



Effects of Donor x Inducer Interaction on the Success of Haploid Induction and Comparison of Haploid Seed Identification Methods in the In vivo Maternal Haploid Technique in Maize

Fatih KAHRIMAN^{a*} , Umut SONGUR^a , Abdullah DIŞBUDAK^b , Sezgin KIZIK^b , Berk VURAL^b 

^aÇanakkale Onsekiz Mart University, Faculty of Agriculture, Department of Field Crops, Çanakkale, TURKEY

^bAGROMAR Inc., Bandırma Yolu 1. Km., Karacabey, Bursa, TURKEY

ARTICLE INFO

Research Article

Corresponding Author: Fatih KAHRIMAN, E-mail: fkahriman@hotmail.com

Received: 10 November 2020 / Revised: 16 May 2021 / Accepted: 20 June 2021 / Online: 01 September 2022

Cite this article

KAHRIMAN F, SONGUR U, DIŞBUDAK A, KIZIK S, VURAL B (2022). Effects of Donor x Inducer Interaction on the Success of Haploid Induction and Comparison of Haploid Seed Identification Methods in the In vivo Maternal Haploid Technique in Maize. *Journal of Agricultural Sciences (Tarım Bilimleri Dergisi)*, 28(3):385-395. DOI: 10.15832/ankutbd.824114

ABSTRACT

This study was conducted to investigate the effect of donor and inducer lines on the haploid induction rate in the in vivo maternal doubled haploid technique in maize and to compare the methods used in haploid seed separation. In the study, three donor materials were used for induction crossing with two inducer lines. Seedling immersion and stem injection techniques were used as chromosome doubling methods. Eye separation, spectral measurements and image processing were used to identify seed classes. Data were modeled using support vector machine method and created models were evaluated over the complexity matrix. The results of

the research revealed that the haploid induction rate varies depending on the donor genotypes and inducers, and the genotype responses against the applied chromosome doubling method were also different. Differences were observed in the success of the three methods compared for haploid and diploid seed separation. The successful classification rates were 87%, 83%, and 79% in visual, spectral and image processing models, respectively. Results showed that both spectral technique and image processing technique can be used to distinguish haploid/diploid seeds in in vivo maternal haploid technique.

Keywords: Seed classification, Spectral analysis, Ploidy, Chromosome doubling, Support vector machine

1. Introduction

Parental line development is one of the strategic steps in hybrid maize breeding. Traditional methods for developing parental lines require selfing over 6-10 generations (Hallauer et al. 2010). In addition, it is theoretically impossible to reach 100% homozygosity by traditional methods; therefore, there is a need to develop alternative methods. Traditional methods are time consuming and labor-intensive process, which are main disadvantages of them. In vitro (Chidzanga et al. 2017) and in vivo (Chalky 1994) techniques developed as alternatives to classical methods used for pure line development in maize breeding eliminated these disadvantages and proved very convenient to breeders. The in vivo doubled haploid technique is widely preferred compared to in vitro methods due to its ease of application and higher success rate.

The in vivo haploid technique is based on obtaining haploid seeds after hybridization of the donor material with a parent called the “inducer”, and duplication of the chromosome number of haploid seedlings with specific chemical agents. In practice, this technique is applied in two ways, paternal and maternal in vivo haploids. The paternal haploid technique relies on using the donor material as male, while in the maternal haploid technique, the donor material is used as the female parent in the induction cross (Röber et al. 2005; Chidzanga et al. 2017). Almost all maternal haploid inducers currently used are generated from an inducer line called Stock 6 (Choe 1959). While the induction potential of Stock 6 was about 2-3%, the induction potential of materials such as RWS, UH400, PHI and MHI developed from this material was increased to 7% to 16% (Kalinowska et al. 2019). As a result, the use of the in vivo haploid technique in maize breeding studies has recently become popular. Regardless of the induction method or induction line used, increasing the success of this method has strategic importance for maize breeding programs all over the world.

Success with the in vivo haploid technique can be attained by three main factors. The first is the genotypic characteristics of the inducer or donor materials, the second is the chromosome doubling method used, and the third is the identification success of haploid seeds. Obviously, the growing conditions and environmental factors cannot be ignored in the success of the in vivo haploid technique.

The most important genotypic effect in the *in vivo* doubled haploid technique is the induction potential of the inducer line used. Not all inducers provide the same success, and they cannot be used in harmony with every donor (female parent) genotype. In fact, the induction rates achieved in studies using different numbers of donor materials showed a wide variation ranging from 7.1% to 12.8% (Cerit et al. 2016; Zararsız et al. 2019). These results demonstrate that donors and inducers play a key role in the *in vivo* doubled haploid technique.

The second issue affecting the success of the *in vivo* doubled haploid technique is the chromosome doubling method. The main techniques used for chromosome doubling are application of an inducing agent (colchicine) to the stem or seedlings; or alternatively, giving ozone or N₂O gas treatment to seedlings (Chaikam et al. 2019). Among these methods, seedling immersion (Prasanna et al. 2012) and colchicine injection to the stem (Vanous et al. 2017) are among the most widely-used techniques. These techniques show differences in terms of application, and affect both the chromosome doubling rates and the number of DH1 seeds that are finally developed.

The third aspect that influences the success of the *in vivo* haploid technique is the correct identification of haploid seeds. The usual approach used in the selection of haploid seeds is by human eye selection. This selection made anthocyanin production by the effects of marker genes alleles namely as R1-nj and P11 (Uliana Trentin et al. 2020). R1-nj allele causes the seed coloration while P11 allele creates colored roots after induction crossing. Selection by haploids/diploids by eye can be performed based on the color change in the embryo and crown area of the seed or root coloration caused by dominant marker genes carried by the inducer lines (Vanous et al. 2017). Since the labor force required to do this task is high with the visual selection method, and the error rate depends on the person performing the classification, alternative methods are needed. For this purpose, it has been shown that by digitizing images taken of the seeds, visual distinction can be carried out more precisely and faster (Altuntaş et al. 2018; Veeramani et al. 2018; Altuntaş & Kocamaz 2019; Altuntaş et al. 2019).

Studies have also shown that it is possible to separate haploid seeds using near infrared spectroscopy (Jones et al. 2012; Lin et al. 2017). It has also been suggested that haploid seed classification is possible depending on the seed oil content and seed color, and that the oil content can be determined by a nuclear magnetic resonance (NMR) device without damaging the seed (Melchinger et al. 2013; Melchinger et al. 2014). Haploid identification can not only determine the seed level in maize but also be carried out based on plant measurements. Choe et al. (2012) developed a method based on the measurement of stomatal cell width at the 2-leaf stage to differentiate true haploid and diploid plants. They verified the results by coloration in the stalk and flow cytometric measurements and emphasized that this classification was successful in the 7-leaf period.

