Plant Breeding, **135**, 439–445 (2016)
doi:10.1111/pbr.12382

Haploid differentiation in maize kernels based on fluorescence imaging

Brett W. Boote[1], Daniel J. Freppon[1], Gerald N. De La Fuente[2], Thomas Lübberstedt[2], Basil J. Nikolau[3] and Emily A. Smith[1,4]

[1]The Ames Laboratory, U.S. Department of Energy, and Department of Chemistry, Iowa State University, Ames, IA 50011-3111, USA; [2]Department of Agronomy, Iowa State University, Ames, IA 50011, USA; [3]Department of Biochemistry, Biophysics, and Molecular Biology, Center for Metabolic Biology, Iowa State University, Ames, IA 50011, USA; [4]Corresponding author, E-mail: esmith1@iastate.edu

With 3 figures and 3 tables

Received August 17, 2015 / Accepted April 16, 2016
Communicated by R. Tuberosa

Abstract

A new fluorescence-based method for inbred haploid differentiation in maize kernels was developed by utilizing the *R1-nj* colour marker in combination with fluorescence microspectroscopy and imaging. Seven inbred lines with varying *R1-nj* expression were used in this study. The fluorescence response of the diploid kernels at the embryonic dye spot was shown to simultaneously exhibit lower intensity and occur at a higher wavelength than the fluorescence of the dye-lacking haploid embryos. Intensity and area thresholds were applied to fluorescence images to sort the haploids from mixed sample populations, and sorting efficiencies of greater than 80% were achieved in all seven inbred lines (with values greater than 90% for five lines). The potential for high-throughput sorting when fluorescence imaging is combined with existing technologies for seed handling as well as high sorting efficiency may make fluorescence a viable and promising alternative to current sorting methods for some inbred lines.

Key words: maize — haploid — diploid — fluorescence — sorting

The development of new hybrids of maize (*Zea mays* L.) with desirable traits (e.g. high yield, pest tolerance) is of considerable importance to the global economy, with the 2014 worldwide production estimated at 1.02 billion metric tons (FAO 2014). In modern commercial maize breeding programmes, new hybrids are generated by crossing homozygous inbred lines that have been produced via doubled haploid (DH) line development (Murigneux et al. 1993, Röber et al. 2005, Prasanna 2012). Traditional inbred line development via manual self-pollination requires 6–8 generations to produce a highly homozygous line, whereas DH line development only requires 2–3 generations for a 100% homozygous line. This reduction in line development time allows breeders not only to produce more inbred lines in less time, but also to react to new selection targets quickly. In the DH process, the haploid kernels need to be collected from the mixed population as only the haploids are suitable for the next generation of inbreeding.

Most methods for haploid kernel selection rely on pioneering work by Nanda and Chase, who used purple embryo marker (PEM) stock as the male parent to produce the anthocyanin marker *R1-Navajo* (*R1-nj*) in the progeny kernels (Nanda and Chase 1966). The dominantly inherited *R1-nj* marker causes a xenia effect resulting in pigmentation of the embryo and on the cap of the aleurone of the seed. As is the case in many angiosperms, sexual reproduction occurs through the double fertilization process where fusion of one sperm and the egg constitutes the union of genetic information that produces the diploid embryo, while union of the second sperm and the central cells develops into triploid endosperm. The DH system takes advantage of the phenomenon of haploid induction. In this case, it is known as maternal haploid induction because haploid embryos contain the cytoplasm of the female (donor) parent and the inducer line is the male in the cross. On average, 10% of double fertilizations will fail in this cross, producing seed with a healthy, triploid endosperm and a haploid embryo. This is due to an unsuccessful union of sperm and egg, and these kernels are referred to as haploid kernels. In a haploid seed developed using the *in vivo* maternal induction method (most common), genetic information contained within the embryo is derived only from the female (donor) plant. Ideally, the female does not carry the *R1-nj* marker gene and thus produces a white (haploid) embryo. In all cases where fertilization is successful between sperm and egg, the *R1-nj* gene from the male parent is inherited and expressed in the embryo due to its dominant inheritance and xenia effect, generating a visible red-purple hue. All seeds with red-purple coloration in the embryo are diploid (or hybrid) as they contain genetic information from both the male and female parents, and thus are not suitable for the next generation of DH line development. It is important to note that the only difference between haploid and diploid classes of kernels is their respective haploid and diploid embryos. Both classes contain triploid endosperm and diploid pericarp from the fertilization process.

There are three main categories of sorting methods based on the *R1-nj* kernel marker: manual sorting, machine optical sorting or spectroscopic sorting. The *R1-nj* marker is not consistent in the level of its expression in different lines of maize, which can complicate established sorting strategies. The shape of kernels also plays a role in the visualization of the embryo marker as large, flat kernels more clearly show the marker than small, round kernels due to the separation of the pericarp from the embryo. Sorting by eye is the most popular method, but it is subject to human interpretation of possibly subtle colour differences and there may be high labour and shipping costs with this method. Machine optical sorting typically uses colour optical images of seeds moving through an automated path (conveyor, gravity-fed tube) and a mechanism for directing the two outcomes of the sort [i.e. an air jet (Affleck and Affleck 1990) or mechanical trap door (Becker et al. 2013)]. Near-infrared (NIR) spectroscopy is becoming widely used for non-destructive analysis of various seed crops (Jiang et al. 2007, Davrieux et al. 2010, Cayula Sánchez et al. 2013, Fox et al. 2013) and has been shown to be suitable for haploid/diploid differentiation. Jones and co-workers report sorting accuracies of 85% for haploid and 100% for hybrid (diploid) kernels for 33 total kernels containing

