AgroLife Scientific Journal - Volume 14, Number 1, 2025
ISSN 2285-5718; ISSN CD-ROM 2285-5726; ISSN ONLINE 2286-0126; ISSN-L 2285-5718

EVALUATION DIFFERENCES BETWEEN HAPLOID AND DIPLOID MAIZE SAMPLES AND CUT-OFF POINTS DETERMINATION FOR SEPARATION HAPLOIDS USING DECISION TREE TECHNIQUE

Talha TUNÇ, Fatih KAHRIMAN

Çanakkale Onsekiz Mart University, Faculty of Agriculture, Department of Field Crops, Çanakkale, Türkiye

Corresponding author email: fkahriman@hotmail.com

Abstract

This study aims to assess the discrimination success of these measurements on seeds, germinated seeds, and seedlings of haploid and diploid maize samples, using cytogenetic verification by acetocarmin staining. This study was conducted in the Laboratory of Field Crops, Çanakkale Onsekiz Mart University. The decision tree method was used to determine which of the measured traits was effective in differentiating haploid/diploid samples. The results of study showed that the success of visual discrimination according to Navajo marker was 85% for haploid samples and 96% for diploid samples. Differences between haploid and diploid samples varied by donor material, with hybrid donors showing more pronounced disparities. Decision tree analysis revealed that the root length and seed circumference as effective traits to distinguish haploids from diploids. In conclusion, visual differentiation based only seed coloration can be misleading, highlighting the importance of multiple measurements for accurate haploid identification using the in vivo doubled haploid technique.

Key words*: ploidy level, homozygous line, machine learning, regression, Zea mays.*

INTRODUCTION

Maize holds significant economic value owing to its multifaceted applications in human nutrition, industry, and livestock feed, underpinned by its superior yield potential and extensive genetic and physiological diversity (Moldovan & Bărbos, 2019). One of the basic stages in hybrid maize breeding is the development of the parent lines (Cerit et al., 2016). With classic plant breeding approaches, the development of parent lines takes a long time, the desired homozygotic levels may not be obtained and it requires high labor power and money. As a result, plant breeders use new breeding technologies to achieve the desired levels of homozygotes, to save time and reduce costs when developing parent lines (Cengiz & Korkut, 2016). In maize breeding studies, techniques developed as an alternative to classic methods may be classified as *in vitro* and *in vivo* techniques.

With the *in vitro* method, the technique to obtain the haploid is based on laboratory studies and is completed with a variety of methods like applying heat to the anther, microspore or pollen (Mathur et al., 1980), radiation treatment of pollen samples (Mathur et al., 1976), or chemical treatment of ear silks (Zuoyo & Mingguang, 1984). However, these methods did not achieve high success due to genetic effects (Cerit et al., 2016). The in vivo doubled haploid technique provides more reliable outcomes in a shorter duration. As a result, interest in the *in vivo* doubled haploid technique is higher. Studies about obtaining homozygote lines using the *in vivo* haploid technique date back decades (Chase, 1969). Application of this technique is based on the use of special lines called inducers. The *in vivo* doubled haploid technique allows the opportunity to create haploid seeds on pollinated plants. According to the form of pollination, the *in vivo* doubled haploid technique may be applied in two ways: maternal and paternal. With the maternal haploid method, the inducer line is used as a pollinator, while in the paternal haploid method, the inducer line is used as a donor (Coe, 1959; Kermicle, 1969). In terms of haploid inducer rate, the *in vivo* maternal haploid technique is more successful compared to the paternal haploid technique (Lashermes & Beckert, 1988). In both methods, four steps are applied. These are induction

hybridization, haploid selection, chromosome doubling, and obtaining DH0 seeds. Among these steps, haploid selection has the most significant effect on success of the technique and different methods are used to separate haploid samples from other samples.