Studies on the *in vivo* doubled haploid technique in the current literature are constantly increasing. Investigating the interaction of inducer and donor materials, and developing methods for successful haploid/diploid seed classification are among the most studied topics. However, the fact that the techniques used in studies on haploid/diploid seed classification were carried out according to a single method (eye separation, image processing, spectral techniques, etc.) leaves an important gap in the scientific literature. A detailed comparative study is needed to clearly identify the advantages of one method over another. Furthermore, a comparison of haploid induction rates and the success of tropic and temperate inducers are still open to research.

Accordingly, this study was carried out i) to investigate the reactions of donor materials to different inducers and chromosome doubling methods, and ii) to compare the effectiveness of different techniques used in haploid seed classification.

2. Material and Methods

2.1. Plant material

In this study, 3 donors and one inducer (RWK76/RWS) belonging to AGROMAR Inc. Donors are F2 materials from AGROMAR maize genetic stocks. Other inducer line (CIM2GTAIL-P2) was used with permission from The International Maize and Wheat Improvement Center (CIMMYT) for scientific studies in the Field Crops Department of Çanakkale Onsekiz Mart University (Turkey), Faculty of Agriculture (Table 1).

Table 1- Donor and inducer materials used in this study

| <i>Code</i> | <i>General Features</i> | <i>Source</i> |
|--------------|--------------------------------|---------------|
| DNR1 | Donor material | AGROMAR Inc. |
| DNR4 | Donor material | AGROMAR Inc. |
| DNR5 | Donor material | AGROMAR Inc. |
| CIM2GTAIL-P2 | Second generation inducer line | CIMMYT |
| RWK76/RWS | Inducer hybrids | AGROMAR Inc. |

2.2. Induction crossing

Induction hybridization was carried out under greenhouse conditions in this study. Each donor was planted in pots with 60-70 plants. Transplantation was carried out in September 2019. When the plants reached the flowering stage, pollination was carried

out with at least 10 plants belonging to each donor genotype by using each inducer line as the pollen source (male parent). The controlled pollination method suggested by Kahrıman (2016) was used in the pollination process. The harvest was carried out by hand and DH0 seeds obtained from induction crosses were kept at +4 °C for use in the next steps.

2.3. Seed classification methods

Three different classification methods were used in the study, and haploid identification was performed based on observation by eye, spectral data and image data, respectively. At least 200 seeds (100 haploids and 100 diploids) were separated for each donor material and totally 600 seeds were used for performing seed identification methods. Detailed information about the identification methods used is described below.

The first separation of the seed samples (with the eye) was named "Initial Classification" in the study, and classification was made according to the coloration of the embryo and crown area in DH0 seeds. If there was coloration in the crown area and without coloration in embryo of the seed examined, this seed was considered as haploid, and if there was coloration in both the crown area and the embryo region, this seed was evaluated as diploid. A seed-based labeling method was used in order to prevent confusing seed samples classified by eye separation with the stage of the field evaluation. Seed samples were kept at +4 °C in the labeled seed trays until other measurements were carried out.

Image classification at the single seed level was performed by making some revisions to the method suggested by Altuntaş et al. (2018). The images of the seeds, which were separated by the eye, were recorded by the embryo side in the desktop scanner with jpeg extension. The background of the seed images was black and this was removed before the extraction of image features. A total of 6 different features were extracted to be used in modeling studies based on the image data. These were the pixel values of the R, G, B channels of the whole seed image and the pixel values of the R, G, B channels of the embryo region. Segmentation and feature extraction operations were carried out with the R package program (R Core Team 2019). These data were kept as an excel file to be used in model development studies.

Spectral data (1200-2400 nm interval) were taken from each seed that was classified by eye and image processing (Figure 1). Data were collected in terms of the embryo side of the seeds, and for this purpose a special sample cup for single seed measurement was used in the NIR spectroscopy device (Spectrastar 2400D, Unity Scientific, USA). Spectra acquisition was carried out in the stationary measurement mode. The obtained spectral data were recorded with the label given to the individual seed sample and converted into an excel file to create the spectral models.

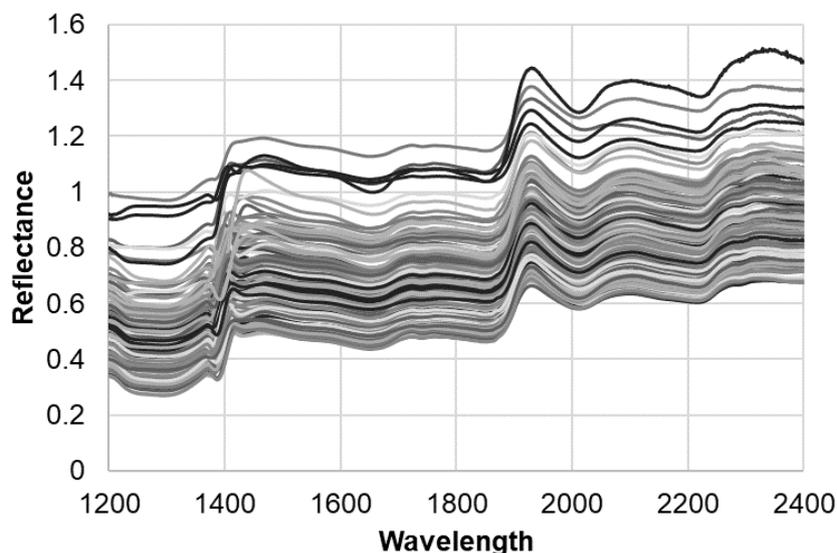


Figure 1- Spectral data obtained from single seed NIR measurements

2.4. Chromosome doubling methods

Two different chromosome doubling methods were used in the study. One of these methods is seedling immersion, which is widely used in the scientific literature (Deimling et al. 1997). The first 50 of the 100 haploid seeds for each donor × inducer cross were germinated in the way that the seed labels were given at the classification stage. After germination, the root tips and coleoptiles (2-3 cm) were cut with a scalpel and kept in a colchicine tank for 12 hours in a solution containing 0.06% colchicine and 0.05% DMSO (Dimethyl sulfoxide). At least 50 seeds per donor remaining from the first method were germinated and the doubling chemical solution was injected into the stem according to the stem injection method proposed by Zabirowa et al. (1996). In the injection, a doubling solution consisting of 100 microliters 0.125% colchicine and 0.5% dimethyl sulfoxide (DMSO) was

injected to stem with a 1 mL sterile syringe at the stage of 2-3 collared leaves (Vanous et al. 2017). After the injection, the plants were kept under greenhouse conditions until they reached the stage to be transplanted in the field. The number of plants that remained alive at the stages of germination, seedling and transplanting to the field was recorded. In both methods, the transplantation process was performed by germinating the diploid seeds without chromosome doubling treatment.

