Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize

Steve A Quarrie¹⁴, Vesna Lazić-Jančić², Dragan Kovačević², Andy Steed¹ and Sofija Pekić³

¹ John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK
² Maize Research Institute, Slobodana Bajica 1, Zemun Polje, 11080 Belgrade-Zemun, Yugoslavia
³ Faculty of Agriculture, University of Belgrade, PO Box 127, Nemanjina 6, 11081 Belgrade-Zemun, Yugoslavia

Received 10 May 1998; Accepted 26 March 1999

Abstract

The usual method to locate and compare loci regulating quantitative traits (QTLs) requires a segregating population of plants with each one genotyped with molecular markers. However, plants from such segregating populations can also be grouped according to phenotypic expression of a trait and tested for differences in allele frequency between the population bulks: bulk segregant analysis (BSA). The same probes used for making a genetic map (e.g. isozyme, RFLP, RAPD, etc) can be used for BSA. A molecular marker showing polymorphism between the parents of the population and which is closely-linked to a major QTL regulating a particular trait will mainly co-segregate with that QTL, i.e. segregate according to the phenotype if the QTL has a large effect. Thus, if plants are grouped according to expression of the trait and extreme groups tested with that polymorphic marker, the frequency of the two marker alleles present within each of the two bulks should deviate significantly from the ratio of 1:1 expected for most populations. As chromosomal locations of many molecular markers have now been determined in many species, the map location of closely-linked QTLs can therefore be deduced without having to genotype every individual in a segregating population. This has been used successfully with composite populations of maize to locate QTLs associated with yield under severe drought. An inbred line derived from one of the populations selected for higher drought yield has been crossed with a drought-susceptible inbred line to produce a mapping population for QTL analysis of physiological and developmental traits likely to regulate yield under drought. Future work to identify traits having QTLs with flanking markers showing significant allele frequency differences in the BSA studies will indicate those traits likely to be important in determining yield under drought.

Key words: Bulk segregant analysis (BSA), drought resistance, genetic maps, maize, molecular markers, Zea mays (L.).

Introduction

Improving the drought resistance of crop plants is a major objective for breeders in tropical, sub-tropical and warm temperate parts of the world where evaporative demands are very high and rainfall is either insufficient or unpredictable in its timing and quantity. Maize is often the preferred crop in these regions despite having less drought resistance than more traditional cereal crops such as sorghum and millet. Improving the drought resistance of a crop is difficult for a breeder because yield usually has a relatively low heritability even under ideal conditions and an unpredictably variable water supply reduces heritability even further. In consequence, physiologists have been trying to identify specific traits that contribute to increasing yield under drought as these, having a higher heritability than yield, should be easier to select for. Until recently physiologists have had to rely upon phenotypic correlation analysis using genetic stocks differing in expression of a particular trait to test the association between the trait and yield. The recent development of molecular marker technologies has changed all this. A wide array of DNA-based molecular marker systems is now available to enable dense genetic maps to be produced of any crop species. With restriction fragment
length polymorphism (RFLP) and many PCR-based marker methods now developed, it is possible for such maps to be made in only a few months, depending on the total map length. The preparation of a genetic map using progeny (F₁s), doubled haploids and recombinant inbreds are the most usual) from a cross between parents differing in expression of a particular trait will allow the number and location of genes of large effect regulating that trait to be determined. As most traits are determined quantitatively (i.e. several genes contributing in varying proportions to expression of the phenotype), the process is termed quantitative trait locus (QTL) analysis. The topic has been reviewed recently (DNA marker aspects by Paterson et al., 1991; Quarrie, 1996; statistical aspects by Kearsey, 1998; and in relation to plant physiology by Prioul et al., 1997).

This article will describe a modification of QTL analysis, bulk segregant analysis (BSA), that has been shown to work well with genes having major effects and that obviates the need for constructing detailed genetic maps. It is demonstrated how BSA can be used to help identify traits important in determining the drought resistance of maize and how DNA markers identified by BSA can be used to help incorporate such traits into a breeding programme to improve drought resistance.