In the *in vivo* doubled haploid technique, markers linked to anthocyanin colors are used for visual separation of haploid samples (Cengiz & Korkut, 2016). If seeds obtained after induction hybridization only have endosperm coloring, the seed sample is haploid; if there is embryo and endosperm coloring, it is accepted as diploid. Seed coloring is controlled by the R1-nj gene and aleurone and scutellum cause deep coloring (Geiger & Gordillo, 2009). This coloring provides the opportunity to easily and rapidly identify haploid seeds in the seed stage during induction hybridization (Chaikam et al., 2015). Haploid-diploid seed differentiation performed based on color markers by eye can be completed along with different methods like measuring pollen sizes and determining chromosome numbers in karyotypic studies with a flow cytometric device (Röber et al., 2005; Cerit et al., 2016). The direct classification level with these methods has different advantages and disadvantages in terms of practicality or cost in practice. In the scientific literature, there is a variety of research using one or more methods for haploid differentiation in maize studies.

In a study by Battistelli et al. (2013), using 6 different hybrids with the KEMS inducer line and F2 generations of these hybrids as material, the aim was to differentiate haploid/diploid plants with the aid of Navajo color marker, flow cytometry and molecular markers. They pointed out that the use of flow cytometry and SSR markers was effective in verifying ploidy levels of samples. A study performed by Chaikam et al. (2017) researched the differentiation potential in terms of germination features of haploid and diploid seeds obtained with the in vivo maternal haploid technique. In the study, measurements were collected for root length, coleoptile length, and number of side roots. The research findings indicated that germination features may be used for differentiation of haploid samples, independent of the genetic marker system. Ribeiro et al. (2018) performed a study with the aim of distinguishing haploid and diploid with morphological, cytogenetic, molecular and flow cytometry techniques for samples obtained from hybridization of a single hybrid maize and an inducer line. In the study, they found that morphological measurements may give misleading results, and that molecular marker and flow cytometry measurements provided reliable results. Research by Molenaar et al. (2019) used stomal length and flow cytometry measurements to classify samples in the early period with the in vivo doubled haploid technique. In the study, they considered whether haploid, diploid and hybrid seeds could be differentiated from each other or not based on these measurements. The results of the study determined that there may be overlaps between hybrid seed measurements and haploid seed measurements.

Machine learning methods classified under artificial intelligence, and they are used for different purposes including medicine, agriculture etc. There are various techniques, but tree-based methods have some advantages over other methods. Decision tree method a supervised learning algorithm that is used for classification and regression modelling, have a potential to determine threshold values for classification studies (Breiman et al., 1984). This technique has a potential to determine effective variables for classification as well as threshold values for classification tasks. Indeed, this method has been used to develop supervised models based on image data for discrimination of haploids (Dönmez, 2020; Kahrıman et al., 2023). To the best of our knowledge, previous studies have inferred haploid and diploid samples based on manual measurements, but machine learning methods have not been directly used to identify the features that are effective in discriminating haploids from others. Decision tree methods can be used for this purpose effectively.

This study aimed to investigate the difference in haploid and diploid samples based on morphological, physiological, and cytological measurements for the in vivo doubled haploid technique. Additionally, a secondary aim was to identify features that may be used efficiently for differentiation of haploid samples among the investigated features using decision tree method which is one of machine learning technique.

MATERIALS AND METHODS

Plant materials

As material in the study, an inducer line and 2 donor genotypes were used. General information related to this material is presented in Table 1. The inducer line (CIM2TAIL-P2) was obtained from CIMMYT for scientific research purposes. The material used as donors were lines used in breeding studies completed in Çanakkale Onsekiz Mart University (ÇOMÜ), Faculty of Agriculture, Department of Field Plants. These donors used in this study because they have different coloration levels for Navajo marker.

