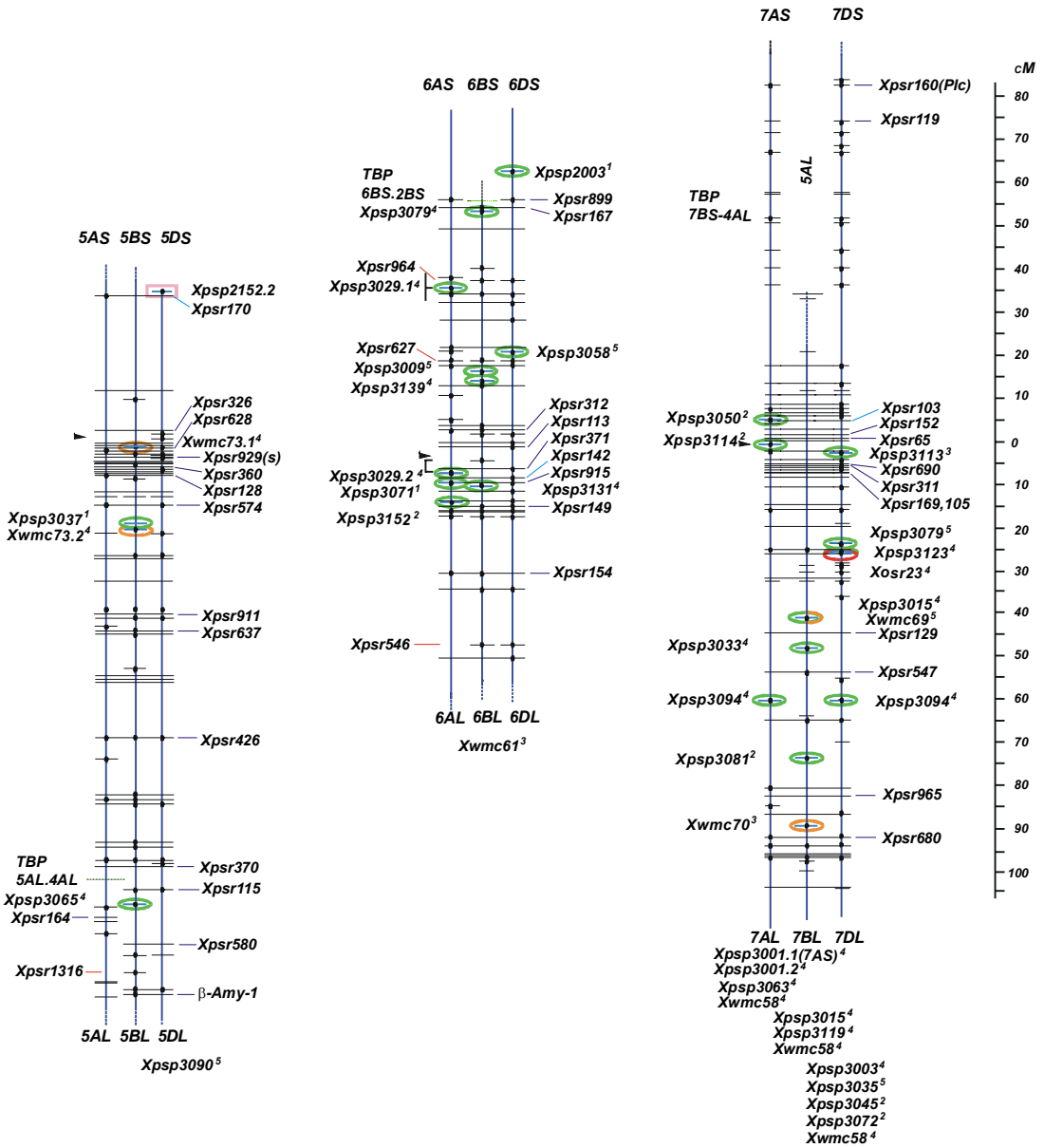

Fig. 1. John Innes Centre hexaploid wheat map (cont.).

We have developed and characterized more than 150 microsatellite markers in wheat [5] and mapped more than 50 loci [6]. Consideration of multiple band PCR profiles of some microsatellites has raised the issue of quality of individual microsatellite markers. Microsatellites, or more specifically the primers used to identify them, vary considerably in their ease of use. This is likely to be a very important issue in large scale screening applications and multiplexing. Based on the degree of stuttering observed on sequencing gels and the number of distinct loci generated by each primer pair, which is a particular problem in polyploids such as wheat, we have devised a scale of 1 (best) to 5 (useful but complex). The scale is explained by example in Figure 2. A standard set of 42 primer pairs, which give unique and predictable PCR patterns, has been selected for distribution. While we are unable to release the primer sequences we supply sufficient primer to allow for at least 100 reactions per pair. The map locations are indicated in Figure 1. PCR conditions and product sizes are also supplied with each set.

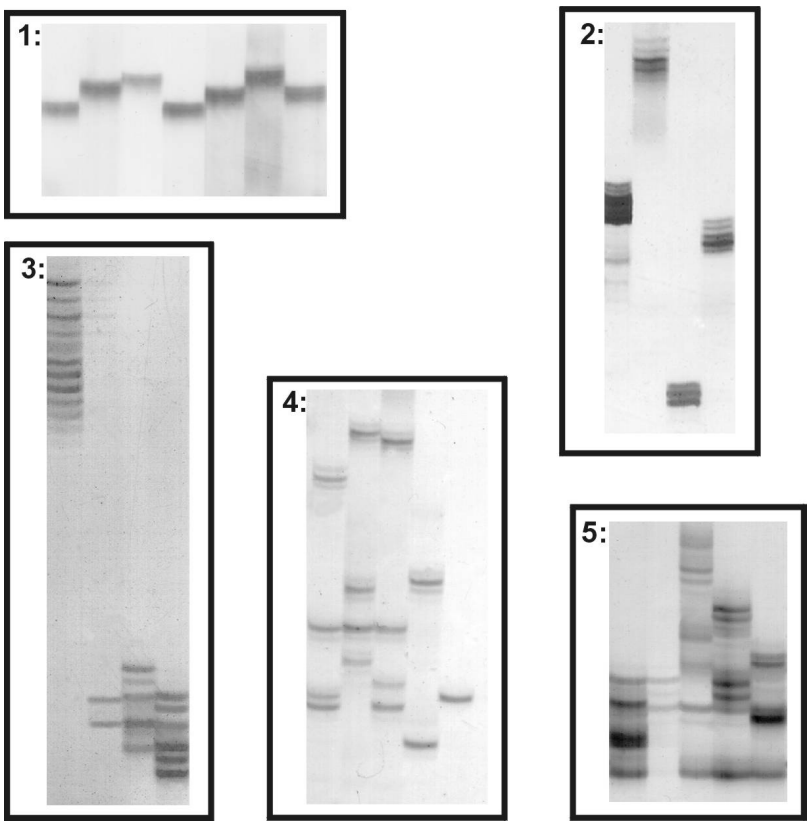

Fig. 2. Examples of microsatellite primer amplification products in different wheat varieties. Markers have been classified with a quality score which we feel best describes their usefulness: 1, single product recognizing a single locus, little or no stuttering; 2, single product recognizing a single locus, some stuttering, however easy to score; 3, single locus, profound and /or irregular stuttering, sometimes difficult to use; 4, multiple products and multiple loci, easy to score but sometimes difficult to relate polymorphic loci to those previously characterized and mapped; 5, multiple products and multiple loci, difficult to relate bands between genotypes.

2.2. Barley RFLP probes

Barley RFLP probes developed at JIC are designated PSB (Table 1) and are *Pst*I genomic fragments from *Hordeum vulgare* 'Betzes' cloned in the *Pst*I site of pUC18. The map shows the locations of JIC probes mapped in barley (Figure 3). Probes connected to chromosomes by short horizontal lines are those mapped in the 'Igri' × 'Triumph' cross [7]. Other probes were mapped in other crosses [8, 9] and their *approximate* positions are shown by reference to common markers.

2.3. Millet: Pearl millet (*Pennisetum*) probes and primers, foxtail millet (*Setaria*) and finger millet (*Eleusine*) probes

From the results of pearl millet mapping projects at JIC [10, 11] a core set of probes has been selected for distribution either individually or as a complete set (Fig. 4). Many of the millet DNA probes have been end-sequenced and primers prepared for use in sequence-tagged-site (STS) analysis, as a result 52 primer pairs are now available for distribution.

A detailed foxtail millet map (Fig. 5) has been developed [12,, 13] and *Pst*I genomic DNA probes are available upon request. Similar work is ongoing to develop a finger millet map and materials will be made available soon.

A millet genes database in ACEDB format has been developed and is now available on the WWW (http://jiio5.jic.bbsrc.uk:8000/cgi-bin/ace/search/millet). It contains information on probes, end-sequences, RFLP/STS polymorphisms, genetic maps, etc.

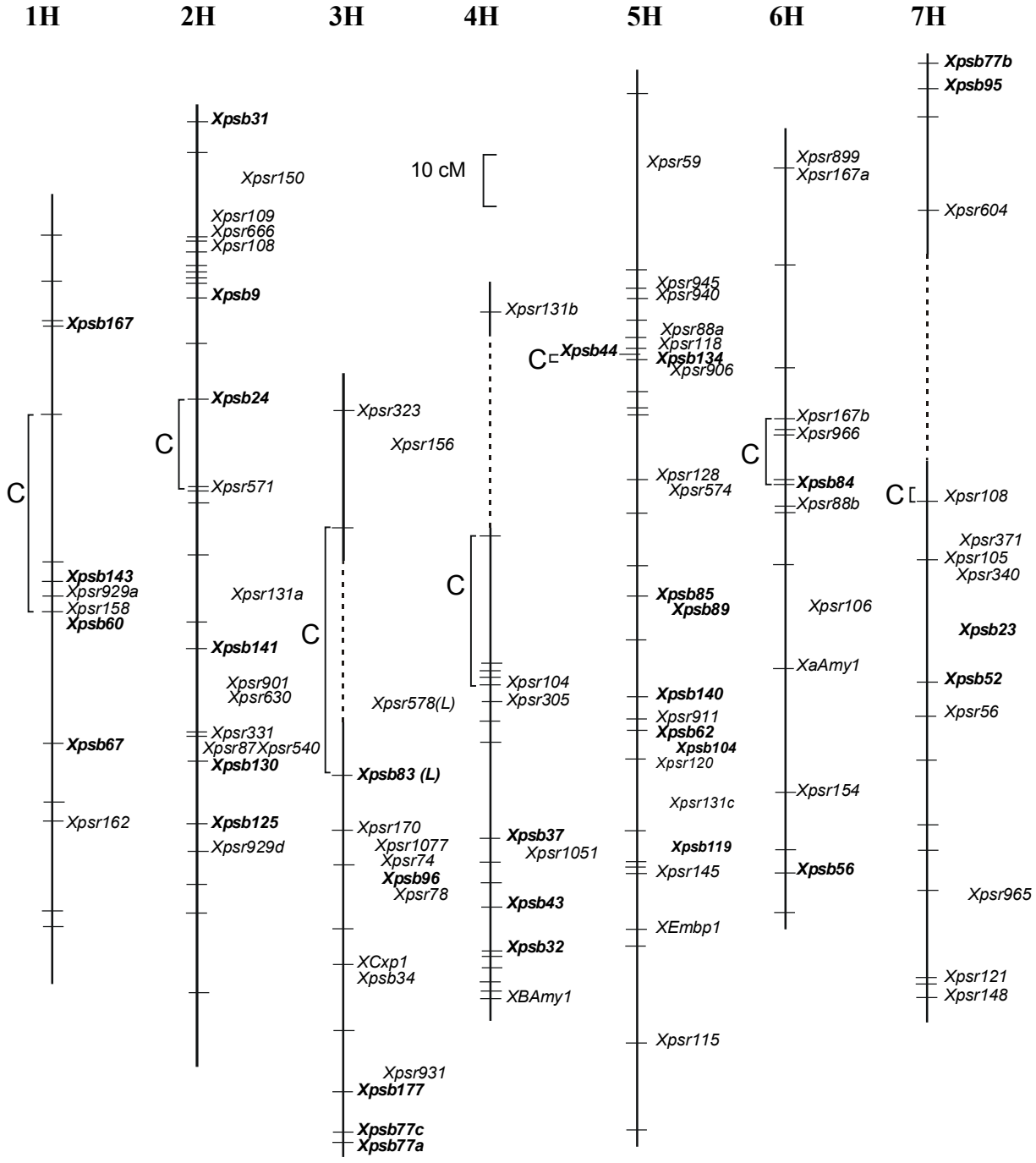

Fig. 3. John Innes Centre barley probe locations, September 1999.

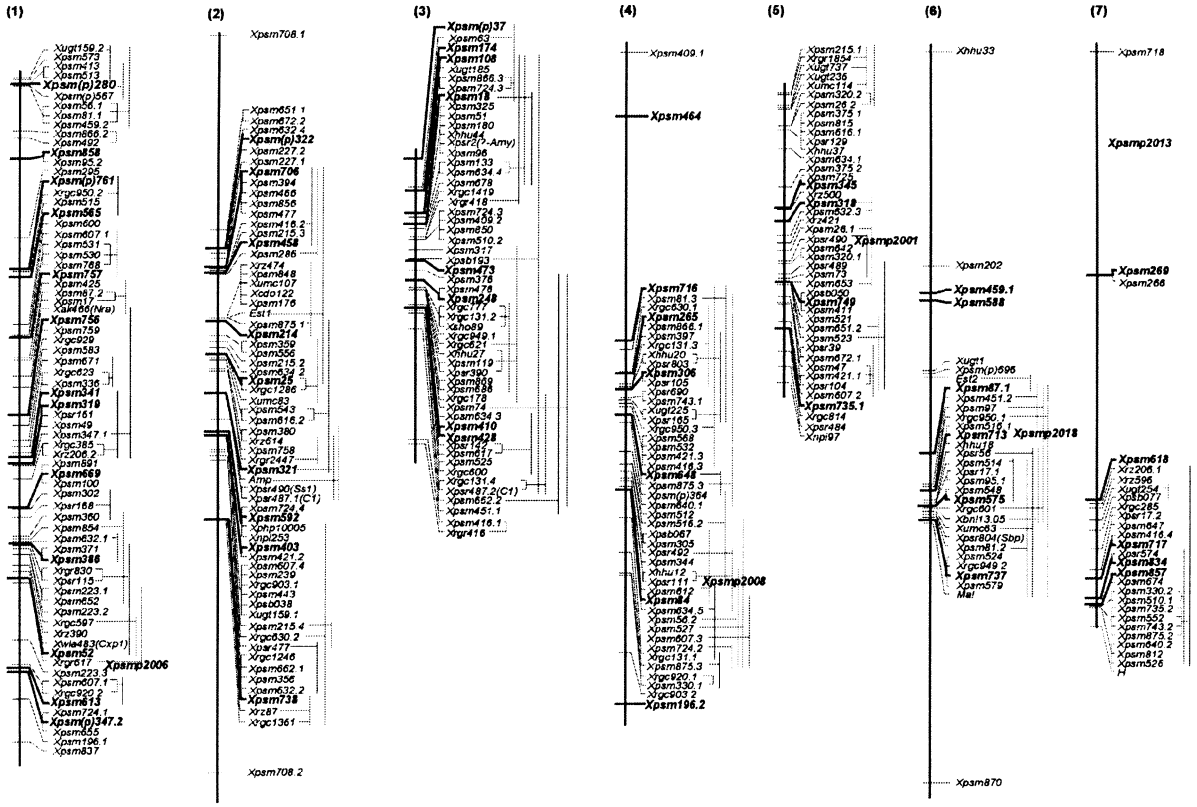

FIG. 4. Pearl millet consensus map, John Innes Centre, September 1999. RFLPs generated by using anchor probes are bold, Xpsimp: microsatellite markers.

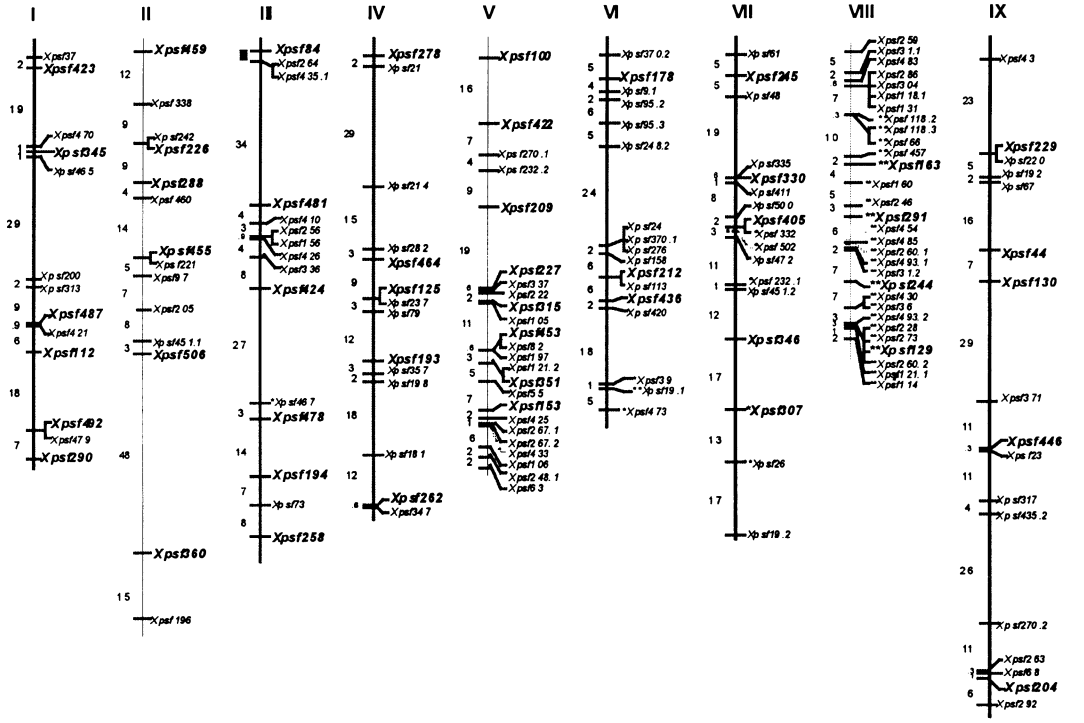

FIG. 5. John Innes Centre foxtail milletmap. Markers generated by using anchor probes are bold.

TABLE 1. PSB PROBE INFORMATION

Clone	Copy number	Addition line data	Hybridization to wheat
PSB9	several	2HS[*1]	Weak
PSB23	several	1HL, 2HS, 3HS, 3HL, 4HL, 5HS, 6HS, 6HL, 7HS	Weak
PSB24	low	2HS, 4H, 6HL	Moderate
PSB31	low	2HS[*2]	No
PSB32	low	4HL	Weak
PSB34	low	3HL	Weak
PSB37	low	4HL	Moderate
PSB43	low	4HL	Moderate
PSB44	low	5HS	Moderate
PSB52	low	7HL	Weak
PSB56	low	6HL	Moderate
PSB60	low	1HL	Moderate
PSB62	low	5HL	Weak
PSB67	low	1HL, 6H	Moderate
PSB77	several	3H, 7H	Weak
PSB83	low	3HL	Strong
PSB84	low	6HL	Strong
PSB85	low	5HL	Strong
PSB89	low	5HL	Moderate
PSB95	low	6H, 7HS	Moderate
PSB96	low	3HL	Moderate
PSB104	high	1HS, 5HL[*3]	Weak
PSB119	low	5HL	Weak
PSB125	low	2HL	Weak
PSB130	low	2HL	Moderate
PSB134	low	5HL	Weak
PSB140	low	5HL	Weak
PSB141	low	2HL	Weak
PSB143	low	1HL	Repeat
PSB167	low	1HS	Weak
PSB177	low	3HL	Moderate

Copy number: Low, one to four bands; Several, four to ten bands; High, more than ten bands or very intense bands indicating high sequence copy number.
Addition line data: Hybridization to EcoRI digests of Chinese Spring/'Betzes' telosomic addition lines.
Hybridization to wheat: "Strong" and "moderate" probes can be used for mapping in wheat although the latter may need longer exposure in order to give good results. Use of "Weak" probes in wheat is not recommended.
[*1] Several polymorphic bands all map to one 2HS locus.
[*2] Two closely linked loci on 2HS.
[*3] Mapped location is one of the minor bands.

3. APPLICATIONS

3.1. Comparative genome analysis

Compared to PCR-based DNA markers, RFLPs are particularly useful for comparative genome studies. Genetic mapping of wheat, maize, and rice and other grass species with common DNA probes has revealed remarkable conservation of gene content and gene order [2]. The consensus grass map aligning ten different genomes-diploid oats, *Triticeae*, the two genomes of maize, pearl millet, sorghum, sugarcane, foxtail millet and rice-was described by 29 rice linkage blocks. This consensus map can be used to rapidly construct maps of other grass species by using a set of evenly spaced anchor probes, and to predict the locations of key genes for adaptation from one crop species to another.

TABLE 2. PSR WHEAT ANCHOR PROBES MAPPED ONTO OTHER GRASS GENOMES

PSR	*Wheat*	Barley	Rice	Rye	Pearl Millet	Foxtail Millet
1	7ABD					
65	7ABD			7R		
102	2ABD	2H	4	2R		
103	7ABD	7H		7R		
104	4ABD	4H		7R	5	
105	7ABD				4	
107	2ABD			2R		
112	2ABD		7	2R		
113	6ABD		2			
115	4A 5BD	5H	1	7R	1	
119	4A 7AD	6H		4R		
126	2ABD	2H				
128	5ABD	5H				
129	7ABD	7H	6	7R	5	IV
130	2ABD			2R		
131	2ABD	2H		2R		
137	2ABD					
138	3ABD					
142	6ABD				3	
144	4ABD	4H		4R	4	
149	6ABD	6H		6R		
151	2ABD			2R	5	
152	7ABD		8	4R		
154	6ABD	6H	8	6R		
156	3ABD	3H		3R		
158	1ABD	1H		1R		
160	4A 7AD	7H	6	6R		
161	1ABD	5H	12	1R	1	III
162	1ABD	1H	5	1R		
163	4ABD	4H		7		
164	4BD 5A	4H		5R		
167	6ABD	6H	9	4R		
169	7ABD			7R		
170	3ABD	3H		3R		
305	3ABD	3H	12	3R		
311	7ABD			7R		
312	6ABD	3H 6H		6H		
326	5ABD	5H	12	5H		
331	2ABD	2H		2R		
360	5ABD	5H		5R		
370	5ABD	5H 2H		5R		
371	6ABD	6H		6R		
388	2ABD	2H		3R		
390	2ABD			2R	3	
391	1ABD	1H		1R		
394	3ABD	5H	1			
426	5ABD	5H		5R		
454	3B			6R		
540	2ABD	2H		2R		
543	3ABD	3H		4R		
544	1ABD	1H		1R		
547	7ABD	7H		7R		
571	2ABD					
574	5ABD		9	5R	7	II
578	3ABD		1	3R		
580	5BD		3	7R		
596	1ABD	3H		1R		
598	3ABD	3H	1	3R		
601	1ABD		9	1R		
604	4A 7AD					
609	2ABD	2H		2R		

TABLE 2. (cont.)