2.5. Growing DH0 plants in the field and determination false classified seeds

At this stage, the colchicine-treated seedlings of haploids and untreated diploid seedlings were transplanted into the field. In the transplanting process, the seedlings obtained from haploid and diploid seeds were transferred to the field in 2-row sub-plots of 5 meters according to the seed labels. One label was used per seedling and the observations made in the field were taken according to these labels, which show the original seed code.

In order to verify the success of the chromosome doubling treatment and the methods used in seed classification, the haploid/diploid plants were subjected to "Final Classification". This classification was made based on the visual state of the plants investigated. Vigorous plants with stem colorization were accepted as diploid, while the plants with erect leaves were classified as haploid. In addition, the success of chromosome doubling process was controlled by checking the fertile plants under field conditions. Records were kept on the labels given to the seeds throughout the whole process, from the beginning of the study to the field stage.

2.6. Statistical analysis

From the calculations made for the first aim of the study, the haploid induction rates and the success of the doubling process according to the state of DH1 plants were determined according to the equations 1, 2, 3, 4, and 5 suggested by Zararsiz et al. (2019).

$$\text{Haploid Induction Rate (HIR)} = \left(\frac{\text{Number of haploid seeds}}{\text{Number of total seeds}} \right) \times 100 \quad (1)$$

$$\text{Haploid Germ. Rate (HGR)} = \left(\frac{\text{No. of haploid seeds} - \text{No. of unger. seeds}}{\text{Num. of total haploid seeds}} \right) \times 100 \quad (2)$$

$$\text{Surviving Seedlings Rate (SSR)} = \left(\frac{\text{Number of transplanted seedlings}}{\text{Number of colchicine-treated seedlings}} \right) \times 100 \quad (3)$$

$$\text{Surviving Plant Ratio (SPR)} = \left(\frac{\text{Number of surviving plants}}{\text{Number of transplanted seedlings}} \right) \times 100 \quad (4)$$

$$\text{Chromosome Duplication Rate (CDR)} = \left(\frac{\text{Number of fertile plants}}{\text{Number of selected plants}} \right) \times 100 \quad (5)$$

For the second aim of the study, the evaluation statistics were calculated over the confusion matrix between the results obtained from the initial classification made by eye and the final classification performed in the field. Initial and final classification data were used as dependent variables in developing classification models based on images and spectral data. In order to develop spectral models, two separate classification models (Spectral Model 1 and Spectral Model 2) were created by using the support vector machines (SVM) method. In these models, spectral data between 1200-2400 nm were used as the predictive variable, and initial classification and final classification data as the predicted variable.

For the purpose of model development based on image data, 6 different features extracted from the images were taken as predictive variables, and two separate classification models (Image Model 1 and Image Model 2) were created using SVM method. Additionally, the Classification and Regression Tree (CRT) method was used in order to determine the features that were effective in haploid/diploid seed classification and the limit values of the features extracted from image data. All analyses were made in the R program (R Core Team 2019). For comparison of the created estimation models, the calculations made over the confusion matrix (Table 2) were used.

Table 2- Confusion matrix template used in calculations for classification of haploid and diploid seeds

| Prediction | Actual | |
|------------|---------|---------|
| | Haploid | Diploid |
| Haploid | TP | FP |
| Diploid | FN | TN |

The false discovery rate (FDR) and false negative rate (FNR) were determined using the confusion matrix using Equations 6 and 7 suggested by Melchinger et al. (2013). In addition, the calculations for the classification success of the models were followed by the sensitivity, specificity, prevalence, positive predictive value, negative predictive value, detection rate, detection prevalence, balanced accuracy, and accuracy values over the confusion matrix, calculated according to the equations 8-16.

$$FNR = \frac{FN}{TP+FN} \quad (6)$$

$$FDR = \frac{FP}{TP+FP} \quad (7)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (8)$$

$$Specificity = \frac{TN}{FP+TN} \quad (9)$$

$$Prevalence = \frac{TP+FN}{TP+FN+FP+TN} \quad (10)$$

$$Pos. Pred. Val. (PPV) = \frac{TP}{TP+FP} \quad (11)$$

$$Neg. Pred. Val (NPV) = \frac{TN}{TN+FN} \quad (12)$$

$$Detection Rate = \frac{TP}{TP+FN+FP+TN} \quad (13)$$

$$Detection Prevalence = \frac{TP+FP}{TP+FN+FP+TN} \quad (14)$$

$$Balanced Accuracy = \frac{Sensitivity+Specificity}{2} \quad (15)$$

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (16)$$

In addition to these calculations, Spearman Rank Correlation was applied in the R program (R Core Team 2019) in order to determine the relationship between model predictions and the actual case. The correlation coefficients obtained are shown in the evaluation table of the model statistics.

3. Results and Discussion

3.1. Donor x inducer interaction on success of haploid induction

The haploid induction rates in this study varied between 9.2% and 16.1% (Table 3). Since HIR calculations were made on all seeds obtained from induction cross, they were not calculated separately according to chromosome doubling methods. It was observed that DNR1 and DNR5 provided higher HIR with RWK76 / RWS inducer line and DNR4 with CIM2GTAIL-P2 inducer line than other induction crosses. After randomly separating two subgroups of haploid seed samples for each chromosome doubling methods, haploid germination rates (HGR) varied between 64% and 100% in seedling immersion method and 72% to 93% in stem injection method. The stem injection method had a higher average for the surviving seedling rate (SSR) than the seedling immersion method. This continued in the plants surviving rate (PSR) in the field and stem injection method was also found to be advantageous compared to seedling immersion method. Chromosome doubling rates of the donors used were found to be higher in seedling immersion method for all donors except for the DNR4xCIM2GTAIL-P2 cross (Table 3).