BSA methodology

Quantitative trait locus analysis is usually associated with a mapping population of plants, each of which has to be (a) genotyped with all the markers selected to cover the genome, and (b) phenotyped for the traits of interest. Genotyping a large mapping population is tedious and relatively costly in consumables. By grouping plants according to either high or low expression of a particular trait and extracting DNA from these two bulks, the process of genotyping the plants is reduced to only two DNA samples to be analysed instead of having to analyse DNA separately from each plant. Two variants of the BSA technique are possible depending on whether these plants are derived from a cross between two parental lines or from a population of plants with diverse genetic backgrounds (e.g. variety mixes or composite populations). The former is illustrated in Fig. 1, and was first described for use in plant genetics by Michelmore et al. who studied disease resistance using bulks screened with random amplified polymorphic DNAs (RAPDs) (Michelmore et al., 1991). They discussed the practicalities of such an approach to identify markers linked to particular genes using either RAPDs or RFLPs, and described the advantages of BSA over other genetic techniques for gene ‘tagging’, such as use of near isogenic lines produced by repeated backcrossing (Michelmore et al., 1991).

For BSA of the trait of interest, parental lines are chosen that differ in their expression and crossed, and, as with QTL analysis, F₂, doubled haploid or recombinant inbred populations are generated which will segregate for the trait. Although Michelmore et al. were successful in identifying markers linked to a resistance gene in lettuce using F₂ plants, for traits controlled quantitatively or by a single recessive gene, doubled haploids or recombinant inbreds will increase the probability of locating markers linked to gene(s) controlling the trait (for example, compare schematic hybridization results for F₂ and Fₙ bulks in Fig. 1b). The population is then phenotyped to identify individual plants or lines having high or low expression of the trait. Two DNA bulks are prepared, one from the ‘high’ individuals and the other from the ‘low’ individuals (for example 10 ‘high’ and 10 ‘low’ from a population of 200 individuals), and analysed for allele frequency with molecular markers. In the work of Michelmore et al., 1991, the number of individuals comprising each bulk varied from 14 to 20 plants. With RAPDs, which usually produce dominant polymorphisms (i.e. a band which is either present or absent), only a few individuals are required in each bulk. The probability of an unlinked locus being polymorphic between two bulks of 10 individuals was calculated to be $2 \times 10^{-6}$ (Michelmore et al., 1991). With dominant markers such as RAPDs, this would apply whether the individuals came from a single segregating population or from pools of genetically diverse individuals, such as variety mixtures (Virk et al., 1996) or composite populations such as maize (see below).

However, when using codominant markers (such as RFLPs) with pools of genetically diverse individuals, where several marker alleles may be present, more individuals would need to be combined to ensure that each allele was represented in the resulting DNA bulks at the same frequency as in the population as a whole. In the work described below with maize, DNA from at least 50 individuals was combined to make each bulk.

Because genomes of all the major crops are now well-characterized with molecular markers, it is usually possible to select probes with RFLPs or PCR-based methods, such as STSs (sequence tagged sites) and SSRs (simple sequence repeats or microsatellites), that are well-dispersed around the genome. Other PCR-based marker systems, such as random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP™), are likely to give polymorphisms specific to a particular mapping population and which cannot be given a chromosomal location without subsequent analysis.