PSR	*Wheat*	Barley	Rice	Rye	Pearl Millet	Foxtail Millet
627	6ABD	6H		6R		
628	5ABD			5R		
630	2ABD	2H		2R		
637	5ABD	5H 7H		5R		
641	2ABD	2H				
653	1ABD					
666	2ABD	2H 6H		2R		
690	7ABD			7R	4	
754	3ABD		5	3R		
838	1ABD					
899	6AD	6H		4R		
901	2ABD	2H		2R		
902	3ABD			3R		
911	5ABD	5H		5R		
915	6ABD			6R		
928	2AD			7R		
929	5ABD	1H		5R		
931	3ABD	3H				
934	2ABD	2H 6H		2R		
949	1ABD			1R		
953	1ABD			1R		
957	1ABD			1R		
965	7ABD			6R		
1051	4ABD			7R		
1060	3ABD			3R		
1077	3ABD	3H		3R		
1149	3ABD	3H		3R		
1196	3ABD	3H		3R		
1205	3ABD			6R		
1316	4A 5B	3H				
1318	4ABD			7R		
1327	4A 1AD					

Our PSR probes have been used to hybridise with other grass species, e.g. barley, rice, rye, pearl millet and foxtail millet (Table 2). Most probes have been mapped to homologous regions in barley, rye and rice, some have been mapped on the pearl and foxtail millet genomes. Similarly, several pearl millet probes have been mapped in foxtail millet. The two sets of anchor probes played an important role in the alignment of the ten grass species genomes.

Many PSR probes from a bread wheat cDNA library have been either end-sequenced or completely sequenced and STS primers have been designed, converting RFLP markers into more easily assayed PCR-based markers. By using 'BLAST', homologous sequences have been identified, many of which were of known function (Table 4). Some of these probes are included in the core sets of wheat probes but any that are not can be supplied individually. The sequences of these probes are also available upon request.

3.2. Genotyping of wheat and barley with SSRs

SSRs are abundant in most genomes, and their ease of use and high information content have ensured that they have largely replace RFLP as a mapping tool. In addition to genetic mapping, SSRs are being used increasingly for genotyping, measuring the genetic diversity of breeding materials and cultivars, surveying genetic resources and gene bank collections. On account of their high informativeness and relative ease of use microsatellites are particularly

useful for such studies. As part of a genetic diversity assessment of UK wheat and barley, microsatellite profiles from 65 wheat and 135 barley genotypes have been generated. The wheat and barley varieties were chosen to represent these crops as they were grown in the UK over the past 70 years. It was shown that there was no significant narrowing of the overall diversity during the time period studied. This indicated that plant breeding has resulted in a qualitative rather than a quantitative shift in genetic diversity [14].

TABLE 3. PEARL MILLET ANCHOR PROBES MAPPED TO THE FOXTAIL MILLET GENOME

PSM probes	Pearl millet	Foxtail millet
18	3	
25	2	
37	3	
52	1	
84	4	
87	6	
108	3	I,VIII
174	3	I
196	1,4	VIII
214	2	
248	3	VII
265	4	III
269	7	
280	1	
306	4	
318	5	
319	1	III
321	2	IV
322	2	
341	1	
345	5	
347	1	
386	1	
403	2	IV
410	3	
428	3	VII
458	2	
459	6	
464	4	
473	3	
565	1	VIII
575	6	
588	6	
592	2	
613	1	V
618	7	
648	4	
669	1	
706	2	IX
713	6	V
716	4	III
717	7	
735	5	
737	6	
738	2	
749	5	
756	1	VI
757	1	VIII
761	1	
834	7	
857	7	II
858	1	

TABLE 4. HOMOLOGIES OF SEQUENCED ANCHOR PSR PROBES AND ITS FUNCTIONS

Probes	Homology (best scoring sequence)
100	Rice cDNA (D15704)
101	*Hordeum vulgare* chloroplast photosystem I PSK-I subunit mRNA (L12707)
102	Rice (O24215), barley (Q42829), maize (O24575), s-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50)
104	Wheat (P09195), fructose-1,6-bisphosphatase
105	Maize (P43188), adenylate kinase, chloroplast (EC 2.7.4.3) (ATP-amp transphosphorylase)
106	Rice cDNA; maize cyclophylin (D29701)
107	Tomato (P12360), Arabidopsis (X71878), chlorophyll a-b binding protein
108	Maize cDNA (T20385)
109	Wheat ribulose-1,5-bisphosphate (RUBISCO) gene (M37477)
115	Alfalfa (P51137), glycogen synthase kinase-3 homolog msk-1 (EC 2.7.1.-); Arabidopsis (X75432) ASK-alpha mRNA protein kinase
116	Pea mRNA for P protein (part of glycine cleavage complex) (X59773)
117	Spinach chloroplast carbonic anhydrase mRNA (M27295)
118	Rice chloroplastic aldolase gene (D13513)
119	Rice (P41344), ferredoxin--nadp reductase
121	Wheat (1,3;1,4) beta glucanase (Z22874)
128	Arabidopsis (P27521), chlorophyll a-b binding protein; pine mRNA encoding Lhca4 protien (light harvesting complex of photosystem I) (Z16409)
129	*Synechocystis* sp. (strain pcc 6803) Q55700. Cell division protein ftsh homolog 1 (EC 3.4.24.-)
131	Trypanosoma (Q06084), procyclic form specific polypeptide b-alpha precursor (procyclin); wheat mRNA for proline rich protein (X52472)
135	Pea precursor of the oxygen evolving complex 17kDa protein mRNA (X87435)
137	Tobacco (P30361), cytochrome b6-f complex iron-sulfur subunit 1 precursor (EC 1.10.99.1)
138	Wheat (P46285), sedoheptulose-1,7-bisphosphatase, chloroplast precursor (EC 3.1.3.37)
141	Wheat mRNA for cytosolic phosphoglycerate kinase (X15232)
142	Wheat (P26302), phosphoribulokinase precursor (EC 2.7.1.19)
143	Barley Lhbc mRNA for type III LHII CAB precursor protein (X63197)
144	Barley gene for CP29 precursor for core chlorophyll a/b binding protein of PSII (X63052)
145	Human xanthine dehydrogenase (S66928)
147	Rice chlorplastic aldolase gene (D13513)
148	Maize catalase isozyme 1 (CAT-1) mRNA (M33104)
150	Rice heat shock protein 82 (HSP 82) (Z11920)
151	Tobacco (P09043), glyceraldehyde 3-phosphate dehydrogenase a, chloroplast precursor (EC 1.2.1.12); maize (X07157) mRNA for subunit A of chloroplast GAPDH(GapA) glyceraldehyde-3-phosphate dehydrogenase
152	Human (P13662), nuclear transport factor 2 (ntf-2); rice cDNA (D23112)
154	Barley (Q40070), photosystem II 10 kd polypeptide precursor
155	Barley rubisco activase genes (M55449)
156	Spinach (P12629), 50s ribosomal protein l13,chloroplast precursor (cl13)
158	Myxococcus xanthus (O30612), ATP-dependent clp protease proteolytic subunit (EC 3.4.21.92)
159	Barley Psah mRNA for 10.2kDa photosystem I polypeptide
160	Barley (P08248), plastocyanin precursor
161	Garden pea (Q02028), chloroplast stroma 70 kd heat shock-related protein precursor
164	Arabidopsis (P25857), Maize (M95076), glyceraldehyde 3-phosphate dehydrogenase b, chloroplast precursor (EC1.2.1.12)
165	Rice succinate dehydrogenase K7 (D10414)
167	Cucumber (P13443), glycerate dehydrogenase (EC 1.1.1.29)
168	Arabidopsis mRNA for high mobility group protein (D13491)
169	Barley (P40880), carbonic anhydrase, chloroplast precursor (EC 4.2.1.1)
170	Maize (P08440), fructose-bisphosphate aldolase, cytoplasmic isozyme (EC 4.1.2.13)
171	Flaveria mRNA for P subunit of glycine decarboxylase multi-enzyme complex (Z25857)
172	Maize mRNA for light harvesting chlorophyll a/b binding protein (X55892)
371	Yeast Q00416, trna-splicing endonuclease positive effector
653	Lithospermum erythrorhizon Q40153, lec14b protein

4. NEW DEVELOPMENTS

4.1. Pearl millet SSRs

We are currently developing SSR markers in pearl millet. SSR sequences were obtained from small insert libraries which were enriched for SSR sequences by use of streptavidin-coated paramagnetic beads/biotinylated microsatellite motifs (such as $(GT)_{15}$, $(CT)_{15}$). Approximately 20 to 30% of the clones in the library contained SSR sequences after a single round of enrichment, and more than 50% of the designed primer pairs produced good PCR products in our preliminary survey. So far, 10 SSRs have been generated, five of which have been mapped (Figure 4).

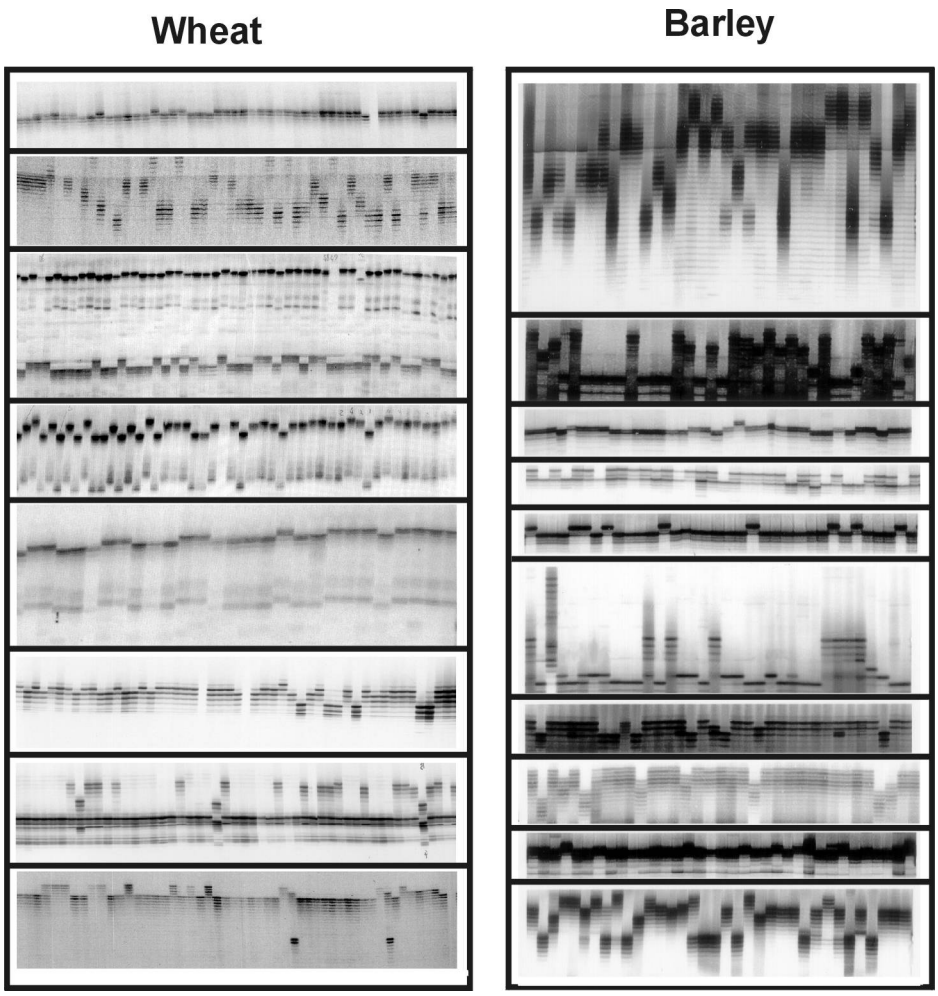

Fig. 6. Genotyping wheat and barley germplasm with SSRs.

A method for the generation of SSR markers from bacterial artificial chromosomes (BACs) has been developed using a pearl millet BAC library developed at JIC that contains approximately 5.8 haploid genome equivalents. Flanking sequences were obtained by 'suppression PCR' with SSRs anchored primers, e.g. $(AC)_{10}T$ and a specific adapter-primer. Preliminary results have demonstrated that all seven PCR primer pairs designed so far produce specific PCR amplicons from the originating BAC clones, from the variety used for BAC library construction and from three other pearl millet lines. Moreover, polymorphism was detected among 20 pearl millet lines. It is predicted that application of this method will yield hundreds of SSR markers. These markers will anchor BACs to the genetic maps and

will provide a useful tool for genome studies. All primer pairs will be made publicly available via MilletGenes (*http://jiio5.jic.bbsrc.ac.uk:8000/cgi-bin/ace/search/millet*). The Department for International Development (DFID, UK) provided funding to develop and map more than 300 SSRs in the pearl millet genome.

4.2. ITEC: International *Triticeae* EST Consortium

Many research groups around the world are collaborating in an effort to put into the public domain at least 40 000 cereal EST sequences by July 2000. As part of this initiative JIC has submitted 1000 sequences from wheat pericarp, wheat endosperm and barley immature inflorescence cDNA libraries. Information on the status of the ITEC database, protocols, etc. can be accessed via the web on *http://wheat.pw.usda.gov/genome.*

REFERENCES

[1] GALE, M.D. and DEVOS, K.M. Comparative genetics in the grasses. Proc. Natl. Acad. Sci. USA 95 (1998) 1971–1974.

[2] GALE, M.D. and DEVOS, K.M. Plant comparative genetics after ten years. Science 282 (1998) 656–659.

[3] DONINI, P. *et al.* The potential of microsatellites ofr high throughput genetic diversity assessment in wheat and barley. Genetic Resources and Crop Evolution 45 (1998) 415–421.

[4] GALE, M.D. *et al.* Genetic maps of hexaploid wheat. In: Li ZS, Xin ZY (eds) Proc 8th Int Wheat Genet Symp. China Agricultural Scientech Press, Beijing, (1995) pp 29–40.

[5] BRYAN, G.J. *et al.* Isolation and characterisation of microsatellites from hexaploid bread wheat. Theor. Appl. Genet. 94 (1997) 557–563.

[6] STEPHENSON, P. *et al*. Fifty new microsatellite loci for the wheat genetic map. Theor. Appl. Genet. 97 (1998) 946–949.

[7] LAURIE, D.A. *et al.* RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (*Hordeum vulgare* L.) cross. Genome 38 (1995) 575–585.

[8] BEZANT, J.H. *et al.* Marker regression mapping of QTL controlling flowering time and plant height in a spring barley (*Hordeum vulgare* L.) cross. Heredity 77 (1996) 64–73.

[9] LAURIE, D.A. *et al.* The distribution of RFLP markers on chromosome 2(2H) of barley in relation to the physical and genetic location of 5S rDNA. Theor. Appl. Genet. 87 (1993) 177–183.

[10] DEVOS, K.M. *et al.* Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. Theor. Appl. Genet. (1999) (in press).

[11] LIU, C.J. *et al.* An RFLP-based genetic map of pearl millet (*Pennisetum glaucum*). Theor. Appl. Genet. 89 (1994) 481–487.

[12] DEVOS, K.M. *et al.* Comparative genetic maps of foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). Theor. Appl. Genet. 96 (1998) 63–68.

[13] WANG, Z.M. *et al*. Construction of RFLP-based maps of foxtail millet, *Setaria italica* (L.) P.Beauv. Theor. Appl. Genet. 96 (1998) 31–36.

[14] DONINI, P. *et al.* Temporal trends in the diversity of UK wheat. Theor. Appl. Genet. (1999) (in press).

DEVELOPMENT, APPLICATION AND DISTRIBUTION OF DNA MARKERS AND GENETIC INFORMATION FOR SORGHUM AND MAIZE IMPROVEMENT

M. LEE
Department of Agronomy,
Iowa State University,
Ames, Iowa, United States of America

Abstract

This final report summarizes the progress made towards the enhancement and distribution of genetic resources (e.g. genetic stocks, seed and DNA clones) used for basic and applied aspects of the genetic improvement of maize and sorghum. The genetic maps of maize and sorghum were improved through comparative mapping of RFLP loci detected by 124 maize cDNA clones and through the development of a new mapping population of maize. Comparative mapping between maize and sorghum and maize and rice, using the set of 124 maize cDNA clones (and other clones) in each study, substantiated previous observations of extensive conservation of locus order but it also provided strong evidence of numerous large-scale chromosomal rearrangements. The new mapping population for maize (intermated B73 × Mo17, 'IBM') was created by random intermating during the first segregating generation. Intermating for four generations prior to the derivation of recombinant inbred lines (RILs) increased the frequency of recombinants at many regions of the maize genome and provided better genetic resolution of locus order. Expansion of the maize genetic map was not uniform along the length of a linkage group and was less than the theoretical expectation. The 350 IBM RILs were genotyped at 512 loci detected by DNA clones, including 76 of the 124 supported by this contract. The production of the sorghum mapping population of RILs from the cross CK60 × PI229828 has been delayed by weather conditions that were not conducive to plant growth and seed development. Seed of the IBM RILs have been distributed (approximately 5000 RILs in total) to 16 research organizations in the public and private sector. The DNA clones have been distributed (1,206 in total) to nine research labs. Further distribution of the seed and clones will be managed by curators at stock centers in the public domain.

1. INTRODUCTION

The options for understanding and enhancing the genetic basis of crop improvement have increased in recent years. One source of new knowledge and technology is the ability to conduct more detailed genetic studies of important traits in one species and then compile the results in such a way that they may be cross-referenced or compared with other studies of the same species and of unrelated species. Such genetic analyses, when coupled with other approaches at the biochemical and physiological level, will reveal the biological basis of important traits and will suggest rational approaches for their modification.

One of the foundations of this new approach to basic and applied biological research in crop plants is the development and utilization of common and shared material and information. Such material and information may take many forms (e.g. data bases, seed, software) and may be used for a wide range of investigations and applications (e.g. marker-assisted selection, assessments of genetic diversity and relationships). For genetic studies, important enabling components have included standard sets of progeny for genetic mapping (mapping populations), DNA clones or sequences used to detect loci on genetic maps and data bases of phenotypic and genotypic information for several species. Subsequently, these resources have been complemented with data bases of DNA sequences of thousands of known

and unknown genes. Collectively, such resources allow researchers to treat groups of sexually isolated species as single genetic systems that may be explored and exploited for crop improvement [1,2].

The need for and utility of such genetic resources depends on many variables related to the forces of nature and actions of human society. This is especially true for the focal species of this project, maize and sorghum. The decades of research in maize and the significant industrial interests have created a vast amount of information that is often difficult to resolve. Specifically, the mapping populations typically used for genetic mapping in maize are represented by a small sample of progeny (40–60 individuals) derived from a population at maximum linkage disequilibrium [3, 4]. The limited opportunities for recovering a recombinant in a given chromosomal region hinders the ability to resolve the order of loci on a genetic map. Thus one of the objectives of this project was to produce a mapping population of maize suitable for higher resolution genetic mapping and other investigations.

In contrast to maize, the basic information and resources for sorghum research are rather limited. However, experiments in comparative genetic mapping with common sets of DNA clones have revealed that large regions of the maize and sorghum genomes have been conserved regarding their gene order and content [5, 6]. Similar analyses have been extended to include several grass species with rice (*Oryza sativa*) as the reference, model genome for the grasses. Such conservation should permit the sharing and transfer of information between the data-rich and data-poor species and research communities. Thus, another objective of the project was to improve the integration of the sorghum and maize genetic maps to facilitate such comparisons and exchange of information derived from investigations of the rice genome. The overall goal of the project was to develop and distribute genetic stocks (seed), DNA clones and sequences, and information which would enable comparative analyses of crop genomes for the purposes of crop improvement.

2. MATERIALS AND METHODS

2.1. Population development

Populations of segregating progeny were developed for maize, sorghum and rice. The maize populations were created by crossing two inbred lines followed by one or more generations of self-pollination to produce recombinant inbred lines (RILs). The maize mapping populations (inbred 1 × inbred 2) used in this project were T × 303 × CO159, T232 × CM37 and B73 × MO17 (intermated; IBM). The first two populations were developed and distributed by Ben Burr [3]. The IBM population was developed at Iowa State University. The intermated population was derived from the single cross hybrid of inbreds B73 (female in this cross) and Mo17. One F1 plant was self-pollinated to produce the F2 generation. In the F2, plants were used once, as male or female, in a cross with another plant so that 250 pairs of plants were mated. A single kernel was taken from each ear and bulked with the seed of the other ears to form the F2 Syn 1 [7]. The procedure was repeated with the F2 Syn 1 plants and for three additional generations to produce the F2 Syn 4.

A set of 370 recombinant inbred lines (RILs) was derived from the F2 Syn 4 generation of the IBM population. Production of these lines was initiated at ISU and completed at Pioneer Hi-Bred International's winter nursery at Puerto Rico. The RILs were derived through single seed descent without intentional selection. Initially, 420 F2 Syn 4 plants were self-pollinated to create F3 lines. Beginning in the F3 generation, independent lineages were

maintained through each generation of subsequent and continuous self-pollination. To advance to the F4 generation, a single F3 plant per line was self-pollinated. Kernels from that selfed ear were planted to grow the F4-generation plants. This was repeated for three successive generations to produce the F7:8 generation kernels (a.k.a. F8-generation kernels tracing to one F7-generation plant). Each of the 370 F7:8 RILs trace to a different plant in the F2 Syn 4 generation. During the derivation of the RILs, 50 lineages of the original 420 were eliminated at some stage of inbreeding for various reasons (e.g. male or female sterility, susceptibility to a disease, extreme lack of vigour) that prevented them from producing seed at some generation. Data at RFLP loci identified an additional 25 RILs that had been contaminated (e.g.. seed mixtures or nonparental pollen) during their development; thus, those RILs were eliminated. Consequently, the maize IBM populations consists of 345 RILs. Adequate seed supplies have been produced for most of those RILs.

The sorghum and rice populations used in this project have been described. The sorghum population is a group of 78 F2 generation plants of a cross between inbred lines CK60 and PI229828 [5]. The rice population is a group of BC1F1 plants of the cross, *Oryza sativa* [8, 9] (cultivar BS125)/*O.longistamata*/*O.sativa*).

2.2. Collection of segregation data at DNA marker loci

The genetic data and maps presented in this report were collected using sets of DNA clones used to detect restriction fragment length polymorphism (RFLPs). The 124 DNA clones characterized and specified under this research contract are maize cDNA clones derived from mRNA isolated from seedling roots [5]. The protocols for using these clones as probes in Southern hybridizations have been described in detail for sorghum [5], maize [10] and rice [9].

2.3. Construction of genetic maps

Construction of genetic maps was facilitated with the software MAPMAKER 3.0 [11] following the procedures described in [4] and [12].

3. RESULTS AND DISCUSSION

3.1. Maize

The 350 RILs were genotyped at 512 RFLP loci. The segregation data are being verified and edited but a preliminary analysis has been conducted. This includes 75 loci detected by the maize cDNA clones. The IBM genetic map comprises 16 linkage groups and over 3000 cM. Normally, a maize genetic map with this number of loci would consist of 10–11 linkage groups with a total of 2000 cM. Apparently, the intermating has enhanced the frequency of recombinants at many regions of the maize genome. A more detailed analysis of two groups of IBM progeny, one group of 95 individuals before intermating and a second group of 90 individuals after five generations of intermating has been conducted with 156 common RFLP loci. The analysis revealed a similar degree of expansion of the genetic map (compared to the full set of 350 RILs) after intermating. However, the degree of map expansion was not uniform along the length of a linkage group.

3.2. Sorghum

Further integration of the maize and sorghum maps was achieved. Previously, the 124 maize cDNA clones had been used to detect RFLP loci in sorghum but they had not been used for genetic mapping in maize [5]. In this project, the 124 clones were used to detect RFLP loci in the maize genome using the mapping populations Tx303 × CO159 and T232 × CM37. Generally, the data collected with the additional 124 clones strengthened the observations of the initial investigation [5]. Each maize linkage group, all of which have been clearly assigned to maize chromosomes, usually contains groups of loci that represent two linkage groups of sorghum. Within a region of a maize linkage group, locus order appears to be highly conserved relative to a given region of a sorghum linkage. However, an adjacent region of the same maize linkage group will often exhibit a high degree of colinearity with a different sorghum linkage group. These observations suggest that the maize and sorghum genomes may be distinguished by several, large-scale segmental rearrangements. Similar patterns of rearrangements have been detected between the maize and rice genetic maps.