Table 3- HIR (%), HGR (%), SSR (%), SPR (%) and CDR (%) values for donor materials according to chromosome doubling methods

| Chromosome Doubling Method | Induction Cross | HIR [†] | HGR | SSR | SPR | CDR |
|----------------------------|-------------------|------------------|-------|-------|-------|------|
| Seedling Immersion Method | DNR1xCIM2GTAIL-P2 | 9.20 | 80.0 | 56.0 | 88.0 | 46.7 |
| | DNR1xRWK76/RWS | 10.50 | 77.0 | 85.0 | 100.0 | 50.0 |
| | DNR4xCIM2GTAIL-P2 | 13.80 | 100.0 | 58.0 | 60.0 | 46.2 |
| | DNR4xRWK76/RWS | 10.40 | 98.0 | 63.0 | 82.0 | 39.5 |
| | DNR5xCIM2GTAIL-P2 | 9.70 | 64.0 | 47.0 | 87.0 | 42.9 |
| | DNR5xRWK76/RWS | 16.10 | 93.0 | 68.0 | 100.0 | 45.9 |
| | Mean | 11.62 | 85.3 | 62.8 | 86.2 | 45.2 |
| Stem Injection Method | DNR1xCIM2GTAIL-P2 | 9.2 | 85.0 | 100.0 | 100.0 | 35.7 |
| | DNR1xRWK76/RWS | 10.50 | 88.0 | 88.8 | 100.0 | 45.7 |
| | DNR4xCIM2GTAIL-P2 | 13.80 | 76.0 | 100.0 | 100.0 | 47.2 |
| | DNR4xRWK76/RWS | 10.40 | 93.0 | 100.0 | 85.0 | 32.3 |
| | DNR5xCIM2GTAIL-P2 | 9.70 | 88.0 | 100.0 | 90.0 | 39.6 |
| | DNR5xRWK76/RWS | 16.10 | 72.0 | 100.0 | 92.0 | 38.0 |
| | Mean | 12.10 | 83.7 | 98.1 | 94.5 | 39.7 |

[†]: HIR rate is the same for both chromosome doubling methods due to the HIR calculation based on haploid/diploid seed classification at the beginning of the study

HIR values for the in vivo maternal haploid technique varied between 2.8% and 12.8% in the scientific literature (Eder & Chalyk 2002; Prigge et al. 2012; Cerit et al. 2016; Zararsız et al. 2019). It was emphasized that the HIR varies according to the donors and inducer lines used. The results obtained in the current study are also in agreement with this. The germination rates of haploid seeds are relatively low compared to diploids. In our study, the results obtained during the germination stage are similar to those obtained in previous studies. It was observed that the chromosome doubling treatment had a significant effect on plant viability. Seedling immersion method was found to be disadvantageous in terms of the number of seedlings and plants surviving both in the seedling and field stage compared to stem injection method. In other studies, the number of seedlings and plants surviving with the seedling immersion method were found to be between 57.9% and 72.9%, respectively (Zararsız et al. 2019). These results can be attributed to the adverse effects of root tip cutting on plant growth and survival rate.

Chromosome doubling rates (CDR) are ultimately an important parameter in the in vivo maternal haploid technique. In our study, chromosome doubling rates ranged from 32.3% to 50.0% and these results were similar to the values reported in the current literature (Prasanna et al. 2012; Zararsız et al. 2019). On the other hand, it has been observed that there are significant differences between induction crosses. According to all assessments, the interaction of donor and inducer lines, as well as their reactions to chromosome doubling methods used in the in vivo maternal haploid technique, displayed significant differences.

3.2. Comparison of seed classification techniques

The confusion matrix of the seed classification results is presented in Table 4. In this matrix, haploids are taken as positive groups and diploids as negative groups. The results for initial classification and final classification showed that eye classification had higher rate of misclassification in haploid seeds than diploids (Table 3). Based on initial classification results, 55 of the 245 haploid seeds and 13 out of 300 diploid seeds were misclassified. Considering the numbers in the final classification, 300 diploid and 245 haploid seeds allocated at the initial classification were determined as 342 diploid and 203 haploid seeds. Dang et al. (2012) attributed the increase of misclassification in the eye method to the fact that the donor material did not adequately show the pigmentation resulting from the inducer line or had a genetic structure of donor material that prevented pigment formation. In our study, the emergence of haploid plants from diploids can be attributed to the adequate selection of pigmentation in the endosperm/embryo region or to spontaneous haploid formation. On the other hand, it is understood that the formation of pigmentation in the embryo region of haploid seeds cannot be distinguished by the eye, and as a result, it is possible to classify diploid seeds as haploid. As an alternative method in seed classification, spectral and image processing data are treated as predictive, and the initial and final classification data made by the eye are taken as dependent variables, whereby FP values are higher than visual classification (Table 4). Similarly, FN values are higher than the erroneous seed numbers in classifications made by eye.

Table 4- Confusion matrix values for eye, spectral and image classification

| <i>Seed Identification Method</i> | <i>FN</i> | <i>FP</i> | <i>TP</i> | <i>TN</i> |
|-----------------------------------|-----------|-----------|-----------|-----------|
| Eye | 55 | 13 | 190 | 287 |
| Spectral Model 1 | 65 | 23 | 180 | 277 |
| Spectral Model 2 | 69 | 28 | 134 | 314 |
| Image Model 1 | 81 | 39 | 164 | 261 |
| Image Model 2 | 79 | 32 | 124 | 310 |
| Mean | 70 | 27 | 158 | 270 |

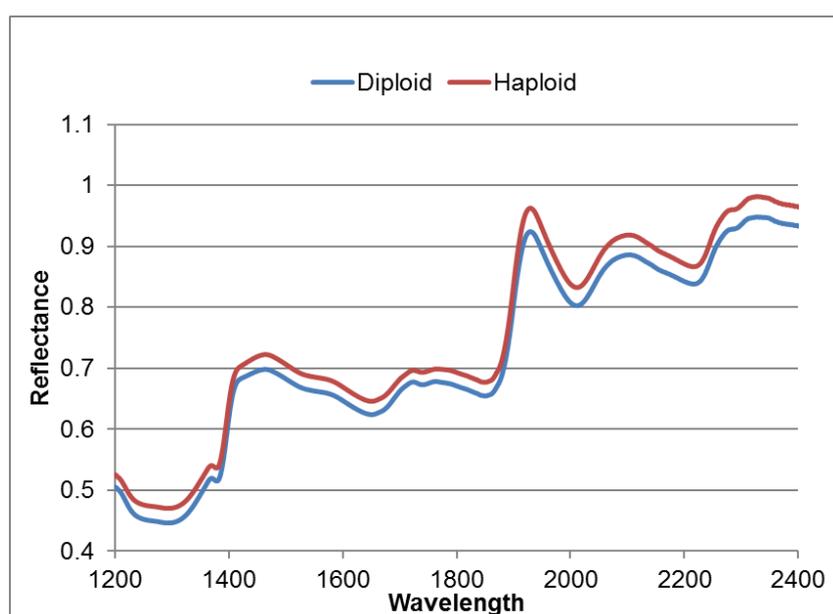
FP: False positive, FN: False negative, TP: True positive, TN: True negative