Table 1. Genotypes used in current study

Genotype name/code	General features	Obtained from[1]
CIM2GTAIL-P2	Navajo marker, inducer line	CIMMYT
B73 (G1)	Yellow, Dent	ÇOMÜ
Hya x B73 (G2)	Yellow, Experimental hybrid	ÇOMÜ

[1]CIMMYT: International Maize and Wheat Improvement Center, ÇOMÜ: Çanakkale Onsekiz Mart University

Induction hybridization was completed in trials in Çanakkale Onsekiz Mart University, Faculty of Agriculture Farm, Plant Research and Implementation Unit (40.07° N, 26.36° E) in 2021. Seeds used in the thesis study were obtained from breeding studies completed in Çanakkale Onsekiz Mart University, Faculty of Agriculture, Department of Field Plants in 2021. In these studies, two parent lines (Hya, B73) and an inducer line (CIM2GTAL-P2) were planted in field conditions. When flowering time arrived, controlled pollination methods were used (Kahrıman, 2016). With this aim, cob tassels of donor materials were protected before emerging. Seed materials used in this study were created by hybridizing pollen collected from the inducer line with protected cob tassels from the donor material.

In the study, seed samples used as initial material were firstly classified as haploid and diploid by eye. This classification separated seed samples as diploid based on anthocyanin coloring of the crown and embryo portions, and as haploid with coloring only at the crown section. This process was performed by same technician to minimize the error due to person difference. For each induction hybrid, at least 100 haploid and 100 diploid seed samples were separated. Later, these seed samples had the following measurements and observations performed.

Measured traits

Seed width (mm), seed length (mm) and seed circumference (mm) measurements were taken related to seed morphology. Among these measurements, measurements were obtained with the SmartGrain program, apart from single seed weight (Tanabata et al., 2012). Single seed weight was determined with a laboratory-type sensitive scale.

Measurements about root features were completed with the method proposed by Chaikam et al. (2017). With this aim, seeds with completed morphological measurements were left for germination. Measurements were performed for features related to the germinated seeds (root length, root count, root color, plant height, leaf count, SPAD value). Root tips of germinated seed samples were cut at 1-2 cm and left in 70% ethanol for use in chromosome analysis. Two root tips were taken from each germinated sample with the aim of taking precautions again any negative events that may be encountered during chromosome analyses.

Germinated seed samples were first transplanted to planters containing 5 x 9 pots filled with turf. Then when plants had emerged and reached the 2-3-leaf stage, they were transferred to small pots. Transplanted plants were raised in greenhouse conditions and watered according to need. Morphological measurements were completed for all plants with measurements performed by sampling the lower section of the leaf for stomal measurements. These samples were imaged with a trinocular microscope with 10 x and 40 x magnification capacity and digital images were recorded as TIFF files from 3 separate points on each slide (Figure 1).

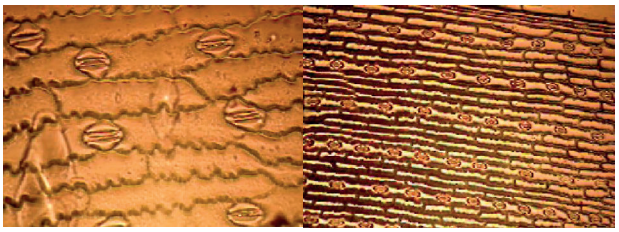

Figure 1. Example of digital images used in stomata measurements (40x at left, 10x at right)

These files were saved for use in image analyses. Digital images obtained from stomata samples were analyzed with the ImageJ program. With this aim, firstly calibration slide images were taken to be able to measure on the micro-images. After calibrating the ImageJ program using these images, stomata measurements were completed. Stomata count and stomata size measurements were manually performed with the ImageJ program.

The root tips were first washed in cold water and then left in 1% acetocarmine dye for at least 12 hours for staining, with samples for each plant dyed in separate tubes. Then, these samples were placed in tubes containing 70% ethanol and digital images were recorded with a 100x magnification capacity trinocular microscope. Root tips taken from germinated haploid and diploid samples were used with the aim of chromosome imaging in the cell metaphase stage. Cytogenetic analyses were completed according to the method recommended by Sekiya et al. (2020).

Statistical analysis

Data obtained in the study were analyzed with the R packet program (R Core Team, 2019).