three genotypes using NIR spectroscopy and soft independent model of class analogy (SIMCA) (Jones et al. 2012). NIR spectroscopy typically requires multiple scans per sample, and in this case, each kernel required 1.0–1.3 min to analyse. To put this into context, a typical cross usually requires at least 100 new lines to evaluate. If a 10% haploid pollination success rate followed by a 10% chromosomal doubling rate is assumed, at least 10 000 kernels need to be sorted, which would equate to over 166 h of data collection. This is much slower than visual selection. An improved high-throughput and selective sorting method will greatly aid maize inbred and hybrid development. Seed sorting is also useful for quality control in areas such as crop diseases [e.g. mould, rot (Berardo et al. 2005)] and crop components [e.g. oil content, total fibre (Hacisalihoglu et al. 2010, Jiang et al. 2007)].

It should be noted that not all haploid identification methods rely on the *R1-nj* marker system. Recent work has shown that oil content of the kernels is a viable sorting method when high oil inducers are used (Melchinger et al. 2013, 2014, 2015). This technique has an advantage in lines with the C1 inhibitor gene, which masks *R1-nj* coloration, making sorting via *R1-nj* impossible (Ford 2000). Additionally, an NMR method based on oil content has been described for sorting haploids as well (Liu et al. 2012). Still, the most widely used method in DH programmes is the *R1-nj* system.

As fluorescence emission generally offers greater selectivity over absorbance-based (e.g. colour) optical signals and maize kernels possess intrinsic fluorescence, fluorescence may be a potential, yet heretofore untested, sorting mechanism. If there is a change in fluorescence intensity or shift in the maximum emission wavelength related to either the haploids or diploids, automated and high-throughput sorting of haploids may be possible. Herein, a combination of fluorescence spectroscopy and imaging of maize kernels is reported to investigate whether such sorting is a viable method and to test instrumentation that may be suitable for sorting. Spatially resolved fluorescence spectra show changes in the fluorescence emission maxima, and multiple kernels imaged using a commercial fluorescence biomolecular imager show the spatially resolved fluorescence intensity of haploids and diploids. A sorting mechanism that uses a combination of fluorescence intensity and spatial thresholds on fluorescence images is shown to be a viable method to sort haploids from a mixed sample. Strategies to automate and generate high-throughput sorts are discussed.

Materials and Methods

Fluorescence spectra of haploid and diploid kernels: Fluorescence microspectroscopy was performed on haploid and diploid kernels using a 532-nm Sapphire SF laser (Coherent, Santa Clara, CA, USA) illuminating a DM IRBE inverted light microscope (Leica Microsystems, Buffalo Grove, IL, USA) fitted with a 10× (0.25 NA) objective. The kernels were placed embryo side down on glass cover slides to analyse the embryo of the kernels. The epi-luminescence was passed into a HoloSpec spectrometer (Kaiser Optical Systems, Ann Arbor, MI, USA) equipped with a Newton 940 CCD camera (Andor Technology, Belfast, UK). Single acquisition fluorescence spectra of the upper embryo were acquired with the laser set to 50 mW (2.5×10^6 W/cm^2), and the acquisition times were 50 ms. The data were plotted using IGOR (WaveMetrics, Portland, OR, USA). This instrumentation allows for full spectra to be collected at spatially resolved locations on the kernel.

The fluorescence microscope was fitted with a motorized XY precision translation stage (Prior Scientific, Rockland, MA, USA) interfaced to a desktop computer using LabView (National Instruments, Austin, TX, USA), which was used to collect sequential fluorescence spectra across the length of the kernel (hereafter called line scans). To collect line scans, the sample stage was rastered one-dimensionally across each kernel from tip cap to the top of the seed coat using a 25-μm step size. The sample was illuminated with a laser power of 0.50 mW (2.5×10^4 W/cm^2), and the acquisition time was 250 ms at each spot. The luminescence maxima were measured from the spectral data using a single-peak Gaussian fit performed in IGOR. For the Gaussian fits, constant baselines were used, and the wavelength data range was limited to 540–847 nm. Full spectra were also plotted as false-colour images in MATLAB (Mathworks, Natick, MA, USA).

Optical images of haploid and diploid kernels were collected with a Cyber Shot DSC-W80 digital camera (Sony Corporation, Tokyo, Japan). The kernels were placed on white paper and imaged from a distance of 6 inches using camera flash and incandescent lighting.

Fluorescence imaging of haploid and diploid kernels: High-spatial-resolution fluorescence images were acquired with the Typhoon FLA 9500 biomolecular imager (GE Healthcare Bio-Sciences, Pittsburg, PA, USA). The kernels were placed embryo side down directly on the scanner bed. The built-in 532 nm laser was used for excitation, and either a 575-nm long-pass (LPG), 570 ± 10-nm band-pass (BPG) or 665-nm long-pass (LPR) filter was used for data collection. Images were collected at spatial resolutions of 25, 50 or 100 μm with the PMT detector set at 300 V. A threshold was applied to the resulting images in ImageJ to highlight pixels with an intensity above the cut-off value, then an area threshold of these regions of interest (ROIs) was applied. These data were plotted as area histograms to identify the appropriate threshold for differentiating diploid and haploid kernels, as further discussed below.