According to the statistics calculated over the confusion matrix (Table 5), eye classification had a higher accuracy (87.50%) than the accuracy from the spectral models (Spectral Model 1= 0.839, Spectral Model 2= 0.822) and image models (Image Model 1= 0.780, Image Model 2= 0.796). For all models, sensitivity values (Eye= 0.776, Spectral Model 1= 0.735, Spectral Model 2= 0.660, Image Model 1= 0.669, Image Model 2= 0.611), which show the proportion of correctly classified seeds separated as haploid, were lower than specificity values (Eye= 0.957, Spectral Model 1= 0.923, Spectral Model 2= 0.918, Image Model 1= 0.870, Image Model 2 = 0.906), which are related to the proportion of correctly classified diploid seeds (Table 5). This was confirmed by the FNR and FDR figures calculated on the confusion matrix, and the FNR values were found to be higher than the FDR values for all classification models. Considering the rank correlations between the actual and predicted classes, eye classification ($r = 0.753$) was observed to have higher similarity than the classification based on spectral models (r for Spectral Model 1= 0.667, r for Spectral Model 2= 0.612) and image analysis (r for Image Model 1= 0.552, r for Image Model 2= 0.553).

Table 5- Evaluation/statistics based on confusion matrix for eye, spectral and image techniques

| Statistics | Eye | Spectral Model 1 | Spectral Model 2 | Image Model 1 | Image Model 2 |
|---------------------------|-------|------------------|------------------|---------------|---------------|
| Sensitivity | 0.776 | 0.735 | 0.660 | 0.669 | 0.611 |
| Specificity | 0.957 | 0.923 | 0.918 | 0.870 | 0.906 |
| Positive Prediction Value | 0.936 | 0.887 | 0.827 | 0.808 | 0.795 |
| Neg. Prediction Value | 0.839 | 0.809 | 0.820 | 0.763 | 0.797 |
| Prevalence | 0.450 | 0.450 | 0.372 | 0.450 | 0.372 |
| Detection Rate | 0.349 | 0.330 | 0.246 | 0.301 | 0.228 |
| Detection Prevalence | 0.372 | 0.372 | 0.297 | 0.372 | 0.286 |
| Balanced Accuracy | 0.866 | 0.829 | 0.789 | 0.770 | 0.759 |
| Accuracy | 0.875 | 0.839 | 0.822 | 0.780 | 0.796 |
| False Discovery Rate | 0.064 | 0.113 | 0.173 | 0.192 | 0.205 |
| False Negative Rate | 0.224 | 0.265 | 0.340 | 0.331 | 0.389 |
| Rank Correlation | 0.753 | 0.667 | 0.612 | 0.552 | 0.553 |

When the classification methods based on spectral measurement and image analysis (used as alternative methods) were examined, it was observed that the spectral models give more successful results. This result obtained from the spectral models can be attributed to the characteristic features in the spectral data from haploid and diploid seeds. It is seen that the average spectra of haploid seeds are higher than the spectra from diploid seeds throughout the scanned spectral range (1200-2400 nm) (Figure 2). One of the main reasons for this difference is the existence of anthocyanin coloration in the embryo region of the haploid and diploid seeds. In our study, spectral data were collected by the embryo side that is distinctive for haploid/diploid seed classification based on anthocyanin coloration. While coloration is seen in the embryo region in diploid seeds, it does not occur in haploid seeds. Therefore, some of the light energy sent from the spectroscopy device absorbed in the diploid seeds is likely to be higher and lower in haploids. As a result of this, the average reflectance of haploid seeds was higher than that of the diploids.

**Figure 2- Mean spectra obtained from haploid and diploid seed classes**

There are various studies in the literature that support these findings. In a study conducted by Lin et al. (2017), it was reported that haploid seeds have higher reflectance values than diploids in the spectral range of 900-1700 nm. Our results also substantiate these findings. The true classification rates of spectral estimation models were found to be relatively lower than other studies conducted for similar purposes. For example, Liu et al. (2017) found the true classification rate to be 95% based on creating a spectral model using SVM. In a different study, it was reported that over 90% success was achieved in the classification of haploid seeds with NIR spectroscopy (Cui et al. 2019).

The reason for the difference between results can be attributed to several issues. Spectral pre-treatment was applied in the study of Liu et al. (2017). In our study, spectral models were created without any data pre-treatment. On the other hand, there are differences between the measurement specialities of the devices used in these studies. The device used in our study is not suitable for contact measurement and can obtain spectral data in closed measurement mode. Therefore, it is not possible for the light energy sent to the seed surface during measurement to provide regular reflection, as in contact measurements. However, according to the results obtained in our study, the device used achieved an acceptable success (~ 82%) in the haploid/diploid seed classification.

The image analysis method used in the study was the method with the lowest success among the classification techniques compared. One of the main reasons for this may be the use of a standard desktop scanner for collecting the seed images to develop image models. Although this device offers a practical solution in image collection, it is observed that there are unexpected changes in the color channels (R, G, B) of the seed images, except for seed coloration (Figure 3). The low accuracy in models based on these data can be attributed to measurement methodology.