The development of sorghum RILs for the CK60 × PI229828 population has been delayed by adverse weather conditions. Simple sequence repeats (SSRs) are being mapped in the population CK60 × PI229828 (78 F2 plants) in collaboration with the University of Milan (G. Taramino and E. Pe) and the USDA/ARS (S. Kresovich, now at Cornell University). Seven SSR loci have been mapped with the Univ. of Milan and those data have been integrated into the RFLP map for that population. Those and related results have been published [13]. Segregation data at 32 SSR loci have been collected by the USDA and those data will establish additional loci.

The CK60 × PI229828 population is being converted into a set of recombinant inbred lines suitable for widespread distribution. In 1996, self-pollinated seed in the F4 generation was produced on 180 F3 plants tracing to 180 F2 plants. The F4 generation seed has been sent to winter nursery to produce the F5 generation seed. In the summer of 1998 in Ames, we produced F8 generation seed for 110 recombinant inbred lines. This population should be suitable as a common mapping for sorghum because of its relatively high frequency of DNA polymorphism and the ease with which inbred progeny are derived. Prior to the distribution of the seed, the lines must be purified in the field on phenotypic and genotypic bases, identified by PCR-based DNA markers and genotyped with a subset of the DNA markers used to make the genetic map. Unfortunately, adverse weather conditions (cool temperatures and excess precipitation) were unsuitable for sorghum growth and development in our field nursery in Ames, Iowa in 1999. Thus, the final stage of seed production and purification has been delayed. These steps will be repeated in the year 2000 in Ames or in an environment better suited for the production of sorghum seed.

3.3. Comparative genetic mapping among maize, sorghum and rice

Of the 124 maize cDNA clones used to map RFLPs in maize and sorghum, 71 also detected RFLPs in rice. These data were added to a larger data set in collaboration with a group at Cornell University [9]. Collectively, the data sets established 182 new loci that have been mapped in the maize and rice genomes as common reference points (previously, only 146 loci had been comparatively mapped between rice and maize).

Comparative genetic analysis revealed over 20 chromosomal rearrangements in maize relative to rice. The changes included telomeric fusions between, and nested insertion of rice

linkage groups, intrachromosomal inversions and a nonreciprocal translocation. A progenitor maize genome of eight chromosomes was inferred.

3.4. Distribution of information, DNA clones and sequences, and seed

The IBM RIL mapping population is becoming a standard, widely used genetic resource in the maize research community. Seed of the IBM maize RILs have been distributed to the USDA-ARS at the University of Missouri, the USDA-ARS at North Carolina State University, the USDA-ARS at Iowa State University, University of Wisconsin, Texas Tech University, University of Georgia, Cornell University, the Maize Genetic Stock Center at the University of Illinois (USDA-ARS), University of Paris, Limagrain, DeKalb Seed Company, Monsanto, Keygene, Pioneer Hi-Bred International, DuPont and Novartis. There have been no requests for the seed of the sorghum mapping population, CK60 × PI229828.

The complete set of segregation data (RFLP loci) are being prepared for deposit into the public domain, available through the internet, at MaizeDB.

Nine sets of 134 cDNA clones have been distributed. The sets were sent to two labs in Brazil, two labs at Texas A&M University, a research institute in California, the National Center for Genome Resources (Santa Fe, New Mexico), Cornell University, Pioneer Hi-Bred and the USDA-ARS Maize Genome Center at the University of Missouri. Further distribution of the clones will be managed by the USDA-ARS at the University of Missouri.

ACKNOWLEDGEMENTS

I thank N. Sharopova, W. Woodman, B. Wilson and S. McCouch for sharing their analyses of segregation data prior to publication. This project was partially supported by IAEA contract 8494/R1 and USDA-NRI Grant 97-35300–4939.

REFERENCES

[1] GALE, M.D. DEVOS, K.M. Plant comparative genetics after 10 years. Science 282 (1998) 656–659.
[2] BENNETZEN, J.L. FREELING, M. The unified grass genome: synergy in synteny. Genome Research 7 (1997) 301–306.
[3] BURR, B. *et al.* Gene mapping with recombinant inbreds in maize. Genetics 118 (1998) 519–526.
[4] DAVIS, G.L. *et al.* A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics 152 (1999) 1137–1172.
[5] PEREIRA, M. G. *et al.* Construction of an RFLP map in sorghum and comparative mapping in maize. Genome 37 (1994) 236–243.
[6] BENNETZEN, J.L. *et al.* Grass Genomes. Proc. Natl. Acad. Sci. USA 95 (1998) 1975–1978.
[7] COVARRUBIAS-PRIETO, J. HALLAUER, A.R., LAMKEY, K.R. Intermating F2 populations of maize. Genetika 21 (1999) 111–126.
[8] CAUSSE, M. *et al.* 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138 (1994) 1251–1274.

[9] WILSON, W. A. *et al.* Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize, and the domesticated panicoids. Genetics 1999 (in press).
[10] VELDBOOM, L., Lee R.M., Woodman, W.L. Molecular-marker-facilitated studies in an elite maize population: I. Linkage analysis and determination of QTL for morphological traits. Theor. Appl. Genet. 88 (1994) 7–16.
[11] LANDER, E.S. *et al.* MAPMAKER: An interactive computer package for maps of experimental and natural populations. Genomics 1 (1987) 174–181.
[12] LUI, S.-C. *et al.* Genome-wide high resolution mapping by recurrent intermating using *Arabidopsis thaliani* as a model. Genetics 142 (1996) 247–258.
[13] TARAMINO, G. *et al.* Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor. Appl. Genet. 95 (1997) 66–72.

DISTRIBUTION AND USES OF LEGUME DNA CLONE RESOURCES

N.D. YOUNG
Departments of Plant Pathology and Plant Biology,
University of Minnesota,
St. Paul, Minnesota, United States of America

Abstract

Since 1990, my lab has developed and distributed various DNA clone resources for the legumes. In the first several years, the focus was on members of the tropical genus, *Vigna*, including the widely cultivated species, mungbean (*V. radiata*) and cowpea (*V. unguiculata*). Both of these grain legumes play key roles in agriculture in developing countries of Asia (mungbean) and Africa (cowpea). Moreover, because there is substantial genome conservation among legumes [1], these genetic resources have also been utilized by a wide range of researchers in other crop species. In 1997, my lab began to focus on the development and distribution of a new generation of DNA clone resources; Bacterial Artificial Chromosomes (BAC). A library of these clones was constructed in soybean (*Glycine max*) the most important legume species worldwide in terms of economic value. Again, the library has become a valuable resource for the legume research community and has been widely used in studies of legume genomics.

1. INTRODUCTION

1.1. Genome studies in *Vigna*

The first DNA marker maps for mungbean and cowpea were published in 1993 [2, 3]. The mungbean map consisted of 171 restriction fragment length polymorphism (RFLP) markers, while the cowpea map consisted of 83 RFLPs. Since that time, the number of RFLPs placed on the mungbean has increased to more 260. Many of these markers, especially those derived from mungbean or cowpea, were developed in my lab at the University of Minnesota. The RFLP clones found on these maps and developed in my lab are available to researchers worldwide who are interested in *Vigna* or other legume crops.

1.2.Genome studies in soybean based on the BAC library

BAC libraries have several applications in the growing field of genomics. Because they have very large inserts, BACs can be used for chromosome walking and positional cloning. This is one of the most powerful methods for isolating genes known only by map position today. BAC libraries form the basis of physical mapping. In the next decade, complete physical maps for many important crop species, including soybean and perhaps other legumes, will probably be completed using BAC libraries. Finally, BAC libraries are also well suited for comparative genomics, especially for questions of microsynteny. In these studies, the question is one of comparing genomes of different organisms at the micro-level. There is a growing body of evidence that gene order is often maintained, even at the kilobase level of resolution.

2. MATERIALS AND METHODS

2.1. Construction of a *Vigna* RFLP library

A DNA library suitable for RFLP analysis was constructed from mungbean and cowpea genomic DNA. To prepare these libraries, DNA was digested with *PstI* (a methylation-sensitive enzyme) and separated according to size by sucrose gradient centrifugation. The fraction between 500 and 3000 base pairs was collected and ligated into pUC19 by standard methods [4].

2.2. Construction of a soybean BAC library

Ten day-old soybean plants, cultivar Faribault, were used for preparation of high molecular weight DNA in agarose plugs. These DNA samples were partially digested with *EcoR*I and separated on a 1% low melting point agarose gel using a clamped homogeneous electric field device. The isolated DNA was ligated into vector, pECSBAC4 (Dr. Richard Michelmore, University of California, U.S.A.) and transformed into *E. coli* by electroporation. White recombinant BAC clones were picked manually and transferred to 384-well plates. For long-term storage, the library was replicated three times. A total of 78 000 BAC colonies were isolated with an average insert size of 120 kilobase pairs [5].

3. RESULTS AND DISCUSSION

3.1. Clone distribution

3.1.1. Vigna Clone Distribution

Since 1992, when the first set of *Vigna* clones were distributed, we have responded to a total of 35 requests for biological materials. Approximately 2000 clones have been sent out to fill these requests, which came from researchers in eleven different countries including, United States, Republic of China, Australia, Nigeria, Germany, India, Italy, United Kingdom, Japan, Korea, and Israel. Twenty-two of the requests were filled before 1995, while 13 have been filled since.

3.1.2. BAC Library distribution

In 1998, the entire BAC library for soybean has been on deposit at the Clemson Genomics Center under the direction of Dr. Rod Wing. This facilities provides libraries, high density filters, and support for numerous plant BAC libraries. Requests to use the library are still routed through my lab at the University of Minnesota, but the distribution of physical resources comes from Clemson University. The BAC library consists of approximately 78 000 clones each with an insert of soybean DNA approximately 120 000 base pairs in length. Thus, the library provides approximately seven-fold coverage of the soybean genome.

There have already been 11 requests for the BAC library, high density filters derived from the library, or pools of BAC clones suitable for PCR screening. These requests have come primarily from scientists in the United States, along with one request from China.

3.2. Clone distribution packages

3.2.1. Vigna clone distribution package

The current package of *Vigna* clones that is sent upon request includes the following:

- Seventy-three single-copy DNA clones from mungbean and cowpea.
- Two moderately repetitive, highly polymorphic mungbean clones.
- Genomic DNA from mungbean, cowpea, soybean, common bean, and pigeon pea.
- Spreadsheet datafile with: locus name, cloning vector, insert size, map position, mapping enzyme, chromosomal locations in common bean and soybean, if known.
- Current RFLP linkage maps for mungbean and cowpea.
- DNA transformation, plasmid miniprep, and PCR amplification protocols.

3.2.1 Soybean BAC library distribution package

- Five high density filters containing all 78 000 soybean BAC clones spotted in duplicate on nylon filters. These filters are suitable for hybridization analysis with radiolabeled DNA probes.
- Pools of BAC clones that make it possible to screen the entire library using Polymerase Chain Reaction (PCR) technology with only 100 total PCR reactions.
- Upon request, specific clones from the library are prepared and distributed, either in the form of bacterial suspensions or purified clone.

3.3. Typical applications of clone resources

3.3.1 Typical Applications of Vigna Clones

The research on *Vigna* clones has focused in areas such as: 1) comparative genome mapping with related legume crop species, 2) mapping and characterization of quantitative trait loci (QTLs), and 3) mapping and tagging genes involved in disease and pest resistance. Table 1 briefly describes the scientists who have worked with the clones and the types of research projects involved.

3.3.2. Typical Applications of Soybean BAC Clones

Research with the soybean BAC clones has focused on two areas: 1) positional cloning of important disease resistance genes and 2) comparative genomic analysis aimed at understand microsyntenic relationships among legume genera. Table 2 briefly describes the scientists who have worked with the clones and the types of research projects involved.

TABLE 1. RESEARCH GROUPS THAT HAVE RECEIVED *VIGNA* RFLP CLONES AND BRIEF DESCRIPTION OF THEIR RESEARCH ACTIVITIES (1992–1998)

Principle investigator	Location	Research Application
E. Vallejos	U. Florida	Comparative mapping
G. Kuo	Taiwan	Disease resistance
J. Manners	CSIRO	Mungbean mapping
R. Shoemaker	Iowa State	Comparative mapping
D. Pignone	Bari, Italy	*In situ* hybridization in
P. Heslop-Harr.	Norwich, UK	cowpea and relatives
J. McCallum	New Zealand	Disease resistance in pea
N.-S. Kim	Korea	Mapping in Korean cultivars
L. Kumar	India	Disease resistance
S. N. Raina	India	Taxonomic relations in *Vicia*
S.R. Rangasamy	Tamil Nadu	Disease resistance
G. Kochert	Univ. Georgia	Peanut mapping
S. Abbo	Rehovot, Israel	QTLs in chickpea
A. Karasawa	Aoba-ku, Japan	CMV resistance in cowpea
C. Lambrides	CSIRO	Mungbean mapping and breeding
R. K. Sahu	India	Bruchid resistance
C. Mendenez	U. Cal., Davis	Seed weight QTL in cowpea
M. Ishimoto	Japan	Disease resistance
C. Liu	CSIRO	*Lablab* mapping
A. Paterson	Texas A&M	Peanut mapping
C.-S. Chen	Taiwan	Disease resistance
S. Chao	Taiwan	Disease resistance
S. Lee	U. Georgia	Comparative genomics
B. Sharma	India	*Vigna* mapping
J. Specht	U. Nebraska	Soybean mapping
V. Sant	India	Disease resistance
S. Lakhanpaul	India	*Vigna* mapping
M. Timko	U. Virginia	*Striga* resistance in cowpea

TABLE 2. RESEARCH GROUPS THAT HAVE RECEIVED SOYBEAN BAC CLONES AND BRIEF DESCRIPTION OF THEIR RESEARCH ACTIVITIES (1998-PRESENT)

Principle investigator	Location	Research Application
E. Vallejos	Florida	Comparative mapping
R. Bolla	Missouri	Disease resistance
P. Cregan	Maryland	Sequence polymorphism search
R. Innes	Indiana	Disease resistance
P. Keim	Arizona	Disease resistance
H. Knap	S. Carolina	Disease resistance
D. Lightfoot	Illinois	Disease resistance
F. Liu	P.R. China	Genome organization
R. Shoemaker	Iowa	Genome organization
L. Vodkin	Illinois	Genome organization
R. Wing	S. Carolina	BAC library characterization

REFERENCES

[1] BOUTIN, S. *et al.* Genome conservation among three legume genera detected with DNA markers. Genome 38 (1995) 928–937.

[2] FATOKUN, C., DANESH, D., YOUNG, N. "RFLP linkage map for cowpea (*Vigna unguiculata* (L.) Walp.)", Genetic Maps 1992 (O'BRIEN, S.J., Ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1993) 6.256–6.258.

[3] MENANCIO-HAUTEA, D., KUMAR, L., DANESH, D., YOUNG, N. "RFLP linkage map for mungbean (*Vigna radiata* (L.) Wilczek)", Genetic Maps 1992 (O'BRIEN, S.J., Ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NewYork (1993) 6.259–6.260.

[4] YOUNG, N. *et al.* D. RFLP mapping of a major bruchid resistance gene in mungbean (*Vigna radiata*, L. Wilczek). Theor. Appl. Genet. 84 (1992) 839–844.

[5] DANESH, D. *et al.* A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. Theor. Appl. Genet. 96 (1998) 196–206.

ARBITRARILY AMPLIFIED DNA: NEW MOLECULAR APPROACHES TO PLANT BREEDING, ECOLOGY AND EVOLUTION

G. CAETANO-ANOLLÉS
Department of Biology,
University of Oslo,
Oslo, Norway

Abstract

Several DNA fingerprinting techniques that use arbitrary primers to characterize, scan and tag genomic DNA were optimized and used to study plants and microbial pathogens. The generated arbitrarily amplified DNA (AAD) profiles could be tailored in their complexity and polymorphic content, allowing analysis of closely related organisms, such as vegetatively-propagated horticultural crops or clonal fungal populations. AAD markers were used in cultivar and strain identification, map-based cloning, and marker-assisted breeding, sometimes as sequence-tagged sites. Phenetic analysis using parsimony, cluster, and numerical methods was applied successfully to the identification of genetic relationships in turfgrass species such as bermudagrass, woody plants such as dogwoods, and floricultural species such as petunia and chrysanthemum. AAD profiles were used to measure for the first time a genome-wide mutation rate, directly in a plant. Mutation rates in vegetatively propagated bermudagrass were comparable to those in human, mice, fruit flies, and worms. In combination with established tools used in molecular systematics (e.g. rDNA sequence analysis), AAD markers tracked the introduction of exotic dogwood anthracnose-causing fungi in North America. As part of a breeding effort to combat dogwood diseases, AAD was used in pseudo-testcross mapping of the tree at the intra-specific level. Markers were efficiently generated despite the close relatedness of parental dogwood material. Finally, DNA markers and tags were also generated in soybean, and were used to construct high density maps and walk towards defined genomic regions in the positional cloning of the supernodulation *nts-1* symbiotic gene.

1. INTRODUCTION

In recent years, a number of molecular techniques have been recruited to complement traditional methods for the evaluation of genomes and biodiversity [1]. Most of these techniques are based on the analysis of information-rich nucleic acid molecules and provide reliable estimators of relatedness, phylogeny and inheritance of genetic material. Nucleic acid markers are by far the most powerful and widely used, because they directly "depict" nucleic acid sequence composition. These markers restrict analysis of the typically 10^6–10^{10} bp of a genome to selected nucleic acid regions representing only 1–10^4 bp of sequence. Many of them use hydrogen-bonding interactions between nucleic acid strands ('hybridization') and oligonucleotide-driven enzymatic accumulation of specific nucleic acid sequences ('amplification') to uncover polymorphism, and therefore sequence variability between individual genomes. Nucleic acid marker methods can be divided into four groups [2]: (i) hybridization-based analysis [e.g. restriction fragment length polymorphism (RFLP) analysis], (ii) amplification-based nucleic acid scanning (e.g. AFLP®), (iii) amplification-based nucleic acid profiling [e.g. polymerase chain reaction (PCR) amplification of microsatellites], and (iv) sequence-targeted techniques (e.g. oligonucleotide arrays). Because different applications demand the use of different DNA marker systems [1], the researcher is usually confronted with the difficult task of choosing a suitable technique. The selection of a marker system depends on a number of technical factors, including throughput and speed, equipment and operator skills, cost and automation, sensitivity, and overall reliability.

Selection also depends on the ability to differentiate the sampled material. Profiling methods differ in their ability to detect DNA polymorphisms in a given population (informativeness) and in the number of loci simultaneously targeted per experiment (multiplex ratio). DNA markers can also be dominant or co-dominant and can exhibit different expected heterozygosities (a measure of the number of alleles that are detected). Informativeness and co-dominance are important factors in genetic mapping and trait tagging. Desirable markers for these applications should be highly polymorphic and exhibit multiple co-dominant alleles (e.g. microsatellites). Alternatively, robust multilocus-fingerprinting techniques with high multiplex ratios (e.g. AFLP) can be used efficiently despite low expected heterozygosities. When markers are used to estimate genetic diversity and build molecular phylogenies, the taxonomical level of analysis becomes relevant. Nucleic acid scanning techniques are here useful for distinguishing closely related organisms (usually at or below the species level). The neutrality and high allele number of microsatellite markers make them ideally suited for the study of populations. Finally, phylogenetic analysis can be well accomplished by direct sequencing of PCR products or by using RFLP markers.

The amplification-based nucleic acid scanning techniques that use arbitrarily amplified DNA (AAD) fall within one special class of nucleic acid typing that does not require prior knowledge of nucleotide sequence or cloned and characterized hybridization probes [3–5]. They are versatile and universal as demonstrated by the many applications and wide range of organisms studied and reported in thousands of publications (for recent reviews see [6, 7]). These techniques have been used in genetic and physical mapping, general fingerprinting, population biology, taxonomy and systematics, and map-based cloning of genes. High throughput has facilitated marker-assisted selection, trait introgression, and study of multigenic or quantitative traits. AAD methods produce characteristic amplification signatures from virtually any nucleic acid template and use one or more short (typically 5–32 nt) oligonucleotides of arbitrary or semi-arbitrary sequence to target a multiplicity of anonymous sites. These signatures (fingerprints) are composed mainly of amplification products of varying length. Extensive mismatching between the oligonucleotide primer and the template can be tolerated [8]. However, primer hybridization to perfect or partially complementary sites in the target nucleic acid requires that the first 5–6 nt from the 3’ terminus of the primer faithfully match those in the template. Primers as short as 5-mers can fingerprint nucleic acids [8, 9]. However, primers can harbour an arbitrary sequence of only 3 nt, if an extraordinarily stable mini-hairpin is attached at their 5’ termini [10]. These mini-hairpin primers have been used effectively in the fingerprinting of small templates, such as PCR products (0.2-1 kb), plasmids (2–5 kb), DAF fingerprints (15–25 kb), and cloned genomic fragments (50–250 kb).

Molecular markers can help in the breeding and management of plants, by evaluating genetic differences in populations, exotic varieties and wild species and guiding the introgression of new and important traits into elite germplasm. Markers can also determine levels of heterozygosity and homozygosity in breeding stock. Their use in genetic mapping can pin-point genes controlling important traits and identify gene combinations that are most favourable in breeding strategies. Finally, markers can facilitate the study of the distribution, diversity and evolution of plant pathogens and pests, by identifying co-evolution strategies that can aid in forecasting epidemics and designing management practices. We have concentrated efforts in the study of markers systems that can detect differences at the genetic level between closely related organisms. These markers were used in a number of applications that include the positional cloning of nodulation genes in legumes, tagging of soybean nematode resistance genes, generation of a genetic map in a model legume (*Lotus japonicus)*,

the study of biodiversity in turfgrass, floricultural crops, woody plants, and in population biology of fungal pathogens. Here we summarize some of our results in the context of the development of new marker systems and experimental approaches.

2. MATERIALS AND METHODS

2.1. Biological materials and DNA extraction

Plant material was grown, kept vegetative, and harvested using standard techniques. Cultivars of horticultural crops were obtained from commercial sources. Dogwood anthracnose resistant lines were selected from asymptomatic trees in Catoctin Mountain Park (Maryland). Tufgrass cultivars were obtained from breeder stocks, and off-types were sampled in golf courses and fields throughout the US. In mutation rate experiments, bermudagrass cultivars were grown in pots under open daylight from single sprigs planted during spring. Samples consisting of individual leaf blades were harvested in concentric rings equidistant from the initial sprig planting. Fungi were collected from dogwood anthracnose lesions. Fungal isolates were grown and maintained in potato dextrose V8 juice agar. Compatibility of hyphal anastomosis was tested in agar plates. Isolates were paired in replicated bifactorial designs, by laying 7 mm plugs 20 mm apart from each other with the mycelium in contact with the agar surface. Mycelial interfaces were examined for compatibility and hyphal lysis using 40 × magnification. DNA was extracted from leaf tissue and mycelia ground to a fine powder with polyvinylpolypyrrolidone under liquid nitrogen, using a commercial kit (Gentra Systems, Research Triangle Park, NC).