DNA extraction: Polymorphic markers that could be used to identify six of the seven inbred lines and the inducer used for pollination were determined to positively confirm the haploid or diploid identity of the kernels analysed by fluorescence. As all kernels studied here contain hybrid endosperm, a seed chipping approach is not useful. DNA extraction was performed after the fluorescence analysis was complete. Kernels were planted and grown to the 23 leaf stages for DNA extraction from leaf tissue. (For the DNA analysis, the '78371A' kernels were not available; thus, only data from the remaining six lines are presented). The DNA was extracted with CTAB as described elsewhere (Stewart and Via 1993) and diluted to 20 ng/μl for PCR identification of polymorphic markers.

Germplasm: The characteristics of the maize inbred lines of interest for this study are listed in Table 1. '78371A' is a yellow dent developed from a cross between a private inbred line and the Iowa Long Ear Synthetic (BSLE) (Cortez-Mendoza and Hallauer 1978). All other lines are solely developed by the listed organizations. All lines are available for public use.

Seed was produced in the summer of 2012 at Iowa State University Agronomy and Agricultural Engineering Research Farm near Ames, IA. The maize inbred lines of interest in this experiment were obtained from a larger experiment that included 120 diverse lines. The inbred lines were planted in single row 5.48-m plots on 76.2-cm row spacing at a density of 59 799 plants per hectare. When plots reached approximately 50% silking, all plants were pollinated using bulk pollen from the maternal haploid inducer RWS/RWK-76 (Röber et al. 2005). All ears were bulk harvested, dried and shelled at maturity. After shelling, seed was visually sorted using the *R1-nj* marker system to categorize haploids and diploids (Nanda and Chase 1966).

Results

The goal of this work is to determine the feasibility of using fluorescence signals for the differentiation of haploid and diploid maize kernels. Fluorescence intensity (i.e. the number of photons emitted), wavelength (i.e. the energy of the emitted photons) or

Table 1: Characteristics of the inbred lines studied, including developer, heterotic group, qualitative marker expression, dent type and other unique characteristics of the kernels

Inbred line	Developer	Heterotic group	Marker expression size, intensity	Dent type	Special characteristics
'78371A'	DeKalb-Pfizer	Non-stiff stalk	Broad, dark	Yellow dent	Large kernels
'PHR36'	DuPont Pioneer	Non-stiff stalk	Narrow, dark	White semi dent	High aleurone expression
'PHT77'	DuPont Pioneer	Non-stiff stalk	Small, very light	Yellow dent	Small kernels
'PHK35'	DuPont Pioneer	Stiff stalk	Variable	Yellow dent	Slight red embryo tint
'PHB47'	DuPont Pioneer	Stiff stalk	Broad, medium	Yellow dent	Large kernels
'NK792'	Northrop, King	Stiff stalk	Small, light	Yellow dent	Red tinted embryo
'MS198'	Michigan State University	Non-stiff stalk	Small, dark	Yellow semi dent	Small kernels

spatial dependence (i.e. localized variation across the kernel) may provide the necessary contrast for a successful selection, and all three properties were investigated. All kernels were first visually sorted for the presence of anthocyanin expression in the embryo and categorized as haploid or diploid. Second, high-spatial-resolution laser scanner fluorescence images were collected to measure the spatially resolved fluorescence response across 'MS198' kernels (Fig. 1) with 532-nm excitation and 575-nm long-pass emission. The kernels were placed embryo side down (towards the detector), as the diploids and haploids have identical fluorescence signals from the seed coat. Within each kernel, the tip cap, germ, embryo and seed coat are well defined in the fluorescence image due to their unique intensity profiles. A markedly lower fluorescence intensity in the diploid embryo (Fig. 1a) compared with the haploid embryo (Fig. 1b) is evident. The reduced fluorescence intensity is measured across the upper embryo, or an area of approximately 0.09 cm^2 for 'MS198' kernels.

Fluorescence spectra were collected from a 2-μm^2 region of the upper embryo of haploid and diploid '78371A' kernels. The maximum fluorescence intensity, average and standard deviation for 10 diploid and 10 haploid kernels are summarized in Table S1. The average diploid intensity was 1.0 ($\pm$0.4) $\times$ 10^4 counts, while the average haploid intensity was 5.6 ($\pm$0.6) $\times$ 10^4 counts. While there is some kernel-to-kernel variability in the fluorescence intensity, there is a statistically significant difference (based on one standard deviation) in intensity among the haploid and diploid kernels. In an effort to visualize the fluorescence response at the cellular level, microscopy images of diploid and haploid '78371A' kernels were acquired using 540 $\pm$ 15-nm excitation and 620 $\pm$ 25-nm emission filters (Fig. S1). The average intracellular fluorescence decreased by 85% in the diploids, indicating an intracellular explanation for the decreased diploid florescence.