Four markers are illustrated in Fig. 1 located on a hypothetical chromosome containing a gene having a significant effect on the trait of interest. Because recombination is more frequent the further away the marker is from the gene, a marker (polymorphic between parents A and B) that is tightly-linked to the gene (M2 in Fig. 1a) will show highly-significant differences in allele frequency between the two DNA bulks (Fig. 1b), and allele frequencies for the marker within each of the two bulks will
Fig. 1. Schematic representation of bulk segregant analysis (BSA) with two bulks selected from both F2 plants and recombinant inbred lines (F1n) derived from the same F2 plants. M1, M2, M3, and M4 indicate markers located on a hypothetical chromosome containing a gene of interest. (a) Schematic representation of segregation of markers and target gene in ‘high’- and ‘low’-bulks of progeny from the cross A × B. (b) Schematic representation of hybridizations with codominant probes for M1–M4 to DNA from ‘high’- (Bk1) and ‘low’- (Bk2) bulks of both F2s and recombinant inbreds (F1n). The intensity of shading of each band indicates the frequency of occurrence of the allele.
complement each other and will deviate significantly from the expected ratio of 1:1 for an unlinked QTL. Thus, Fig. 1b shows allele B to be absent in bulk 1 and allele A to be absent in bulk 2 in hybridizations at the F₅ generation with marker M2. Other markers in the example of Fig. 1 would show varying proportions, approaching 1:1 on average with large pool sizes, of both alleles within each of the two DNA bulks.

The second variant of BSA applies when using pools of genetically diverse individuals, such as variety mixtures or composite populations of outbreeding species, and differs from the first only in the number of alleles likely to be present at any marker locus. For example, RAPD markers have been used to analyse diverse rice germplasm (Virk et al., 1996) and markers associated with a number of agronomic characters, such as flowering time and panicle length have been identified. In such cases, plants or genotypes could be analysed individually as well as in bulk to measure the frequency with which particular alleles occur.

Bulk segregant analysis offers the molecular biologist opportunities for testing candidate genes for QTL effects, for example R genes for disease resistance. Using a population segregating for disease resistance, by hybridizing DNA from resistant and susceptible bulks with the R gene cDNA or its regulatory sequences as probes, the presence of a significant allele frequency difference for the gene between the two DNA bulks would support the gene’s role in disease resistance. Strategies for validating a candidate gene have recently been discussed (Prioul et al., 1999) in relation to genes for carbohydrate metabolism in maize. As an example, they measured kernel starch content in 46 diverse lines and the presence of an Sac1 restriction site in the Sh2 gene, which codes for a subunit of ADP-glucose pyrophosphorylase. Only eight lines contained the restriction site and if BSA had been used with the ten highest and the ten lowest starch-containing lines, only the high-starch bulk would have shown the fragment corresponding to the Sac1 restriction site (five of the ten lines).

Considerable effort is currently being spent on functional analyses of stress-induced genes and their products, in the expectation of identifying one or other genes having a significant effect on improving stress resistance, so far with limited success. Transformation to modify expression patterns is currently the most favoured technique. However, BSA with cDNAs of stress-induced genes used as probes hybridized to populations selected on the basis of differences in stress responses or resistance could be used as a simpler alternative to transformation (Quarrie et al., 1996).

In the next sections, BSA studies on composite populations of maize, which were analysed before and after selection for yield under drought, and the application of the results to help improve drought resistance in maize, are described.

### BSA for yield in droughted maize composites

Two sets of maize composites derived initially from CIMMYT, Mexico (Tuxpeño Sequia, TS, and Drought Tolerant Population, DTP) were selected on the basis of yield under droughted conditions for several cycles. After eight cycles of selection in TS (C8), yield under severe drought was about 50% greater than in the unselected population (C0) (Bolanños and Edmeades, 1993a). After two rounds of selection of the DTP population on the basis of agronomic characters (C2), followed by three cycles of selection under drought conditions, the DTP population was selfed at C5 and S1 families transferred to Zambia where they were tested under severe drought (M Vidaković, unpublished data). The top 10% of lines for yield were recombined to make C6. The distribution for yield for 223 S1 families in this trial is shown in Fig. 2.