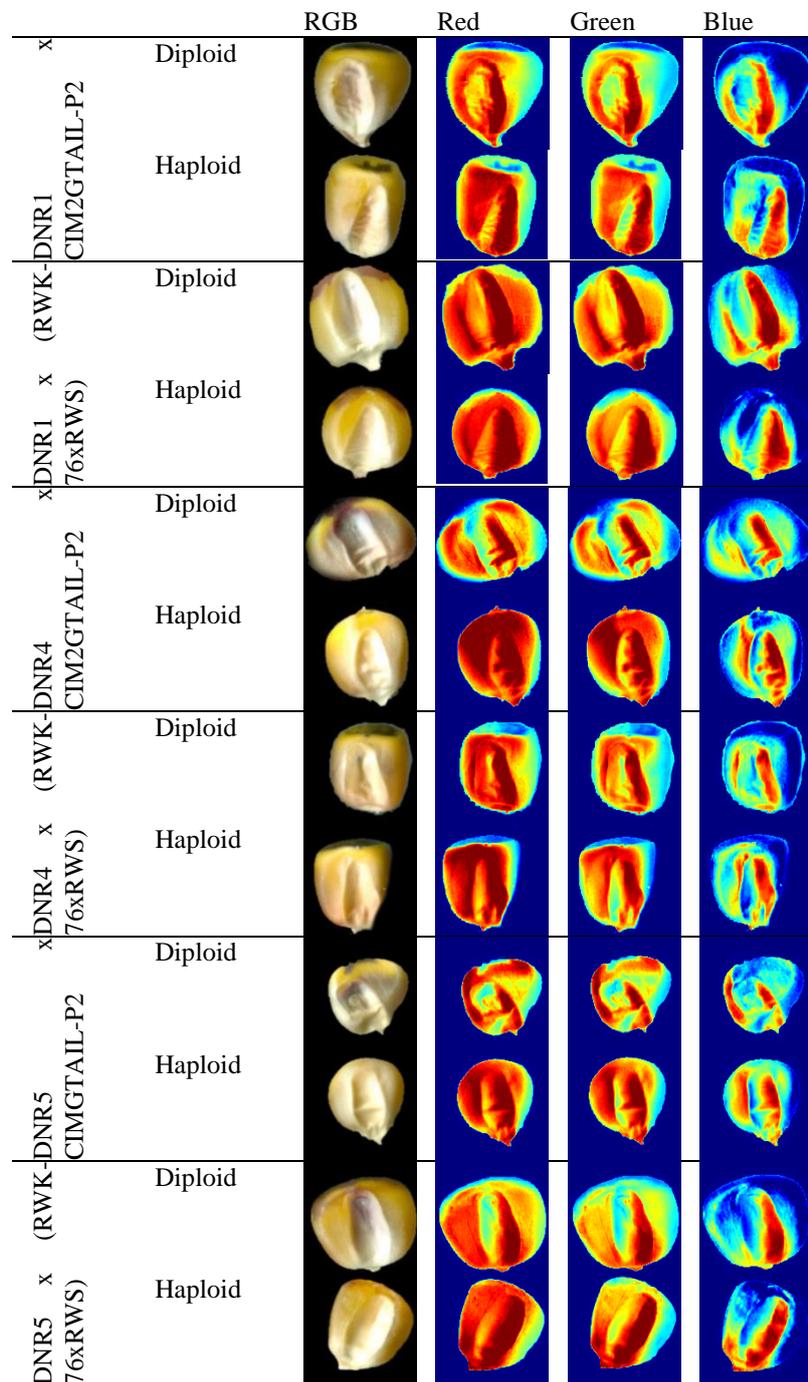


Figure 3- Examples of haploid and diploid seed images of donor materials with RGB, R, G, and B channels

Various studies may be found in the scientific literature on haploid/diploid seed classification with the help of image processing. Veeramani et al. (2018) used SVM, Random Forest (RF) and Logistic Regression (LR) methods to distinguish haploid and diploid seeds by using images taken by the embryo side of seed samples. Researchers achieved 87.6% success in seed classification with the SVM method. In another study in which image data and RGB channels were used in haploid/diploid seed classification, the classification success was 98.04% and 94.4% for haploid and hybrid seeds, respectively (Zhang et al. 2013). It is seen that the correct classification rates of the models created in our study are relatively low when compared with previous studies. This is due to differences in the image acquisition devices used and the method for processing image data. In

our study, data obtained from a desktop scanner was used while high resolution imaging systems were used in other studies. This may have caused a difference in classification success.

On the other hand, it has been reported that automatic classification of haploid/diploid seeds is possible with machine-assisted image processing (Zhang et al. 2013). Integration of models created for image data input into a specific software or machines suitable for automation is possible. To do this, it is important to determine the effect of color channels and the variables that should be included in the models to be created. In our study, the initial classification and final classification data were modeled with CRT analysis to examine the effect of the data of R, G, B color channels taken from the whole seed and embryo region on the classification of haploid/diploid seeds.

CRT output created with the initial classification results, 13 different nodes were formed (Figure 4). Of these, nodes 1, 2, 6, and 12 show the current path to differentiate diploid seeds based on image processing data, while nodes 3, 7 and 13 show the current path to differentiate haploid seeds (Figure 4). According to the CRT graph, it is seen that haploid and diploid seeds can be classified with the help of this model according to the embryo red band value (around 0.9) and the blue band value of the embryo region (0.73). Seed samples were classified under 29 separate nodes in the CRT graph (Figure 5) created based on the final classification data. Of these, nodes 1, 2, 6, 14, and 28 were used for the classification of diploid seeds, and nodes 3, 7, 15 and 29 for the classification of haploid seeds (Figure 5). Considering the effect of the variables of color channels on classification, it is seen that the red band value of the embryo region, the blue band value of the whole seed and the embryo region are effective variables to discriminate haploid/diploid seeds. According to the CRT analysis, it can be said that the red band value and the blue band value of the embryo region are effective features in distinguishing haploid/diploid seeds based on both the initial and final classification. In the “Final Classification”, the red band value of the whole seed is the main feature for distinguishing the seed classes.