In addition to fluorescence intensity and location, the wavelength of the emission spectrum is a third parameter that may be

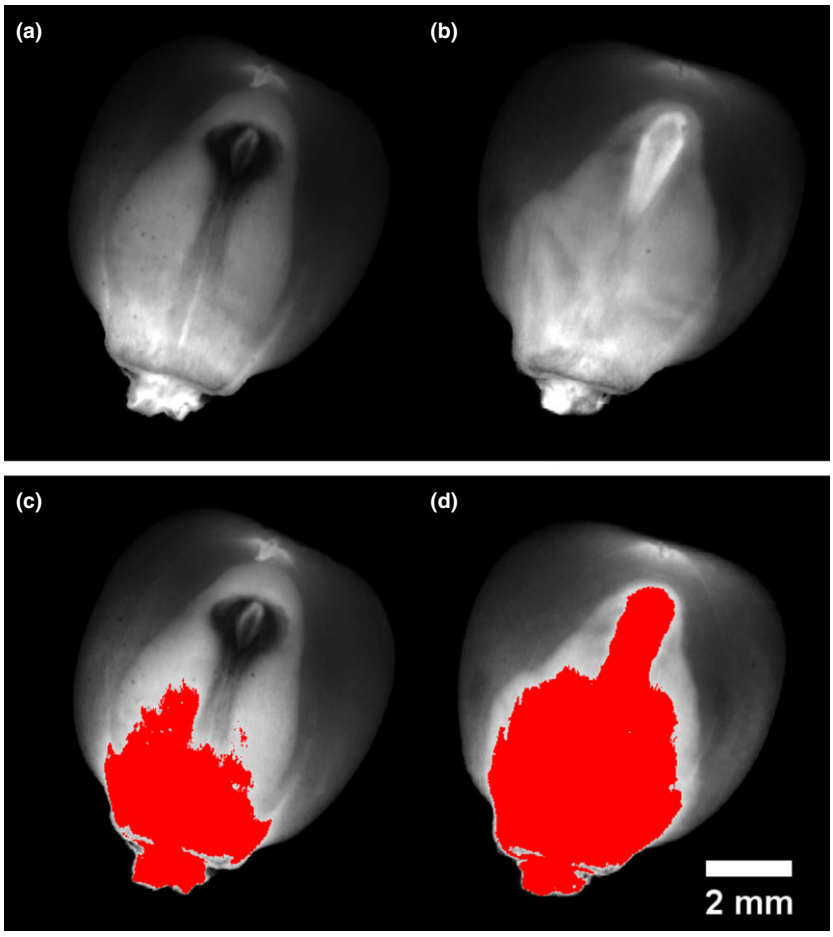

Fig. 1: High-resolution 25 μm/pixel laser scanner fluorescence image of (a) diploid and (b) haploid kernels from '78371A'. The lower image (c, d) shows the intensity and area threshold (red color) used for haploid sorting of this inbred line. Note the large contrast difference in the central germ area; this is the feature useful for fluorescence-based sorting of haploid kernels from mixed samples

used in the haploid selection process. Fluorescence line scans showing the emission wavelengths as a function of location on the kernel were collected (Fig. 2). The red lines indicate the sampling path, and the false-colour images plot the fluorescence intensity as a function of the two parameters: location in the kernel and wavelength. The wavelength of maximum fluorescence intensity as a function of distance from the tip cap is quantified in Fig. 2c. The plots represent an average of five diploid and five haploid kernels of inbred lines '78371A' and 'MS198' for lengthwise scans across the kernel analogous to the collection method used to generate the data shown in Fig. 2a,b. Starting from the tip cap, the wavelength of maximum fluorescence intensity is approximately 620 nm in both the diploid and the haploid and in both inbred lines. Moving farther into the embryo, the diploid kernel shows a red-shifted fluorescence maximum at locations where the anthocyanin dye is visually observed in the seed. The red shift associated with the dye spot of the diploid kernels is measured across an approximately 4 mm distance in both inbred lines and shows a gradual onset from the lower germ to the centre of the embryo. The diploid fluorescence maxima from both lines shift to approximately 650 nm on average, which may be the result of similar anthocyanin composition. The extent of the red shift is highest approximately 7 mm from the tip cap, which is where the highest concentration of dye is visually observed.

To show the relative differences in kernel size and dye expression for different lines, optical images and corresponding laser scanner fluorescence images of representative haploid and diploid kernels from '78371A' and 'MS198' were collected (Fig. S2). The '78371A' kernels are approximately 20% larger, tend to have more rounded embryos and have large dye spots that cover over half the germ area. The haploid kernels also show high fluorescence throughout the embryo. In contrast, the 'MS198' kernels and dye spots are smaller, with the latter confined to the upper embryo, and the haploid fluorescence is quite variable throughout the germ.

Establishing thresholds in fluorescence intensity, wavelength and spatial location is the basis for discriminating haploids from diploids using fluorescence signals. The seed-to-seed variation within each inbred line and appropriate threshold values were established using images of up to 10 diploid and 15 haploid kernels of each of the seven inbred lines using the biomolecular imager. For example, high-pass intensity thresholds of 7380 and 7640 counts were used for '78371A' and 'MS198' kernels, respectively, to define regions of interest in the image. The regions of interest corresponding to diploid and haploid kernels for '78371A' and 'MS198' were plotted as histograms, and area thresholds of 0.160 cm^2 for '78371A' and 0.125 cm^2 for 'MS198' were applied to select for haploids (Fig. 3). The area threshold for haploid selection is plotted as the dashed line: A haploid area that falls below the threshold will produce a false negative, while a diploid area that is higher than the threshold will produce a false positive. Kernel images and histograms of the sorts for the remaining five lines are shown in Figs S3–S7. A sorting statistics table summarizing all seven lines is shown in Table 2 for the fluorescence method and Table S2 for visual and genetic sorting methods.

