Plant Molecular Biology Reporter
15 (2) 1997

pages 154–162

Protocols

Use of STSs and SSRs as Rapid and Reliable Preselection Tools in a Marker-Assisted Selection-Backcross Scheme

Jean-Marcel Ribaut,[1] *Xueyi Hu, David Hoisington, and Diego González-de-León*

E-mail: jribaut@cimmyt.mx
Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Int., Lisboa 27, A.P. 6-641, México 06600 D.F., México

Key Words: DNA isolation, maize, MAS, microsatellites, multiplexing, PCR screening, sequence-tagged sites

Abstract: We describe a new approach for using suitable STS and SSR markers as a powerful molecular tool for screening segregating populations involved in backcross schemes for marker-assisted selection, as a preselection step. Since it can be applied to very large populations, this preselection strategy allows one to increase substantially the pressure of selection at each backcross generation. The technique is fast and reproducible, and can be made even more efficient and cost-effective by simultaneous DNA amplification from different primer pairs. In the example illustrated here, three suitable PCR-based markers were used to complete the selection of 300 individuals out of 2300 in less than one month with two people working on the project.

During the last ten years, the development of new molecular marker systems has been one of the most dynamic areas in applied molecular genetics. The recent increase in the number of available PCR primers for STSs, developed principally from RFLP probes (MaizeDB, at http://www.agron.missouri.edu/top.html), and SSRs or

---

[1]Author for correspondence

Abbreviations: BC, backcross; BME, β–mercaptoethanol; CTAB, alkyltrimethyl-ammonium bromide; MAS, marker-assisted selection; PCR, polymerase chain reaction; QTL, quantitative trait locus; RFLP, restriction-fragment-length polymorphism; SSR, simple sequence repeat; STS, sequence-tagged site; Taq, *Thermus aquaticus*.

microsatellites for species, such as maize (e.g., Taramino and Tingey, 1996), has opened new doors for genome mapping and especially for developing new PCR selection strategies based on markers. Given the typical length of STS and SSR primers (18–30 bp), the products of PCR amplification are highly reproducible and can be used reliably as molecular markers of plant genomes whenever they reveal polymorphisms. Such markers should map to specific loci, irrespective of which segregating population is used for mapping. If such loci are linked to genes involved in the expression of a trait of interest, it is possible, therefore, to use the corresponding markers for selection of that trait. Before the availability of suitable PCR-based markers, marker-assisted selection (MAS) schemes had to be designed within the constraints imposed by the most reliable marker system, RFLPs. This technique imposed a practical (both financial and logistical) limit on the number of individuals that could be genotyped efficiently. Here, we present a simple strategy that can be used for marker-based selection on populations that are at least ten times larger than is currently practical with RFLPs. This approach will be illustrated with a practical example of preselection, using three PCR markers at the BC_2 level, a part of the MAS project conducted at CIMMYT to improve drought tolerance in tropical maize.

Procedures

The preselection strategy

The basic goal of a BC MAS scheme consists of the transfer of one or more pertinent genomic segments from a donor to a target variety. By screening a large BC population with a reliable PCR-based marker for a particular locus, all the individuals not having the desired allele (statistically half of the population) are discarded. Those retained are then assayed at a second locus to repeat the preselection cycle, and so on, for as many loci as is practical. The population can thus be quickly reduced in size to one that can then be handled more efficiently using RFLPs for genotyping at further loci of interest, and to evaluate the remaining percentage of the donor parent genome in the different preselected individuals. Here we describe the implementation and application of this powerful preselection tool, using a specific example of preselection that employs three PCR-based markers on 2300 BC_2F_1 individuals.

Sample harvest

Approximately 15 cm (about 350 to 500 mg) of a leaf tip of each segregat-

ing BC maize plant were harvested, placed in a 15-mL plastic tube, kept on ice until arrival at the laboratory, and then stored directly at -80°C. Harvesting early in the morning is recommended to avoid the dehydration of leaf tissue that might be encountered later in the day.