2.2. DNA amplification

In the course of this decade we developed several AAD methodologies, including the original DNA amplification fingerprinting (DAF) technique [9]. Some of these strategies, such as minihairpin primed DAF (mhpDAF) [10], template endonuclease cleaved multiple arbitrary amplicon profiling (tecMAAP) [11] and arbitrary signatures from amplification profiles (ASAP) [12], increase detection of polymorphic DNA and are therefore capable of analyzing a same set of templates at different taxonomical levels. Since AAD techniques depend on many amplification factors, they required careful optimization. We used Taguchi methods [13–15] to optimize different AAD techniques [15 16]. Taguchi-optimized DAF reactions (10–20 μL) contained 8-9 μM of primer and 0.1 ng/µl of template DNA, and were generally amplified for 35 cycles of 0.5 min at 96°C, 1 min at 50°C, and 1 min at 74°C, depending on the thermal cycling unit. ASAP reactions contained high primer concentrations (9 μM) and were assembled and amplified as previously described [12] with minor modifications. Amplification products were electrophoresed in polyester-backed 0.45 mm-thick polyacrylamide gels, silver stained, and preserved by drying at room temperature.

2.3. Measurement of genome-wide mutation rates

ASAP reactions were used to measure mutation rates according to the procedure described by Caetano-Anollés [17]. Total genomic mutation rates (μ) were calculated over several time scales from estimates of nucleotide sequence divergence (*D)* within arbitrary collections of DAF products. Mutation rates were given as rate of change per unit DNA (genome or bp) per unit time (generation or a). Mutation rates per generation were either expressed per nucleotide (μ) or per total genome (μ_g). Sequence divergence was estimated

from the number of ASAP polymorphisms (p) and the total number of ASAP products (n), was expressed as nucleotide changes (%) observed within the amplified sequence of monomorphic DAF fragments, and was calculated according to the Eq. (1).

$$D = \frac{p}{2fn} 100 \quad (1)$$

The value f is the length of the arbitrary sequence targeted by annealing of the mini-hairpin primers (f = 5.5 bp; cf. [18]). Since D and p were linearly correlated ($t < 0.05$), sequence divergence could be calculated by extrapolation.

2.4. Genetic mapping, bulked segregant analysis and artificial chromosome libraries

Data from segregating populations were subjected to maximum likelihood analysis using the MAPMAKER program. Mapping data obtained using soybean recombinant inbred lines (RIL) was placed in the Utah and Iowa soybean databases. Wild-type Bragg and EMS-induced mutants defective in the autoregulatory control of nodule mass and number (nts382) or nodulation [nod49 (*rj1*) and nod139 (*rj6*)] were crossed with *Glycine soja* PI468.397 to produce F_2 and F_3 segregating populations. Bulked segregant analysis (BSA) [19] was used to find markers linked to simple and complex traits with the sole knowledge of phenotype. Markers associated with particular traits or genomic regions were identified using pooled DNA samples. Individuals in a segregating population that either expressed or fail to express a symbiotic phenotype or RFLP pattern were pooled, and DNA polymorphisms between the pooled samples were identified using AAD techniques. Adequately large pools ensured the random assortment of all genetic variability in the parental material through chromosome segregation and recombination at meiosis, except for the control sorting of the chromosome region linked to the desired phenotype. DNA polymorphisms were converted into landmarks for genetic and physical mapping, by direct isolation of AAD bands from silver stained gels, cloning and sequencing. These sequence-characterized amplified regions (SCARs) [20] were used, for example, as landmarks in the assemblage of contigs in the positional cloning of the *nts* locus or as tags in marker-assisted breeding of soybean cyst nematode resistance. Selected SCARs were also used in the study of primer-template interactions during AAD amplification.

The construction of a soybean yeast artificial chromosome (YAC) library was initiated. High molecular weight DNA was isolated from agarose-embedded and lysed soybean mesophyll protoplasts, size-selected (>250 kb) by pulsed field gel electrophoresis (PFGE), and ligated into the pYAC4 vector. Ligated DNA was transformed into yeast, and selected by colony colour and uracyl auxotrophy. Southern analysis of PFGE karyotypes showed strong hybridization to total genomic DNA in about half of clones, probably due to high abundance of repeated DNA sequences. Weakly hybridizing signals were assumed to result from inserts containing predominantly unique DNA. About 7% of clones contained chloroplast DNA. At present, the partial soybean YAC library represents about 30% of the soybean genome, with inserts being stable and averaging 220 kb in size (range 50–960 kb). Endclones from YAC insert DNA have been isolated using vectorette PCR. When endclones were hybridized to restricted DNA from *G. max* and *G. soja*, they detected with equal frequency either long and interspersed repeated sequences or unique fragments, validating their use in chromosome walking. A soybean bacterial artificial chromosome (BAC) library with 4 × genome coverage and cloned inserts averaging 105 kb (range 50–80 kb) was also constructed.

2.5. Phenetic, phylogenetic and RNA structure analysis

AAD products were scored as unordered genetic characters (loci) and analyzed using NTSYS-pc (*Numerical taxonomy and multivariate analysis*; v. 1.8, Exeter Software, Setauket, NY) or PAUP* (*Phylogenetic analysis using parsimony*; v. 4.0, Sinauer Assoc., Sunderland, MA). Phenetic relationships were evaluated using Jaccard similarity coefficients and cluster (UPGMA) and neighbor-joining (NJ) algorithms, and confirmed using principal co-rdinate analysis (PCO). PCR-amplified internal transcribed sequences (ITS) of rDNA were aligned using the ClustalX program (v. 1.68β), and phylogenies reconstructed using PAUP*, PHYLIP (*Phylogenetic inference package*; v. 3.75c), and PUZZLE (*Quartet puzzling*; v. 3.1). Phylogenetic reliability was tested by the nonparametric bootstrap method. Pairwise distances were computed from nucleotide sequences based on the two-parameter nucleotide substitution model of Kimura in PHYLIP, and were used to assess differences in substitution rates within spacer regions. RNA secondary structure of spacer regions was predicted by the free-energy minimization method using the MFOLD program (v. 3.0). The ITS2 secondary structure was inferred from recurrent RNA folding patterns, 5.8S rRNA constraints and the yeast spacer model [21].

3. RESULTS AND DISCUSSION

3.1. AAD: optimization and identification of important primer-template interactions

3.1.1. AAD optimization using Taguchi methods

The simplicity and proven success of Taguchi methods [13] in industrial process design, offered a cost-effective strategy for optimization of DNA amplification of plants and microbes [14 15]. For example, we used L_9 (3^4) and L_{18} (3^8) orthogonal arrays to study the interaction of amplification components and thermal cycling parameters in DAF [15 16]. Analysis of variance (ANOVA) decomposed the contribution of individual amplification factors to the responses of amplification yield or product number, while verification experiments established that optimum conditions were predictable, verifiable and reproducible. Several amplification components (primer, magnesium and enzyme) conditioned amplification. However, annealing temperature and time were the only important thermal cycling contributing factors. The Taguchi approach defined a robust and transportable amplification protocol based on high annealing temperatures (typically 48°C) and primer concentrations (typically 8 μM), which can be applied to the fingerprinting of a wide range of DNA templates of plant and fungal origin.

3.1.2. Worldwide web resource

A worldwide web site has been constructed with detailed protocols, general information on molecular markers, useful tips, and an extensive collection of links [22]. Protocols cover several AAD techniques (RAPD, DAF, AFLP, ASAP), DNA separation, silver staining, and the optimization of DNA amplification.

3.1.3. Primer-template interactions in sequence tags from arbitrarily amplified DNA

The generation of arbitrarily amplified DNA is primarily determined by the sequence and concentration of the arbitrary primer that anneals to short and complementary inverted

repeats closely spaced in the nucleic acid template [8]. Amplification can accommodate extensive primer-template mismatching events [8, 23–25], depending mainly on primer length and template complexity. Since not all targeted sites are efficiently amplified [8], mismatching could influence the competitive ability of many amplification products and ultimately the kinetics of the overall amplification reaction. In previous studies, we found evidence to suggest that amplification products are capable of forming hairpin loops because of their palindromic termini [8, 10]. Assignment of amplification products generated with mini-hairpin primers to amplicons expected in bacterial plasmids showed the existence of physical interaction between annealing sites, probably during amplification of first-round products [10]. The ability of the primer to displace these first-round template structures appears an important element during the template "screening" phase of the amplification process, whereby the rare primer-template duplexes are stabilized by primer extension and later transformed into accumulating amplification products. We explored the existence of template-template interactions during the nucleic acid scanning reaction by analyzing the sequence of amplicons from SCARs in soybean, rice, lettuce, bean, apple, strawberry, tomato, burrowing nematodes, zebrafish and mosquito. The extent of duplex formation between the termini of the SCAR products was evaluated. Figure 1 shows results obtained from mosquito SCARs. Chi-square analysis indicated that while base pair matching values fitted those expected from average GC or AT ratios for the individual SCARs with high level of confidence, there were significant departures that could only be explained by the selective interaction of sequence positions internal to the amplified products. This observation suggests that sequence-tagged sites are selected not only by the sequence of the primers but also by the sequence of the targeted regions.

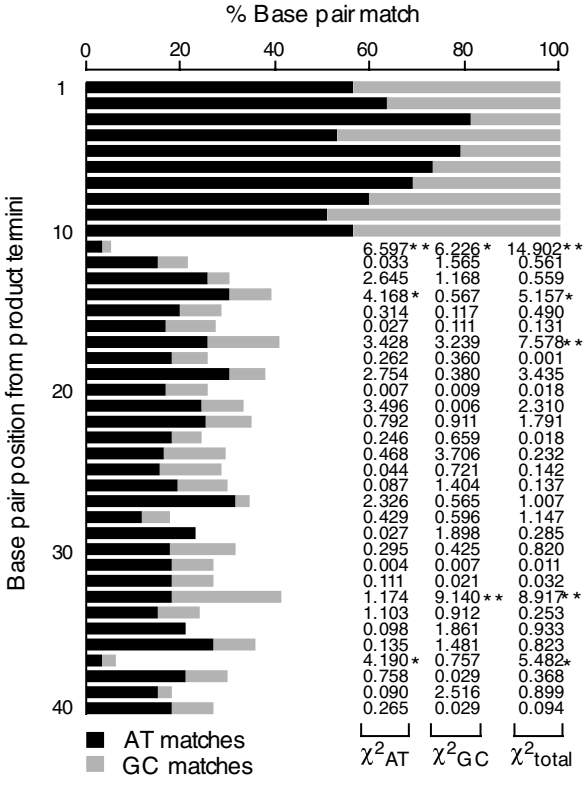

Fig. 1. Duplex interactions between termini of arbitrarily amplified DNA products used as sequence-tagged sites in the mosquito genetic map [26]. Calculated chi square values for AT, GC and total base pair matches are given. One and two asterisks show values that reject a null hypothesis of no significant differences to random duplex formation at the $p<0.05$ and $p<0.005$ confidence levels.

3.2. AAD markers as characters for identification, ecology and evolution of plants

3.2.1. AAD analysis in floriculture

Cultivar identification and pedigree verification are important for the protection of intellectual property and royalty income, and for the development of "essentially derived varieties". Varietal identification has been particularly important for the floriculture industry. The production of bedding plants is one of the fastest growing segments in horticulture, with annual sales exceeding billions of dollars. Floricultural crops like petunia contribute importantly to this growth. However, cultivars are closely related and are difficult to differentiate at the genetic level. We have used AAD markers successfully for the characterization of cultivars of petunia [27], chrysanthemum [18, 28], geranium [29] and carnation [30], clarifying in some cases their origin or establishing genetic relationships. In particular, the ASAP technique permitted the clear identification of somatic mutants and radiation-induced sports that are genetically highly homogeneous [18, 29], facilitating future marker assisted breeding and protection of plant breeders rights of varieties or cultivars.

3.2.2. Analysis of rates of radiation-induced and somatic mutation in chrysanthemum

Chrysanthemum (*Dendranthema grandiflora*) was probably cultivated before 500 BC and is endemic to Asia. It was introduced to Europe in 1789 and to the US after the 19th century. These earlier plants bare little resemblance with currently cultivated cultivars, which are usually developed from a single progenitor either spontaneously or by radiation-induced mutagenesis (sports). Because of their close genetic relationship, there is a need to differentiate vegetatively derived accessions. The chrysanthemum cultivars Dark Charm, Salmon Charm, Coral Charm and Dark Bronze Charm are either radiation-induced mutants or spontaneous sports of cultivar Charm and constitute a family or series of plants that primarily differ in flower colour. These cultivars, which were difficult to differentiate genetically by DAF analysis [28], were easily identified by ASAP analysis. In one study [18], genomic DNA was first amplified with 3 standard octamer primers, all of which produced monomorphic profiles. Products from each of these DNA fingerprints were subsequently reamplified using 4 mini-hairpin decamer primers. The 12 primer combinations produced signatures containing about 37% polymorphic character loci, which were then used to estimate genetic relationships between the cultivars of the series. This number of ASAP polymorphisms detected provided an estimate of DNA changes in the mutant cultivars, ranging 0.03–4% of nucleotide changes within an average of 18 kb of arbitrarily amplified DAF sequence. Mutation levels were therefore comparable to those reported in *Drosophila*, *Transdescantia chiensis* and *Anopheles gambiae*. No differences in mutation rates were observed between the individual mutants (one-way ANOVA, $p > 0.963$), indicating that radiation-induced and somatic mutations occur at similar paces in chrysanthemum.

3.2.3. Analysis of off-types, cultivar instabilities and mutation rates in turfgrass

DNA analysis has been used profusely to characterize turfgrass species, cultivars and accessions (reviewed in [31]). We invested effort in the study of bermudagrass (*Cynodon*). The genus *Cynodon* comprises nine species and ten varieties that constitute a diverse group of important warm-season perennial sod-forming grasses, most of which have been used as turf, pasture and fodder throughout warm temperate and tropical regions of the world [32]. Only few clonally propagated lines bred during the later part of the century are currently being

widely cultured. These include sterile triploids ($2n = 3x = 27$) resulting from the interspecific hybridization of tetraploid *C. dactylon* var. *dactylon* and diploid *C. transvaalensis*. Such is the case of 'Tifway', 'Tifgreen' and 'Tifdwarf'. The narrow genetic base of these cultivars poses a risk for extensive damage from virulent or introduced pests [32], an apparent vulnerability that must be counteracted by assessing and broadening the genetic diversity of plant material used in breeding programmes. We recently used DAF to establish the levels of genetic variation within and between selected species and interspecific hybrids of bermudagrass [33]. DAF was also coupled with phenetic analysis to examine the origin of "off-type" derivatives of cultivar Tifway that exhibit variant morphology and performance [34]. Studies showed that Tifway was intrinsically stable and that off-types originated from contamination, were diverse, and were probably interspecific hybrids. In contrast, subsequent analysis of cultivars Tifgreen and Tifdwarf using mini-hairpin primers (ASAP and mhpDAF) indicated that off-types in these cultivars resulted predominantly from genetic instabilities arising from somatic mutation [35]. These instabilities did not derive from gross chromosomal rearrangements.

3.2.4. Genome-wide mutation rates in vegetatively propagated bermudagrass

An accumulation of point mutations within plants that are multiplying vegetatively at high rates by specialized means (stolons and rhizomes) could have important consequences on biological fitness and could be the cause of the high incidence of genetic off-types in these bermudagrass cultivars. Kondrashov [36] has developed a general model whereby mutation load in a population exhibiting obligate vegetative reproduction appears substantially higher than under sexual or asexual reproduction, explaining the rarity of this reproductive mode throughout evolution. In line with this proposal, we were able to show that genetic instabilities detected by phenetic analysis in bermudagrass sterile hybrids resulted from increased genome-wide mutation levels in vegetative culture [17]. Genome-wide somatic mutation rates were measured using the ASAP cascade amplification strategy. In ASAP [12], a single arbitrary primer is used to produce a collection of AAD products which then serve as template for a second round of DNA amplification that is generally directed by one or more mini-hairpin primers. This second ASAP amplification reaction can render allelic signatures that are characteristic of the sequence of the initial DAF products to be compared, provided initial AAD profiles are monomorphic. DNA polymorphisms detected by ASAP provided a measurement of sequence divergence within invariant DAF profiles in bermudagrass, chrysanthemum and *Discula destructiva* fungi. This also allowed a direct estimation of mutation rate at the whole genome level. The approach was used to study mutations that were induced by irradiation in Tifway II, those arising from genetic instabilities in the Tifgreen-Tifdwarf bermudagrass complex, and those appearing as the result of normal vegetative growth in Tifdwarf [17]. Mutations studied were essentially nucleotide substitutions and strand inversions that occurred within monomorphic DAF products. They were not the result of insertions or deletion events, the activity of transposable elements, major rearrangements in the genome, artifacts produced by the DNA amplification process itself, or the existence of non-orthologous bands in DAF products.

The method was first used to measure radiation-induced mutation in bermudagrass. DNA sequence divergence between cultivar 'Tifway' and its gamma radiation-induced mutant 'Tifway II' ($D = 0.70 \pm 0.66\%$) was comparable (though 2–3 times lower) to estimates in radiation-induced mutants and spontaneous sports of chrysanthemum [18]. In these experiments, sequence divergence was indicative of nucleotide changes induced by irradiation in the mutant, because ASAP using this same set of primers was unable to detect polymorphisms in the analysis of several Tifway accessions with different histories of culture.

A similar divergence in sequence (D = 0.95 ± 0.20%) was observed in the pairwise comparison of 17 non-disjunctive 'Tifgreen' and 'Tifdwarf" accessions. Despite the different culture histories, somatic mutations appear to have been arising in golf green bermudagrass to cumulative levels comparable to those induced by irradiation. A similar observation was reported for chrysanthemum sports [18]. Finally, mutation during normal Tifdwarf vegetative growth was evaluated by planting sprigs and sampling their offspring. Somatic sequence divergence levels (D = 0.004 ± 0.007%) resulted in a mutation rate of 1.05 × 10–^{8} per nucleotide per generation, assuming that a bermudagrass sprig constitutes a generation of growth. This rate is comparable to those found in human, *Drosophila melanogaster, Caenorhabditis elegans* and the mouse (ranging $\mu = 0.4 \times 10^{-8}$ -2 × 10^{-8}) (reviewed in [37]). This observation supports the contention that rates of sequence evolution in eukaryotes are cell-or-organism generation dependent rather than time-dependent.

The bermudagrass vegetative mutation rate was strikingly congruent with a long-term rate measured across accessions and indicative of the accumulation of mutations in Tifgreen-Tifdwarf populations (1.02 × 10^{-8} per nucleotide per generation), suggesting absence of evolutionary constraints in the sampled genomic regions. Therefore, most mutations detected by ASAP accumulate freely, appearing populationally neutral. Mutation rates calculated from across-accesions divergence estimates (D = 5.18 ± 0.53%) indicated that plant material was evolving 100 times faster (3.8 × 10^{-7} changes per nucleotide per a) than a molecular clock rate estimate for grasses, probably resulting from the compound effect of clonal growth and life span of the hybrid plant material.

Mutation rates have never been measured "directly" in plants before but were proposed to occur at levels of one mutation per diploid genome per generation [37]. The relatively high genomic mutation rate (μ_g = 10 per triploid genome) during bermudagrass vegetative growth results in effective μ_{eg} rates (U = 1) higher than calculated deleterious mutation rates using chlorophyll-deficient lethals (U = 0.003–0.074) but consistent with rates in self-fertilizing annual plants (U = 0.2–0.9). The high deleterious mutation rate compares well with recent estimates in hominids (U = 1.2–1.7) [38]. The high incidence of deleterious mutations, with rate estimates (U) comparable to those in plants subjected to inbreeding depression, casts doubt on the long-term success of the interspecific hybrids, if mutational effects on fitness were to combine in a multiplicative manner during clonal growth. Further research is therefore needed to evaluate the effect of mutation accumulation on vegetative culture.

3.2.5. Genetic diversity and breeding for anthracnose resistance in dogwood

There are about 65 species of dogwood (*Cornus* sp.) distributed primarily throughout the temperate regions of the northern hemisphere. Most species are either small trees or shrubs and have woody rhizomes and bracts. A number of species are commercially cultivated, predominantly for their ornamental characteristics, including the large-bracted Pacific (*C. nuttalli*), flowering (*C. florida*.), and Chinese (*C. kousa*) dogwood trees. By far, the most important North American species is the flowering dogwood, a relatively small understory tree native of the eastern deciduous forests. A number of varieties have been developed and commercialized, primarily based on differences in morphological traits such as floral bract shape, size and colour, variegated foliage, and growth habit. Over 100 named cultivars have been selected and maintained mainly through vegetative propagation by grafting or root cuttings. However, many cultivars are so phenotypically similar that identification has become very difficult.

The North American dogwoods have been recently compromised by a devastating anthracnose disease outbreak that has destroyed widely disseminated native stands and is caused by filamentous coelomycetous fungi of the genus *Discula* [39]. At the University of Tennessee, we initiated a genetic linkage mapping and marker-assisted selection (MAS) effort, initially targeted to control anthracnose and powdery mildew in dogwood. We first used DAF to determine the genetic relationships of representative cultivars and hybrids of flowering dogwood, evaluate the extent of plant hybridization, and generate markers in pseudo-testcross mapping of dogwood at the intraspecific level [16]. Most cultivars and hybrids could be uniquely identified. However, they were relatively conserved at the genetic level when compared with other plants [27, 28, 33]. Phenetic analysis identified cultivar hybridization in the F_1 progeny in the absence of phenotypic or physiological markers. The study showed a larger contribution of the female to the genetic make-up of each individual hybrid. Several cultivars grouped according to their recorded ancestry, such as the descendents of 'Cherokee Chief' ('Pink Sachet' and 'Cherokee Brave'), and others were highly similar ('Barton' and 'Cloud 9') as suggested in previous studies [40]. North Appalachian anthacnose-resistant lines grouped separately from those of South Appalachian origin, suggesting that genetic diversity may be influenced by a bio-geographical component. Finally, the DAF protocol was also tested in pseudo-testcross mapping of dogwood and showed it was very efficient in generating markers segregating at 1:1 ratios in the F1 progeny (3 per primer), despite the close relatedness and intra-specific nature of parental cultivars. Thousands of dogwood hybrids are now being evaluated.

3.3. AAD markers in population biology of exotic fungal pathogens

North American forest trees have been ravaged by a number of diseases with severe ecological and societal impact, many of them caused by exotic fungal invaders with greatly enhanced virulence. The recent dogwood anthracnose epidemic is one recent example [39]. *D. destructiva* and *Discula* sp. (with low incidence) cause the disease. The exotic invader hypothesis in *Discula* is supported by the severity and rapid onset and spread of the disease, the absence of records of disease symptoms prior to the 1970s, the lack of detection of conspecific fungi in herbarium specimens, the apparent lack of plant resistance, and the AAD-defined genetic homogeneity [41] and fine population structure [42] of the pathogen. Despite a proposal that the pathogen was introduced in infected *C. kousa* from eastern Asia, its origin remains unknown as well as the long-term effects of the disease. We characterized a cross-section of fungi that cause anthracnose in broadleaf temperate trees using a combination of sequence and secondary structure analysis of rDNA spacers, DAF, ASAP and compatibility of hyphal anastomosis. The study clarifies the taxonomy and phylogeny of *Discula*, raises the possibility of horizontal transfer of pathogenicity, and reveals an unusually rapidly evolving ITS sequence in an exotic and highly clonal fungal population.