To explore the difference in the fluorescence signal produced by analysing different emission wavelengths, images of five diploid and haploid kernels from '78371A' and 'MS198' inbred lines were collected with the biomolecular imager using three filters to selectively pass the wavelengths that reach the detector: a 575-nm long-pass (LPG, allows all wavelengths longer than 575 nm to reach the detector), 570 ± 10-nm band-pass (BPG) and a 665-nm long-pass (LPR). Figure S8 shows three spectra from a diploid '78371A' kernel and the wavelength ranges of collection for the three filters that were studied. As before, an intensity threshold was applied to the images, and the average

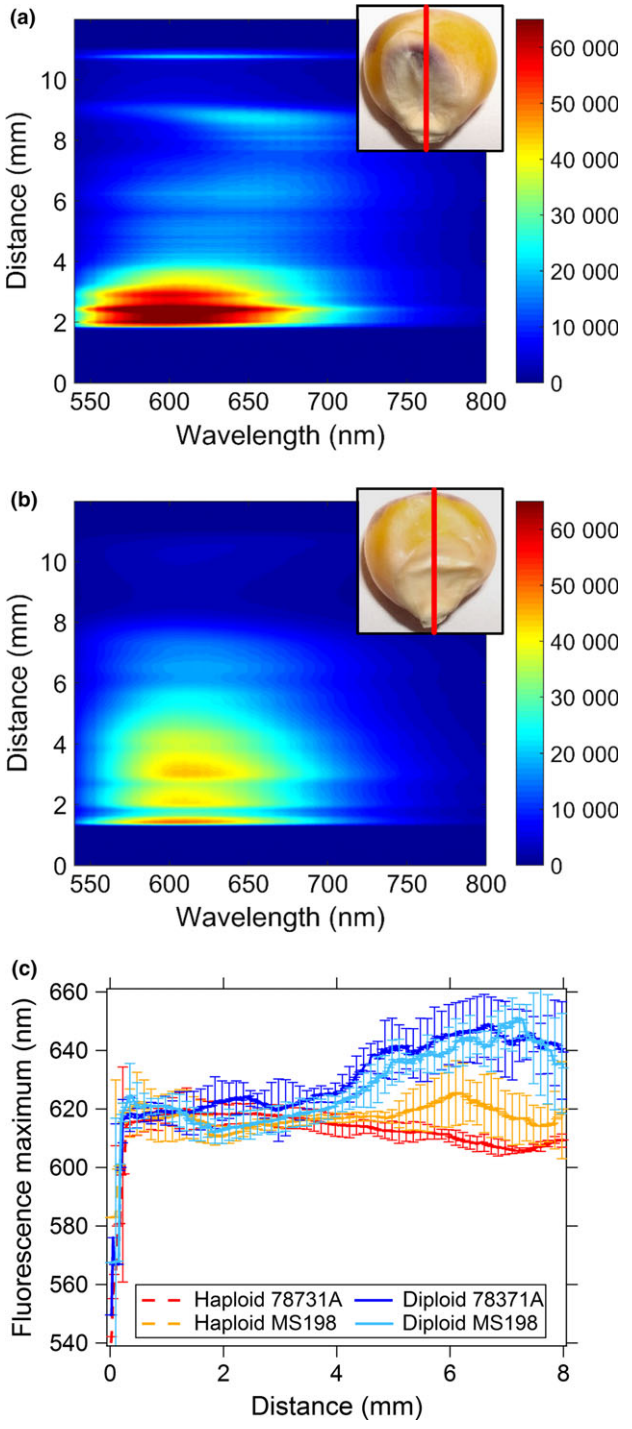

Fig. 2: Digital photograph and false-color fluorescence line scans of a (a) diploid and (b) haploid '78371A' kernel, and (c) maximum fluorescence wavelength measured across the kernel. The wavelength of maximum fluorescence is red-shifted in the dye-rich area of the diploid kernel. The excitation wavelength was 532 nm operating at 0.50 mW. For (a) and (b), each point acquisition was 250 ms, and the step size was 25 μm. The diploid kernels show a notable red shift that corresponds to the anthocyanin-rich area of the germ, whereas the dye-lacking haploids show little-to-no shift. For (c), the traces are an average of 5 kernels of each line listed in the legend, and the error bars denote one standard deviation of the average