Reagents

extraction buffer: 0.1 M tris pH 8.0, 0.15 M EDTA pH 8.0, 2.1 M NaCl, 2% PVP, 2% CTAB, 1% 14 M BME

DNA extraction for PCR amplification

- Extract leaf juice from frozen tissue using a sap extractor (Clarke et al., 1989). As the tissue is pressed, add 1.2 mL of extraction buffer, and collect the extract, about 1 mL, at the tip of the rollers in 2-mL tubes.
- Incubate the extracts in a water bath or an oven at 65°C for 20–40 min; mix twice by inversion during this incubation. Remove the tubes from the heat, and let cool down for 5–10 min.
- Extract the samples with 1 mL of octanol-chloroform (1:24). Mix by inversion for 5 min and spin in a table-top centrifuge at 3200 rpm for 10 min.
- Transfer 600 µL of the aqueous supernatant containing the DNA to 1.5-mL Eppendorf tubes.
- Add 75 µL of 5 M NaCl and precipitate DNA with 700 µL of cold absolute ethanol.
- Spin DNA down, decant ethanol, and dry under a weak vacuum for 30 min.
- Resuspend in 500 µL TE, pH 8.0.
- Quantify DNA of a small number of randomly chosen samples, using a fluorometer, as a rough guide for adjusting sample DNA concentration.
- One person can easily handle 100 to 150 samples per day.

Selection of PCR primers

Primers for both STS and SSR markers were tested. A marker is considered of interest when its theoretical map position fits within the two flanking markers of a genomic segment (typically resulting from the detection of a QTL for a given trait) that is to be transferred. In the case of STSs, many terminal sequences of RFLP probes have been published, and thus appropriate primers can be designed. For maize, many sequences are available from the Maize Genome Database (MaizeDB), and primers can be designed using various computer programs available commercially or in the public domain. In the case of maize SSRs, primer

sequences are directly available on the Internet (MaizeDB) or from published papers (e.g., Taramino, and Tingey, 1996; Maize Genetics Cooperation Newsletter No. 70, 1996). After amplification of parental DNA, polymorphic products can be mapped using a small number of segregating individuals to confirm the locus position in the map of the segregating population under study.

PCR protocol and identification of polymorphisms

In order to increase the screening efficiency of markers, more than one pair of primers can be used for multiplex amplification, using the same PCR conditions (Table I). This implies that care must be taken in designing or choosing primer pairs for which the annealing temperatures are reasonably compatible, and for which there is no overlap in the size of the amplification products. PCR reactions were run in 96-well thermal cyclers (ERICOMP, Inc.) under the following conditions: 2 min at 93°C; 30 cycles of 1 min at 93°C, 2 min at 56°C, and 2 min at 72°C; and finally 5 min at 72°C. Different reaction volumes can be used according to the final number of gel wells that might need to be loaded for separation of the amplification products (Table I). Different gel systems can be used according to the sizes of the amplification products and clarity of the polymorphisms: normal agarose gels (1.5%, e.g., SeaKem LE, FMC BioProducts, Rockland, USA) for larger fragments, high-resolution aga-

Table I. Composition of PCR reactions as a function of the number of primer pairs that can be used for simple or multiplex amplification.

Component	Stock	Volume in µL for 1 primer pair	2 primer pairs	3 primer pairs
ddH_2O		9.4	4.4	8.2
Primer 1.1	2 µM	2.0	2.0	2.0
Primer 1.2	2 µM	2.0	2.0	2.0
Primer 2.1	2 µM	-	2.0	2.0
Primer 2.2	2 µM	-	2.0	2.0
Primer 3.1	2 µM	-	-	2.0
Primer 3.2	2 µM	-	-	2.0
Taq buffer	10	2.0	2.0	3.0
dNTP mix	2.5 mM each	1.6	1.6	2.4
$MgCl_2$	50 mM	0.8	0.8	1.2
Template DNA	10 ng/µL	2.0	3.0	3.0
Taq polymerase	5 U/µL	0.2	0.2	0.2
Total reaction vol.		20	20	30