Eukaryotic ribosomal genes are arranged in tandem repeats with the 5.8S coding region flanked by internal transcribed spacers (ITS) regions. ITS-inferred phylogenies rejected the null hypothesis of only one lineage by defining four monophyletic and well differentiated groups corresponding to *Discula* sp., *D. quercina*, *D. umbrinella* and *D. destructiva*, with the last two species sharing a common and recent ancestor. Results support a species concept based on lineages with monophyletic groups representing phylogenetically-derived species groups. For the most part, each group shared a same ecological niche defined by host-specificity and interspecific vegetative incompatibilities. We also showed the appearance of rare dogwood anthracnose-causing isolates that were phylogenetically related to the white oak pathogen *D. quercina* though distinct from *D. destructiva* by DAF analysis. This suggests that

pathogenicity determinants may be transmitting horizontally in *Discula*. A low incidence of dogwood anthracnose disease associated with *Discula* sp. could be similarly explained by lineage cross-talk mechanisms. One possibility is the transfer of genetic elements by cytoplasmically replicating double-stranded RNA viruses common in fungi, present in *D. destructiva* [43], and usually transmitted through hyphal anastomosis. This possibility is supported by the compatibility of *Discula* sp. and *D. quercina* in anastomosis experiments. The possibility of horizontal transfer of virulence determinants is of great concern for the future control of the dogwood disease in North America and warrants rigorous experimental challenge.

DAF analysis confirmed the high variability of *D. umbrinella* and *Discula* sp. and remarkable homogeneity of *D. destructiva* at the genetic level. Despite the almost clonal nature of *D. destructiva*, the ITS2 region of *D. destructiva* was found evolving at an unusually rapid pace (3.1×10^{-3} substitutions per nucleotide per year). These nucleotide substitution rates are 6 orders of magnitude higher than average levels in eukaryotic nuclear genomes and 5 times higher than a genome-wide mutation rate in *D. destructiva* (6.3×10^{-4} per nucleotide per a) measured using the ASAP technique. This feature was conspicuously absent in ITS1. The punctuated diversification phenomenon here described appears uniquely fast and highly selective. It follows that of populations subjected to a perturbed adaptive landscape under the stochastic effects of mutation, fluctuating selection and random genetic drift [44] and ultimately relates to the debated issue of whether evolutionary change proceeds as discontinuous transitions within the framework of punctuated equilibrium. Accumulating evidence supports the existence of rapid co-evolution of interspecific interactions that are ecologically important and occur during a time-span of less than 100 a [45]. The selective founder effect diversification in ITS2 sequence apparently involves the optimization of some aspect of protein synthesis, the major energy utilizing process in the cell, that is crucial to the fitness of the pathogen and documents rapid gene evolution during clonal dispersal and adaptation of *D.destructiva* to new hosts.

3.4. AAD markers in genetic mapping and positional cloning

3.4.1. Mapping of the supernodulation nts gene in soybean: a paradigm

The symbiosis between leguminous plants and rhizobia is a complex interplay between organisms that leads to the formation of a nitrogen-fixing organ, the root nodule [46]. Nodulated legumes, such as soybean, have the ability to fix atmospheric nitrogen for direct utilization of the plant. However, legumes seldomly fulfill their symbiotic potential because applied or residual nitrate in the soil inhibit the symbiosis. Research on the role of the plant on nodulation and nitrogen fixation necessitates a direct approach to gene isolation in order to understand the underlying developmental processes curving nodulation. As the result of a concerted genetic and physiological approach, a number of unique soybean mutants that nodulate in the presence of fixed nitrogen (*nts*, nitrate-tolerant symbiosis mutants) or are blocked in the early stages of nodule development and fail to nodulate at all (homologous to *rj1* and *rj6*) were isolated (reviewed in [46, 47]). Since no association of individual mutations to biochemical steps or cloned nodulin genes has yet been identified, positional cloning was the strategy of choice.

We searched for molecular markers tightly linked to symbiotic loci and constructed YAC and BAC libraries covering the roughly 1100 Mb of the soybean genome. A regional genetic map containing RFLP, SCARs and other markers was defined around the *nts-1* locus

in linkage group H of the Iowa soybean map (Fig. 2). Initially, several F_2 populations from crosses between *G. soja* PI468397 (carrying the wild-type *nts-1* allele) and supernodulating lines nts382 and nts1007 were analyzed for marker co-segregation with the mutant gene. One sequence-tagged marker (pUTG-132a) was found tightly linked to the *nts* locus (about 0.3 cM away), its sequence being highly conserved in other soybean varieties (Peking, Enrei, Minsoy, Noir 1 and DPS3589). Several RFLP markers of the bean map were syntenic in soybean and mapped close to the *nts-1* region. The *nts*-linked probes were also mapped on the Utah immortal soybean map of recombinant inbred lines (RIL). pPV-7 was mapped on linkage group U23 near pA-381 (8.4 cM distant). To enrich for markers in the region, we used AAD markers and BSA. DNA pools from segregants homozygous for pUTG-132a and flanking markers pA-381 and pA-36 were analyzed by scanning with 301 unstructured, 64 mini-hairpin decamer and 256 mini-hairpin undecamer primers. Several AAD polymorphisms were identified and converted into SCARs, and 4 bonafide markers were mapped. Furthermore, ASAP with several mini-hairpins and simple-sequence repeat primers revealed additional markers tightly linked to *nts-1*. Similarly, tecMAAP produced several markers that were polymorphic between wild-type Bragg and EMS-derived allelic mutants nts382 and nts1007 [11]. These markers will facilitate the construction of a high density map and genome walking in the region. Our overall results show that despite the extremely conserved nature of the *nts-1* region, BSA and AAD techniques can be successfully coupled in search for *nts*-associated amplification markers.

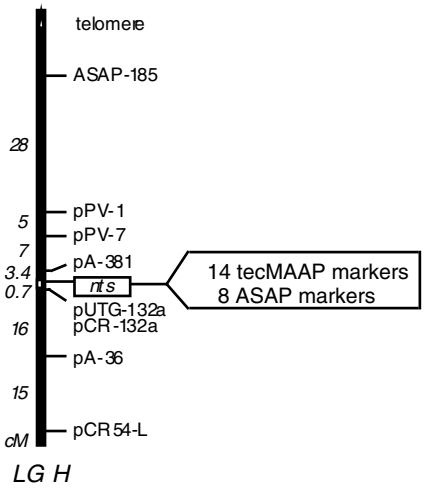

Fig. 2. The soybean supernodulation nts-1 region. The map was determined from 113 F_2 plants using MAPMAKER program. Inclusion of other mapping data from different crosses gave a nts-1 to pUTG-132a distance of about 0.3 cM. pPV-1 . pPV-7 are random genomic clones derived from Phaseolus vulgaris, that map to linkage group J of the Florida bean map(E. Vallejos, Univ. of Florida, Gainesville). Physical mapping of the pA-36 marker region showed an equivalence of 550 kb per cM, suggesting the overall region is recombnogenically average.

Soybean YAC and BAC libraries have been analyzed for sequences that are complementary to *nts*-linked marker regions. Two BAC pools of 350 clones contained sequences detected by primers for pUTG-132a, and several YAC clones hybridized to pUTG-132a. Three BACs were isolated (50, 75 and 165 Kb) and their endclones amplified and sequenced. One endclone had similarity to components of a retrotransposon, and another was common in two of the isolated BACs. Endclones are being converted into PCR-derived markers and will be used in the walking towards the *nts-1* gene.

3.4.2. Mapping of the non-nodulation rj1 and rj6 loci

The *rj1* and *rj6* non-nodulation traits segregate as single recessive Mendelian characters, and are necessary for nodule meristem development (*rj1*) and nodule initiation (*rj6*). A large group of mini-hairpin and standard primers were used to screen bulked DNA from F_2 segregating populations. In BSA of *rj1*, 10 862 DAF loci were generated by amplification, 0.2% of which were polymorphic between the bulks and 22 of them were potentially interesting. In BSA of *rj1*, 11 351 loci were generated, 0.24% of which were polymorphic and 27 were potentially interesting. However, when identified markers were studied for segregation in the F_2 population, most of them showed high recombination frequencies with the nodulation trait. As with *nts*, the symbiotic regions appear highly conserved.

In order to explain the lack linked markers in these BSA experiments, we studied the effect of product competition during the amplification of mixed samples of DNA. DNA from several soybean cultivars was mixed at various ratios with DNA from bermudagrass or from the ancestral soybean *G. soja*. The extent of dilution tolerated by each amplified band was then recorded. These competition experiments showed that not all amplification products were equally competitive. Rather, they exhibited different "amplification potentials". An interesting observation was that, in general, bermudagrass and *G. soja* amplification products were more competitive that soybean products. One conclusion drawn from these experiments is that genome size is not a relevant factor in competition. Instead, differences in competition resulted from difference in sequence variability expressed as genetic diversity within templates. In this regard, a number of studies have identified *G. soja* as harboring a panel of more diverse genetic loci than soybean, including microsatellites. These competition experiments also showed that each amplification product will have its own amplification threshold allowing expression of polymorphism according to its own amplification potential.

4. CONCLUSIONS

We have concentrated efforts in the study of marker systems that are powerful in their ability to detect genetic differences between closely related organisms. Markers such as those generated in ASAP and tecMAAP analyses were of great value in our efforts of plant breeding, variety identification and map-based cloning. Improved AAD methods produced numerous markers close to defined loci in the positional cloning of a soybean symbiotic gene. These same techniques were valuable in phylogenetic analysis, especially when coupled with PCR-based amplification of informative genomic regions such as rDNA, and in the identification of vegetatively propagated plant lines. They also permitted the evaluation of an important evolutionary parameter, mutation rate. AAD methods generated sequence tags for genetic and physical mapping applications. The selection of these tags was dependent on the context of their amplification and on sequences other that those targeted by the oligonucleotide primers. Finally, we used Taguchi methods very effectively to optimize DNA amplification and produce robust protocols of analysis without extensive effort. These optimization tools have great potential for the development of improved techniques in molecular biology.

ACKNOWLEDGMENTS

Research in horticultural crops, dogwood anthracnose and turfgrass involved R.N. Trigiano, M.K. Windham, S. Schlarbaum, M.C. Scott, A. Ament, T. Cerny, T. Starman, K. Weaver and L.M. Callahan (University of Tennessee, Knoxville, USA), and E. Carlsen and G. E. Caetano-Anollés (University of Oslo, Oslo, Norway). Positional cloning and mapping efforts involved J. Padilla, F. Ghassemi, D. Landau-Ellis, R. Funke, R. Prabhu, S. Abbitt, A. Kolchinsky and P.M. Gresshoff (Knoxville). Many others provided materials and technical help. Programmes were supported by the Tennessee Agricultural Experiment Station and the Racheff Endowment (Knoxville), the Biology Department (Oslo), and the International Atomic Energy Agency (Vienna).

REFERENCES

[1] KARP, A. *et al.* Molecular technologies for biodiversity evaluation: opportunities and challenges. Nature biotechnology 15 (1997) 625–628.

[2] CAETANO-ANOLLÉS, G., TRIGIANO, R.N. Nucleic acid markers in agricultural biotechnology. AgBiotech News & Info 9 (1997) 235N-242N.

[3] LIVAK, K.J. *et al.* Process for detecting polymorphisms on the basis of nucleotide differences. US Patent 5,126,239 (1992).

[4] BASSAM, B.J., CAETANO-ANOLLÉS, G., GRESSHOFF, P.M. Method for profiling nucleic acids of unknown sequence using arbitrary oligonucleotide primers. US Patent 5,413,909 (1995).

[5] McCLELLAND, M., WELSH, J.T., SORGE, J.A. Arbitrarily primed polymerase chain reaction method for fingerprinting genomes. US Patent 5,467,985 (1996).

[6] SCHIERWATER, B. *et al.* "Arbitrarily amplified DNA in ecology and evolution", DNA Markers: Protocols, Applications and Overviews (CAETANO-ANOLLÉS G, GRESSHOFF, P.M., Eds.), John Wiley & Sons, New York (1997) 313–330.

[7] CAETANO-ANOLLÉS, G. "Arbitrary oligonucleotides: primers for amplification and direct identification of nucleic acids, genes and organisms", Molecular Approaches to Ecology and Evolution (DeSALLE, R., SCHIERWATER, B., Eds.), Birkhäuser Verlag, Basel (1998) 107–123.

[8] CAETANO-ANOLLÉS, G., BASSAM, B.J., GRESSHOFF, P.M. Primer-template interactions during *in vitro* amplification with short oligonucleotides. Mol. Gen. Genet. 235 (1992) 157–165.

[9] CAETANO-ANOLLÉS, G., BASSAM, B.J., GRESSHOFF, P.M. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Bio/technology 9 (1991) 553–557.

[10] CAETANO-ANOLLÉS, G., GRESSHOFF, P.M. DNA amplification fingerprinting using arbitrary mini-hairpin oligonucleotide primers. Bio/technology 12 (1994) 1011–1026.

[11] CAETANO-ANOLLÉS, G., BASSAM, B.J., GRESSHOFF, P.M. Enhanced detection of polymorphic DNA by multiple arbitrary amplicon profiling of endonuclease digested DNA: identification of markers linked to the supernodulation locus of soybean. Mol. Gen. Genet. 241 (1993) 57–64.

[12] CAETANO-ANOLLÉS, G., GRESSHOFF, P.M. Generation of sequence signatures from DNA amplification fingerprints with mini-hairpin and microsatellite primers. Biotechniques 20 (1996) 1044–1056.

[13] TAGUCHI, G. Introduction to quality engineering. Asian productivity organization. UNIPUB, New York, New York (1986).
[14] COBB, B.D., CLARKSON, J.M. A simple procedure for optimizing the polymerase chain reaction (PCR) using modified Taguchi methods. Nucleic Acids Res. 22 (1994) 3802–3805.
[15] CAETANO-ANOLLÉS, G. DAF optimization using Taguchi methods and the effect of thermal cycling parameters on DNA amplification. Biotechniques 25 (1998) 472–480.
[16] CAETANO-ANOLLÉS, G., SCHLARBAUM, S.E., TRIGIANO, R.N. DNA amplification fingerprinting and marker screening for pseudo-testcross mapping of flowering dogwood (*Cornus florida* L.). Euphytica 106 (1999) 209–222.
[17] CAETANO-ANOLLÉS, G. High genome-wide mutation rates in vegetatively propagated bermudagrass. Mol. Ecol. 8 (1999) 1211–1221.
[18] TRIGIANO, R.N., SCOTT, M.C., CAETANO-ANOLLÉS, G. Genetic signatures from amplification profiles characterize DNA mutation in somatic and radiation-induced sports of chrysanthemum. J. Amer. Soc. Hort. Sci. 123 (1998) 642–646.
[19] MICHELMORE, R.W., PARAN, I., KESSELI, R.V. Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. Proc. Natl. Acad. Sci. USA 88 (1991) 9828–9832.
[20] PARAN, I., MICHELMORE, R.W. Development of reliable PCR markers linked to downy mildew resistance genes in lettuce. Theor. Appl. Genet. 85 (1993) 985–993.
[21] YEH, L-.C.C., LEE, J.C. Structural analysis of the internal transcribed spacer 2 of the precursor ribosomal RNA from *Saccharomyces cereviceae*. J. Mol. Biol. 211 (1990) 699–712.
[22] DNA KAFFE. A forum for DNA marker methodologies. URL http: //biologi.uio.no/FellesAvdelinger/DNA_KAFFE/default.html (1998).
[23] PARKER, J.D., RABONOVICH, P.S., BURMER, G. (1991) Targeted gene walking polymerase chain reaction. Nucleic Acids Res. 19 (1991) 3055–3060.
[24] VENUGOPAL, G. *et al.* Multiple mismatch annealing: basis for random amplified polymorphic DNA fingerprinting. Biochem. Biophys. Res. Commun. 197 (1993) 1382–1387.
[25] BERTIOLI, D.J. *et al.* An analysis of differential display shows a strong bias towards high copy number mRNAs. Nucleic Acids Res. 23 (1994) 4520–4523.
[26] DIMOPOULOS, G. *et al.* Integrated map of *Anopheles gambiae*: use of RAPD polymorphisms for genetic, cytogenetic and STS landmarks. Genetics 143 (1996) 953–960.
[27] CERNY, T.A. *et al.* Molecular phylogeny and DNA amplification fingerprinting of *Petunia* taxa. Theor. Appl. Genet. 92 (1996) 1009–1016.
[28] SCOTT, M.C., CAETANO-ANOLLÉS, G., TRIGIANO, R.N. DNA amplification fingerprinting identifies closely related Chrysanthemum cultivars. J. Amer. Soc. Hort. Sci. 121 (1996) 1043–1048.
[29] STARMAN, T.W., ABBITT, S. Evaluating genetic relationships of geranium using arbitrary signatures from amplification profiles. HortScience 32 (1997) 1288–1291.
[30] TRIGIANO, R.N., CAETANO-ANOLLÉS, G. Laboratory exercises on DNA amplification fingerprinting for evaluating the molecular diversity of horticultural species. HortTechnology 8 (1998) 413–423.
[31] CAETANO-ANOLLÉS, G. DNA analysis of turfgrass genetic diversity. Crop Sci. 38 (1998) 1415–1424.
[32] TALIAFERRO, C.M. Diversity and vulnerability of bermuda turfgrass species. Crop Sci. 35 (1995) 327–332.

[33] CAETANO-ANOLLÉS, G. *et al.* DNA amplification fingerprinting analysis of bermudagrass (*Cynodon*): genetic relationships between species and interspecific crosses. Theor. Appl. Genet. 91 (1995) 228–235.
[34] CAETANO-ANOLLÉS, G., CALLAHAN, L.M., GRESSHOFF, P.M. The origin of bermudagrass (*Cynodon*) off-types inferred by DNA amplification fingerprinting. Crop Sci. 37 (1997) 81–87.
[35] CAETANO-ANOLLÉS, G. Genetic instability of bermudagrass (*Cynodon*) cultivars 'Tifgreen' and 'Tifdwarf' detected by DAF and ASAP analysis of accessions and off-types. Euphytica 101 (1998) 165–173.
[36] KONDRASHOV, A.S. Mutation load under vegetative reproduction and cytoplasmic inheritance. Genetics 137 (1994) 311–318.
[37] KONDRASHOV, A.S. Measuring spontaneous deleterious mutation. *Genetica* 102/103 (1998) 183–197.
[38] EYRE-WALKER, A., KEIGHTLEY, P.D. High genomic deleterious mutation rates in hominids. Nature 397 (1999) 344–347.
[39] DAUGHTREY, M.L. *et al.* Dogwood anthracnose. Understanding a disease new to North America. Plant Dis. 80 (1996) 349–358.
[40] TRIGIANO, R.N., WINDHAM, M.T. Are Barton and Cloud 9 the same cultivar of *Cornus florida*? H. Environ. Hort. 16 (1998) 163–166.
[41] TRIGIANO, R.N. *et al.* DNA amplification fingerprinting provides evidence that *Discula destructiva*, the cause of dogwood anthracnose in North America, is an introduced pathogen. Mycologia 83 (1995) 633–642.
[42] CAETANO-ANOLLÉS, G., TRIGIANO, R.N., WINDHAM, M.T. Sequence signatures from DNA amplification fingerprints reveal fine population structure of the dogwood pathogen *Discula destructiva*. FEMS Microbiol. Lett. 145 (1996) 377–383.
[43] McELREATH, S.D. *et al.* Double-stranded RNA in isolates of *Discula destructiva* from the eastern United States. Curr. Microbiol. 29 (1994) 57–60.
[44] ELENA, S.F., COOPER, V.S., LENSKI, R.E. Punctuated evolution caused by selection of rare beneficial mutations. Science 272 (1996) 1802–1804.
[45] THOMPSON J.N. Rapid evolution as an ecological process. Trends Ecol. Evol. 13 (1998) 329–332.
[46] CAETANO-ANOLLÉS, G. Molecular dissection and improvement of the nodule symbiosis in legumes. Field Crop Res. 53 (1997) 47–68.
[47] CAETANO-ANOLLÉS, G., GRESSHOFF, P.M. Advances in the positional cloning of nodulation genes in soybean. Plant & Soil 186 (1996) 1–7.

UPDATE ON THE USE OF RANDOM 10-mers IN MAPPING AND FINGERPRINTING GENOMES

R.M. SINIBALDI
Operon Technologies,
Alameda, California, United States of America

Abstract

The use of Randomly Ampliflied Polymorphic DNA (RAPDs) has continued to grow for the last several years. A quick assessment of their use can be estimated by searching PubMed at the National Library of Medicine with the acronym RAPD. Since their first report in 1990, the number of citations with RAPD in them has increased from 12 in 1990, to 45 in 1991, to, 112 in 1993, to, 130 in 1994, to 223 in 1995, to 258 in 1996, to 236 in 1997, to 316 in 1998, to 196 to date (August 31) 1999. The utilization of 10-mers for mapping or fingerprinting has many advantages. These include a relatively low cost, no use of radioactivity, easily adapted to automation, requirement for very small amounts of input DNA, rapid results, existing data bases for many organisms, and low cost equipment requirements. In conjunction with a derived technology such as SCARs (sequence characterized amplified regions), it can provide cost effective and thorough methods for mapping and fingerprinting any genome. Newer methods based on microarray technology may offer powerful but expensive alternative approaches in determining genetic diversity. The costs of arrays should come down with time and improved production methods. In the meantime, RAPDs remain a competent and cost effective method for genome characterizations.

1. INTRODUCTION

1.1. Randomly Amplified Polymorphic DNA (RAPD)

The use of Randomly Amplified Polymorphic DNA (RAPDs) has remained relatively steady for the last several years. A quick assessment of their popularity can be measured by searching PubMed at the National Library of Medicine with the acronym RAPD. Since their first report in 1990, the number of citations with RAPD in them has gone from 12 in 1990, to 45 in1991, to, 112 in 1993, to, 130 in 1994, to 223 in 1995, to 258 in 1996, to 236 in 1997, to 316 in 1998, to 196 to date (August 31, 1999). Operon RAPD 10-mer Kits contain 10-base oligonucleotide primers for use in genetic mapping [1] and DNA fingerprinting [2]. Operon Technologies presently has 1,200 different 10-base primers in stock. The primers have been used to generate polymorphic markers in many species including humans, mice, lettuce and various bacteria. These primers are packaged in kits of 20 sequences each and are designated “Kit A” through “Kit Z,” “Kit AA” through “Kit AZ,” and “Kit BA” through “Kit BH.” The sequences were selected randomly, with the requirement that their (G+C) content is 60% to 70%, and that they have no self-complementary ends.

1.2. Principle of technique

A single 10-base oligonucleotide primer is used to amplify genomic DNA. A DNA amplification product is generated for each genomic region that happens to be flanked by a pair of 10-base priming sites (in the appropriate orientation), which are within 5000 base pairs of each other. Amplification products are analyzed by electrophoresis. Genomic DNA from two different individuals often produce different amplification fragment patterns. A particular DNA fragment which is generated for one individual but not for another represents a

DNA polymorphism and can be used as a genetic marker. These markers are inherited in a Mendelian fashion (1). In mapping studies, the segregation of these markers among the progeny of a sexual cross can be used to construct a genetic map. In fingerprinting studies, the banding patterns are compared directly to allow strain determination, usually without the need to correlate band differences with particular properties.

2. METHODS

Each Operon 10-mer sample tube contains enough primer for approximately 1000 amplification reactions. Each 10-mer sample tube should be resuspended in sterile water. For long term stability, we recommend that you subdivide each 10-mer sample into several aliquots, dry each aliquot, and store at -20°C. Use clean disposable plasticware for all transfers. The actual methods have not changed significantly in the last several years.

