Development of a seed DNA-based genotyping system for marker-assisted selection in maize

Shibin Gao · Carlos Martinez · Debra J. Skinner ·
Alan F. Krivanek · Jonathan H. Crouch · Yunbi Xu

Abstract Leaf collection from the field, labeling and tracking back to the source plants after genotyping are rate limiting steps in leaf DNA-based genotyping. In this study, an optimized genotyping method using endosperm DNA sampled from single maize seeds was developed, which can be used to replace leaf DNA-based genotyping for both genetic studies and breeding applications. A similar approach is likely to be suitable for all plants with relatively large seeds. Part of the endosperm was excised from imbibed maize seeds and DNA extracted in 96-tube plates using individuals from eight F_2 populations and seven inbreds. The quality of the resultant DNA was functionally comparable to DNA extracted from leaf tissue. Extraction from 30 mg of endosperm yields 3–10 µg DNA, which is sufficient for analysis of 200–400 agarose-gel PCR-based markers, with the potential for several million chip-based SNP marker analyses. By comparing endosperm DNA and leaf DNA for individuals from an F_2 population, genotyping errors caused by pericarp contamination and hetero-fertilization were found to average 3.8 and 0.6%, respectively. Endosperm sampling did not affect germination rates under controlled conditions, although under normal field conditions the germination rate, seedling establishment, and growth vigor were significantly lower than that of non-sampled controls for some genotypes. However, careful field management can compensate for these effects. Seed DNA-based genotyping lowered costs by 24.6% compared to leaf DNA-based genotyping due to reduced field plantings and labor costs. A substantial advantage of this approach is that it can be used to select desirable genotypes before planting. As such it provides an opportunity for dramatic improvements in the efficiency and selective gain of breeding systems based on optimum combinations of marker-assisted selection and phenotypic selection within and between generations.

Keywords Seed DNA · Genotyping ·
Marker-assisted selection · Hetero-fertilization ·
Maize

S. Gao · C. Martinez · D. J. Skinner ·
A. F. Krivanek · J. H. Crouch · Y. Xu (✉)
International Maize and Wheat Improvement Center
(CIMMYT), Apartado Postal 6-641, Mexico 06600,
DF, Mexico
e-mail: y.xu@cgiar.org

S. Gao
Maize Research Institute, Sichuan Agricultural
University, Sichuan 625014, China

Introduction

As molecular markers associated with agronomically important traits have increasingly become available in maize, marker-assisted selection (MAS) has been incorporated into breeding programs, particularly in the private sector, to improve the efficiency of

selection (as reviewed for various crops by Xu 2003; Miklas et al. 2006; Ragot and Lee 2007; Dwivedi et al. 2007; William et al. 2007). However, there are several major constraints that hinder wide application of MAS, particularly in public sector plant breeding programs. Although collection of leaf tissue and DNA extraction are often considered one of the most significant rate limiting factors in MAS systems, there have been few attempts in the literature to provide solutions (Xu and Crouch 2008). Leaf DNA-based genotyping requires growing all candidate plants in the field or greenhouse, collecting leaf tissue from the plants, and tracking back to the desirable plants after genotyping. Seed DNA-based genotyping is an important alternative that could reduce costs and dramatically increase the scale of uptake and efficiency of molecular breeding. A non-destructive sampling method, allows germination of the sampled seed, permits selections to be carried out in advance of planting, avoiding the complexity of leaf sampling, saving field space and providing the possibility of working with larger effective populations for complex agronomic traits. Samples can be tracked easily between seed and genotype, by retaining all steps of the process in 96-well formats, thus reducing errors. Working with seeds also allows flexibility to conduct MAS in off-seasons and out-stations where needed. In addition, the seed samples can be handled and stored at room temperature, saving refrigeration space and costs normally associated with the use of leaf tissue. Ground seed or excised sections of seeds can also be shipped between labs and field stations more easily with fewer quarantine-related issues than leaf tissue. Therefore, seed DNA-based genotyping could be outsourced more easily, when in-house genotyping is impossible or less efficient.

Genetic testing and detection of transgenes using DNA extracted from multiple seeds using a destructive protocol has been previously reported (Papazova et al. 2005; van Deynze and Stoffel 2006; http://gmo-crl.jrc.it/statusofdoss.htm). Efforts have also been made to extract DNA from single seeds of various crops such as rice (Chunwongse et al. 1993), wheat and soybean (Hee et al. 1998), maize (Sangtong et al. 2001), barley (von Post et al. 2003), and peanut (Chenault et al. 2007). These methods are appropriate for genetic analysis in relatively small-scale screens. Seed tissues that can be used for DNA extraction include endosperm (most monocots including maize) and cotyledons (most dicots including soybean). In order to make seed DNA-based genotyping

applicable in large-scale MAS, it should meet three challenges. First, the whole process including DNA extraction and genotyping should be easily automated for high-throughput applications. This particularly requires improvement in the efficiency of endosperm/cotyledon sampling. Next, a sufficient quantity of DNA must be obtained from individual seeds so as not to require the pooling of multiple seeds while the nature and extent of the sampling should not significantly influence seed germination or seedling establishment. This is particularly important for MAS applications, such as simultaneous foreground and background selection with molecular markers where a large number of markers have to be analyzed. Finally, the DNA quality should be adequate for genotyping with any type of genetic marker without modification of the protocol and use low cost extraction buffers and other component solutions.

Maize is a good candidate for seed (endosperm) DNA-based genotyping due to its large seed size relative to other crop plants, although there are differences between various types of maize in terms of seed size, texture and shape which may influence the utilization of this approach across diverse maize germplasm. However, concerns associated with DNA quality and quantity, potential genotyping errors that might be caused by pericarp contamination, triploid endosperm, and hetero-fertilization (Sprague 1929), plus the germination and seedling establishment capacity of sampled seeds, have hitherto not been investigated. Therefore, the main objectives of this study were to (1) develop an efficient and generally applicable method to sample endosperm from single maize seeds, (2) develop suitable DNA extraction protocols to obtain high quality and quantity DNA from endosperm samples, (3) establish a system that can be easily scaled-up and automated for all steps including DNA extraction, marker analysis and sample tracking, (4) investigate potential errors resulting from the sampling of different seed tissues (endosperm, pericarp and embryo), and (5) evaluate the germination capability of sampled seed under different conditions and the seedling establishment rates under field conditions. The study reported here used elite lines from CIMMYT breeding programs that are important for traits such as drought tolerance and protein quality. The resulting seed DNA-based genotyping system should make MAS more attractive and applicable for maize as well as other crops with relatively large seeds.

Materials and methods

Plant materials

A wide range of genotypes were used during development of the grinding and extraction methods. However, results for sampling, grinding, DNA extraction, and germination and seedling establishment tests are described for 15 representative maize genotypes that differed in seed texture (dent vs. flint), size (14–36 g for 100 seeds), color (yellow vs. white) and shape (Table 1). These included both inbred lines and segregating populations. Pedigrees of four lines used in F_2 populations were abbreviated as follows, HGA = Resistant synthetic HGA-61-2-2-1, LPS = La Posta Seq C7-F64-2-6-2-1-B-B, DTP = DTPWC9-F104-5-4-1-1-B-B, and CLQ = (CL04368 × CML264Q)-B-2-4-1-1-2. All seeds were obtained from the CIMMYT Genebank or Maize Program with storage time from half a year to 1 year.

Sampling and grinding of seed endosperm

Seeds were soaked in water, either bulked in a container or arrayed in 48-well plates with adequate well size to hold large seeds (Falcon, NJ, USA or Nunc, CA, USA), and were soaked for approximately

24 h. Wet seeds were transferred to dry tissue for removal of surface water and single seeds were placed embryo side up on a plastic board. A small piece was cut from the endosperm end using a scalpel, carefully avoiding damage to the embryo, and limiting the amount of removed endosperm to 20–60 mg. To aid grinding, this piece was cut again into two to four smaller pieces, and immediately transferred into individual 1.1 ml tubes in a 96-tube plate (12 rows each with eight linked tubes, Neptune, CA, USA). Two 48-well plates together stored the sampled seeds that corresponded to one 96-tube plate for the endosperm samples and resulting DNA. The sampled seeds were allowed to air dry for at least 2 days at room temperature (RT, about 23°C), and the plates were covered and stored under two different temperatures (RT or 4°C), depending on trials, in the dark for subsequent screening and planting.