rose gels (3.0%, e.g., 2% MetaPhor agarose, FMC BioProducts, Rockland, USA, and 1% SeaKem LE agarose) for a wide range of fragment sizes, or polyacrylamide gels (6%) for small fragments and tight polymorphisms. For both normal and high-resolution agarose gels, we routinely run 20x25-cm gels with eight 30-well combs for a total of 240 samples. Expensive high-resolution agarose can be reused at least three times after running off the previous samples from the gel and remelting it. An increase in the use-efficiency of acrylamide gels can also be achieved by loading additional sets of samples after the first set has been electrophoresed for a period of time.

Results and Discussion

Three key issues were considered during the development of each step of the preselection protocol: high reproducibility and reliability of the amplification products of large numbers of samples, relative rapidity of the operation, and cost effectiveness.

Efficiency of DNA extraction

The harvest of predetermined lengths of leaf-tip tissue into pre-labeled plastic tubes can be very fast, and samples do not need to be weighed. Since little material is needed for DNA extraction, harvesting can be performed at an early stage of plant development. Direct extraction of frozen samples using a sap extractor obviated the need for lyophilization and/or laborious grinding procedures, which would be too time consuming for a large number of samples. The DNA extraction protocol is not the fastest one (see Rogers et al., 1996), but the chloroform-partition step assures good DNA quality and more uniform amounts. High molecular DNA is obtained with relatively low partial degradation (about 30%), and could be amplified by PCR. No effect on PCR amplification was detected in the presence of RNA, and therefore an RNAse treatment was not necessary. Even with the variation in the initial fresh weight among leaf samples, the range of DNA concentrations at the end of the extraction never exceeded a factor of four. This variation in the final DNA concentration was found not to affect the proper amplification of the DNA samples. Consequently, with a DNA quantification of only a few randomly chosen samples (e.g., 30) the final resuspension volume of TE could be established. In our example, the mean DNA concentration chosen was 10 ng/μL, obtained with a final resuspension volume of 500 μL of TE. This amount of DNA can be used for more than a hundred PCR amplifications.

PCR primer selection criteria

The two kinds of PCR-based markers used in our strategy, STSs and SSRs, have different advantages and weaknesses. In both cases, their number in publications and databases is increasing rapidly, thereby facilitating the choice of those most appropriate for a given purpose. In the case of STS markers, their genomic location can be directly associated with the position of the corresponding RFLP locus detected by the probe from which the STS primers were derived. The chance of detecting directly a polymorphism with a pair of STS primers is, however, very low compared to the corresponding RFLP probe. In effect, the latter will usually reveal polymorphisms based on changes in restriction sites within or flanking the sequence homologous to the probe, and therefore over a relatively long DNA sequence (typically up to 20 kb). Polymorphisms revealed by an STS will be strictly limited to sequence variation within the limits of the amplified sequence, which is much shorter (generally 0.6 to 3 kb). Often, a restriction digest of the amplified product may be needed to increase the chance of revealing a polymorphism between two individuals. By contrast, since SSRs are hypervariable regions, these detect much higher levels of polymorphism. Moreover, the initial use of SSR primers is faster, since the primer sequences are publicly available, while for STSs, primers may have to be designed from published sequence information. SSRs generally yield much smaller amplification products and thus polymorphisms are more difficult to resolve (e.g., higher resolution gel systems may be needed), whereas STS products are longer and polymorphisms can be resolved using cheaper conventional gel systems, as suggested in the **Procedures** section. The quality of selected PCR-based markers is crucial for successfully achieving the preselection protocol, since the same primers will be reused at the successive levels of BC selection and over a very large number of samples. So far, we have found that all the maize SSR and STS primers tested in our laboratory reveal loci that map at the expected positions.