Equal leaf weights of at least 50 plants from each of these four populations (TS C0 and C8, DTP C2 and C6) were combined to provide four bulks which were extracted for DNA using a standard CTAB method (Murray and Thompson, 1980). DNA was restricted with five enzymes, size-separated and blotted onto membranes, essentially as described previously (Sharp et al., 1988). Maize cDNA and gDNA probes were labelled with 32P, hybridized with the membranes (Sharp et al., 1988) and visualized either on X-ray film or with Phosphorimager screens (Molecular Dynamics). Probes for RFLP hybridizations were largely from the California State University (csu) collection with others coming from the University of Missouri (umc) collection and drought-induced cDNA probes from

---

**Fig. 2.** Frequency distribution of yield per plant for 223 S1 families of DTP cycle 5 trialled in Zambia. Darkly hatched individuals were recombined to produce C6. The yields of S1 families, DTP12 and DTP79 are indicated by arrows.
Hybridization results were scored according to the number of restriction enzymes with which the hybridized probe showed clear visual differences in allele frequencies between unselected and selected populations. DNA from 36 individuals from each of the TS C0 and C8 populations was also screened with some multi-copy probes to help assign alleles to each locus.

Figures show that markers showing consistent differences in allele frequency between populations unselected and selected for higher drought yield were widely distributed around the maize genome. QTLs for yield specifically under droughted conditions (hatched areas in Fig. 4) have been identified previously (Agrama and Moussa, 1996; Ribaut et al., 1997b), and several of these QTLs were in regions of the genome showing major differences in allele frequency: in particular around umc11 (chromosome 1), and csu149 and umc126b (chromosome 5). The major effect present on chromosome 6 at csu94 is in a region of the genome reported to contain a QTL for ear length (Austin and Lee, 1996). This is consistent with the findings of Bolanos and Edmeades (1993b) who showed that ear biomass at anthesis in C8 plants was significantly greater than in C0 plants. The location of csu94 on chromosome 6 is also very close to a gene for sucrose phosphate synthase (SPS) and a QTL for adult leaf carbohydrate content (glucose plus fructose) identified by Prioul and co-workers (Prioul et al., 1999; Causse et al., 1995). In the work of Causse et al., SPS structural genes and SPS activity were both shown to be coincident with QTLs for growth in young maize plants (Causse et al., 1995). Furthermore, several regions of the maize genome identified to carry structural genes or QTLs for aspects of carbohydrate metabolism (Prioul et al., 1999; Causse et al., 1995) appear coincident, not only with differences in allele frequency shown in Fig. 4, but also with QTLs for yield under drought stress shown hatched in Fig. 4: around umc11 and hhu504 (chromosome 1) and csu149/csrl37b and umc126b (chromosome 5).

Other regions of the genome showing differences in allele frequency may indicate the location of QTLs for other drought resistance traits identified in the TS C8 population, such as anthesis-silking interval and tassel biomass (Bolanos and Edmeades, 1993b) and root biomass in the upper soil layer, reflected in root pulling resistance (Bolanos et al., 1993). A significant QTL for root pulling resistance has been identified in a maize F2 population (Lebreton et al., 1995) near csu133 on chromosome 2, a marker showing significant differences in allele frequency in our work and in another study with the TS C0 and C8 populations (Ribaut et al., 1995, 1997a).
characterized for their responses to drought stress (Bolaños and Edmeades, 1993a, b; Bolaños et al., 1993). However, little is known so far of the physiological, morphological and developmental responses of the DTP populations to drought stress. To identify some of the traits associated with the better yield under drought of S<sub>1</sub> families used to make the DTP C6 population (Fig. 2), two inbred lines (DTP12 and DTP79) were extracted from the S<sub>1</sub> families, used in Zambia to create the C6 population, by selfing for several further generations. The North American inbred lines B73 and B84 were selected as possible sources of drought susceptibility, with the intention of creating a mapping population for QTL analysis of drought responses by crossing a drought resistant DTP line with a susceptible line. These four lines were compared in rainfed field trials at the Maize Research Institute, Belgrade and Centre for Agricultural and Technological Research, Zaječar, Serbia and in soil column experiments at the John Innes Centre, Norwich to study rooting behaviour.