After the cut pieces (endosperm samples) were dried at RT for at least 1 day, two 4 mm diameter steel balls were put into each tube, the tubes were sealed with caps, and before grinding the whole plate was stored at −20°C for at least half an hour to reduce the risk of overheating during grinding. Two plates of endosperm pieces were ground simultaneously in a QIAGEN Tissuelyser (Retsch, Haan,

Table 1 Maize materials used in DNA extraction, genotyping and germination tests

Plant material	Grain characteristics	100-Grain weight (g)	Storage time (year)
Inbred lines			
CML451	Yellow and flint	31.45	0.5
CML491	White and flint	16.89	0.5
CML494	White and dent	19.33	0.5
CML498	White and flint	22.66	0.5
CML502	White and flint	14.10	0.5
CL02450	Yellow and dent	23.21	0.5
CML454	Yellow and flint	24.55	1
F2 populations			
CML492 × CML494	White and dent	26.62	1
CML460 × CML461	Yellow and dent	28.63	0.5
CML312 × CML451	Yellow and flint	29.82	0.5
CML 418 × CML 312	White and flint	36.10	0.5
HGA × CML491	White and flint	25.49	0.5
CML491 × LPS	White and dent	26.19	0.5
CML491 × DTP	White and dent	31.10	0.5
CLQ × CML492	White and dent	31.68	0.5

Germany) at 30 strokes per second for 3–5 min depending on seed texture. This step was repeated as required to ensure finely ground powder. In preparation for DNA extraction, the caps of the tubes were tapped gently and the tubes shaken to ensure that the powder collected in the bottom of the tube before tipping out the steel balls.

DNA extraction

A set of five extraction buffers for endosperm DNA extraction was developed on the basis of a leaf-tissue based DNA extraction protocol established at CIMMYT (CIMMYT Applied Molecular Genetics Laboratory 2003) and endosperm extraction protocols (Larkins Lab, http://ag.arizona.edu/research/larkinslab/) (Table 2). For these five buffers, the extraction protocol was performed as follows, taking care to close the lids of the tubes securely at each step, particularly after the addition of chloroform, to avoid cross-contamination, and using multichannel pipettes for maximum efficiency. A volume of 400 µl of extraction buffer, suitable for 20–50 mg of endosperm powder, was added to the powder in 96-tube plates. The buffer and powder were homogenized very well by inverting the plate strongly several times, and then mixed continuously in a rotary mixer at RT for 30 min. Next, 400 µl of phenol: chloroform (1:1) was added to the tubes and mixed gently with continuous inversion in a rotary mixer at RT for 10 min. Plates were centrifuged at 3,500 rpm at RT for 10 min to generate an aqueous phase and an organic phase. Approximately 300 µl of the aqueous phase was transferred into a new plate and 0.5 volume ice-cold 100% isopropanol (2-propanol) was added for precipitation. The solution was then mixed very gently and incubated at −20°C for at least 1 h to precipitate the nucleic acids. Plates were centrifuged

at 3,500 rpm at RT for 10 min, and the resulting DNA pellet was washed twice with 400 µl of 70% ethanol. The pellet was dried at RT and resuspended in 100 µl of TE (10 mM Tris—8.0, 1 mM EDTA pH 8.0) or double-distilled water. DNA quality and quantity were tested by electrophoresis in 1.0% agarose gels, and were also measured with a spectrophotometer (Nanodrop, ND-1000, Wilmington, DE, USA) for A260:280 and A260:230 nm ratios. Two replicate readings were measured for each sample.

In addition, another quick extraction method, 'NaOH–Tris method', modified slightly from von Post et al. (2003), was also attempted for endosperm powder. The protocol was carried out as follows: 200 µl of 0.15 M NaOH was mixed with approximately 30 mg endosperm powder, and incubated at 55°C for 10 min or heated twice in a 700 W microwave oven for 1 min at 10% power. A total of 750 µl containing 0.03 M Tris–HCl, pH 8.0, and 1 mM EDTA pH 8.0 was added and the samples left to settle at 4°C overnight. Samples were centrifuged at 3,500 rpm at RT for 10 min and 800 µl of supernatant transferred into a storage plate. It was found that mixing of the powder in the initial NaOH solution required extensive vortexing or other agitation methods, to ensure efficient DNA extraction.

Leaf DNA used as controls in each comparative experiment was extracted using a DNA isolation protocol developed for leaf tissue at CIMMYT (CIMMYT Applied Molecular Genetics Laboratory 2003). Briefly, ground lyophilized leaf tissue in 1.5 ml tubes was incubated with 400 µl of buffer containing 100 mM Tris pH 7.5, 700 mM NaCl, 50 mM EDTA pH 8.0, 1% CTAB and 140 mM β-mercaptoethanol at 65°C for 60 min. And then 400 µl of 24:1 chloroform:octanol was added and mixed for 10 min. Following centrifugation at

Table 2 Five conventional extraction buffers used for endosperm DNA extraction

Component	CTAB	SDS	Sarcosyl	Sarcosyl + CTAB	Sarcosyl + SDS
Tris–HCL PH 7.5 (mM)	100	100	100	90	90
NaCl (mM)	700	700	700	630	630
EDTA PH 8.0 (mM)	50	50	50	45	45
CTAB (%)	1	–	–	0.9	–
Sarcosyl (%)	–	1	–	1	1
SDS (%)	–	–	1	–	0.9

3,500 rpm for 10 min, the supernatant was transferred to fresh tubes and the DNA precipitated with 70% ethanol. DNA was resuspended in 100 µl of TE pH 8.0.

PCR amplification and SSR genotyping

A volume of 15 µl PCR reactions contained 1× Taq buffer (20 mM Tris–HCl, pH 8.4, 50 mM, KCl), 2.5 mM MgCl2, 150 µM of each dNTP, 1U Taq enzyme, 0.25 µM of forward and reverse primers and approximately 50 ng of genomic DNA. To improve specificity and efficiency of amplification, touchdown temperature cycles were used: one cycle of step 1 (94°C for 2 min), seven cycles of step 2 (94°C for 1 min, 60°C for 1 min with decreasing 1°C per cycle, 72°C for 1 min), 35 cycles of step 3 (94°C for 1 min, 57°C for 1 min, 72°C for 1 min), one cycle of step 4 (72°C for 5 min), 10°C "forever". SSR markers used for PCR and genotyping were selected from Maize Genetics and Genomics Database (http://www.maizegdb.org). Amplification products were run on 8% polyacrylamide gels (29 acrylamide: 1 bisacrylamide) or 3% agarose gels.

Germination test

As a simple test for germination under laboratory conditions, 1,800 seeds each of two segregating populations (CML460 × CML461) F_2 and (CML312 × CML451) F_2 were used to study the effect of soaking times (24, 36 and 48 h) and temperatures (RT and 4°C), and storage times (0, 15, 30 days) and temperatures (RT and 4°C) after endosperm excision. Each treatment had 100 seeds with 100 unsoaked normal seeds as a control. An average of 41.3 mg endosperm was cut from each single seed. A total of 100 sampled seeds for each treatment were wrapped on wet filter paper in plastic plates at RT and watered each day for approximately 7 days. The germination rate was assessed when the shoots had emerged to approximately 2 cm in length.

Under greenhouse conditions, 260 sampled seeds from (CML492 × CML494) F_2 population were planted in individual pots in sterile soil. When the fourth leaf emerged, the germination rate was assessed and compared to the corresponding germination rate of 300 control seeds. In addition, leaf tissue was harvested from the plants which grew from

sampled seed, for DNA extraction to compare the results from leaf and endosperm DNA by SSR markers.

For field germination and seedling vigor tests, sampled seeds from seven genotypes, including two inbred lines (CML454 and CML494) and five F_2 populations derived from the crosses CML418 × CML312, HGA × CML491, CML491 × LPS, CML491 × DTP, and CLQ × CML492, were used. A total of 150 seeds from each genotype were soaked and differing proportions of the endosperm excised. The weight of the seed and corresponding piece of excised endosperm were recorded individually. Then sampled seeds were classified into small-sampled group (SS) and large-sampled group (LS) according to sample weight as a proportion of seed weight, and stored at RT for 3 weeks prior to planting. The field experiment was designed as a split plot design with three replications of 25 seeds each. The seven genotypes were assigned to the main plots, and three treatments (Control, SS, and LS) were assigned to the subplots. The seeds were processed with normal field practices, including fungicide treatment before planting. Hand sowing of one seed per 20 cm in 4.8 m length rows and base fertilization and appropriate watering were performed as per normal field management. Seeds were planted on July 6, 2007 at El Batan field station, Texcoco, Mexico. Germination rate and seedling survival rate (seedling establishment) were recorded at first visible leaf and 12-leaf stages, respectively. In addition, the growth vigor of seedlings was measured by the Normalized Difference Vegetative Index (NDVI) (Teal et al. 2006) using a Greenseeker Hand Held optical sensor (NTech, Model 505, Ukiah, CA) on August 15, 2007, from which an assessment of plant growth and biomass was estimated.