For DNA amplification, the following conditions are those originally recommended by Williams *et al*. [1]: Amplification reactions are performed in a volume of 25 µL containing 10 mM Tris-Cl, pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 0.001% gelatin, 100 µM each of dATP, dCTP, dGTP, and dTTP, 5 picomoles [resuspend in 1 ml of water, and use 1 µL per reaction] of a single 10-base primer, 25 ng of genomic DNA, and 0.5 units of Taq DNA polymerase (Perkin Elmer Cetus). Amplification should be performed on a top quality thermal cycler programmed for 45 cycles of 1 minute at 94°, 1 minute at 36°, and 2 minutes at 72°. Amplification products are analyzed by electrophoresis in 1.4% agarose gels and detected by staining with ethidium bromide. Other published methods recommend an annealing temperature between 33° and 35°.

Nadeau *et al.* [9] have recommended slightly different conditions than those of the original procedure. They recommend the use of 25 picomoles of primer [resuspend in 200 µl of water and use 1 µl per reaction] and 100 ng template (a 5-fold and 4-fold increase over Williams' recommendations [1], respectively). These conditions may improve the reproducibility of the results of this method.

3. RESULTS AND DISCUSSION

3.1. Number of amplification products

The number of different amplification products for each reaction depends upon the primer sequence, the genomic sequence, and the genome size. Assuming that the priming sites are randomly distributed throughout a genome, probability theory predicts that the number of amplification products will be approximately $2.5 \times 10^{-9} \times G$, where G is the size of the haploid genome in base pairs. For example, lettuce has a haploid genome size of approximately 2×10^9 base pairs, so the above formula predicts that a typical reaction should yield between 5 and 10 bands, depending on the extent of genetic heterozygosity. This prediction is in close agreement with actual results in lettuce obtained by Michelmore *et al*. [5]. However, for much smaller genome sizes, such as that of *E. coli* ($G = 4 \times 10^6$ base pairs), the above formula predicts that most primers should generate no bands at all. Nevertheless, several laboratories have reported multiple amplification products from prokaryotic DNA. Such results can only be explained on the basis of mismatch between the primer and the DNA template [1].

The polymorphisms are caused by one to a few nucleotide changes in one of the priming sites, insertions or deletion of DNA between priming sites, or rearrangements of the DNA encompassing the amplified region. The RAPD technique usually calls for the use of one 10-mer for each reaction, and this works well for most eucaryotic genomes. Some investigators have been successful in utilizing two different 10-mers for the generation of distinctive markers, and this is particularly true in organisms with small genomes (i.e. bacteria). However, multiplexing with many 10-mer primers has not been very fruitful in generating great numbers of polymorphic bands per reaction with most genomic DNAs.

3.2. Naming of primers and markers

In order to prevent ambiguity in the naming of primers from different sources, Operon attaches the prefix "OP" to the names of all of its primers. For example, the fourth sequence in Kit H is labeled "OPH-04." To refer to a specific polymorphic amplification product, we recommend the convention used by Paran *et al*. [3], which is to add a subscript denoting its size to the primer name. For example, an 800 bp amplification product produced by primer OPAC-01 would be called "OPAC-01_{800}."

3.3. Troubleshooting

Not all amplifications products arise from perfect pairing between primer and DNA template. Amplification products arising from mis-pairing may still be reproducible and may be useful genetic markers. However, we suspect that these mismatched markers are more sensitive to slight changes in the temperature cycle, so we strongly suggest using identical amplification conditions when comparing results.

The DNA amplification method described above is unusual in that it uses very short (10-base) primers, which have less specificity than longer primers. As a result, this method is quite sensitive to small variations in the temperature cycle, particularly the annealing temperature. Since the actual temperatures delivered to the tubes by different thermal cyclers may differ significantly, it is often necessary to refine the recommended temperature program in order to optimize this method to your particular thermal cycler. If no amplification products are seen, it may be necessary to adjust temperatures downward. If too many products are seen, it may be necessary to adjust temperatures upward.

Occasionally, a "smear" of amplification products is observed, and this may be converted to discretely sized bands by adjusting the concentration of the polymerase, the primer, or the genomic DNA.

It is important that your genomic DNA is relatively free of single strand breaks since such breaks will prevent amplification. Treat your genomic DNA very gently to prevent shearing. In particular, avoid repeated boiling or freeze-thawing of your genomic DNA samples.

Although we have chosen oligonucleotide sequences, which do not have self-complementary ends, some of our 10-mers are still capable of forming complicated secondary structures, which can lead to the production of amplification artifacts. These artifacts tend to appear only if one intentionally omits the genomic DNA to test a "reagent blank." These artifacts have now been directly observed [10]. Fortunately, most artifacts are not produced when genomic DNA is included.

3.4. Uses in genetic mapping

Of those kits already tested, our customers report getting useful genetic markers from about 50% to 98% of our 10-mer sequences, depending on the species. A variety of mapping and fingerprinting strategies, which employ this technique have appeared in the scientific literature and are listed below [1–10]. This technology has been reviewed by Rafalski *et al*. [6]. The 10-mers have also been useful in generating information that can be utilised for other mapping technologies. For example, diagnostic 10-mer amplification products can be deciphered by DNA sequencing and this sequence information can be used for designing longer sequence characterised primers that can be used in additional PCR reactions to generate sequence characterised amplified regions (SCARs).

Other commonly used mapping procedures include restriction fragment length polymorphism (RFLP), simple sequence repeats (SSR) and amplified fragment length polymorphism (AFLP). RFLP is limited to yielding very few data points per reaction and the original procedure involved the use of radiolabeled probes. SSR requires prior sequence knowledge and yields a small number of data (3–10) per reaction. AFLP does not require prior sequence knowledge and generates a higher number of data points per reaction. However, the cost of an AFLP reaction is estimated to be $ 7.50 and the equipment needed to effectively use this technology is in excess of $ 80 000.00. RAPD technology generates 4 to 15 data points per reaction, requires no prior sequence knowledge; but is limited to detection of dominant markers. However, the cost of a RAPD reaction is substantially lower ($ 1.50) than an AFLP reaction and does not require a large investment in capital equipment.

Other genetic methods that have been recently developed based on microarray technology raise the hope of relatively cost effective methods for detecting single nucleotide polymorphisms (SNPs), genotyping and strain identification. For a review of the microarray technology see the supplement to Nature Genetics [11]. Identification of novel DNA variants can be performed by printing high-density arrays composed of oligonucleotides. One can construct a ‘tiling’ array to scan for a target sequence for each mutation by using overlapping sequences. Each overlapping 25-mer in the sequence is covered by four complementary oligonucleotide sequences that differ only by having A, C, G or T substituted in the central position. An amplified product containing the expected sequence will hybridize best to the expected target, whereas a sequence variation will alter the hybridization pattern. Such tilling arrays have been used to detect variants in specific genes like p53 as well as viral geneomes such as HIV. The approach has also been used for larger segments of DNA. For example more than 2 Mb of genomic DNA was scanned for SNPs by using a set of more 100 tilling arrays [12]. This offers powerful but not perfect methods to scan large segments of DNA for polymorphisms. High throughput methods using these tilling arrays will detect homozygous variants but may miss heterozygotes, so are by no means foolproof in detecting variants. Other array designs can be used for performing genotyping analysis. Microarrays are powerful but currently very expensive, however, mass production of arrays can substantially lower their costs.

4. SUMMARY

The use of 10-mers for mapping or fingerprinting has many advantages. These include a relatively low cost, no use of radioactivity, easily adapted to automation, requirement for very small amounts of input DNA, rapid results, existing data bases for many organisms, and low cost equipment requirements. The RAPD technique is not an end-all solution for mapping and

should be thought as one of the tools than can be used. In conjunction with a derived technology such as SCARS, it can provide cost effective and thorough methods for mapping and fingerprinting any genome. Newer methods based on microarray technology may offer powerful but expensive alternative approaches in determining genetic diversity. The costs of arrays should come down with time and improved production methods. In the meantime, RAPDs remain a compotent and cost effective method for genome characterizations.

REFERENCES

[1] WILLIAMS, J.G.K. *et al.* DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. 18 (1990) 6531–6535.

[2] WELSH, J., McCLELLAND, M. Fingerprinting genomes using PCR with arbitrary primers. Nucl. Acids Res. 19 (1990) 303–306.

[3] PARAN, I., KESSELI, R., MICHELMORE, R. Identification of RFLP and RAPD markers linked to downy mildew resistance genes in lettuce using near-isogenic lines. Genome 34 (1991) 1021–1027.

[4] MARTIN, G.B., WILLIAMS, J.G.K., TANKSLEY, S.D. Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. Proc. Natl. Acad. Sci. USA 88 (1991) 2336–2340.

[5] MICHELMORE, R.W., PARAN, I., KESSELI, R.V. Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. Proc. Natl. Acad. Sci. USA 88 (1991) 9828–9832.

[6] RAFALSKI, J.A., TINGEY, S.V., WILLIAMS, J.G.K. RAPD markers–a new technology for genetic mapping and plant breeding. AgBiotech News and Information 3 (1991) 645–648.

[7] REITER, R.S. *et al.* Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc. Natl. Acad. Sci. USA 89 (1992) 1477–1481.

[8] WOODWARD, S.R., SUDWEEKS, J., TEUSCHER, C. Random sequence oligonucleotide primers detect polymorphic DNA products which segregate in inbred strains of mice. Mammalian Genome 3 (1992) 73–78.

[9] NADEAU, J.H. *et al.* Multilocus markers for mouse genome analysis: PCR amplification based on single primers of arbitrary nucleotide sequence. Mammalian Genome 3 (1992) 55–64.

[10] HIGUCHI, R., FOCKLER, C., DOLLINGER, G., WATSON, R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Bio/Technology 11 (1993) 1026–1030.

[11] THE CHIPPING FORECAST. Nature Genetics 21 (1999) supplement.

[12] WANG, D.G. *et al.* Large-scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. Science 280 (1998) 1077–1082.

DEVELOPMENT AND APPLICATION OF SEQUENCE-TAGGED MICROSATELLITE SITE (STMS) MARKERS IN CHICKPEA (*CICER ARIETINUM*), BANANA (*MUSA* SPP.) AND THEIR MAJOR PATHOGENS, *ASCOCHYTA RABIEI* AND *MYCOSPHAERELLA FIJIENSIS*

P. WINTER, D. KAEMMER, T. PAFF, J. GEISTLINGER, C. NEU, G. KAHL
Plant Molecular Biology, Biocentre, Johann Wolfgang Goethe-University,
Frankfurt am Main, Germany

Abstract

DNA markers of various kinds have found widespread application in many facets of plant breeding and plant pathogen control. Yet another marker type, sequence-tagged microsatellite (STMS) markers, provides the markers of choice for nearly every crop because of their co-dominant nature, reliability, ease of application and high polymorphic information content. We report here on the development of a whole set of STMS markers and the respective, selected primer sequences for two important crops, chickpea (*Cicer arietinum* L.) and banana (*Musa acuminata*), and for their most devastating fungal pathogens, *Ascochyta rabiei* and *Mycosphaerella fijiensis*, respectively. These markers were generated either by direct screening of size-selected genomic libraries with microsatellite-complementary oligonucleotides, or by enrichment of DNA fragments containing microsatellite sequences. A total of 69 markers for chickpea, 15 markers for *M. acuminata*, 19 markers for *A rabiei* and 11 markers for *M. fijiensis*, selected on the basis of their high information content and ease of use are presented here. These can be applied for mapping of the respective genomes, for various population studies, and cultivar and isolate identification. We further demonstrate that several of these markers can potentially be applied across species boundaries and thus could increase the marker repertoire also for other species of the genus *Cicer*, *Musa* and for *Ascochyta*-type pathogens of bean, and potentially also of lentil and pea.

1. INTRODUCTION

In recent years, plant breeding has profited enormously from the advent of molecular marker technologies. Molecular markers, and especially those based on selectively neutral DNA polymorphisms, facilitate the reliable identification of breeding lines, cultivars, clones and hybrids, allow the monitoring of introgression of alien DNA into cultivated germplasm, and the estimation of genetic diversity. Moreover, advanced high-density DNA marker maps are now available for most important crops, providing a basis for marker-assisted selection of useful traits, pyramiding of resistance genes, and the isolation of these and other agronomically important genes via map-based cloning (reviews in [1, 2]).

Set their importance for genome analysis of plants aside, molecular markers also catalysed the characterisation of plant pathogens. The unequivocal identification of pathogenic fungi, for example, which in former times required a skilled pathologist, can now be performed routinely. Molecular markers allow to rapidly and reliably identify fungal races and pathotypes, and to monitor their population structure [3, 4].

The requirement for more and more easily applicable markers has led to the development of a plethora of different marker techniques (review in [2]). A particular type of markers, i.e. those based on microsatellite variability has gained popularity in the past years, thanks to the many advantages it provides for research and applications. A special type of microsatellite markers are the subject of this article.

The term "microsatellite" [5], also *s*imple *s*equence *r*epeat (SSR), or *s*hort *t*andem *r*epeat (STR) characterises short, reiterated sequence motifs of about 1 to 6 nucleotides that, like classical satellite DNA, are organised in more or less perfect tandem arrays of a few up to even thousands of repeat units. The key feature of SSR repeats that makes them the preferred target for marker development is a strong tendency to change their overall length by slipped-strand mispairing and other less well understood processes (reviews in [6, 7]) leading to *v*ariable *n*umbers of *t*andem *r*epeats (VNTR) [8] and resulting in *s*imple *s*equence *l*ength *p*olymorphisms (SSLPs) [9].

Mutation rates of SSRs are generally high. Consequently, up to 30 alleles were reported for particular loci in plants [10, 11]. Also in fungi, hypervariable SSRs have been observed [12]. However, microsatellite mutation rates vary considerably among loci and organisms. STRs are abundant and usually more or less evenly dispersed throughout eucaryotic genomes but appear to be less frequent in plants as compared to vertebrates [13]. Reported estimates of microsatellite frequencies vary considerably, with average inter-SSR distances ranging from 10 kb to more than 1 Mb, depending on the motif and the organism. For example, in the chickpea genome $[TAA]_n$, $[GA]_n$ and $[CA]_n$ arrays are present at more than 12 000 loci with an average spacing of around 60 kb [14].

In contrast to plants, studies of microsatellite sequences in fungi are scarce [12, 15–18], as are reports on the systematic development and application of markers from microsatellite sequences of these lower eucaryotes [19–22].

1.1. Amplification of single microsatellite loci: STMS markers

Currently the most popular method to exploit SSR variability for the generation of genetic markers uses primers targeted to SSR-flanking sequences to amplify the enclosed SSR. The results are locus-specific amplification products that often exhibit considerable length differences among different individuals or populations of the same species, mostly due to the variable number of tandem repeats within the SSR. These sequence-tagged microsatellite site (STMS) [23] markers are the markers of choice for nearly every organism. They represent single-locus, co-dominant, easy-to-use and reliable markers with high polymorphic information content possessing the potential for automated, non-radioactive detection. Moreover, besides the availability of standard instrumentation (thermocycler and gel apparatus) the knowledge of STMS primer sequences is the only prerequisite for their application. Since these can easily be exchanged between laboratories, STMS markers can be applied by any interested lab without time-consuming preparation and mailing of probes [24, 25].

In spite of the many advantages of STMS, they also have some limitations, which have precluded their widespread application for a while. Main obstacles for an extensive use are the high costs for cloning, sequencing and primer synthesis. Further, standard protocols still use radioisotopes and sequencing gels to detect the amplified SSRs. Finally, the efficiency of primer generation suffers from a number of problems including redundancy of clones and the occurrence of artificial chimeras (recent reviews in [24, 25]).

However, in recent years, the advent of microsatellite enrichment techniques (reviews in [24, 25]) and dropping costs for DNA sequencing and primer synthesis reduced the necessary investment for the generation of large numbers of STMS markers opening an avenue also for application in more orphan species, such as the tropical tree *Simarouba* [26].

Here, we describe the development and sequences of STMS primer pairs for chickpea and banana as well as their major fungal pathogens *Ascochyta rabiei* and *Mycospaerella fijiensis*, respectively.

2. DEVELOPMENT OF STMS MARKERS FOR CHICKPEA, BANANA, *A. rabiei* AND *M. fijiensis*

2.1. STMS markers for chickpea (*Cicer arietinum* L.)

2.1.1. Isolation of microsatellite-containing clones

The STMS primer pairs described here were derived from a size-selected genomic library of 280 000 colonies that represented ~18% of the chickpea genome and was screened for $(GA)_n$, $(GAA)_n$ and $(TAA)_n$ micosatellite-containing clones. The plasmids of 389 positive colonies were sequenced. The majority (~75%) contained perfect repeats. Interrupted, interrupted-compound and compound repeats were only present from 6 to 9%. Microsatellites of the (TAA)-type contained the longest repeats with unit numbers ranging from 9 to 131. For 218 loci, primers could be designed and used for the detection of microsatellite length polymorphisms in 6 chickpea cultivars, *C. echinospermum and C. reticulatum* (chickpea`s closest relatives). Altogether, 174 primer pairs gave interpretable banding patterns, 137 (79%) of which revealed at least 2 alleles on native polyacrylamide gels. Genetic mapping of 120 of these STMS in a population of recombinant inbred lines from an inter-species cross between *C. reticulatum* and the cultivated chickpea line ICC 4958 is described in [27]. A more comprehensive map, spanning around 2000 cM is reported in a forthcoming paper [28]. The table given below contains only those STMS markers that revealed at least 4 alleles. More STMS primer pairs for chickpea can be found in [14, 27].

2.1.2. Detection of microsatellite polymorphisms in chickpea: PCR conditions and gel electrophoresis

Primers were 22 to 32 nucleotides long allowing annealing temperatures of 55°C. Thirty-five cycles of PCR were performed on 50 ng DNA in 20 µl "Silverstar" reaction buffer containing 2 µM primers, 1.5 mM $MgCl_2$, 250 mM nucleotides and 0.4 U "Silverstar" polymerase (Eurogentec, Belgium). The DNA was first denatured for 2 min at 94°C, annealing was at 55°C for 50 sec, and elongation at 60 °C (TAA-repeats) or 72°C (GA and GAA-repeats) for 50 sec. The reduced elongation temperature for TAA-repeats (as compared to the optimal 72°C) was necessary to stabilise the long (TAA)-microsatellites. Between 4 to 12 µl of the reaction mix were separated either on 2% agarose gels for a first analysis of amplification success, or on 8% native polyacrylamide gels for determination of allele length and numbers.

For those familiar with Spanish language, detailed protocols for PCR with STMS primer pairs and gel electrophoresis of amplification products can be found in [29].

TABLE 1. HIGHLY POLYMORPHIC STMS MARKERS FOR CHICKPEA

Locus	Repeat Type	Primer Sequence (5`-3`)	Expected Product (bp)
TA1	P	L TGAAATATGGAATGATTACTGAGTGAC	243
		R TATTGAAATAGGTCAGGCTTATAAAAA	
TA2	I	L AAATGGAAGAAGAATAAAAACGAAAC	175
		R TTCCATTCTTTATTATCCATATCACTACA	
TA5	P	L ATCATTTCAATTTCCTCAACTATGAAT	205
		R TCGTTAACACGTAATTTCAAGTAAAGAT	
TA8	P	L AAAATTTGCACCCACAAAATATG	246
		R CTGAAAATTATGGCAGGGAAAC	
TA11	P	L CATGCCATAAACTCAATACAATACAAC	230
		R TTCATTGAGGACAATGTGTAATTTAAG	
TA13	P	L TAAGTTAAGGGACCAACGAA	243
		R CAAGTTGGAGTCAAACCAAT	
TA14	I	L TGACTTGCTATTTAGGGAACA	250
		RTGGCTAAAGACAATTAAAGTT	
TA18	P	L AAAATAATCTCCACTTCACAAATTTTC	147
		R ATAAGTGCGTTATTAGTTTGGTCTTGT	
TA20	I	L ATTTTCTTTATCCGCTGCAAAT	299
		R TTAAATACTGCCTTCGATCCGT	
TA21	P	L GTACCTCGAAGATGTAGCCGATA	347
		R TTTTCCATTTAGAGTAGGATCTTCTTG	
TA25	P	L AGTTTAATTGGCTGGTTCTAAGATAAC	247
		R AGGATGATCTTTAATAAATCAGAATGA	
TA27	P	L GATAAAATCATTATTGGGTGTCCTTT	241
		R TTCAAATAATCTTTCATCAGTCAAATG	
TA28	I	L TAATTGATCATACTCTCACTATCTGCC	300
		R TGGGAATGAATATATTTTTGAAGTAAA	
TA34	P	L AAGAGTTGTTCCCTTTCTTTT	230
		R CCATTATCATTCTTGTTTTCAA	
TA37	P	L ACTTACATGAATTATCTTTCTTGGTCC	282
		R CGTATTCAAATAATCTTTCATCAGTCA	
TA39	P	L TTAGCGTGGCTAACTTTATTTGC	249
		R ATAAATATCCAATTCTGGTAGTTGACG	
TA42	C	L ATATCGAAATAAATAACAACAGGATGG	209
		R TAGTTGATACTTGGATGATAACCAAAA	
TA43	P	L GGTTGTGTTCTCCAGATTTT	183
		R AAGAGTTGTTGGAGAGCAA	
TA44	P	L ACCGAAATGGAAACAAATAA	193
		R ACAAAACTGGGGGACTAAAT	
TA45	C	L ATGCGTATAAAACCCAGAGA	190
		R TGTTTTTATTGGATTTTCAGTTTCA	
TA53	P	L GGAGAAAATGGTAGTTTAAAGAGTACTAA	249
		R AAAAATATGAAGACTAACTTTGCATTTA	
TA59	P	L ATCTAAAGAGAAATCAAAATTGTCGAA	258
		R GCAAATGTGAAGCATGTATAGATAAAG	
TA64	P	L ATATATCGTAACTCATTAATCATCCGC	239
		R AAATTGTTGTCATCAAATGGAAAATA	
TA71	P	L CGATTTAACACAAAACACAAA	225
		R CCTATCCATTGTCATCTCGT	

TABLE 1. (cont.)

Locus	Repeat Type	Primer Sequence (5`-3`)	Expected Product (bp)
TA72	P	L GAAAGATTTAAAAGATTTTCCACGTTA	256
		R TTAGAAGCATATTGTTGGGATAAGAGT	
TA76s.	I	L TCCTCTTCTTCGATATCATCA	206
		R CCATTCTATCTTTGGTGCTT	
TA78	P	L CGGTAAATAAGTTTCCCTCC	205
		R CATCGTGAATATTGAAGGGT	
TA80	P	L CGAATTTTTACATCCGTAATG	211
		R AATCAATCCATTTTGCATTC	
TA89	I	L ATCCTTCACGCTTATTTAGTTTTTACA	233
		R CAAGTAAAAGAGTCACTAGACCTCACA	
TA93	P	L TTTCTCACACAAATAACAAATTAACTGA	192
		R TCAACATTAATTAACTACTATGATCTGTCA	
TA96	C	L TGTTTTGGAGAAGAGTGATTC	275
		R TGTGCATGCAAATTCTTACT	
TA103	P	L TGAAATATCTAATGTTGCAATTAGGAC	184
		R TATGGATCACATCAAAGAAATAAAAT	
TA106	P	L CGGATGGACTCAACTTTATC	248
		R TGTCTGCATGTTGATCTGTT	
TA110	P	L ACACTATAGGTATAGGCATTTAGGCAA	220
		R TTCTTTATAAATATCAGACCGGAAAGA	
TA113	P	L TCTGCAAAAACTATTACGTTAATACCA	203
		R TTGTGTGTAATGGATTGAGTATCTCTT	
TA114	P	L TCCATNTAGAGTAGGATNTTNTTGGA	298
		R TGATACATGAGTTATTCAAGACCCTAA	
TA116	I	L AATTCAATGACGAATTTTTATAAGGG	182
		R AAAAAGAAAAGGGAAAAGTAGGTTTTA	
TA117	P	L GAAAATCCCAAATTTTTCTTCTTCT	248
		R AACCTTATTTAAGAATATGAGAAACACA	
TA125	P	L TTGAAATTGAACTGTAACAGAACATAAA	235
		R TAGATAGGTGATCACAAGAAGAGAATG	
TA130	P	L TCTTTCTTTGCTTCCAATGT	219
		R GTAAATCCCACGAGAAATCAA	
TA135	P	L TGGTTGGAAATTGATGTTTT	192
		R GTGGTGTGAGCATAATTCAA	
TA140	IC	L TTTTGGCATGTTGTAGTAATCATATTT	180
		R TGAAATGAAAAAGAAAAGGAAAAAGTA	
TA142	P	L TGTTAACATTCCCTAATATCAATAACTT	135
		R TTCCACAATGTTGTATGTTTTGTAAG	
TA144	P	L TATTTTAATCCGGTGAATATTACCTTT	241
		R GTGGAGTCACTATCAACAATCATACAT	
TA146	P	L CTAAGTTTAATATGTTAGTCCTTAAATTAT	161
		R ACGAACGCAACATTAATTTTATATT	
TA180	P	L CATCGTGAATATTGAAGGGT	205
		R CGGTAAATAAGTTTCCCTCC	
TA194	P	L TTTTTGGCTTATTAGACTGACTT	132
		R TTGCCATAAAATACAAAATCC	

TABLE 1. (cont.)