In the field trials, a dry period during July resulted in signs of leaf rolling in the North American inbreds, but
no leaf rolling in the DTP lines. At the end of July (flowering time in B73 and B84, but before flowering in the DTP lines) in the Belgrade experiment, B73 and B84 reached the same degree of stress according to leaf water potentials (−1.1 MPa), while DTP12 and DTP79 remained at only −0.8 and −0.7 MPa, respectively. Leaf relative water contents sampled at a similar time in the Zaječar experiment were lowest in B73 (85.8%) and highest in DTP79 (93.5%), with B84 and DTP12 very similar (91.9% and 92.3% respectively).

In the soil column experiment, plants of the four inbred lines were grown for seven weeks in 70 cm soil columns in 15 cm diameter tubes (three plants per tube) with a control group watered twice a week and a droughted group given water sparingly from 2 weeks after sowing. Shoot and root biomass were recorded, and the total water given to the plants and the water remaining in the soil columns was determined. From these, the water-use efficiency of each line during the experiment was calculated (Fig. 5). Considering both treatments, DTP79 had the highest water-use efficiency of the four lines, and B73 the lowest. In addition, in the droughted treatment, DTP79 plants had the greatest root biomass (1.28 g plant$^{-1}$) and B73 the smallest (0.62 g plant$^{-1}$).

On the basis of these trials, DTP79 and B73 were selected as the lines differing most in their responses to drought stress. An F$_2$ population from the cross DTP79 × B73 was made and mapping work is in progress to carry out QTL analyses of drought responses. Recent results (Quarrie, Lazić-Janič and Steed, unpublished data) using a skeletal map of 70 RFLP markers have given evidence with QTLCartographer (QTL mapping software developed by Basten et al., 1996) for a significant yield QTL with F$_2$ plants grown under drought conditions (average yields about 30% below normal for the site). The only yield QTL to reach the significance threshold was 2 cM from umc11 on chromosome 1. This location coincides not only with QTLs for drought yield found previously (Agrama and Moussa, 1996; Ribaut et al., 1997a, b) but with a marker showing significant differences in allele frequency between the DTP selected and unselected composites (umc11; Fig. 4). Other mapping studies to characterize the DTP79 × B73 population are in progress to compare QTLs for specific drought-related traits, such as rooting development and anthesis-silking interval, with regions of the maize genome showing differences in allele frequency. In this way, the combination of BSA with QTL analysis will allow us to locate regions of the genome regulating yield under drought as well as identifying the physiological, morphological and developmental traits that are important in determining drought yield.

Conclusions

With the advent of molecular marker technologies, new opportunities are available for plant physiologists to study the relationships between traits and their genetic control. While QTL analysis with mapping populations of plants offers the most precise method to locate genes regulating a particular trait, BSA is a valuable alternative approach that avoids the necessity to genotype every member of a population. The technique also allows probes for candidate genes for a trait to be tested for differences in allele frequency between lines selected to differ specifically in that trait. It is, therefore, a useful alternative to the time-consuming and frequently unreliable approach of plant transformation to identify the function of a gene. This study has shown how BSA coupled with physiological studies can help to identify traits important in determining drought resistance in maize and provide molecular markers for those traits. BSA should be a valuable aid for plant physiologists in the future.

Acknowledgements

We acknowledge the kind provision of hhu probes by Dr M Höfer, Heinrich-Heine-University, Düsseldorf, Germany, and csu and umc probes, provided by the University of Missouri, Columbia, USA. SA Quarrie, V Lazić-Janić and S Pekić are grateful to the Royal Society for Fellowships and financial support for exchange visits. D Kovačević gratefully acknowledges a fellowship from UNESCO. The valuable assistance of Radenko Radošević is gratefully acknowledged. This research was supported by a grant-in-aid from the Biotechnology and Biological Sciences Research Council to the John Innes Centre.

References