Data analysis

All data collected in this experiment including DNA quality (A260:A280, A260:A230 nm ratios) and concentration, as well as seed germination rate, number of survived seedlings, and NDVI were analyzed using the General Linear Model Procedure in statistical software SPSS (Version 11.0). Analysis of Variance was first used to evaluate significant effects of each factor in the model, and then Least Significant Differences test was used for multiple comparisons between means of different levels for each factor.

Results

In this study, a relatively large-scale seed DNA-based genotyping system was developed, as shown in Fig. 1, which includes seed soaking, non-destructive sampling and grinding, DNA extraction, PCR-based genotyping, plus data and sample tracking for seed selection. This system can be easily scaled-up through automation of several steps in the process, and modified for different crops and labs based on the requirements and the availability of facilities.

Development of an effective endosperm sampling method

A simple method of endosperm sampling and grinding was developed (Fig. 1). While soft dent kernels could be sampled easily with a razor blade or scalpel, cutting dry flint and small kernels was more difficult and often led to crumbling of the excised endosperm, reducing sampling efficiency. For this reason, seeds were first soaked in water for 24 h, which softened the endosperm sufficiently for effective sectioning. The soaking time depended mainly on the texture, size and physical condition of the seeds. Those seeds with a dent texture and small size were soaked for less time (10 h) and those with flint texture and large size were soaked for more time (24–36 h). Grinding of the excised pieces of endosperm using steel balls in a mechanical shaker was found to be most efficient in 96-tube plates (1.1 ml tube volume), but also worked well in individual 2 ml tubes but with much lower throughput efficiency. The fineness of the powder was an important factor in DNA quality, and it was found that dent seed was generally ground

Fig. 1 Flowchart of large-scale seed DNA-based genotyping system

more finely than flint seed. The 96-tube format reduced the time required for grinding and allowed the subsequent DNA extraction to be performed directly in these tubes using multi-channel pipettes which improved efficiency and reduced errors compared with extraction in single tubes as there was no need to label the tubes or to work with samples individually.

Comparison of different DNA extraction buffers

In order to show that DNA could be successfully extracted from endosperm powder, three DNA extraction methods were compared using seed from five freshly harvested inbred lines, with leaf DNA extraction as a control: the NaOH–Tris method (von Post et al. 2003), a plant DNA extraction kit (Nucleospin) and a method using CTAB buffer. The results showed that DNA extraction from endosperm was achievable, and all methods led to successful amplification of SSR markers despite differences in DNA quality and quantity (Fig. 2).

To further investigate the relative efficiency of different extraction buffers, a base buffer was used with five different combinations of the key extraction ingredients: CTAB; SDS; Sarcosyl; Sarcosyl + CTAB and Sarcosyl + SDS. DNA was extracted from 30 mg of pooled endosperm powder for two genotypes and the DNA quality and quantity were evaluated by agarose gel electrophoresis (Fig. 3) and with a UV spectrophotometer (Table 3). In addition, NaOH–Tris buffer was also used to compare with other buffers (Figs. 2 and 3). As shown by the electrophoresis results, the molecular weights of the endosperm DNA were the same as the leaf DNA, the genomic DNA band showed no obvious degradation (normal size is approximately 25 kb) and there was no obvious differences in quality between DNA samples extracted using different buffers. When DNA samples were obtained from NaOH–Tris buffer, however, no DNA band was observed at the expected position (Figs. 2 and 3), which will be discussed later. Spectrometry gave an indication of DNA concentration and the absorbance ratios at A260:A230 and A260:A280 were used to assess quality (Table 3). For CTAB-extracted leaf DNA and endosperm DNA, the absorbance ratios are usually 1.5 and 2.0, respectively, while the A260:A280 ratio was similar, the A260:A230 ratio was consistently
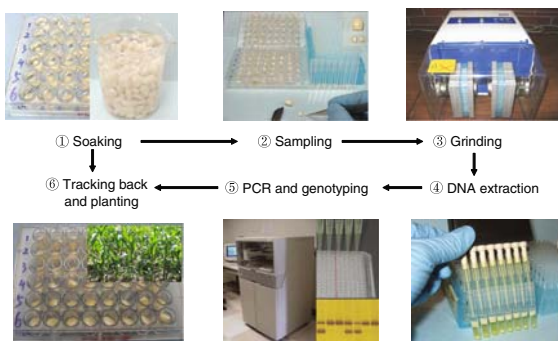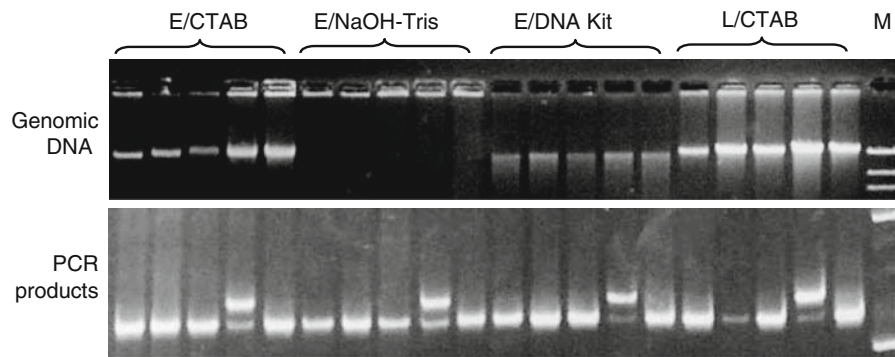
Fig. 2 Comparison of DNA samples extracted using 30 mg endosperm powder (E) and three DNA extraction buffers (CTAB buffer; NaOH–Tris buffer; DNA kit-Nucleospin) compared with leaf DNA extraction (L) (see Materials and Methods for the protocol). Above: 10% of total genomic DNA was used to check DNA quality. Below: PCR products amplified with SSR primer umc1066. Five inbreds were used in the following order from left to right: CML502, CML491, CML498, CML451, CL02450. M: λ/HindIII for genomic DNA, φ X174/HaeIII for PCR

lower ranging from 0.3 to 1.5. Because the ratio of A260:A230 is related to the presence of RNA and other molecular impurities, it was considered an important quality index for endosperm DNA. The low A260:A230 ratio is likely due to contamination with carbohydrates. The results from the two genotypes, as well as other tested genotypes (data not shown), indicate that a specific combination of buffer components could give better quality and quantity of DNA. For CML491, Sarcosyl + SDS performed best, giving the highest A260:A230 ratios, and for CL02450, Sarcosyl + CTAB performed best. Typical DNA yield from 30 mg endosperm powder varied from 3 to 10 µg, depending on the genotype, seed storage time and fineness of the powder. Dent genotypes (CL02450) gave slightly higher DNA amounts than flint (CML491), likely due to the better grinding of the dent genotypes.

Among the methods with different DNA extraction buffers, the most economical and simple extraction method is NaOH followed by neutralization with TE. Figures 2 and 3 showed that the quality of the DNA extracted by this method was lower than that extracted by routine methods. The DNA was retained in the loading well during electrophoresis, probably due to the presence of starch or proteins that had not been effectively removed during the extraction process. However, these DNA extractions still gave reasonable PCR results (as shown in Fig. 2) although DNA quality is lower than that obtained using conventional buffers. As DNA samples extracted from NaOH buffer usually degrade quickly after

extraction, it is suggested that NaOH buffer should be used when genotyping will be carried out immediately following DNA extraction.

Sampling factors influencing DNA extraction

In practice, excised pieces of seed would not be weighed before grinding, and therefore it was necessary to determine the minimum amount of endosperm tissue required for successful DNA extraction. DNA from randomly excised and weighed endosperm pieces was extracted by Sarcosyl + SDS, and then compared with extraction from weighed amounts of powder. The results showed that sufficient DNA could be extracted from as little as 8 mg of endosperm tissue. While there is slight loss of sample due to powder sticking to the steel ball, the amount of DNA extracted was similar to weighed powder (Fig. 4). This means that it should be possible to use a small piece of endosperm for DNA extraction, minimizing the influence of sampling on the germination of seed.