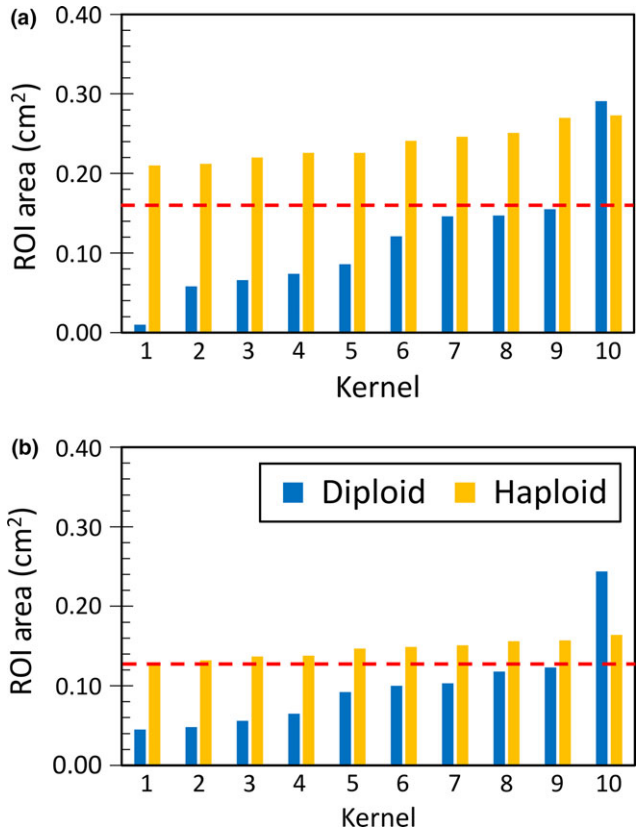

Fig. 3: Kernel sorting histograms for (a) '78371A' and (b) 'MS198'. The fluorescence images are thresholded based on intensity, which generates regions of interest (ROIs) of varying area. These ROI areas are used to differentiate haploid kernels from diploid seeds. In each lot, 10 of 10 haploids are correctly identified with 1 false positive

Table 2: Sort statistics of all maize lines studied.

Line	Haploid ID[a]	False negatives	False positives
'78371A'	10/10	0	1
'MS198'	10/10	0	1
'PHR36'	9/10	1	1
'PHT77'	5/6	1	5
'PHK35'	8/8	0	1
'PHB47'	9/10	1	3
'NK792'	4/5	1	1

[a] All kernels were first visually classified as haploid or diploid, then the true ploidy was determined by DNA marker analysis (except for '78371A' which was unavailable for DNA analysis). The maximum numbers shown here are the positive haploids from DNA analyses.

areas for each kernel type (inbred and ploidy) and filter type were calculated. The results are shown in Table 3.

Discussion

The total fluorescence intensity of the maize kernels decreased in locations of the embryo where anthocyanin dye is present as shown in Table S1 and Figs 1 and 2. The average magnitude of this decrease varied by inbred line. There is also a simultaneous shift in the spectrum of diploid kernels that is not present in the haploid kernels. As the fluorescence response of anthocyanins is highly pH dependent and many substitutions are possible on the cyanidin backbone, a thorough understanding of the molecular constituents of maize anthocyanins may aid in understanding the

Table 3: Emission filter comparison of area-sorting statistics using 575-nm long-pass (LPG), 570 ± 10 nm band-pass (BPG) and 665 nm long-pass (LPR) filters

	Filter	Ploidy	Avg. area (cm^2)	Difference (cm^2)	Improvement (%)
'78371A'	LPG	Haploid	0.268	0.173	–
		Diploid	0.095		
	BPG	Haploid	0.278	0.197	14
		Diploid	0.081		
	LPR	Haploid	0.267	0.012	−93
		Diploid	0.256		
'MS198'	LPG	Haploid	0.174	0.074	–
		Diploid	0.099		
	BPG	Haploid	0.171	0.091	22
		Diploid	0.080		
	LPR	Haploid	0.184	0.000	−100
		Diploid	0.184		

source of this red-shifted fluorescence (Drabent et al. 2007, Rakić et al. 2015). Potential explanations for the lower intensity in the diploids include displacement of a more highly fluorescing species by the anthocyanins or quenching of the fluorescent species by the anthocyanins, which have been reported to have some absorbance at the emission wavelengths used to generate the data in Fig. 2 and Table S1 (Rakić et al. 2015).

A total of seven inbred lines were analysed for this study; the sorting principle is the same for all lines. The analysis of these inbred lines reveals that the selection based on fluorescence intensity and area thresholds should be optimized for each inbred line due to inherent kernel, embryo and dye spot variability. The seed-to-seed and inbred line-to-line variability must be understood to develop a sorting protocol based on the fluorescence intensity, wavelength and/or area.

To sort a large population of kernels, a small subset (e.g. 10–100 kernels depending on the qualities of the line) should be visually sorted into haploid and diploid populations. Subsequently, fluorescence measurements should be performed to generate the intensity and area thresholds that allow selection of haploids from diploids for that line or variety. There are a few ways the threshold could be established; for example, haploids and diploids could be sorted by eye based on the presence of the dye spot or they could be genetically tested to definitively establish the properties of the haploid and diploid fluorescence signals. These preliminary measurements provide the information necessary to automate the sort of a large quantity of seed. An intensity threshold of greater than 7080 counts was applied to the images in Fig. 1 to select the embryo area; the applied threshold generates a smaller area in the diploid due to the low fluorescence in the embryo that results from *R1-nj*-expressed anthocyanin dye in the diploids. The fluorescence intensities collected from outside the embryo region are statistically similar between diploids and haploids, so selection of haploids and diploids by fluorescence intensity is only feasible when analysing the embryo region.

For both '78371A' and 'MS198' inbred lines, the area thresholds are plotted as dashed lines over area histograms of diploid and haploid kernels (Fig. 3). In general, the area threshold for each line was selected so that there will statistically be few (if any) false negatives. The justification for this is that the haploids are rare and false positives are more tolerable than false negatives as the haploid kernels make up about 10% of the progeny kernels, even when using lines with a high haploid induction rate (Prasanna 2012). Thus, selecting all the haploid kernels is worth having a small false-positive population. The thresholds that were selected

for '78371A', 'MS198' and 'PHK35' inbred lines identified all 10 haploids with one false-positive diploid (Table 1). Ignoring the false positive, the '78371A' kernels produce a larger area difference between the correctly sorted highest diploid area and the lowest haploid area and are, therefore, easier to sort than the 'MS198' and 'PHK35' kernels. This was expected given the overall larger size of the dye spots for '78371A'. 'PHR36' and 'NK792' inbred lines also had one false-positive diploid and misidentified one of the haploid kernels. For the remaining two lines, the false-positive rate was higher. For all lines studied, 80% or greater of the haploids were correctly identified with, at most, five false positives. 'PHT77' and 'NK792' were particularly difficult to sort, both visually and using the fluorescence method, due to their uncharacteristically small and obscured dye spots. Visual sorting misclassified nearly half the kernels for these lines (Table S2) and led to uneven populations of haploid and diploid kernels.