The results for ploidy level identified after verification processes performed with cytogenetic analyses and the ploidy level obtained with visual discrimination were converted into a complexity matrix. Reliability statistics were calculated for this matrix with the following formulae (1, 2, 3, 4, 5). In the formulae, TP is true diploid classified as diploid, TN is true haploid classified as haploid, FP is diploid samples classified as haploid and FN is haploid samples classified as diploid.

$$\text{Sensitivity: } TP/(TP+FN) \quad (1)$$

$$\text{Specificity: } TN/(TN+FP) \quad (2)$$

$$\text{PPV: } (\text{Sensitivity} \times \text{Prevalence}) / (((\text{Sensitivity}) \times \text{Prevalence}) + ((1-\text{Specificity}) \times (1-\text{Prevalence}))) \quad (3)$$

$$\text{NPV: } \text{Specificity} \times (1-\text{Prevalence}) / (((1-\text{Sensitivity}) \times \text{Prevalence}) + ((\text{Specificity}) \times (1-\text{Prevalence}))) \quad (4)$$

$$\text{Accuracy: } (TP+TN)/(TP+FN+FP+TN) \quad (5)$$

Comparisons in terms of features measured in samples differentiated as haploid and diploid used the T test. The decision tree method was used to differentiate haploid and diploid samples using different criteria simultaneously. The rattle package (Williams, 2011) in R was used to create model development. No pruning was applied, and maximum depth, minimum split, minimum bucket parameters settled as 20, 20 and 7, respectively.

RESULTS AND DISCUSSIONS

Cytogenetic verification

Verification procedures in cytogenetic analyses were completed based on images in the metaphase stage of mitosis division at the root tips. For these analyses, it is notable there were variations in chromosome staining efficiency and image quality in diploid and haploid samples (Figure 2). Samples verified as diploid were observed more clearly in the division stage of chromosomes.

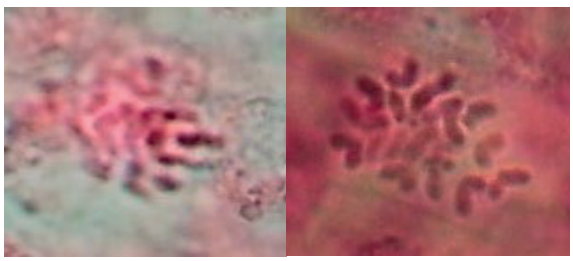

Figure 2. Example images taken for chromosome counts in haploid (on left) and diploid (on right) samples

The seed counts for ploidy level assumed for seeds differentiated by eye and according to true ploidy classes after cytogenetic analyses in the study are presented in the following table (Table 2).

Table 2. Variation in haploid/diploid counts verified with cytogenetic analyses compared to donor material

	B73		Hya x B73		Combined	
	D	H	D	H	D	H
D	23	1	70	1	93	2
H	6	20	10	34	16	54

D: Diploid, H: Haploid.

In total, there were 165 samples, of which 95 were diploid and 70 were haploid according to the verification procedure, and the majority of samples identified as diploid were observed to be truly diploid. Contrary to this, there was 22% misclassification for haploid samples (16 misclassified haploid) (Table 2). According to these results, initial classification according to seed coloring is understood to involve a notable error rate for haploid samples.

A variety of studies about the success of visual differentiation according to the R1-nj marker have been performed and, in these studies, a variety of tools like cytological measurements, flow cytometry and molecular markers were used as verification tools (Couto et al., 2013; Ribeiro et al., 2018). In these studies, success in haploid differentiation based on seed coloring induced by the R1-nj gene displayed variability. In our study, it is notable that more than 70% of haploid samples from both genotypes were accurately classified, while in the study by Ribeiro et al. (2018), this rate was only reported as 1.41%. However, in their study, only 16 measurements were made on tumor cells assumed to be haploids. This situation may be due to differences in donor material and inducer lines. Ribeiro et al. (2018) reported that donor materials with horse tooth endosperm structure displayed anthocyanization more clearly and differentiation success may increase, linked to this. In our study, both genotypes had horse tooth/horse tooth-hard endosperm structure.