In addition, fresh seed stored only for a short time gave higher DNA yields and quality than old seed. Good quality DNA was acquired from all genotypes stored for less than 2 years at RT, which should be sufficient for meeting the requirements of all practical MAS applications. Some genotypes could be stored up to 6 years in the Genebank without affecting DNA quality, while others did not give good quality DNA when stored for a long time (data not shown).
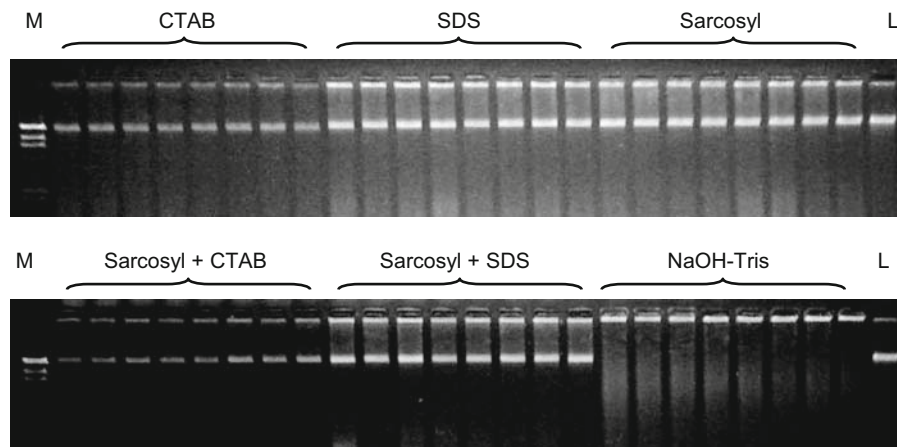
Fig. 3 Quality comparison of the DNA samples extracted from endosperm of maize inbred CML 491 using six different DNA extraction buffers (eight samples each) with leaf DNA (L) as control

Table 3 Comparison of quality and quantity of DNA samples extracted with different buffers

Plant material[a]	Extraction buffer	Concentration (ng/µl)[b]	Absorbance ratios[b]	
			A260:A230	A260:A280
CML491	CTAB	$29.88^C \pm 0.83$	$0.97^D \pm 0.02$	$1.93^{ns} \pm 0.65$
	SDS	$32.82^{BC} \pm 1.14$	$1.10^C \pm 0.32$	$1.88^{ns} \pm 0.55$
	Sarcosyl	$32.57^{BC} \pm 0.49$	$1.04^{CD} \pm 0.22$	$1.94^{ns} \pm 0.01$
	Sarcosyl + CTAB	$34.22^B \pm 0.54$	$1.20^B \pm 0.02$	$1.92^{ns} \pm 0.34$
	Sarcosyl + SDS	$37.74^A \pm 0.62$	$1.49^A \pm 0.02$	$1.93^{ns} \pm 0.17$
CL02450	CTAB	$76.12^b \pm 2.53$	$1.08^B \pm 0.03$	$2.06^C \pm 0.01$
	SDS	$76.60^b \pm 3.82$	$1.07^B \pm 0.34$	$2.12^B \pm 0.13$
	Sarcosyl	$89.30^a \pm 5.82$	$0.29^C \pm 0.02$	$2.24^A \pm 0.02$
	Sarcosyl + CTAB	$90.80^a \pm 4.84$	$1.32^A \pm 0.36$	$2.04^C \pm 0.06$
	Sarcosyl + SDS	$85.50^a \pm 2.22$	$1.17^B \pm 0.32$	$2.07^C \pm 0.01$

[a] Averages for CML491 and CL02450 were derived from 8 and 5 samples, respectively

[b] Multiple comparisons are tested at $P = 0.05$ (shown with lower case letters) and $P = 0.01$ (shown with capital letters) levels, respectively; the numbers followed by the same letter are not significantly different. Extracted DNA was dissolved in total of 100 µl TE

DNA quality and quantity confirmed by PCR amplification

To confirm that the quality of extracted endosperm DNA was suitable for amplification of PCR-based SSR markers, a set of 24 randomly selected SSR primers were screened across endosperm DNA extracted by different methods (Fig. 5). The results showed that the endosperm DNA extracted with sarcosyl, CTAB and the combination of the two worked as well as leaf DNA. DNA extracted using SDS-based buffers showed similar results (data not shown). Using traditional extraction buffers to extract

DNA from 30 mg endosperm yielded 100 µl (approximately 100 ng/µl) of high-quality DNA solution, which can be used for 200 PCR reactions (15 µl). Meanwhile, endosperm DNA from 22 extra inbred lines selected randomly from the CIMMYT Genebank, extracted with a combination of SDS and sarcosyl buffers, was amplified successfully by two SSR markers (Fig. 6), which suggested that there was no genotype limitation for endosperm DNA extraction.

In addition, different amounts of the DNA solution (2, 4, and 6 µl of total 800 µl), extracted by NaOH–Tris from 30 mg endosperm powder were tested in a
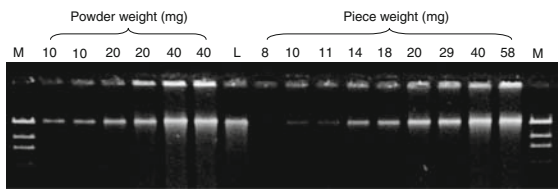
Fig. 4 Quality comparison of DNA samples extracted from different amounts of endosperm powder and pieces, with leaf DNA (L) as control. 10% of total DNA was loaded. M: λ/ *Hind*III
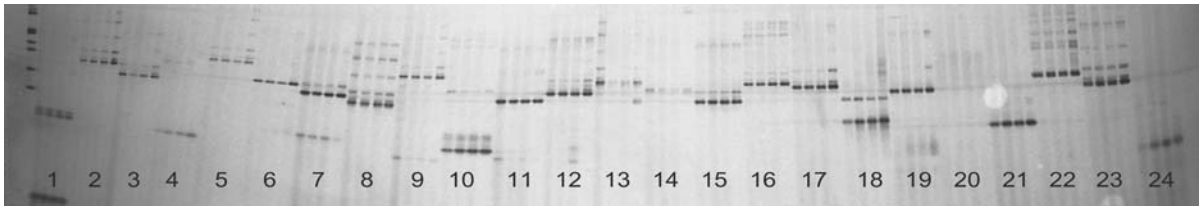
form the embryo and the central cell to form endosperm or, possibly, when the egg and polar nuclei have different genetic constitutions and fuse with identical sperm. This phenomenon could lead to incorrect inference of the embryo genotype from the endosperm genotype (Sprague 1932).

If the leaf genotype is homozygous, and the endosperm genotype is the same, this suggests that there has been no pericarp contamination or hetero-fertilization. When the leaf genotype is homozygous

Fig. 5 PCR amplification of DNA samples extracted from endosperm of maize inbred CML491 using three different extraction buffers (from left to right: sarcosyl, sarcosyl + CTAB, CTAB, plus leaf DNA as control in the end) and 24 randomly selected SSR primers: 1—umc1561, 2—bnlg1812, 3—bmc1792, 4—bnlg1811, 5—bnlg1867, 6—

umc1492, 7—umc1658, 8—umc1562, 9—bnlg2235, 10—umc1505, 11—umc1650, 12—umc1572, 13—bnlg2244, 14—bnlg2204, 15—umc1551, 16—bnlg1839, 17—bnlg2305, 18—umc1494, 19—umc1644, 20—umc1706, 21—umc1497, 22—bnlg1908, 23—bnlg1879, 24—umc1723). Amplification products were run on 8% polyacrylamide gels

15 µl PCR reaction to determine an optimal amount of DNA template for PCR amplification. The results showed that 2 µl of DNA solution performs better than larger amounts, likely due to keeping the concentration of PCR inhibitors present in the solution, such as EDTA and cellular debris, at a lower concentration. A total of 800 µl of DNA solution was extracted from 30 mg endosperm powder using NaOH–Tris buffer, which is sufficient for 400 PCR reactions (each 15 µl).