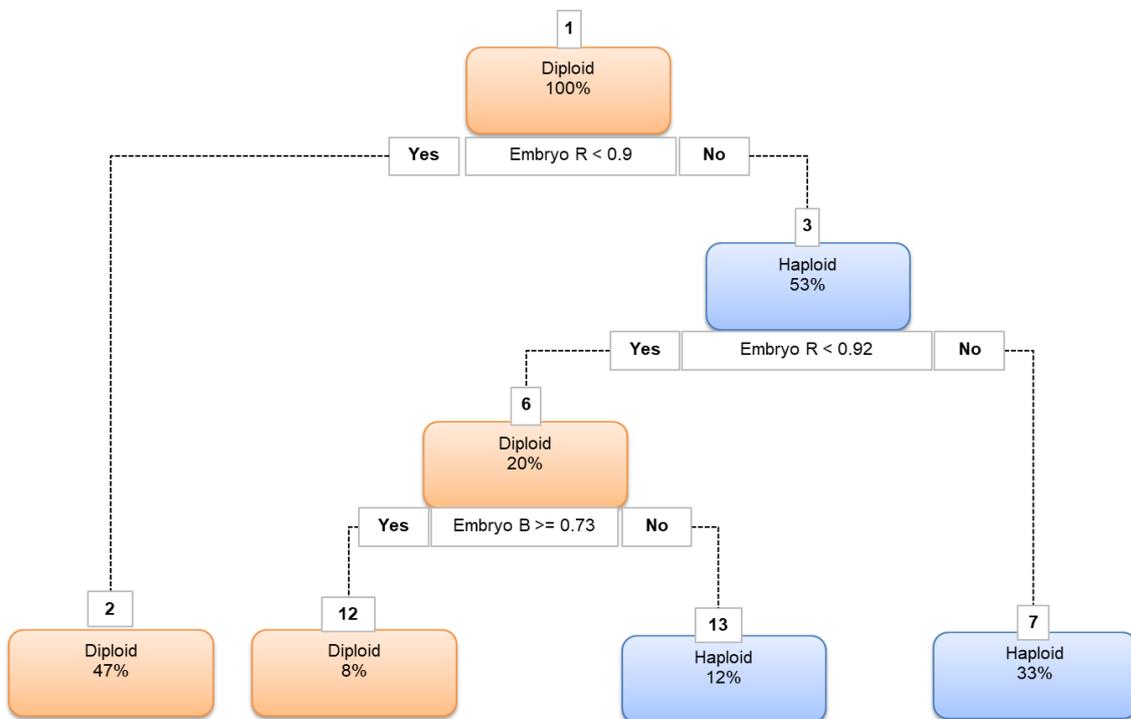


Figure 4- CRT plot generated from seed images based on initial classification of haploid and diploid seed samples obtained from induction crosses

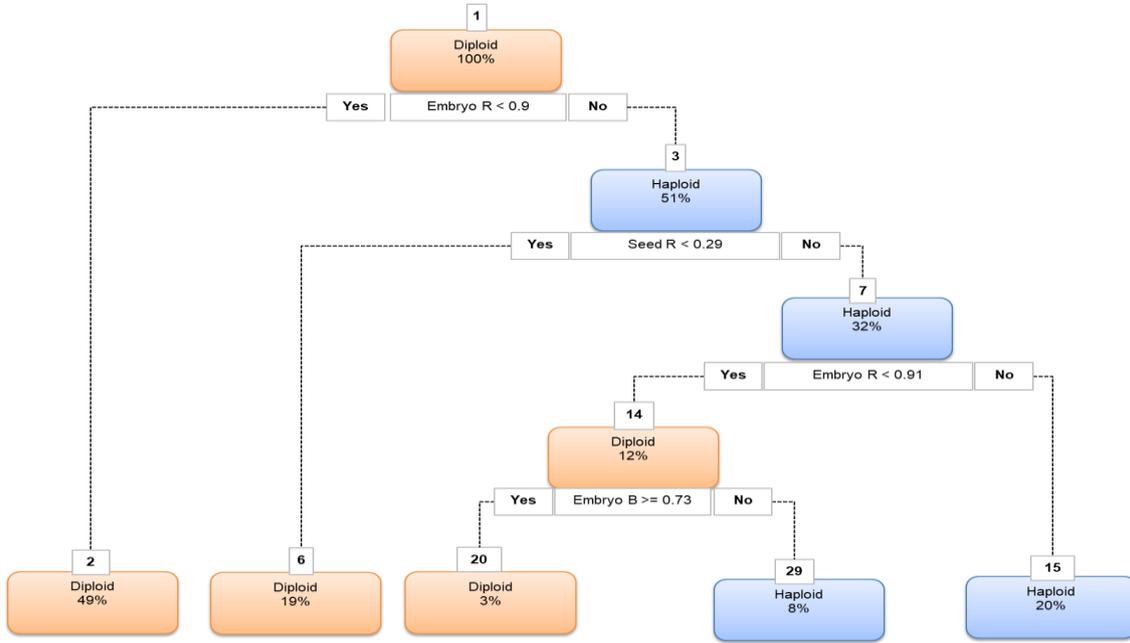


Figure 5- CRT plot generated from seed images based on final classification of haploid and diploid seed samples obtained from induction crosses

4. Conclusions

The results of this study revealed that the donor material, inducer line and chromosome doubling methods have important effects on the success of using the in vivo maternal haploid technique in maize. According to the haploid induction rates and viability levels of the donors used in the study, DNR4 gave more successful results than the others. Among the inducer lines, RWK76xRWS shows good performance compared to CIM2GTAIL-P2. Regarding the chromosome doubling methods compared, the seedling immersion method had a negative effect on the viability of the induction crosses. On the other hand, the number of fertile plants was found to be higher with the seedling immersion method than the stem injection method. The results of the current study are beneficial for researchers who may prefer these methods in further studies. Also, it should be considered that colchicine treatment requires the special setup or controlling areas in both methods. Results indicates that cutting of root and coleoptile parts has a negative effect on the success of seedling immersion method.

Significant differences were observed in the classification success of haploid/diploid seed classification methods compared in the study. The average accuracy of eye separation (87.5%) was higher than the spectral method and image processing methods. Despite this, it was observed that the models created based on spectral and image processing analyses give satisfactory results. It appears possible to develop more successful models by increasing the number of materials/samples used in future studies using powerful spectral and image analysis devices from which the measurements are taken and applying chemometric techniques to the statistical analysis methods used.

Acknowledgements

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) (Project number: 119O619) The authors thank TÜBİTAK for their financial support.

Abbreviations and Symbols

| | |
|-----|------------------------------------|
| CRT | Classification and Regression Tree |
| DH | Doubled Haploid |
| SVM | Support Vector Machine |
| NIR | Near Infrared Reflectance |
| NMR | Nuclear Magnetic Resonance |
| RGB | Red, Blue, Green |

References

Altuntaş Y & Kocamaz AF (2019). Renk momentleri ve destek vektör makineleri kullanarak haploid mısır tohumlarının tanımlanmasında renk uzaylarının sınıflandırma performansına etkisinin karşılaştırılması. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi* 31(2): 551-560. <https://doi.org/10.35234/fumbd.585312>