To quantify the sorting efficiency, differences between the average area generated by the intensity threshold for the haploid and diploid of two of the inbred lines and three filter types were found. Subsequently, the percentage improvement in the area differences was calculated with respect to the 575-nm long-pass filter (Table 2). The 570-nm band-pass filter provided a better sort in both inbred lines as the red-shifted fluorescence of the dye spots produces better discrimination over the range of wavelengths passed by the band-pass filter (Fig. S8). The 665-nm long-pass filter performed poorly, and the areas of the diploids and haploids were nearly identical. This is because this filter rejects the majority of the fluorescence, and there is little discrimination in the red edge of the emission peaks. This shows that the sorting will be most effective using band-pass filtration to take advantage of the red-shifted emission spectrum in the diploid to select a narrow range of wavelengths that offer the largest difference in intensity between the diploid and haploid. The optimal filters for an effective sort do not significantly vary across the inbred lines that were studied. While the intensity and area thresholds may vary from one line to the next, as discussed above, the results show the same instrumentation can be used to obtain fluorescence data to successfully sort several inbred lines.

In theory, any instrument or device capable of measuring the fluorescence of a solid sample could be used for the sort. A GE Typhoon FLA 9500 laser fluorescence biomolecular imager was chosen for this study as it provides the ability to measure the fluorescence signal of many kernels simultaneously, which may be beneficial for high-throughput sorts. The instrument has a 40 × 46 cm scanning area, so approximately 2400 kernels can be analysed in a single 12-min scan (0.3 s per seed). Additionally, if combined with existing seed manipulation strategies such as pneumatic bursts (Affleck and Affleck 1990), suction (Becker et al. 2013) or robotics (Song et al. 2010), no manual handling of the kernels would be necessary (although these strategies were not available in this study).

Conclusion

A fluorescence imaging-based haploid sorting procedure was developed using the biomolecular imager as the instrument of choice. The method applies intensity thresholds and subsequently area thresholds to fluorescence images to sort the haploids that have large areas of high-fluorescence intensity. The applied intensity and area thresholds should be determined for each inbred line and also for the instrumentation that is used to take the fluorescence measurements. This work is not meant to be a comprehensive analysis of the technology for all possible maize populations to be sorted, rather a demonstration of how fluorescence methods may be applied to a typical *R1-nj*-based sort. The benefits of this sorting scheme are high selectivity (i.e. based on qualities of fluorescence), a rigorously testable false-positive and/or false-negative rate, the ability to sort large populations with already developed seed handling equipment, and in many cases analysis capabilities that can be performed on-site. Combining high-throughput potential and greater than 80–90% accuracy in the sorts (based on classifications that were determined by DNA marker analysis of derived plants with the exception of '78371A', where DNA analysis was unavailable), this method shows great promise for both large and small maize hybrid and inbred producers.

Acknowledgements

Funding for this work was provided by the National Science Foundation Partnerships for Innovation: Building Innovation Capacity (PFI:BIC) award no. 1237720. TL would also like to thank USDA's National Institute of Food and Agriculture (Project Numbers: IOW04314, IOW01018), as well as the RF Baker Center for Plant Breeding and K.J. Frey Chair in Agronomy at Iowa State University for supporting this work. The authors thank Dr. M. Paul Scott (Agricultural Research Service, USDA) for helpful discussions.

References

Affleck, S., and L. Affleck, 1990: Apparatus for sorting seeds according to color. United States, Patent US4946046 A, August 7, 1990.

Becker, S. M., S. J. Corak, J. L. Hunter, G. L. Jaehnel, H. Lu, and J. W. Stevenson, 2013: Method and computer program product for distinguishing and sorting seeds containing a genetic element of interest. United States Patent US8452445 B2, May 28, 2013.

Berardo, N., V. Pisacane, P. Battilani, A. Scandolara, A. Pietri, and A. Marocco, 2005: Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. J. Agric. Food Chem. **53**, 8128—8134.

Cayula Sánchez, J. A., W. Moreda, and J. M. Garciá, 2013: Rapid determination of olive oil oxidative stability and its major quality parameters using vis/NIR transmittance spectroscopy. J. Agric. Food Chem. **61**, 8056—8062.

Cortez-Mendoza, H., and A. R. Hallauer, 1978: Divergent mass selection for ear length in maize. Crop Sci. **19**, 175—178.

Davrieux, F., F. Allal, G. Piombo, B. Kelly, J. B. Okulo, M. Thiam, O. B. Diallo, and J.-M. Bouvet, 2010: Near infrared spectroscopy for high-throughput characterization of shea tree (*Vitellaria paradoxa*) nut fat profiles. J. Agric. Food Chem. **58**, 7811—7819.