Evaluation of potential genotyping errors caused by pericarp contamination, triploid endosperm, and hetero-fertilization

Retention of the maternal F_1 heterozygous tissue of the pericarp in the cut endosperm pieces could lead to errors in genotyping. In addition, the different maternal and paternal proportions (2:1) of the 3n endosperm DNA complement could cause inaccuracy in PCR amplification and thus the genotyping. Finally this method assumes that the endosperm is of the same genotype as the embryo. However, the phenomenon of hetero-fertilization might occur, when two different male gametes fertilize the egg cell to

but the endosperm genotype is heterozygous, it is possible that pericarp contamination or hetero-fertilization has occurred, which can be further distinguished using extra polymorphic markers. Finally, if the leaf genotype is heterozygous but the endosperm genotype is homozygous, it is likely that hetero-fertilization has occurred.

In order to investigate these concerns, both F_2 seeds and the subsequent germinated plants (from a segregating population derived from a cross of CML492 × CML494) were analyzed using five polymorphic SSR markers (results from marker umc1066 shown in Fig. 7). A total of 173 pair-wise comparisons between endosperm genotype and corresponding leaf genotype were performed, and samples for which both endosperm and leaf were heterozygous were excluded from further analysis. Two markers each detected one case where the endosperm genotype was homozygous and the leaf genotype was heterozygous, which was interpreted as presence of hetero-fertilization. This led to an observed hetero-fertilization rate of 0.6%. There were several cases where the leaf genotype was homozygous and the endosperm genotype was heterozygous. These cases were assumed to be due to
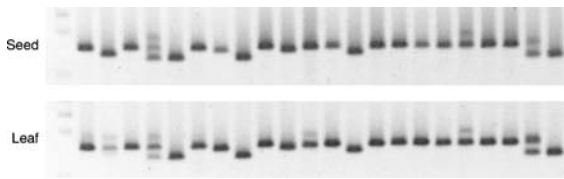
Fig. 6 PCR amplification of DNA samples extracted from endosperm (above) with SDS extraction buffer and leaf (below) with SDS extraction buffer using 22 randomly selected maize inbred lines and SSR marker umc2217

contamination by heterozygous pericarp, as the rate was higher than the hetero-fertilization rates seen above (Table 4). The results show that different primers have different levels of sensitivity to pericarp DNA contamination. For example, the endosperm genotypes revealed by two markers (umc1066 and bnlg1018) were the same as those of corresponding leaf DNA for all test individuals. In contrast, primer bnlg1811 appears to be very sensitive to pericarp DNA contamination as 18 of 107 pair-wise comparisons exhibited a heterozygous genotype in the endosperm but a homozygous CML492 genotype in leaf. This indicated that for some markers pericarp DNA was amplified, and thus it may be necessary to remove the pericarp completely when using such markers. In addition, no obvious PCR amplification differences between alleles were observed when the samples were heterozygous, despite the 2n:1n ratio of parental genomes in the endosperm (Fig. 7).

Germination and seedling establishment capacity of sampled seed

Presoaking the seeds (which may lead to initiation of germination processes), removing a portion of the endosperm, and re-drying the seed prior to planting could all be expected to affect seed germination and seedling establishment. Therefore, it is important to test whether the sampled seeds are significantly different from the controls in these regards.

The result of germination tests under laboratory conditions showed that the control germination rate (98% for both genotypes tested), and the average germination rates across all sampled seeds (96.83 and 96.22% for the two genotypes), were not significantly different from each other. When treatment types were compared statistically, there were no differences among the three soaking times (24, 36 and 48 h),

the two soaking temperatures (4°C and RT), or the three storage times (0, 15, and 30 days after sampling) for both genotypes (Table 5). The only exception to this is that the storage time of 15 days showed a significant decrease in germination rate at $P = 0.05$ as compared with 0 and 30 days of storage in F_2 seeds of cross CML312 × CML451. Overall, the results suggested that soaking treatment could be used for improving the sampling efficiency and that the germination rate was not very sensitive to soaking time which allows flexibility during endosperm sampling. However, it is better to decrease the soaking time and store at low temperature to minimize the potential influence of endosperm sampling on seed germination.

When the germination rate was tested in pots under greenhouse conditions using a segregating population (CML492 × CML494, F_2), seeds with sampled endosperm showed a 94% ($n = 264$) germination rate, as compared to a 95% ($n = 100$) germination rate of control seeds. The result indicates that seed germination under controlled conditions, with no water or temperature stress, is not unduly affected by the sampling process or reduced endosperm.

As an initial test of the applicability of this method in breeding programs, germination rates, and seedling establishment and vigor were tested under field conditions using five F_2 populations and two inbred lines. The experiment included two cutting treatments, one with a larger proportion of endosperm sampled (LS) and the other with a smaller proportion of endosperm sampled (SS). The average amount of sampled endosperm for the five F_2 populations was 43.2 mg for LS and 23.2 mg for SS (Table 6). The average germination rate and seedling establishment rate of sampled seeds for five F_2 segregating populations was 89.9 and 86.9% for LS, and 87.7 and 85.1% for SS, respectively, which was slightly lower than, but not statistically significantly different from, corresponding control seeds (92.8 and 90.9%). Across F_2 genotypes, only one cutting treatment on one F_2 row showed a statistically significant lower germination rate and seedling survival rate compared to the control. Experiments with two inbreds indicated significant differences in germination rates and seedling survival rates between the seeds with endosperm partially sampled versus the control seeds for CML494, but no significant difference for
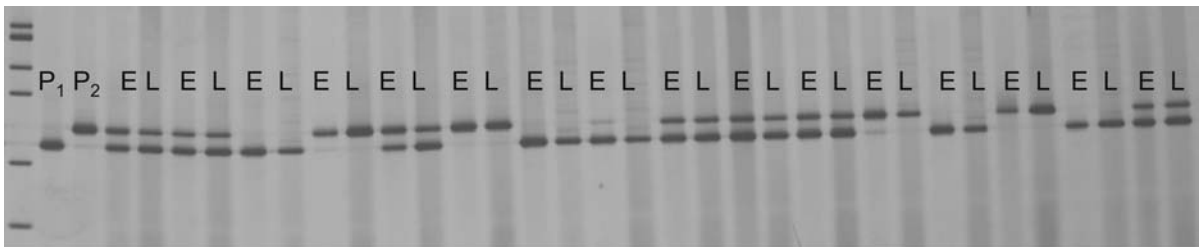
Fig. 7 An example of genotyping using one of the five SSR markers tested, umc1066, and DNA samples extracted from endosperm (E) and leaf (L) tissues from 16 representative F$_2$ individuals between CML492 (P$_1$) and CML494 (P$_2$). The left lane is size marker φ X174/*Hae*III

Table 4 Evaluation of pericarp contamination using a segregating population (CML492 × CML494) F$_2$ using five SSR markers

Comparison result	Genotype (E vs. L)[a]	umc1066	bnlg1018	umc1071	bnlg1867	bnlg1811	Average
Identical	CML492 vs. CML492	48	51	66	38	50	
	CML494 vs. CML494	35	34	29	39	39	
Different	HET vs. HOM	0	0	1	1	18	
Total pairs of comparison		83	85	96	78	107	
Pericarp contamination (%)		0.0	0.0	1.0	1.3	16.8	3.8

[a] Pericarp contamination is determined where endosperm genotype is heterozygous (HET) while the corresponding leaf genotype is homozygous (HOM)

Table 5 Germination rates under lab conditions for sampled seeds from two F$_2$ populations[a]

Soaking before sampling			Storage after sampling		
Treatment	Average rate (%)		Treatment	Average rate (%)	
	CML460 × CML461	CML312 × CML451		CML460 × CML461	CML312 × CML451
Time (h)			Time (days)		
24	98.17	96.83	0	96.17	98.50[a]
36	94.17	95.50	15	96.33	94.50[b]
48	96.33	98.17	30	96.17	97.50[ab]
Temperature			Temperature		
RT	95.00	96.78	RT	95.00	96.78
4°C	97.44	96.89	4°C	97.44	96.89

[a] Multiple comparisons are tested at $P = 0.05$ (shown with lower case letters) and $P = 0.01$ (shown with capital letters) levels, respectively; the numbers followed by the same letter are not significantly different. No comparison was significant at $P = 0.01$. RT, room temperature

CML454 (Table 6). This might be because that CML494 had a much smaller seed size compared to the F$_2$ populations.

In general, the results suggest that the remaining nutritional capacity of the endosperm might be sufficient for seedling establishment, particularly for the genotypes with relatively large seeds when 20–56 mg of endosperm was sampled (Table 6). Importantly, there was no significant effect on germination or survival rate between large versus small amount of endosperm sample and this was consistent across all F$_2$ genotypes. Thus, there appears to be much flexibility regarding how much endosperm can be sampled for DNA extraction. In contrast, the germination and seedling survival rates of inbred lines were significantly affected, with CML494 having the worse rate of 57.3 and 52.0%, respectively. In addition, the amount of endosperm sampled also

had a significant effect. This is likely to be due to the smaller seed size of inbred lines leading to less available endosperm for nutrition after sampling.