Increased efficiency of polymorphism screening through multiplex PCR amplifications and gel loading

For both STSs and SSRs, primer length is at least 18 bases. PCR amplification using such primer lengths is highly sequence specific and reproducible. Although each primer pair may have an optimum temperature for annealing to the template DNA, the sequence specificity of the amplification reaction due to high primer length allows some degree of flexibility in selecting a single annealing temperature for two or more

pairs of primers. This means that more than one target sequence could be amplified, under the same PCR conditions, in the same reaction tube (Table I). Since STSs and SSRs usually produce very different amplification product lengths, they often can be combined. It is often possible to resolve the two polymorphisms simultaneaouly on the same gel (Figs. 1 and 2). This strategy, which in some cases might be extended to more than two markers, significantly reduced selection costs and screening time. Note in Table I that the same amount of Taq polymerase (by far the most expensive item in the reaction) could be used irrespective of the number of primer pairs included in the reaction.

A practical example

The use of PCR-based markers as a preselection step was successfully tested for the transfer of quantitative QTLs involved in the expression of a favorable trait for drought tolerance in tropical maize. This MAS experiment is in progress at CIMMYT. Based on selected QTL intervals, several STS and SSR markers were tested. After confirmation of the map position of loci identified by several markers, three different primer pairs were selected for screening the BC_2F_1 individuals, and thus reduce the large initial population size by approximately a factor of eight. Selection was based on the quality of the polymorphism identified and their

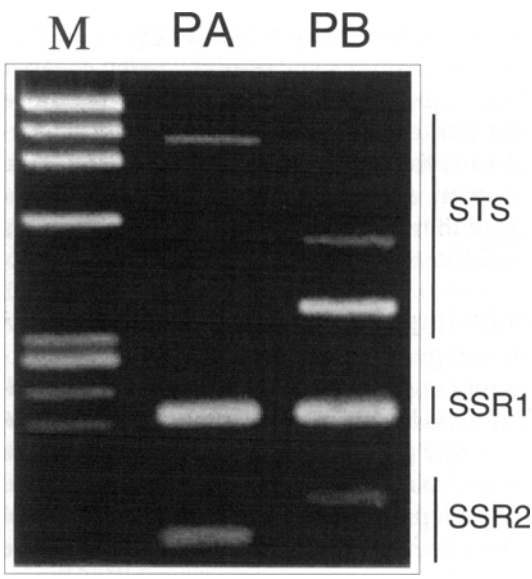

Fig. 1. Multiplex PCR amplification of DNA of two parental maize lines that gave rise to the BC population. (PA = donor line Ac7643, PB = recurrent line CML247). The three PCR primer pairs used were for an STS (umc67) and two SSRs (1: phi071; 2: MAG.IA01). The molecular weight marker (M) was ϕx174-*Hae* III.