Drabent, R., B. Pliszka, G. Huszcza-Ciolkowska, and B. Smyk, 2007: Ultraviolet fluorescence of cyanidin and malvidin glycosides in aqueous environment. Spectrosc. Lett. **40**, 165—182.

Food and Agricultural Organization (FAO) of the United Nations, 2014: Food Outlook: Biannual Report of Global Food Markets. FAO, Rome.

Ford, R. H., 2000: Inheritance of kernel color in corn: explanations and investigations. Am. Biol. Teach. **62**, 181—188.

Fox, G., A. Wu, L. Yiran, and L. Force, 2013: Variation in caffeine concentration in single coffee beans. J. Agric. Food Chem. **61**, 10772—10778.

Hacisalihoglu, G., B. Larbi, and A. M. Settles, 2010: Near-infrared reflectance spectroscopy predicts protein, starch, and seed weight in intact seeds of common bean (*Phaseolus vulgaris* L.). J. Agric. Food Chem. **58**, 702—706.

Jiang, H., Y. Zhu, L. Wei, J. Dai, T. Song, Y. Yan, and S. Chen, 2007: Analysis of protein, starch and oil content of single intact kernels by near infrared reflectance spectroscopy (NIRS) in maize (*Zea mays* L.). Plant Breeding **126**, 492—497.

Jones, R. W., T. Reinot, U. K. Frei, Y. Tseng, T. Lübberstedt, and J. F. McClelland, 2012: Selection of haploid maize kernels from hybrid

kernels for plant breeding using near-infrared spectroscopy and SIMCA analysis. Appl. Spectrosc. **66**, 447—450.

Liu, J., T. Guo, P. Yang, H. Wang, L. Liu, Z. Lu, X. Xu, J. Hu, Q. Huang, and S. Chen, 2012: Development of automatic nuclear magnetic resonance screening system for haploid kernels in maize. Trans. Chin. Soc. Agric. Eng. **28**, 233—236.

Melchinger, A. E., W. Schipprack, T. Würschum, S. Chen, and F. Technow, 2013: Rapid and accurate identification of in vivo-induced haploid seeds based on oil content in maize. Sci. Rep. **3**, 2129.

Melchinger, A. E., W. Schipprack, H. Freidrich, and V. Mirdita, 2014: In vivo haploid induction in maize: identification of haploid seeds by their oil content. Crop Sci. **54**, 1497—1504.

Melchinger, A. E., W. Schipprack, X. Mi, and V. Mirdita, 2015: Oil content is superior to oil mass for identification of haploid seeds in maize produced with high-oil inducers. Crop Sci. **55**, 188—195.

Murigneux, A., D. Barloy, P. Leroy, and M. Beckert, 1993: Molecular and morphological evaluation of doubled haploid lines in maize. 1. Homogeneity within DH lines. Theor. Appl. Genet. **86**, 837—842.

Nanda, D. K., and S. S. Chase, 1966: An embryo marker for detecting monoploids of maize (*Zea Mays* L.). Crop Sci. **6**, 213—215.

Prasanna, B., 2012: Doubled haploid (DH) technology in maize breeding: an overview. In: Prasanna B., Chaikam V. and Mahuku G. (ed.), Doubled Haploid Technology in Maize Breeding: Theory and Practice, 1—8. CIMMYT, Mexico.

Rakić, V. P., A. M. Ota, M. A. Skrt, M. N. Miljković, D. A. Kostić, D. T. Sokolović, and N. E. Poklar Ulrih, 2015: Investigation of fluorescence properties of cyanidin and cyanidin 3-O-β-glucopyranoside. Hem. Ind. **69**, 155—163.

Röber, F. K., G. A. Gordillo, and H. H. Geiger, 2005: *In vivo* haploid induction in maize – performance of new inducers and significance of doubled haploid lines in hybrid breeding. Maydica **50**, 275—283.

Song, P., J. Zhang, Y. Xun, X. Chen, and W. Li, 2010: Developement of automatic inspection system of corn seeds. Trans. Chin. Soc. Agric. Eng. **26**, 124—127.

Stewart, C. N. J., and L. E. Via, 1993: A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques **14**, 748—750.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Fluorescence images of (a) diploid and (b) haploid embryo cells from '78371A'.

Fig. S2. Optical images and corresponding fluorescence images of representative diploid and haploid maize kernels showing the local fluorescence intensity within the germ.

Fig. S3. Sorting histogram and two diploid (left) and two haploid (right) kernel images for 'PHR36'.

Fig. S4. Sorting histogram and two diploid (left) and two haploid (right) kernel images for 'PHT77'.

Fig. S5. Sorting histogram and two diploid (left) and two haploid (right) kernel images for 'PHK35'.

Fig. S6. Sorting histogram and two diploid (left) and two haploid (right) kernel images for 'B47'.

Fig. S7. Sorting histogram and two diploid (left) and two haploid (right) kernel images for 'NK792'.

Fig. S8. Plot of selected fluorescence spectra from a diploid '78371A' kernel with the three filters used for the comparison (575-nm long pass, LPG; 570 $\pm$ 10-nm band pass, BPG; 665-nm long pass, LPR).

Table S1. Maximum fluorescence intensities of 10 diploid and haploid '78371A' kernels analyzed at the embryo, showing much lower fluorescence in the diploid seeds.

Table S2. DNA marker analysis of available inbred lines studied, compared to the original visual ploidy classification.