Significant differences in seedling growth under field conditions were observed associated with different sampling treatments (Fig. 8). This was confirmed with NDVI measurements, which showed significant differences in growth vigor between the different cutting treatments and the control for some genotypes (Table 6). These results suggest that that sampling of the smallest possible amount of tissue from the endosperm, earlier application of nutrition (fertilization) and careful field management are necessary for obtaining the seedling establishment rates for sampled seed that are the same as unsampled controls, particularly for genotypes with relatively small seeds.

Cost and time efficiency of seed DNA-based genotyping as compared with leaf DNA-based genotyping

The major differences between seed DNA-based genotyping and leaf DNA-based genotyping are sample collection, DNA extraction, sample tracking

and field plant management since both approaches share common genotyping procedures. Our comparisons were focused on sample collection, DNA extraction and field plant management. Leaf DNA is normally extracted with a similar throughput protocol that is also performed with 96-tube plates and is more efficient than DNA extraction with individual tubes. The two different genotyping systems were compared under the assumption that leaf tissue needs to be collected from plants in the field, not from seedlings in the greenhouse (Dreher et al. 2003). As some experimental steps in both methods require waiting periods that do not have a labor cost, such as drying leaf tissue in a freeze-drier for 3 days, or soaking seed in water for 24 h, only working time was taken into account. A hypothetical MAS case was used for comparison on the basis of selecting for one trait using two SSR markers, for 1536 samples, which is suitable for 16 plates with 96 wells each for homozygote selection with a 16% selection rate (for convenience of cost comparison, the number of plants selected are assumed to be 96, in-line with the number of samples accommodated by one plate for seed-DNA based genotyping). The comparison showed that the total cost of seed DNA-based

Table 6 Germination and seedling vigor tests for the seeds with endosperm partially sampled for DNA extraction, under field conditions, El Batan, Mexico

Plant materials	Sampled endosperm (mg)		Germination rate (%)			Seedling survival rate (%)			NDVI		
	SS	LS	Control	SS	LS	Control	SS	LS	Control	SS	LS
F₂ population											
CML418 × CML312	27.96	55.59**	90.67	84.00	86.67	90.67	82.67	85.33	0.47	0.45	0.39
HGA × CML491	20.78	34.46**	96.00	82.67	92.00	94.67	81.33	90.67	0.35	0.32	0.31
CML491 × LPS	24.21	42.94**	92.00	92.00	88.00	89.33	86.67	85.33	0.37	0.32	0.31
CML491 × DTP	20.55	44.43**	86.67	94.67	89.33	84.00	92.00	82.67	0.41[a]	0.29[b]	0.34[ab]
CLQ × CML492	22.31	38.59**	98.67[a]	85.33[b]	93.33[ab]	96.00[a]	82.67[b]	90.67[ab]	0.38[a]	0.35[ab]	0.32[b]
Mean	23.16	43.20	92.80	87.73	89.87	90.93	85.07	86.93	0.40	0.35	0.33
Inbred lines											
CML494	16.16	25.30**	77.33[aA]	46.67[bB]	68.00[aAB]	76.00[aA]	45.33[bB]	58.67[bAB]	0.25[A]	0.19[B]	0.19[B]
CML454	22.59	43.96**	86.67	76.00	84.00	85.33	69.33	81.33	0.31	0.31	0.26
Mean	20.28	36.67	84.57	67.62	79.30	83.11	63.93	74.03	0.31	0.27	0.25

NDVI, Normalized difference vegetative index; SS and LS are two treatments with small-sampled and large-sampled endosperm, respectively. Each number in the table is averaged from 75 seeds or seed-derived plants

** The weight of sampled endosperm from LS is significant higher than that from SS at $P = 0.01$ level

Multiple comparisons among control, SS and LS were tested at $P = 0.05$ (shown with low case letters) and $P = 0.01$ (shown with capital letters) levels, respectively; the numbers followed by the same letter are not significantly different
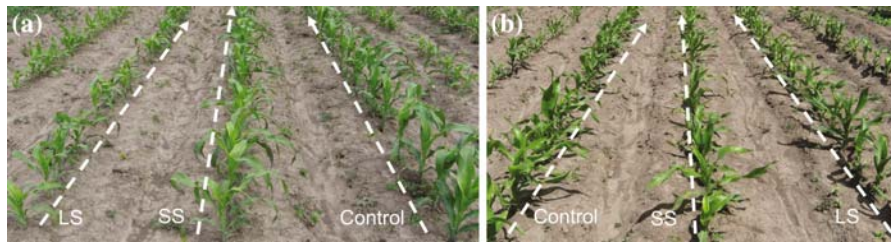
Fig. 8 Germination and seedling establishment tests under field conditions, El Batan, Mexico, for the seeds with endosperm partially sampled for DNA extraction. (**a**) Seeds from (CML418 × CML312)F$_2$, (**b**) Seeds from (CLQ × CML492)F$_2$.

LS, large-sampled group of seeds with a large proportion of endosperm sampled for DNA extraction; SS, small-sampled group of seeds with a small proportion of endosperm sampled for DNA extraction; Control, unsampled seed

genotyping is 24.6% lower than leaf DNA-based genotyping in this test case (Table 7), with a cost of less than $1 per data point for a single marker using seed DNA-based genotyping. With the increasing population size and number of markers while decreasing the rate of selection, more labor and land costs will be saved using seed DNA-based genotyping, compared to MAS after planting using leaf-DNA based genotyping. For leaf DNA-based genotyping, a large proportion of extra cost comes from the labor time required for the sampling process and the extra field management required to maintain undesirable plants prior to genotypic selection. Additionally, some factors that increase the cost for leaf DNA-based genotyping are not considered in this comparison. For example, the refrigerated transport of leaf tissue adds an extra burden which will vary depending on the distance from field station to lab. As a normal practice at CIMMYT, the drying of the leaf tissue using a lyophilizer before grinding is often a limiting factor on throughput as there is usually limited available capacity for lyophilization, whereas

seed pieces are dried in the open without space limitations.

Discussion

The protocol used for sample preparation from single seeds is a critical step affecting DNA extraction efficiency, DNA quality and quantity, plus germination and seedling establishment of sampled seed. Therefore, we have attempted to develop a suitable method for sampling which is versatile for all types of maize seeds, produces high quality and quantity of DNA and has minimal effects on subsequent germination and establishment. Drilling the seed to obtain powder, as reported for barley (von Post et al. 2003) and maize (Sangtong et al. 2001), was not time efficient and led to powder overheating. However, after soaking in water for a suitable time, seeds could be easily cut with a scalpel. However, soaking can lead to pre-germination or physiological changes which may influence

Table 7 Comparative evaluation of time and cost for genotyping 1,536 seed and leaf samples using two SSR markers

Experiment items	Seed				Leaf			
	Labor[a]	Reagent	Supplies	Total	Labor	Reagent	Supplies	Total
Sample preparation	100.00	0.00	119.98	219.98	200.00	0.00	119.04	319.04
Grinding	20.00	0.00	51.73	71.73	10.00	0.00	51.73	61.73
Extraction	80.00	39.90	234.65	354.55	80.00	40.08	234.65	354.73
PCR analysis	320.00	180.48	309.29	809.77	320.00	180.48	309.29	809.77
Field plant cost[b]	–	–	–	26.24	–	–	–	419.91
Total				1482.27				1965.18

[a] Labor cost was calculated as $10 per hour based on the payment for a high level technician at CIMMYT

[b] Field planting cost was calculated based on the 96 plants selected from seed-based genotyping and 1536 plants in leaf-based genotyping with charge of $0.27/plant at Tlaltizapan Station of CIMMYT (Dreher et al. 2003)

subsequent germination in the field. Soaking seeds under low temperature or with the addition of abscisic acid have been considered as potential measures to reduce the impact on subsequent germination based on conventional physiological theory (Bewley 1997). Fortunately, in this study we observed that the germination rate of sampled seeds was similar to control seeds, even after being stored for up to 1 month under room temperature before planting. Thus, it is not necessary to treat maize seeds during soaking, although this may be a valuable approach for other species. The proportion of the endosperm excised was also investigated for its effect on germination and seedling establishment. For relatively large seeds, sample sizes up to 50 mg, or 20% of the seed weight, still allowed highly acceptable germination and seedling establishment rates, while providing plenty of tissue for DNA extraction.