- Altuntaş Y, Cömert Z & Kocamaz A F (2019). Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach. *Computers and Electronics in Agriculture* 163: 1-11. <https://doi.org/10.1016/j.compag.2019.104874>
- Altuntaş Y, Kocamaz A F, Cengiz R & Esmeray M (2018). Classification of haploid and diploid maize seeds by using image processing techniques and support vector machines. In 2018 26th Signal Processing and Communications Applications Conference (SIU) 2-5 May, İzmir, pp. 1-4. <https://doi.org/10.1109/siu.2018.8404800>
- Cerit İ, Cömertpay G, Oyucu R, Çakır B, Hatipoğlu R & Özkan H (2016). Melez mısırs ıslahında in-vivo katlanmış haploid tekniğinde kullanılan farklı inducer genotiplerin haploid indirgeme oranların belirlenmesi. *Tarla Bitkileri Merkez Araştırma Enstitüsü Dergisi* 25: 52-57. <https://doi.org/10.21566/tarbitderg.280162>
- Chaikam V, Molenaar W, Melchinger A E & Boddupalli P M (2019). Doubled haploid technology for line development in maize: technical advances and prospects. *Theoretical and Applied Genetics* 132(12): 3227-3243. <https://doi.org/10.1007/s00122-019-03433-x>
- Chalky S T (1994). Properties of maternal haploid maize plants and potential application to maize breeding. *Euphytica* 79: 13-18. <https://doi.org/10.1007/bf00023571>
- Chidzanga C, Muzawazi F, Midzi J & Hove T (2017). Production and use of haploids and doubled haploid in maize breeding: A review. *African Journal of Plant Breeding* 4: 201-213.
- Choe E H (1959). A line of maize with high haploid frequency. *American Naturalist* 93:381-382. <https://doi.org/10.1086/282098>
- Choe E, Carbonero CH, Mulvaney K, Rayburn AL & Muhm RH (2012). Improving in vivo maize doubled haploid production efficiency through early detection of false positives. *Plant Breeding* 131: 399-401. <https://doi.org/10.1111/j.1439-0523.2012.01962.x>
- Cui Y, Ge W, Li J, Zhang J, An D & Wei Y (2019). Screening of maize haploid kernels based on near infrared spectroscopy quantitative analysis. *Computers and Electronics in Agriculture* 158: 358-368. <https://doi.org/10.1016/j.compag.2019.01.038>
- Dang N C, Munsch M, Aulinger I, Renlai W & Stamp P (2012). Inducer line generated double haploid seeds for combined waxy and opaque 2 grain quality in subtropical maize (*Zea mays*. L.). *Euphytica* 183(2): 153-160. <https://doi.org/10.1007/s10681-011-0423-0>
- Deimling S, Röber F K & Geiger H H (1997). Methodology and genetics of in vivo haploid induction in maize. *Vortr Pflanzenzuchtg* 38: 203-224.
- Eder J & Chalys S (2002). In vivo haploid induction in maize. *Theoretical and Applied Genetics* 104(4): 703-708. <https://doi.org/10.1007/s00122-001-0773-4>
- Hallauer A R, Carena M J & Miranda F J B (2010). *Quantitative Genetics in Maize Breeding*. Springer New York. <https://doi.org/10.1007/978-1-4419-0766-0>
- Jones R W, Reniot T, Frei U K, Tseng Y, Lubberstedt T & McClelland J F (2012). Selection of haploid maize kernels from hybrid kernels for plant breeding using near infrared spectroscopy and SIMCA analysis. *Applied Spectroscopy* 66: 447-450. <https://doi.org/10.1366/11-06426>
- Kahrman F (2016). Mısırd a Polen Etkisi ve Bu Etkinin Kontrolünde Uygulanan Yöntemler. Lambert Academic Publishing Saarbrücken, Almanya.
- Kalinowska K, Chamas S, Unkel K, Demidov D, Lermontova I, Dresselhaus T, Kumlehn J, Dunemann F, Houben A (2019). State-of-the-art and novel developments of in vivo haploid technologies. *Theoretical and Applied Genetics* 132(3):593-605. <https://doi.org/10.1007/s00122-018-3261-9>
- Lin J, Yu L, Li W & Qin H (2017). Method for identifying maize haploid seeds by applying diffuse transmission near-infrared spectroscopy. *Applied Spectroscopy* 72(4): 611-617. <https://doi.org/10.1177/0003702817742790>
- Liu W J, Li W J, Li H G, Qin H & Ning X (2017). Research on the method of identifying maize haploid based on KPCA and near infrared spectrum. *Spectroscopy and Spectral Analysis* 37(7): 2024-2027. [https://doi.org/10.3964/j.issn.1000-0593\(2017\)07-2024-04](https://doi.org/10.3964/j.issn.1000-0593(2017)07-2024-04)
- Melchinger A E, Schipprack W, Würschum T, Chen S & Technow F (2013). Rapid and accurate identification of in vivo-induced haploid seeds based on oil content in maize. *Scientific Reports* 3: 2129. <https://doi.org/10.1038/srep02129>
- Melchinger A E, Schipprack W, Utz H F & Mirdita V (2014). In vivo haploid induction in maize: identification of haploid seeds by their oil content. *Crop Science* 54: 1-8. <https://doi.org/10.2135/cropsci2013.12.0851>
- Prasanna BM, Chaikam V & Mahuku G (2012). Doubled haploid technology in maize breeding: theory and practice. *CIMMYT*.
- Prigge V, Schipprack W, Mahuku G, Atlin G N & Melchinger A E (2012). Development of in vivo haploid inducers for tropical maize breeding programs. *Euphytica* 185(3): 481-490. <https://doi.org/10.1007/s10681-012-0657-5>
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Röber F K, Gordillo G A & Geiger H H (2005). In vivo haploid induction in maize-performance of new inducers and significance of doubled haploid lines in hybrid breeding. *Maydica* 50: 275-283
- Uliana Trentin H, Frei U K & Lübberstedt T (2020) Breeding maize maternal haploid inducers. *Plants (Basel)* 9(5):614. <https://doi.org/10.3390/plants9050614>
- Vanous K, Vanous A, Frei UK & Lübberstedt T (2017). Generation of maize (*Zea mays*) doubled haploids via traditional methods. *Current Protocols in Plant Biology* 2(2): 147-157. <https://doi.org/10.1002/cppb.20050>
- Veeramani B, Raymond J W & Chanda P (2018). DeepSort: deep convolutional networks for sorting haploid maize seeds. *BMC Bioinformatics* 19(9): 289. <https://doi.org/10.1186/s12859-018-2267-2>
- Zabirova E R, Chumak M V, Shatskaia O A & Scherbak V S (1996). Technology of the mass accelerated production of homozygous lines (in Russian). *Kukuruza Sorgo* 4: 17-19
- Zararsız D, Öztürk L, Yanıkoğlu S, Turgut I, Kizik S & Bilgin B (2019). Production of double haploid plants using in vivo haploid techniques in corn. *Journal of Agricultural Sciences* 25(1): 62-69. <https://doi.org/10.15832/ankutbd.539000>
- Zhang J, Wu Z, Song P, Li W, Chen S & Liu J (2013). Embryo feature extraction and dynamic recognition method for maize haploid seeds. *Transactions of the Chinese Society of Agricultural Engineering* 29(4): 199-203. <https://doi.org/10.3969/j.issn.1002-6819.2013.04.025>