Locus	Repeat Type	Primer Sequence (5`-3`)	Expected Product (bp)
TR1	P	L CGTATGATTTTGCCGTCTAT	224
		R ACCTCAAGTTCTCCGAAAGT	
TR3	P	L GAAGTATCAGTATCACGTGTAATTCGT	244
		R CTTACGGAGAACATGAACATCAA	
TR7	P	L GCATTATTCACCATTTGGAT	204
		R TGTGATAATTTTCTAAGTGTTTT	
TR19	P	L TCAGTATCACGTGTAATTCGT	227
		R CATGAACATCAAGTTCTCCA	
TR20	P	L ACCTGCTTGTTTAGCACAAT	172
		R CCGCATAGCAATTTATCTTC	
TR26	P	L TCATCGCAGATGATGTAGAA	215
		R TTGAACCTCAAGTTCTCTGG	
TR29	I	L GCCCACTGAAAAATAAAAAG	220
		R ATTTGAACCTCAAGTTCTCG	
TR44	P	L TTAATATTCAAAAACTCTCTTGTGCAAT	289
		R TTTACAACAGCGCTTGTATTTAGTAAG	
TR56	P	L TTGATTCTCTCACGTGTAATTC	248
		R ATTTTGATTACCGTTGTGGT	
TR59	C	L AAAAGGAACCTCAAGTGACA	174
		R GAAAATGAGGGAGTGAGATG	
TR60	P	L TGAGTCAAAACAAAGAACTTG	250
		R CTACCGGAAATTTCATTGAC	
TS29	P	L AACATTCATGAACCTACCTCAACTTA	342
		R CCATATATGAGTACACTACCTCTCGG	
TS43	P	L AAGTTTGGTCATAACACACATTCAATA	212
		R TAAATTCACAAACTCAATTTATTGGC	
TS53	P	L GATCNTTCCAAAAGTTCATTTNTATAAT	267
		R TTAAAGAACTGATACATTCCGATTATTT	
TS57	IC	L TCAATTTATAATCATAGAGAATCNGAGA	321
		R CCTAAAACAAATAAAATCTTAAATAATA	
TS104	P	L TCAAGATTGATATTGATTAGATAAAAGC	214
		R CTTTATTTACCACTTGCACAACACTAA	
GA 6	P	L ATTTTTCTCCGGTGTTGCAC	221
		R AAACGACAGAGAGTGGCGAT	
GA 16	P	L CACCTCGTACCATGGTTTCTG	247
		R TAAATTTCATCCTCTCCGGC	
GA 26	P	L GATGCTCAAGACATCTGCCA	234
		R TCATACTCAACAAATTCATTTCCC	
GA 34	P	L CCTTTGCATGTATGTGGCAT	133
		R CCGTTTATAAAGGATGTAZGAGAC	

Listed in the Table are only those markers that reveal at least 4 alleles in 6 chickpea accessions and *C. reticulatum* and *C. echinospermum* on native polyacrylamide gels. Additional primer sequences can be found in [14,27]. The locus name and the sequences of the left (L) and right (R) primers are given. It is further indicated whether the microsatellite is of the perfect (P), interrupted (I), compound (C) or interrupted compound (IC) type. The expected sizes of amplification products in chickpea accession ILC 3279, from which the clones were derived, are given in bp. A more detailed description of the microsatellites and position of the respective genomic localisation for many of them can be found in [27].

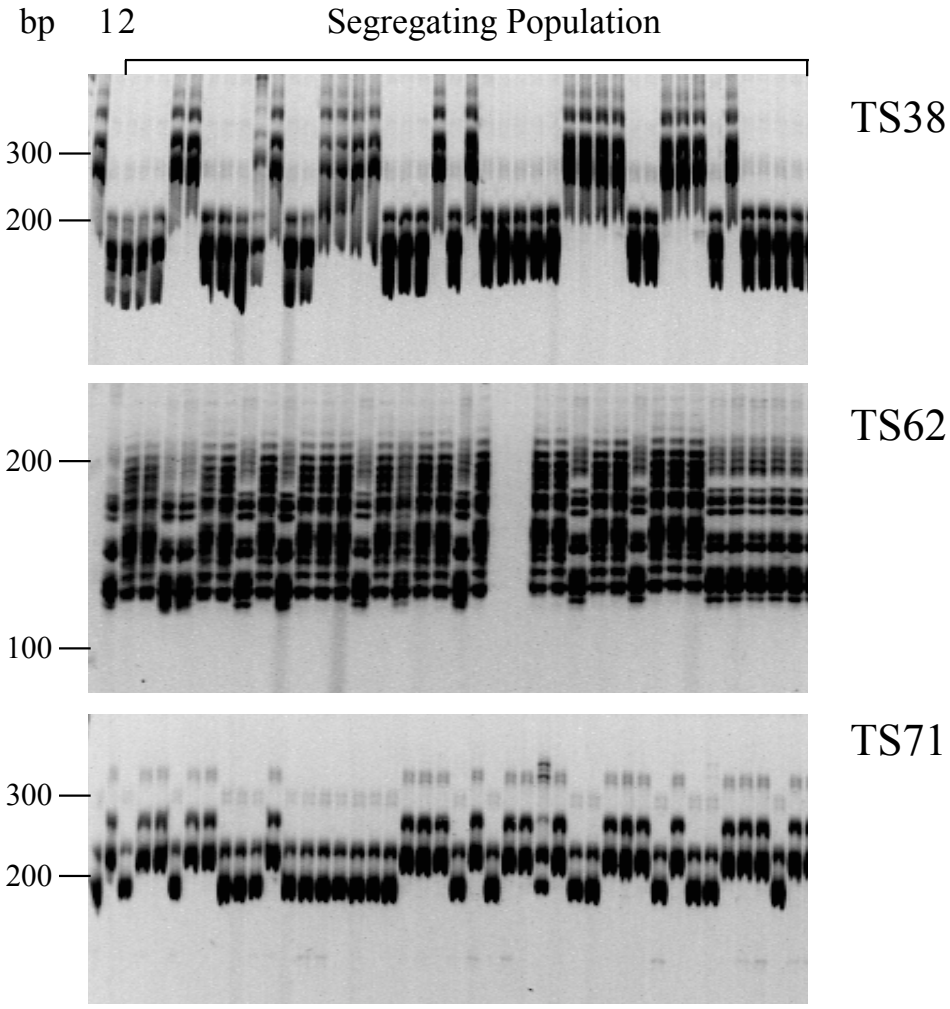

Fig. 1. Mapping of STMS markers in chickpea. Mapping of STMS marker loci TS38, TS62 and TS71 in recombinant inbred lines from a wide cross between a cultivated chickpea accession (ICC4959) and an accession of C. reticulatum (P.I. 498777, slot 2), chickpeas wild, intercrossable relative. Nondenaturing 8% polyacrylamide gel after staining with ethidium bromide (courtesy of Ruth Jungmann, Frankfurt). Numbers on the left indicate molecular weights in base pares. Parental lines ICC 4958 and P.I. 498777 are loaded in slots 1 and 2, respectively. Note, that in all cases several bands are visible that segregate the same way and thus are derived from the same locus

2.2. STMS markers for banana and plantain

2.2.1. Isolation of microsatellite-containing clones

CsCl-gradient purified DNA was isolated from leaf tissue of *M. acuminata* ssp. *malaccensis* [30]. The DNA was restricted with *TaqI*, separated in low melting point agarose gels and fragments in the size range of 300 to 600 bp were eluted from the gel. These were then cloned into the *ClaI* site of pGEM-7Zf (Promega), approximately 8000 clones were plated on Petri dishes, lifted onto nitrocellulose membranes and their DNA fixed on the membrane using standard procedures [31]. The membranes were probed with radiolabled synthetic oligonucleotides $(GT)_{11}$, $(AT)_{11}$ and $(CT)_{15}$ for the presence of microsatellite-containing sequences [30]. After several rounds of screening, the inserts of 30 clones were sequenced. Perfect, compound and imperfect repeats were identified. Primers could successfully be designed for the loci shown in Table 2.

TABLE 2. SELECTED STMS PRIMER PAIRS FROM *Musa acuminata* SSP. *malaccensis*

Locus	Repeat Type	Primer Sequence (5´-3´)	Expected Product (bp)
MaSSR 1	P	L TGAGGCGGGGAATCGGTA	126
		R GGCGGGAGACAGATGGAGTT	
MaSSR 5	P	L AGATGGCGGAGGGAAGAG	120
		R CCGGATCCAAGCTTATCGA	
MaSSR 7	P	L AAGAAGGCACGAGGGTAG	212
		R CGAACCAAGTGAAATAGCG	
MaSSR 8	P	L GGAAAACGCGAATGTGTG	250
		R AGCCATATACCGAGCACTTG	
MaSSR 9	IC	L ATGTCGCTTCGGACCAGA	162
		L R GCAGGACGAAGAACTTACC	
MaSSR 10	P	L ATGATCATGAGAGGAATATCT	127
		R TCGCTCTAATCGGATTATCTC	
MaSSR 11	I	L GGTTGGAACGAAGGTATACTAA	270
		R TCCAAGCTTATCGATCTACG	
MaSSR 12	C	L TGTCGAAGCATCCTACATC	262
		R CTTGGAAACATGAGAAACATAC	
MaSSR 14	P	L TTGAAGTGAATCCCAAGTTTG	131
		R AAAACACATGTCCCCATCTC	
MaSSR 15	P	L TGCTCTTCCACATCTCAAGAAC	247
		R GATTGCACGGAGATTCAACA	
MaSSR 16	I	L ATGGTTAGCTCCGCTTGAAT	294
		R GAGGTGGAAACCCAATCATT	
MaSSR 18	P	L ATGGTTAGCTCCGCTTGAAT	179
		GAGGTGGAAACCCAATCATT	
MaSSR 19	I	L CGTCACAGAAGAAAGCACTTG	144
		R AACCCGGATATTTCATTGTA	
MaSSR 20	P	L GAAATGGAGTTGGAGAAACA	222
		R CACATATCCTTGTCGGAAGT	
MaSSR 24	P	L GAGCCCATTAAGCTGAACA	172
		R CCGACAGTCAACATACAATACA	

The locus name and the sequences of the left (L) and right (R) primers are given. It is further indicated whether the microsatellite is of the perfect (P), interrupted (I), compound (C) or interrupted compound (IC) type. The expected sizes of amplification products in *Musa acuminata* spp. *malaccensis*, from which the clones were derived, are given in bp. A more detailed description of the microsatellites and position of the respective genomic localisation for many of them can be found in [30].

2.2.2. Detection of SSLPs in Musa species

PCR started at 94°C for 4 minutes for initial denaturation followed by 35 cycles of 30 seconds denaturation at 94°C, 30 seconds annealing at T_a (T_a depends on the primer pair, usually 55 °C), 30 seconds extension at 72°C and a final extension at 72°C for 10 minutes.

Reaction volume ranged from 10 to 25 µl. Concentration of the genomic DNA template in the PCR reaction mixture was 2 ng/µl. Final concentrations in the reaction solution were 1.5 mM $MgCl_2$, 150 µM dNTP, 0.2 µM of each primer and 0.2 U/10µl of thermostable polymerase. Before analysis, 1:1 volume 'stop mix' (95% formamide, 0.05% xylene cyanole,

0.05% bromophenol blue, 12.5% sucrose, 10mM NaOH) was added to the aliquots. Denaturing was for 15 minutes at 75°C.

In order to promote the exportation of microsatellite markers to developing countries, we adapted a previously described non-radioactive procedure [32], based on silver stained, denaturing polyacrylamide gels (urea-PAGE), to allele length detection. The procedure starts with the separation of PCR products were separated on denaturing 5% polyacrylamide gels (8 M urea, 1 × TBE, 40 to 60 cm long). After electrophoresis, the gels were soaked in 10% acetic acid for 20 minutes, then washed 3 times for 2 minutes in deionized water and stained for 30 minutes in 0.1% $AgNO_3$ containing 0.06% formaldehyde. After a brief (5–15 seconds) rinse in double-distilled water, signals are developed in 500 ml of 3% Na_2CO_3 with 200 µL of a 10% solution of $Na(S_2O_3)_2$ and 1.5 ml of 38% formaldehyde prepared with double distilled water. After bands become visible, gels are soaked in 10% acetic acid to stop development. The gels are then rinsed with deionized water, and might either be kept in 10% glycerol and dried for 2 hours at 80°C on a gel drier, or dried in a fume hood at room temperature for 24 hours.

That way, up to 20 alleles could be detected in a collection of 69 *Musa* accessions including diploid and various polyploid accessions, plantains and cooking bananas, covering most of the *Musaceae* family [30].

2.3. STMS markers for *Ascochyta rabiei*

Ascochyta rabiei (Pass.) Labrousse is a major fungal pathogen of chickpea. The ascomycete causes severe yield losses in Africa, Asia, Australia, Northamerica, and the Mediterranean basin. The sexual stage of its life cycle, the teleomorph *Didymella rabiei* (Kovachevski) v. Arx has been identified and two compatible *Ascochyta rabiei* mating type isolates were deposited at ATCC (*A*merican *T*ype *C*ulture *C*ollection, Rockville, Md., USA) accession numbers 76501 and 76502. Based on virulence to different chickpea cultivars the pathogen was classified into distinct pathotypes. The pathotype of an isolate may be related to fingerprint patterns obtained from in-gel hybridization of short synthetic oligonucleotides containing microsatellite sequences to restricted genomic DNA as has been shown in an extended survey of *A. rabiei* isolates from Pakistan [4].

2.3.1. Development of STMS markers for Ascochyta rabiei

Our research aimed at developing STMS markers for the characterisation and monitoring of *A. rabiei* populations, and for the construction of a genetic linkage map. For this purpose, DNA was isolated from lyophilized mycelia of ATCC76501 as described [3]. Genomic libraries for the isolation of microsatellite-loci were established from DNA digested separately with either *Alu*I, *Rsa*I, *Hae*I or *Sau*IIIA, size selected on agarose gels (fragment sizes between 300 and 750 bp), ligated into the *Eco*RV- and *Bam*HI-site, respectively, of the pBluescript II SK phagemid (Stratagene) and electroporated into *E. coli* SURE cells. Approximately 5.000 clones were plated on selective 2YT medium, and transferred onto nylon filters for colony hybridization. The filters were hybridized to cocktails of radioactively endlabeled, synthetic single-stranded oligonucleotide probes [$(CA)_{10}$ and $(GA)_{10}$ or $(CAA)_{10}$ and $(GAA)_{10}$] complementary to microsatellite sequences. The dinucleotide mix was hybridized at 43°C and the trinucleotide mix at 48°C, using standard procedures [31]. Microsatellite-containing clones were picked and rescreened by hybridizing the same

cocktails. Plasmids of the 50 positive clones were isolated and inserts sequenced. Primers were designed from microsatellite flanking regions as 20- to 23–mers with a GC-content of at least 45% and a T_m optimum of 56°C. Primer design was possible with 38 of 50 sequences, three clones did not carry a microsatellite at all, but other CA- or GA-rich sequences and nine inserts contained microsatellites too close to the cloning site to allow primer design [22].

2.3.2 Detection of SSLPs in four A. rabiei isolates

PCR was performed in 25 μL volumes containing 1.6 mM $MgCl_2$, 0.2mM dNTPs, 5 pmol of each primer and 0.5 units *Taq* DNA polymerase (Eurogentec, Belgium) as well as the provided 10 × buffer and 10 ng of template DNA. After an initial denaturation (95°C, 20 s) PCR was run for 35 cycles (94°C for 20 s, 53°C for 25 s, 65°C for 23 s) followed by a 20 s final extension step at 65°C in a Perkin-Elmer 2400 thermocycler. Samples were separated on 2% agarose gels and stained with ethidium bromide. Under these conditions 31 of 38 primer pairs worked immediately. Decreasing the annealing temperature to 48°C allowed to amplify a product of the expected molecular weight from six of the remaining seven primer pairs, but three primer pairs displayed more than one band. One primer pair yielded no amplicon at all and was excluded without further testing. More than one band was produced by five primer pairs at an annealing temperature of 53°C. An increase to 57°C produced less bands but still at least two loci were amplified. Major products in the expected molecular weight ranges were observed [22].

Here we present several STMS markers from *Ascochyta rabiei* isolate ATCC76501, and the sequences of the flanking primers. We further investigated, whether these primer pairs would also amplify loci from the related bean pathogen *A. fabae*, and thus could be transferred between *Ascochyta* species [22].

2.4. STMS markers for *Mycosphaerella fijiensis*

M. fijiensis DNA was digested with either the 4 bp blunt end cutter *AluI* or *RsaI* followed by ligation of an adapter allowing PCR amplification of the restriction fragments after enrichment. Microsatellite enrichment was performed with biotinylated oligonucleotides complementary to either $(GA)_{10}$ or a mixture of $(CAA)_8/(GAA)_8/(CA)_{10}$. Biotinylated duplex molecules were fished out of the hybridization solution using streptavidine-coated magnetic beads (Dynal®). After PCR amplification a second enrichment round was performed. PCR products of the second amplification were cloned into a pGEMT vector (Promega®). Single white colonies were selected for colony-PCR which uses a vector-derived primer pair to amplify the cloned insert. All amplified inserts were separated in an agarose gel and thereafter blotted onto a nylon membrane. This Southern blot was hybridized with radioactively labelled $(GA)_{10}$ or a mixture of $(CAA)_8/(GAA)_8/(CA)_{10}$. Positive inserts were sequenced by an automatic fluorescence sequencer (Applied Biosystems ABI 3700®) using a vector-derived nested primer. Primer pairs were designed from the microsatellite-flanking unique sequences using the Primer 3 program (www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi) [20].

TABLE 3. STMS MARKERS FOR *A. rabiei*

Locus	Primer sequence (5'- 3')	Expected size (bp)	Annealing (°C)	Amplicons in *A. fabae*
*ArA*03T	L TAGGTGGCTAAATCTGTAGG R CAGCAATGGCAACGAGCACG	379	53	2
*ArA*05T	L CGAGGATGAAGAGATTCTCAAG R CACTTCTGATGCTACGCTTAC	162	53	2
*ArA*06T	L CTCGAAACACATTCCTGTGC R GGTAGAAACGACGAATAGGG	162	53	1
*ArA*08T	L CAGAGGGGAATTGTTGTTC R ACGACGAGGATGAGGACTTC	249	53	0
*ArA*12T	L AAAGTCAAGTAGACCTGAATACG R GAGAAATTTGACCAAGTGAGAA	175	48	1
*ArR*01T	L AAGAGTCCAGCGTATCGTTT R GCTGTGTCTGTCTCCATCTC	206	57	1
*ArR*04T	L ACGCGTGGAAGAGTCCAGCG R ATGCTCGACAACCTCTCTGGC	262	57	2
*ArH*01T	L CCAAGCTTGGGGACATGGACA R TCAGTTGGCAGACCGTGGTGGC	274	57	1
*ArH*02T	L CTGTATAGCGTTACTGTGTG R TCCATCCGTCTTGACATCCG	365	53	3
*ArH*05T	L CATTGTGGCATCTGACATCAC R TGGATGGGAGGTTTTTGGTA	197	53	1
*ArH*06T	L CTGTCACAGTAACGACAACG R ATTCCAGAGAGCCTTGATTG	167	53	1
*ArH*07T	L CGACAATAACAACCAGTACGAT R CTATGCTGTCTGCCCTCAGT	248	48	0
*ArS*01T	L GAGAAAGAGAAGCGCTATTAAA R GAAGGTATTTCCCTAGCAGAAGA	166	48	2
*ArS*03T	L ATGGAGAAGTCGAGGTCCAT R CTCTTGCGTGGCCTAGAAGG	152	53	2
*ArR*05D	L GTCTAGTTTGCGGAGAGAGTG R CTAGGATGGACACGTTAAAGC	259	53	1
*ArA*02D	L CTATCACCATGCCTCCATCA R TGTTCCCTTTGAGTTGAAGAG	150	53	0
*ArH*02D	L AGAAAGGGGAGATTCGAGAC R AGGTCAGCACGAGATAGCAC	141	53	1

All STMS markers for *A. rabiei* listed here are of the perfect type. The expected size of amplification products in ATCC76501, from which the DNA was derived is given in base pares. The annealing temperature best suited for the respective locus and the number of alleles in two accessions of *A. fabae* are also indicated. More detailes can be found in [22].

The results of three enriched libraries are shown in Tables 4 and 5. GA repeats seem to be more abundant (27%) as compared to CA repeats (15–20%). Trinucleotide repeats are present in the enriched libraries between 4 and 9 percent. Thirty percent (106) of the white colonies (348) contained a microsatellite. However, the sequences were not always good enough to design primer pairs, because the microsatellite was either too close to the cloning site, or it was too short. Also, 10 duplicates were detected in the libraries. From the 26 designed primer pairs, those described in Table 6 have been selected with respect to the number of polymorphisms they detect in the populations.