The variation in genetic characteristic of different seed tissues is an important concern in determining the potential application of this method. The pericarp or seed coat, which is of maternal origin, may cause false heterozygotes in selfed segregating populations. von Post et al. (2003) reported no pericarp contamination in barley seed DNA but this was based on analysis of just one marker. In the study reported here we have used five markers, from which we have observed a variable level of pericarp contamination depending on the marker used (from 0 to 16.7%), and there may also be some genotype-specific effects. Therefore, when this method is used to select for heterozygous genotypes, for example in marker-assisted backcross breeding programs, the problem can be avoided by using the recurrent parent as the female. In this way, the target allele from the donor is never present in the pericarp of the tested seed, eliminating the possibility of false positives. Conversely, for selection in F_2 segregating populations, false genotypic scores may occur when homozygotes are incorrectly genotyped as heterozygotes due to pericarp contamination. This would not cause false positives but lead to a level of false negatives proportional to the level of contamination. In practice, candidate markers that are particularly sensitive to pericarp contamination could be eliminated. Another option is to carefully remove the pericarp before grinding, although this would add time and labor cost to the process, this is feasible with

soaked seeds in maize. In addition, we evaluated the potential cross-contamination that would happen in sampling and DNA extraction, which also might confound pericarp contamination analysis. Three types of seeds with previously known genotypes from a single cross (i.e., P_1, P_2, F_1) were analyzed. Inferred genotypes and corresponding previously known genotypes were compared for several polymorphic SSR markers (data not shown). The results showed that no contamination was detected by these markers. However, careful operation is still required to avoid potential cross-contamination from neighboring samples in the 96-well plate, which is also of equally concern for leaf DNA-based genotyping.

Hetero-fertilization, which was first reported in maize by Sprague (1929), is another concern that could cause errors in genotyping when using endosperm-based DNA. Using morphological markers differentially expressed in endosperm and pericarp, the incidence of hetero-fertilization in maize is on average approximately 1.25%, although there is significant genotypic variation (Sprague 1932). This means that an average of 1.25% of errors would be expected when genotyping is based on endosperm DNA. More recently, Robertson (1984) reported rates up to 5% in diverse germplasm. Although this problem cannot be avoided, low incidence could be tolerated considering the time and cost benefits compared with using leaf-based DNA. However, it is clearly advisable to define the rates of hetero-fertilization for new populations and germplasm sources to ensure that the rates are acceptable for the target breeding program. The seed DNA-based genotyping system developed in this study makes it possible for the first time to carry out large-scale accurate investigations of the incidence of hetero-fertilization.

Seed weight or size is also an important factor that should be considered when designing genotyping systems based on single seeds. Large seed size means that sufficient DNA quantities can be extracted while leaving enough endosperm for seedling nutrition. For this reason, some inbred lines are not suitable for large-quantity DNA extraction, while others such as CML454 performed adequately well in field tests. In general, inbred lines would not be routinely tested using DNA from single seeds, since their seeds can be bulked for DNA extraction because all seeds within a genotype should be identical. However,

some F_2 or BC_1 populations may have individuals with small seeds. In our field experiment, there was no preferential selection for plants with larger seed size in the segregating populations, but in practice, it should be possible to use relatively larger seeds for genotyping that are usually found in the middle area of each ear, potentially minimizing the seed size effect. In addition, large-scale MAS programs are usually conducted at early generations for high efficiency (Hospital et al. 1997; Stam 2003; Liu et al. 2004). In this situation, seed size is usually relatively large because of hybrid vigor. On the other hand, MAS in later generations is usually family-based so DNA from multiple seeds rather than single seeds can be used for genotyping. Finally, for MAS we recommend that the smallest possible amount of endosperm is sampled, since most MAS programs only require small amounts of DNA in order to carry out relatively small numbers of marker assays. We also recommend optimum field management to maximize early seedling growth and survival since the seeds planted are the pre-selected desirable genotype with considerable labor and cost already invested in such seeds. For large-scale MAS with large numbers of seeds selected, a standard machine planting can be adopted as the sampled seeds are still large enough for this approach. In addition, a simple seed-coating technique will not only help with the planting of sampled seeds via a normal planting procedure but is also likely to help improve germination rate and seedling establishment.

As seed DNA-based genotyping can be processed before planting, for example selecting F_2 seeds harvested from an F_1 plant, it is possible to solely plant selected desirable genotypes. This has a potentially large impact on breeding programs, from changing optimum population sizes and selection pressures to allowing differences in field design and MAS strategies. Over several breeding cycles, this is likely to lead to cumulative and accelerated gains in selection pressure and improvements in overall breeding efficiency. Another advantage of seed DNA-based genotyping is that genotyping can be carried out in batch mode until a minimum target number of desirable genotypes has been identified. This means that the number of target genotypes can be closely controlled while avoiding the risk that no desirable genotypes can be found within the available plants in the field or the need to grow out excessively

large populations in order to maximize the probability of identifying a minimum number of desirable genotypes. This offers the opportunity for breeding programs to come much closer to optimum recommendations from simulation and modeling analyses. For example, a theoretical proportion of homozygotes at n target loci in an F_2 population is $(1/4)^n$, and thus for three loci, 1/64 plants in the population will have the desirable genotypes. For leaf DNA-based genotyping, to ensure a 99% probability of obtaining at least one desirable genotype, the minimum number of plants that has to be planted is $\log(1 - 0.99)/\log(1 - 1/64) = 292$. As the number of target loci increases, the minimum number of plants that have to be planted in the field will go beyond the capacity of most current breeding programs. All these factors have significant impacts on the efficiency of procedures, methods and strategies for MAS, not least the inefficiencies introduced by the number of plants in the field rarely coinciding with the optimum number of maximum lab efficiency. In addition, two cycles of MAS can be done in one crop season: the first MAS is based on leaf-DNA and the second based on the seed set on the plants selected. Thus, seed DNA-based genotyping offers opportunities to simplify the entire process and improve breeding efficiency in a designed-led manner. An important next step is a comprehensive modeling and analysis of all aspects of MAS associated with this genotyping process. This was first done in the pioneering study by Lande and Thompson (1990) for leaf DNA-based MAS under the assumption that the selected is made after planting. This also needs to incorporate both negative factors such as hetero-fertilization and potential pericarp contamination and positive factors such as reduced labor time and selection of desirable genotypes before planting.

In theory, the quality and quantity of DNA from endosperm is not expected to be comparable to that of DNA extracted from leaf tissue. However, in practice, MAS is normally conducted using PCR-based markers (e.g., Dussle et al. 2002; Zhang et al. 2006) which do not require large-amounts or high-quality DNA, as reported in barley (von Post et al. 2003) and rice (Collard et al. 2007). As shown in the study reported here, the DNA extracted from 30 mg of endosperm (which is also a safe amount to sample in terms of effects on germination and seedling establishment), is sufficient for 200 PCR reactions

when 7.5–15 µl of PCR reaction volume is used, as in the case of agarose-gel or PAGE-gel based SSR marker analysis. As a result, seed DNA-based genotyping could meet all MAS requirements including foreground selection for target traits and whole genome selection using PCR-based markers. However, seed DNA-based genotyping may not meet the demands of some special cases, such as large-scale application of agarose-gel or PAGE-gel based genotyping for high-density whole genome selection that requires the genotyping of more than five hundred PCR-based markers. Even in such cases seed-based DNA may still have a valuable role in pre-selection which is then followed by leaf DNA-based large-scale applications. However, there is a concern regarding whether MAS using seed DNA-based genotyping can be completed within the time window from harvesting to planting between two consecutive crops when three or four cycles per year are being used to accelerate the breeding process. In this particular case, optimized breeding procedures and integrated seed and leaf DNA-based genotyping may be necessary.