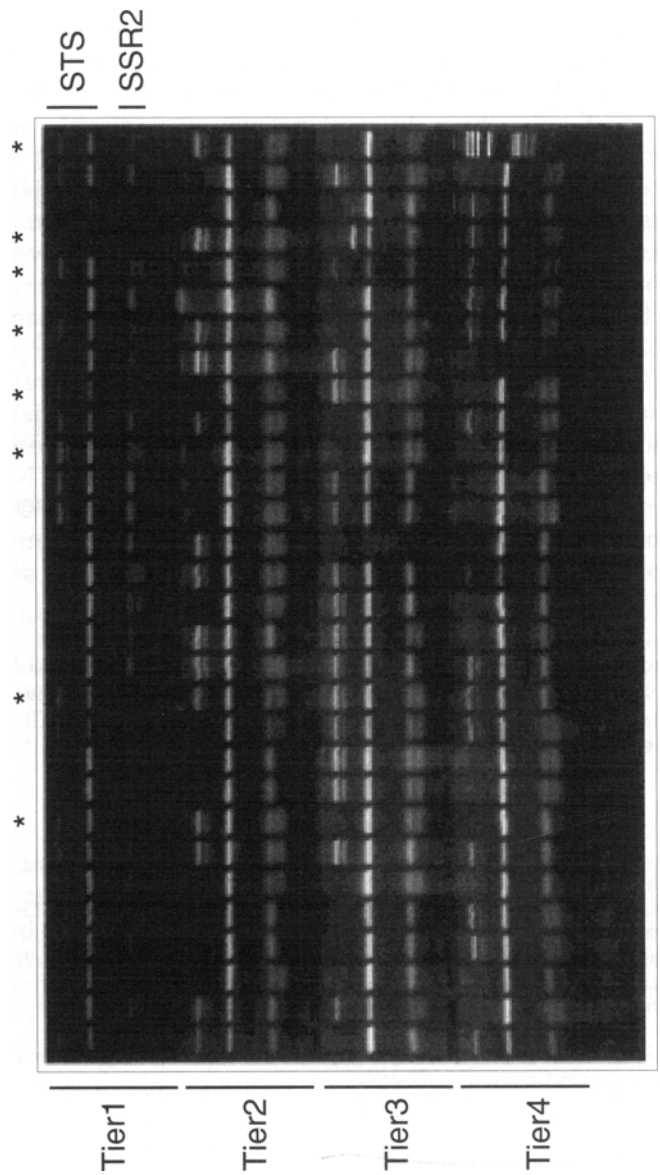

Fig. 2. Screening of BC_2F_1 plants. Two PCR-based markers pre-select individuals heterozygous at the two loci, STS (umc67) and SSR2 (MAG.IA01), with the gel lanes marked by an asterisk (*) for the first tier. Half the Metaphor agarose gel is shown, with 119 individuals arranged in four tiers, the last lane being ϕx174-*Hae III*.

combinability for multiplex amplification. These primers amplified an STS (umc67) on chromosome 1, an SSR (MAG.IA01, DuPont) on chromosome 2, and an SSR (phi071, Pioneer) on chromosome 10. Multiplex amplification for the three different primers was possible, and the amplification products with DNA from the donor (Ac7643) and the recurrent elite line (CML247) (Ribaut et al., 1996) are presented in Fig. 1. Fragments of DNA amplified with phi071 had to be separated on polyacrylamide gels, while for the other two, polymorphisms were clear enough to resolve them simultaneously on the same MetaPhor agarose gel. Results on Fig. 2 show the screening of a sample of BC_2F_1 plants with STS umc67 and MAG.IA01, after multiplex amplification. The percentage of PCR reaction failures was low, less than 3%, demonstrating the high efficiency of the method. The early harvest of plant material for the preselection step allowed one to achieve the total selection process, using PCR-based and RFLP markers, within a growing plant season. No more than a month was needed for the preselection PCR-based step (300 plants selected out of 2300, two technicians), and about three months were needed for the RFLP screening of the whole genome for final selection (300 to 10 plants, one technician).

Acknowledgments. We wish to thank E. Perotti for advice and helpful discussion, as well as E. Huerta and E. Magaña for their excellent technical work. This work was supported by the Swiss Development Corporation (SDC) and the United Nations Development Program (UNDP)

References

Clarke, B.C., L.B. Moran, and R. Appels. 1989. DNA analyses in wheat breeding. Genome 32:334-339.

Ribaut, J.-M., D.A. Hoisington, J.A. Deutsch, C. Jiang, and D. González-de-León. 1996. Identification of quantitative traits loci under drought conditions in tropical maize. 1. Flowering parameters and the anthesis-silking interval. Theor. Appl. Genet. 92:905-914.

Rogers, H.J., N.A.. Burns, and H.C. Parkes. 1996. Comparison of small-scale methods for the rapid extraction of plant DNA suitable for PCR analysis. Plant Mol. Biol. Reptr. 14:170-183.

Taramino, G., and S. Tingey. 1996. Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39:277-287