TABLE 4. EFFICIENCY OF MICROSATELLITE ENRICHMENT IN 3 INDEPENDENT PARTIAL GENOMIC LIBRARIES OF *M. fijiensis* [33]

Genomic Library	Colonies Selected	Positive Signals	Enrichment Factor
1. library	140	CA : 28	CA : 20%
[Pool, Rsa I]		CAA: 10	CAA: 7%
		GAA: 12	GAA: 9%
Total of Positive Clones		49	35%
2. library	112	CA : 17	CA : 15%
[Pool, Alu I]		CAA: 9	CAA: 8%
		GAA: 5	GAA: 4%
Total of Positive Clones		31	28%
3. library	96	26	27%
[GA, Alu I]			
Total of Positive Clones		26	27%
Total of Three Libraries	348	106	30%

TABLE 5. NUMBERS OF SEQUENCES, DUPLICATES AND PRIMER PAIRS FROM 3 INDEPENDENT GENOMIC LIBRARIES OF *M. fijiensis* [33]

Genomic Library	Positive Colonies	Sequences	Duplicates	Primer Pairs
1. Library [pool, *Rsa I*]	49	26	8	12
2. Library [pool, *Alu I*]	31	19	2	8
3. Library [GA, *Alu I*]	26	12	-	6
Total	106	57	10	26

TABLE 6. STMS MARKERS FOR *M. fijiensis*

Locus	Repeat Type	Primer Sequences (5´-3´)	Expected size (bp)	Nig/Mex/ Col
MfSSR-005	P	TCCAAATTCCATCGTTGTCA CGATGATTTGGGTGGTCAAGCTA	158	Mex
MfSSR-025	P	CATGACTGACGTCCTCTTCTCA ATATGGGAAGGGGAAAGGTG	176	Mex/Nig
MfSSR-061	P	TGCAAACTCTGATGCTGGAC TTCAGAGGCTCGTCTTTGGT	124	-
MfSSR-137	P	GGCTCGAAGTGGACTAGCAC CTGGTCGAGGGTCGGG	243	Mex
MfSSR-175	C	AACCTCACATAGGCTGCCAC TATACCTTTCGTTCGGCCTG	286	-
MfSSR-203	P	CTCTGTGGCGTAAGTGGGTG TGATTGCACAGCAGGAAGAG	227	-
MfSSR-230	P	ACAAACTCCCAAGCATCACC GATCGATTCTATTGGCGGAA	265	Nig
MfSSR-244	P	GGCCATTTCATTTGCAAGAC ATGCCACAAAATCTCCATCC	215	Mex/Nig
MfSSR-304	P	TACATACCAGGCCGTCAACA TGACGCATGCATGATACAGA	232	Mex
MfSSR-308	P	TGCAGACTTTCCGATTCCTT TTACGTGGAAAACGCTACCC	152	Mex
MfSSR-316	P	TCCCAGCCAAATCAAAACTC AAGAAATTCGGCATTGAACG	110	Mex/Nig

The outermost left column indicates polymorphisms within or between isolates from Nigeria (Nig), Mexico (Mex) and Colombia (Col) that are visible in agarose gels [33, 36].

3. PERSPECTIVES: A BROAD SPECTRUM OF APPLICATIONS FOR STMS MARKERS

We present here a set of STMS markers for two important crops, chickpea and banana, and for two of their most important fungal pathogens. These can now be used for genome mapping, population and genetic diversity studies in the species from which they have been derived, thereby providing a readily applicable and reliable common basis for data exchange between populations and laboratories. However, first experiments indicate that the range of species for which they can be used is broader than anticipated. For example, chickpea STMS primer pairs amplify loci in the genomes of its wild relatives, markers from *Musa acuminata* can be used in other *Musaceae*, and primers from *A. rabiei* amplify loci from *A. fabae*. These results are portrayed below to encourage the application of these markers also in other species of the respective genus and across genus boundaries. First studies suggest that STMS primers from chickpea, for example, also amplify loci in genomic DNA of both, pea and lentil. However, our results also show, that amplification products from a distinct locus in chickpea are not necessarily allelic to a locus amplified from another species, and thus are not syntenic. Yet, if results are interpreted cautiously, the presented STMS markers may be valuable tools not only for the original species but across species and genus boundaries.

3.1. Transferability of STMS markers from chickpea between species of the genus *Cicer*

It has frequently been observed that SSR-flanking sequences are conserved in closely related species. For example, human STMS amplified corresponding loci from chimpanzees, and bovine STMS alleles from goats and sheep. Transferability of markers was observed in *Canidae*, *Cetaceae*, different species of pines, and *Medicago* (see [34] for references).

One of the aims of our work is the marker-assisted utilisation of the primary and secondary gene pool of chickpea for the improvement of the crop. Therefore, we explored whether and to which extent STMS primers designed for the cultigen could also be applied to genome analysis of wild *Cicer* species. We explored if conservation of microsatellite-flanking sequences reflects the known evolutionary relationship between these species. Further, we exemplarily investigated what underlies the differences in number and size of amplification products derived from the same or different species observed in preliminary experiments [34].

The conservation of 90 microsatellite-flanking sequences from chickpea was investigated in 39 accessions of 8 annual, and one accession of a perennial species of the genus *Cicer*. All primer sequences successfully amplified microsatellites in related species, indicating conservation of microsatellite-flanking sequences in chickpea's relatives. However, conservation of primer sites varied between species depending on their known phylogenetic relationship to chickpea. It ranged from 92.2% in *C. reticulatum*, chickpea's closest relative and potential ancestor, down to 50% for *C. cuneatum*. A phylogenetic tree generated using PAUP revealed a closer relationship between chickpea and the other members of its crossability group to the perennial *C. anatolicum* than to other annual species of the genus. In many cases variation in size and number of amplification products between and within species was observed. Sequence analysis of highly divergent amplification products revealed that differences in their size are either due to large variation in the number of microsatellite repeats of the same allele (in the case of chickpeas closest relatives), or to the amplification of another locus unrelated to the one amplified from chickpea (in the case of more distantly related species). Both, sequence information and bootstrapping suggested that STMS derived from chickpea may efficiently and reliably be used for synteny studies in chickpea's

crossability group including *C. anatolicum*. However, care should be taken when applying these markers to more distantly related species of the genus [34].

3.2. Transferability of STMS markers from *M. acuminata* between species of the genus *Musa*

It is not surprising that STMS primers developed from *M. acuminata* ssp. *malaccensis* (AA) also amplify loci in other subspecies and in the related species *M. balbisiana* (BB). Almost all of the banana and plantain cultivars are either intraspecific (AAA) or interspecific (AAB/ABB) hybrids which can be characterised by the developed STMS markers [30, 35]. Moreover, some of the primer pairs also amplify polymorphic fragments in other *Musa* species and even related taxa, e.g. other members of the family of *Musaceae* [30, 36]. Levels of expected and observed heterozygosity were highest among diploid *M. acuminata* accessions and wild types. Variation among triploid banana cultivars, plantains, and cooking bananas was considerably lower, although the "frozen" alleles in these sterile hybrids allow to trace back potential ancestors [30, 36].

3.3. *Ascochyta rabiei* microsatellite markers for population studies of *Ascochyta* pathogens of chickpea, pea, bean and lentil

Applications of STMS markers to natural field populations detect polymorphic allele sizes that are observed more frequently at loci with comparatively high repeat numbers. The analysed *A. rabiei* microsatellites remain stable during vegetative propagation and mutations have only very rarely been observed after sexual reproduction and are therefore well suited for this purpose [3, 22]. The use of only few of the presented primer pairs flanking microsatellites with highest repetition grades and single-locus performance should result in lineage- or strain-specific allele patterns with high discriminatory potential. Reliable monitoring of *A. rabiei* populations in chickpea-growing regions of the world can now be performed with a fast PCR approach.

It is striking, that the STMS markers suggested for widespread use here, were developed from genome screening with trinucleotide repeats. Although the experiments were performed in parallel and under identical conditions, $[CAA]_n/[GAA]_n$ screening was more efficient [25 markers] than the use of $[CA]_n/[GA]_n$ [only 12 markers]. In addition, trinucleotide-based markers revealed more polymorphisms [about two thirds of the bands were polymorphic on agarose gels] as compared to dinucleotide repeats [only 25% polymorphisms]. The same holds for the transferability. Only one third of the dinucleotide markers were transferable to *A. fabae* as compared to 80% of the primer pairs from sequences flanking trinucleotide repeats. One advantage of the primers flanking dinucleotide repeats is single-locus amplification in *A. rabiei* and *A. fabae* [22]. Furthermore, transferability to other species of the genus *Ascochyta* can be expected as has been exemplified for *A. fabae*. These plant pathogens are strictly host specific, and are highly virulent to leguminous crops such as pea, bean or lentil. The conservation of primer binding sites, especially those flanking trinucleotide repeats, is demonstrated by unique amplification products in most of the cases [22]. It remains to be examined whether these fragments are also polymorphic and contain microsatellites in heterologous species.

3.4. *Mycosphaerella fijiensis* microsatellites for population studies

STMS primers derived from a Nigerian isolate (*Mf*-Nig-862) can successfully be applied to isolates from Latin-American *M. fijiensis* populations. Although the distribution of alleles varied considerably between both continents due to geographic isolation, the markers detected unique haplotypes as well as internal structures of the populations [20]. In some countries (e.g. Colombia) banana are grown in lowlands and highlands. In the highlands a second *Musa* pathogen, *M. musicola*, the causative agent of Yellow Sigatoka is observed. This fungus has a lower temperature optimum than the Black Sigatoka pathogen *M. fijiensis*. On the other hand, *M. fijiensis* is more aggressive because of its shorter life cycle, and could eventually replace *M. musicola* in the lowlands. We are now applying STMS markers from *M. fijiensis* for monitoring both species in Latin America asking if adaptive processes occur in the *M. fijiensis* populations of the lower highlands. If *M. fijiensis* would be adapting to lower temperatures it could replace *M. musicola* even in the highlands.

Another problem caused by these mainly sexually propagating ascomycetes is their resistance to fungicides. First benomyl-resistant isolates have been observed, and the appearance in regions of extensive spraying (e.g. banana plantations in Central America) of strains carrying more than a single resistance factor is only a matter of time. A molecular pathogen forecast system allowing to monitor the population before spraying would help to reduce the risks for more and more resistant strains. Such strains can be characterised by STMS markers and be identified by their unique multi-locus haplotypes to improve plant protection.

ACKNOWLEDGEMENTS

Research of the authors has been funded by German Research Council [DFG Ka 332/15; DFG Ka 332/17], European Community [ERB-IC-18/CT-970192], International Atomic Energy Agency [CRP 302-02-GFR-8148], and Vereinigung von Freunden und Förderern der Johann Wolfgang Goethe-Universität Frankfurt am Main.

REFERENCES

[1] TANKSLEY, S.D., GANAL, M.W., MARTIN, G.B. Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends Genet. 11 (1995) 63–68.

[2] WINTER, P., KAHL, G. Molecular marker technologies for plant improvement. World J. Microbiol. Biotechnol. 11(1995) 438–448.

[3] MORJANE, H., GEISTLINGER, J., HARRABI, M., WEISING, K., KAHL, G. Oligonucleotide fingerprinting detects genetic diversity among *Ascochyta rabiei* isolates from a single chickpea field in Tunisia. Curr. Genetics 26 (1994) 191–197.

[4] JAMIL, F.F. *et al.* Genetic and pathogenetic diversity within *Ascochyta rabiei* (Pass) Lab. populations in Pakistan causing blight of chickpea (*Cicer arietinum* L.) Physiol. Mol. Plant. Pathol. (2000) in press.

[5] LITT, M., LUTY, J.A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44 (1989) 397–401.
[6] McMURRAY, C.T. Mechanisms of DNA expansion. Chromosoma 104 (1995) 2–13
[7] SIA, E.A., JINK-ROBERTSON, S., PETES, T.D. Genetic control of microsatellite instability. Mutation Res. 38 (1997) 361–70.
[8] NAKAMURA, Y. *et al.* Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235 (1987) 516–522.
[9] CHO Y.G. *et al.* Diversity of microsatellites derived from genomic libraries and GeneBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100 (2000) 713–722.
[10] SAGHAI-MAROOF, M.A., BIYASHEV, R.M., YANG, G.P., ZHANG, Q, ALLARD, R.W. Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc. Natl. Acad. Sci. USA 91 (1994) 5466–5470.
[11] WEISING, K., FUNG, R.W.M., KEELING, J., ATKINSON, R.G., GARDNER, R.C. Cloning and characterisation of microsatellite repeats from *Actinidia chinensis*. Mol. Breeding 2 (1996) 117–131.
[12] GEISTLINGER, J., WEISING, K., KAISER, W.J., KAHL, G. Allelic variation at a hypervariable compound microsatellite locus of the ascomycete *Ascochyta rabiei*. Mol. Gen. Genet. 256 (1997) 298–305.
[13] LAGERCRANTZ, U., ELLEGREN, H., ANDERSSON, L. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. 21 (1993) 1111–1115.
[14] HÜTTEL, B. *et al.* Sequence-tagged microsatellite markers for chickpea (*Cicer arietinum* L.). Genome 42 (1999) 1–8.
[15] RODRIGUEZ, R.J., YODER, O.C. A family of conserved repetitive DNA elements from the fungal plant pathogen *Glomerella cingulata* (*Colletotrichum lindemuthianum*). Exp. Mycol. 15 (1991) 232–242.
[16] GROPPE, K., SANDERS, I., WIEMKEN, A., BOLLER, T. A microsatellite marker for studying the ecology and diversity of fungal endophytes (*Epichlöe* spp.) in grasses. Appl. Environ. Microbiol. 61 (1995) 3943–3949.
[17] OSIEWACZ, H.D., HAMANN, A., WERNER, A. Genome analysis of filamentous fungi: identification of a highly conserved simple repetitive sequence in different strains of *Podospora anserina*. Microbiol. Res. 151 (1996) 1–8.
[18] BUCHELI, E., GAUTSCHI, B., SHYKOFF, J.A. Isolation and characterisation of microsatellite loci in the anther smut fungus *Microbotryum violaceum*. Mol. Ecol. 7 (1998) 657–666.
[19] OWEN, P.G., PEI, M., KARP, A., ROYLE, D.J., EDWARDS, K.J. Isolation and characterisation of microsatellite loci in the wheat pathogen *Mycosphaerella graminicola*. Mol. Ecol. 7 (1998) 1611–1612.
[20] NEU, C., KAEMMER, D., KAHL, G., FISCHER, D., WEISING, K. Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis*. Mol. Ecol. 8 (1999) 513–525.
[21] MOON, C.D., TAPPER, B.A., SCOTT, B. Identification of *Epichlöe* endophytes in planta by microsatellite-based PCR fingerprinting assay with automated analysis. Appl. Environ. Microbiol. 65 (1999) 1268–1279.
[22] GEISTLINGER, J., WINTER, P., KAHL, G. Locus-specific microsatellite markers for the chickpea pathogen *Didymella rabiei* (anamorph: *Ascochyta rabiei*). Mol. Ecol. (2000) in press.

[23] BECKMANN, J.S., SOLLER, M. Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Bio/Technology 8 (1990) 930–932.
[24] WINTER, P., HÜTTEL, B., WEISING, K., KAHL, G. "Microsatellites and molecular breeding: Exploitation of microsatellite variability for the analysis of a monotonous genome", Molecular Techniques in Crop Improvement (JAIN, S.M., Ed.), Kluwer Academic Publishers, Dordrecht, The Netherlands (2000). In press.
[25] GUPTA P.K., VARSHNEY, R.K. (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113 (2000) 163–185.
[26] RODRIGUEZ, H., GEISTLINGER, J., BERLYN, G., KAHL, G., WEISING, K. Characterisation of novel microsatellite loci isolated from the tropical dioecious tree *Simarouba amara*. Mol. Ecol. 9 (2000) 489–504.
[27] WINTER, P. *et al.* Characterisation and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. Mol. Gen. Genet. 262 (1999) 90–101.
[28] WINTER, P. *et al.* A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* × *C. reticulatum* cross: Localisation of resistance genes for *Fusarium* races 4 and 5. Theor. Appl. Genet. (2000) in press.
[29] VALADEZ-MOCTEZUMA, E., KAHL, G. Huellas de ADN en Genomas de Plantas : Teoria y Protocolos de Laboratorio, Ediciones Mundi-Prensa, Mexico (2000)
[30] KAEMMER, D. *et al.* Molecular breeding in the genus *Musa*: A strong case for STMS marker technology. Euphytica 96 (1997) 49–63.
[31] SAMBROOK J., FRITSCH, E.F., MANIATIS, T. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
[32] BASSAM, B.J., CAETANO-ANOLLÉS, G., GRESSHOFF, P.M. Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Biochem. 196 (1991) 80–83.
[33] NEU, C. Entwicklung von STMS-Markern und ihre Anwendung in der Genomanalyse von *Mycosphaerella fijiensis*. Diploma Thesis, University of Frankfurt am Main (1998).
[34] CHOUMANE, W., WINTER, P., WEIGAND, F., KAHL, G. Conservation and variability of sequence-tagged microsatellite sites from chickpea (*Cicer arietinum* L.) within the genus Cicer. Theor. Appl. Genet. 101 (2000) 269–278.
[35] KAEMMER, D. *et al. Musa acuminata* ssp. *malaccensis*-derived and chloroplast microsatellites detect different levels of variability in *M. balbisiana* and interspecific AB, AAB, and ABB hybrids. Mns. in preparation.
[36] KAEMMER, D. *et al.* "Microsatellite markers for genome analysis in *Musa* and *Mycosphaerella*", Cellular biology and biotechnology including mutation techniques for creation of new useful banana genotypes, Report of the third FAO/IAEA Research Co-ordination Meeting. Vienna, Austria (2000) in press.

Appendix I

LIST OF SELECTED NUCLEIC ACID ANALYSIS TECHNIQUES

Modified from: KARP, A. *et al.* Nature biotechnology **15** (1997) 625–628.

Hybridization-based techniques

Multilocuc minisatellite fingerprinting
Oligonucleotide fingerprinting
Restriction fragment length polymorphism (RFLP)

Amplification-based nucleic acid scanning techniques

Random amplified polymorphic DNA (RAPD)
Arbitrarily primed PCR (AP-PCR)
DNA amplification fingerprinting (DAF)
Arbitrary signatures from amplification profiles (ASAP)
Amplified fragment length polymorphism (AFLP)
Microsatellite-primed PCR (MP-PCR)
Anchored MP-PCR (AMP-PCR)
Random amplified microsatellite polymorphism (RAMP)
Nucleic acid scanning-by-hybridization (NASBH)
RAPD dot-blot hybridization

Amplification-based nucleic acid profiling techniques

Simple sequence repeat PCR (SSR-PCR)
Minisatellite variant repeat PCR (MVR-PCR)

Sequence-targeted techniques

Cleaved amplified polymorphic sequence (CAPS)
Single strand conformation polymorphism (SSCP)
Allele specific oligonucleotide (ASO) hybridization
TaqMan ASO
Oligonucleotide arrays

Appendix II

COMPARISON OF DIFFERENT MARKER SYSTEMS

TECHNIQUE	RFLP	RAPD/DAF	MICROSATELLITES	SCARS/CAPS	AFLP	MICROARRAY
Description	Endonuclease restriction Southern blotting Hybridization	DNA amplification with random primers	PCR of simple sequence repeats	Polymorphic PCR products or endonuclease restriction of PCR products	Endonuclease restriction followed by selective PCR amplification	DNA amplified, labeled and hybridized to oligos or cDNA
Type of Polymorphism	Single base changes Insertions Deletions	Single base changes Insertions Deletions	Changes in repeat length	Single base changes Insertions Deletions	Single base changes Insertions Deletions	Single base, quantitative
Genomic abundance	High	Very high	Medium	High	Very high	very high
Level of polymorphism	Medium	Medium/High	High	Medium	Very high*	very high
Dominance	Co-dominant	Dominant	Co-dominant	Codominant	Dominant	Codominant
REQUIREMENTS						
Amount of DNA	2-10ug	10-25ng	50-100ng	50-100ng	500ug*	100 ng
Sequence information	No	No	Yes	Yes	No	Yes/No
Radioactive detection	Yes/no	No	No/yes	No	Yes/No	No
Gel system	Agarose	Agarose	Acrylamide/agarose	Agarose	Acrylamide	No gel
IMPLEMENTATON						
Development costs	Medium	Low	High	Medium/high	Medium/High	Very high
Start-up costs	Medium/High	Low	High	High	Medium/High	Very high
Portability - Lab/Crop	High/High	Medium/Nil	High/Low	High/Low	High/Nil	Unknown
Suitable applications	Comparative Mapping Framework mapping	Varietal /hybrid identification Marker-assisted selection	Framework/region specific mapping. Fingerprinting. Marker-assisted selection. Comparative mapping	Framework mapping. Marker-assisted selection. Can be converted to allele specific probes.	Fingerprinting. Very fast mapping. Region-specific marker saturation.	Fingerprinting, sequencing, transcription

** Expressed Sequence Tags (ESTs) may be used in several marker systems in a variety of ways (e.g. a source of DNAclones for RFLPs, DNA sequence information for SCARs and AFLPs and hybridization sites on microarrays)

* Because this technique leads to multiple infomative DNA bands per sample, the generalizations in the table are based on results typically achieved with an entire sample (rather than individual bands).

Appendix III

PLANT GENOME DATABASE CONTACT INFORMATION

DATABASE	CROPS	CURATOR	E-MAIL ADDRESS	DATABASE ADDRESS
AAtDB	*Arabidopsis*	David Flanders	flanders@genome.stanford.edu	http://genome-www.stanford.edu/Arabidopsis/
Alfagenes	Alfalfa (*Medicago sativa*)	Daniel Z. Skinner	Dzolek@ksu.ksu.edu	http://naaic.org/
Bean Genes	*Phaseolus* and *Vigna*	Phil McClean	mcclean@beangenes.cws.ndsu.nodak.edu	http://probe.nalusda.gov:8300/cgi-bin/browse/beangenes
ChlamyDB	*Chlamydomonas reinhardtii*	Elizabeth H. Harris	chlamy@acpub.duke.edu	http://probe.nalusda.gov:8300/cgi-bin/browse/chlamydb
CoolGenes	Cool season food legumes	Fred Muehlbauer	muehlbau@wsu.edu	http://probe.nalusda.gov:8300/cgi-bin/browse/coolgenes
CottonDB	*Gossypium* species	Sridhar Madhavan	msridhar@tamu.edu	http://probe.nalusda.gov:8300/cgi-bin/browse/cottondb
GrainGenes	Wheat, barley, rye and relatives	Olin Anderson	oandersn@pw.usda.gov	http://probe.nalusda.gov:8300/cgi-bin/browse/graingenes
MaizeDB	Maize	Mary Polacco	maryp@teosinte.agron.missouri.edu	http://www.agron.missouri.edu/
MillettGenes	Pearl millett	Matthew Couchman	Matthew.Couchman@bbsrc.ac.uk	http://jiio5.jic.bbsrc.ac.uk:8000/cgi-bin/ace/search/millet.
PathoGenes	Fungal pathogens of small-grain cereals	Henriette Giese	h.giese@risoe.dk	http://probe.nalusda.gov:8300/cgi-bin/browse/pathogenes
RiceGenes	Rice	Susan McCouch	srm4@cornell.edu	http://genome.cornell.edu/rice/
RiceGenome Project	Rice			http://www.staff.or.jp
SolGenes	*Solanaceae*	Molly Kyle	mmk9@cornell.edu	http://genome.cornell.edu/solgenes/welcome.html
SorghumDB	*Sorghum bicolor*	Russel Kohel/Bob Klein	nus6389@tam2000.tamu.edu	http://probe.nalusda.gov:8300/cgi-bin/browse/sorghumdb
Soybase	Soybeans	David Grant	dgrant@iastate.edu	http://129.186.26.94/
TreeGenes	Forest trees	Kim Marshall	kam@s27w007.pswfs.gov	http://dendrome.ucdavis.edu/index.html
National Center for Genome Resources	Various			http://www.ncgr.org/

LIST OF PARTICIPANTS

Caetano-Anolles, G.	Department of Biology, University of Oslo, P.O. Box 1045 Blindern, N-0316 Oslo, Norway
Kaemmer, D.	J.W. Goethe University, Plant Molecular Biology Biocentre, Marie Curie Str. 9, D-60439 Frankfurt/Main, Germany,
Lee, M.	Department of Agronomy, Iowa State University of Science & Technology, Ames, Iowa 50011-1010, United States of America
Qi, X.	John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, United Kingdom
Sinibaldi, R.	Operon Technologies, 1000 Atlantic Avenue, Suite 108, Alameda, California 94501, United States of America
Young, N.D.	Department of Plant Pathology, College of Agriculture, 495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, Minnesota 55108, United States of America
Maluszynski, M.	Plant Breeding & Genetics Section, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, International Atomic Energy Agency, Wagramer Strasse 5, P.O. Box 100, A-1400 Vienna, Austria