With the development of chip-based SNP genotyping systems, DNA extracted from single seeds, as shown in maize in this report, can be used to genotype tens of thousands of SNP markers based on the currently available chip-based SNP genotyping system in maize (250 ng DNA for 1536 markers, Ed Buckler, Cornell University, Personal communication) to several million SNP markers as shown by Genome-Wide Human SNP Array 6.0 developed for human genotyping at Affymetrix, which needs 500 ng DNA for an array containing 1.8 million markers (http://www.affymetrix.com/products/arrays/specific/genome_wide_snp6/genome_wide_snp_6.affx). As a result, it is expected that seed DNA-based genotyping will replace leaf DNA-based genotyping in most cases, for all crops with relatively large seeds. The seed DNA-based genotyping process can be fully automated as has been achieved in some companies in order to operate truly high-throughput DNA extraction, which includes automation systems for sampling of endosperm and tracking samples during DNA extraction, genotyping and MAS. This helps improve the throughput and also avoid the soaking treatment required for facilitating sampling of endosperm. However, for most laboratory uses, particularly small- and medium-throughput

genotyping, manual sampling after seed-soaking is likely to be the most viable option. At present, this method is being used in our laboratory for improvement of MAS for quality protein maize and provitamin A carotenoid traits using simple PCR-based markers. In addition, use with SNP markers is being optimized for large-scale application within the maize breeding program at CIMMYT.

In addition to facilitating MAS, single seed-based genotyping can be used in various aspects of genetics research and breeding programs. For example, identification of recombinant individuals from a large number of segregants, as required for near-isogenic line development and fine mapping in a specific chromosome region using several markers, can be simplified as only target recombinants will be selected for planting and phenotyping (e.g., Blair et al. 2003). Genetic studies of seed traits may not need to plant out any material as the seeds segregating for the target trait can be used for both genotyping and phenotyping. Single seed-based genotyping also provides a unique opportunity for genetic studies of fertilization related processes including abnormal fertilization and apomixis. With the increased application of double-haploid techniques in plant breeding, particularly in large commercial maize breeding programs, the demand on individual-based genotyping during breeding cycles will drastically decrease while the efficiency and accuracy of haploid identification will need to be improved. Single seed-based genotyping may be used to identify double haploid plants based on endopsperm genotype without the use of conventional morphological markers, as shown by leaf-DNA based fingerprinting (Belicuas et al. 2007).

There are several issues to be considered before the seed DNA-based genotyping system developed in this report can be extended to other crops (Xu et al. 2008). First, crops should have relatively large seeds with at least 8–10 mg of endosperm or cotyledon that can be sampled for DNA extraction, particularly for agarose gel-based genotyping. Second, seed texture should be suitable for sampling or the seed should be able to tolerate soaking without significant adverse effects on the rate of germination. Third, the pericarp contamination during seed tissue excision should be at a relatively low level unless the pericarp can be easily removed during the sampling process. Finally, tailoring of the DNA extraction protocol may be required for crops with special seed compositions.

DNA extraction from seeds has been reported for several crops (Chunwongse et al. 1993; Hee et al. 1998; Sangtong et al. 2001; von Post et al. 2003, Chenault et al. 2007). However, the seed DNA-based genotyping system reported here can be easily scaled-up using the automated PCR system developed by Dayteg et al. (2007), and could be modified for all other crops except for those with very small seeds. It can be expected that this approach would be a good alternative to leaf DNA-based genotyping for many crops for applications such as intellectual property protection fingerprinting, transgene detection, genetic testing for varietal purity and hybridity, gene mapping, genetic diversity analysis, and MAS.

References

Belicuas PR, Guimarães CT, Paiva LV, Duarte JM, Maluf WR, Paiva E (2007) Androgenetic haploids and SSR markers as tools for the development of tropical maize hybrids. Euphytica 156:95–102

Bewley JD (1997) Seed germination and dormancy. Plant Cell 9:1055–1066

Blair MW, Garris AJ, Iyer AS, Chapman B, Kresovich S, McCouch SR (2003) High resolution genetic mapping and candidate gene identification at the xa5 locus for bacterial blight resistance in rice (Oryza sativa L.). Theor Appl Genet 107:62–73

Chenault KD, Gallo M, Seib JC, James VA (2007) A non-destructive seed sampling method for PCR-based analyses in marker assisted selection and transgene screening. Peanut Sci 34:38–43

Chunwongse J, Martin GB, Tanksley SD (1993) Pre-germination genotypic screening using PCR amplification of half-seeds. Theor Appl Genet 86:694–698

CIMMYT Applied Molecular Genetics Laboratory (2003) Laboratory Protocols, 3rd edn., Mexico, DF, pp 7–11

Collard BCY, Das A, Virk PS, Mackill DJ (2007) Evaluation of 'quick and dirty' DNA extraction methods for marker-assisted selection in rice (Oryza sativa L.). Plant Breed 126:47–50

Dayteg C, Tuvesson S, Merker A, Jahoor A, Kolodinska-Brantestam A (2007) Automation of DNA marker analsysis for molecular breeding in crops: practical experience of a plant breeding company. Plant Breed 126:410–415

Dreher K, Khairallah M, Ribaut JM, Morris M (2003) Money mattes (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. Mol Breed 11:221–234

Dussle CM, Quint M, Xu ML, Melchinger AE, Tübberstedt T (2002) Conversion of AFLP fragments tightly linked to SCMV resistance genes Scmv1 and Scmv2 into simple PCR-based markers. Theor Appl Genet 105:1190–1195

Dwivedi SL, Crouch JH, Mackill DJ, Xu Y, Blair MW, Ragot M, Upadhyaya HD, Ortiz R (2007) The molecularization of public sector crop breeding: progress, problems and prospects. Adv Agron 95:163–318

Hee WK, Yong GC, Ung HY, Moo YE (1998) A rapid DNA extraction method for RFLP and PCR analysis from a single dry seed. Plant Mol Biol Rep 16:1–9

Hospital F, Moreau L, Lacoudre F (1997) More on the efficiency of marker-assisted selection. Theor Appl Genet 95:1181–1189

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Liu PY, Zhu J, Lu Y (2004) Marker-assisted selection in segregating generations of self-fertilizing crops. Theor Appl Genet 109:370–376

Miklas PN, Kelly JD, Beebe SE, Blair MW (2006) Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. Euphytica 147:105–131

Papazova N, Malef A, Degrieck I, Van Bockstaele E, De Loose M (2005) DNA extractability from the maize embryo and endosperm—relevance to GMO assessment in seed samples. Seed Sci Technol 33:533–542

Ragot M, Lee M (2007) Marker-assisted selection in maize: current status, potential, limitations and perspertives from the private and public sectors. In: Guimarães EP, Ruane J, Scherf BD, Sonnino A, Dargie JD (eds) Marker-assisted selection: Current status and future perspectives in crops, livestock forestry and fish. Food and Agriculture Organization of the Unites Nations, Rome, pp 117–150

Robertson DS (1984) A study of heterofertilization in diverse lines of maize. J Hered 75:457–462

Sangtong V, Mottl EC, Long MJ, Lee M, Scott MP (2001) Serial extraction of endosperm drillings (seed)—A method for detecting transgenes and proteins in single viable maize kernels. Plant Mol Biol Rep 19:151–158

Sprague GF (1929) Hetero-fertilization in maize. Science 69:526–527

Sprague GF (1932) The nature and extent of hetero-fertilization in maize. Genetics 17:358–368

Stam P (2003) Marker-assisted introgression: speed at any cost? In: van Hintum Th JL, Lebeda A, Pink D, Schut JW (eds) Eucarpia leafy vegetables, pp 117–124

Teal RK, Tubana B, Girma K, Freeman KW, Arnall DB, Walsh O, Raun WR (2006) In-season prediction of corn grain yield potential using normalized difference vegetation index. Agron J 98:1488–1494

van Deynze A, Stoffel K (2006) High-throughput DNA extraction from seeds. Seed Sci Technol 34:741–745

von Post R, von Post L, Dayteg C, Nilsson M, Forster BP, Tuvesson S (2003) A high-throughput DNA extraction method for barley seed. Euphytica 130:255–260

William HM, Trethowan R, Crosby-Galvan EM (2007) Wheat breeding assisted by markers: CIMMYT's experience. Euphytica 157:307–319

Xu Y (2003) Developing marker-assisted selection strategies for breeding hybrid rice. Plant Breed Rev 23:73–174

Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Sci 48: 391–407

Xu Y, Gao S, Skinner DJ, Martinez C, Krivanek AF, Crouch JH (2008) Seed DNA-based genotyping for marker-assisted selection in crops: opportunities and challenges. Use of molecular markers for plant breeders workshop, Plant & Animal Genome XVI Conference, January 12–16, 2008, San Diego, California, USA, W313

Zhang ZF, Wang Y, Zheng YL (2006) AFLP and PCR-based markers linked to *Rf3*, a fertility restorer gene for S cytoplasmic male sterility in maize. Mol Gen Genomics 276:162–169