Analytical statistics calculated for the complexity matrix are presented in Table 3. For these calculations, the positive group was assigned as diploid. For both donors, the sensitivity value was 0.70 and above, and the specificity value was observed to be higher than sensitivity. The PPV values were found to be much higher than the NPV values. These findings indicate higher differentiation success for diploid samples for both Hya and Hya x B73 genotypes. Contrary to this, it is understood that haploid samples can be separated by eye with 70-80% accuracy. When classification accuracy is considered in general, Hya x B73 hybrid material had higher classification accuracy than the B73 line (Table 3).

Table 3. Variation in haploid/diploid counts verified with cytogenetic analyses compared to donor materials

	B73	Hya x B73	Mean
Sensitivity	0.79	0.88	0.85
Specificity	0.95	0.97	0.96
PPV†	0.96	0.99	0.97
NPV††	0.77	0.77	0.77
Accuracy	0.87	0.92	0.91

†: PPV indicates positive prediction values, ††: NPV indicates negative prediction values

Morphometric & physiological measurements

The variance analysis results for the 12 features investigated in the study are presented in Table 4. According to the variance analysis results, seed width, seed length, seed circumference, single seed weight, root length, SPAD values, and stomata size had significant genotype effect before and after cytogenetic verification. Among the features investigated according to seed class, differences in seed width, seed circumference, single seed weight, root length, root count and plant height were significant according to differentiation before cytogenetic verification. Contrary to this, after cytogenetic verification, seed width, single seed weight, root length, and plant height were significantly different between haploid and diploid groups. The genotype x class (G x C) interaction only had significant effect in terms of root length and stomata count both before and after cytogenetic verification (Table 4).

According to the results, diploid seed samples were found to have higher numerical averages for germination features than haploid samples for both genotypes. According to mean stomata size and stomata count, differences were observed between genotypes. Mean values for haploid samples of B73 x CIMP2 samples were higher than for diploid samples. For the (HyaxB73) x CIMP2 genotype, diploid samples had higher mean stomata size and stoma count compared to haploid samples. Considering the differences in genotype means, the B73 line of the Hya x B73 genotype appears to have higher mean values both before and after verification in terms of seed measurements. This situation is an

expected result, due to being hybrid material of this genotype. As Hya is a line with relatively smaller seed size, both area and weight values were lower than the hybrid genotype. Considering the mean values for haploid and diploid seed samples, before verification, the difference between class means was statistically significant in terms of other features, apart from seed length. After verification, the differences between classes were significant in terms of seed circumference and single seed weight. Both before and after verification, in terms of mean values for seed measurements of samples, diploid samples had higher means than haploid samples. However, this difference was not identified to be significant in terms of all features. This situation indicates that after cytogenetic verification, the difference between the classes according to seed length and seed width were insignificant. For both genotypes, due to diploid samples generally having higher mean values than haploid samples, the G x S interaction effect was observed to be insignificant (Table 5).

Table 4. The results of variance analysis for investigated traits

	Before cytogenetic verification			After cytogenetic verification		
Feature	Genotype	class	G × C†	Genotype	Class	G × C†
Seed width	<0.001	0.014	0.963	<0.001	0.164	0.685
Seed length	<0.001	0.098	0.381	<0.001	0.613	0.633
Seed circumference	<0.001	<0.001	0.631	<0.001	0.045	0.782
Single seed weight	<0.001	0.001	0.312	<0.001	0.004	0.642
Root length	<0.001	<0.001	0.009	<0.001	0.001	0.025
Root number	0.476	0.01	0.737	0.483	0.236	0.91
Shoot length	0.897	0.124	0.866	0.897	0.215	0.481
Plant height	0.364	0.01	0.951	0.367	0.031	0.975
Leaf count	0.909	0.076	0.104	0.909	0.08	0.658
SPAD	0.05	0.138	0.187	0.052	0.51	0.292
Stomata size	<0.001	0.327	0.118	<0.001	0.042	0.285
Stomata count	0.087	0.364	0.024	0.085	0.131	0.012

†G × C = Genotype × Class effect

Table 5. Differences in haploid and diploid samples for investigated traits before and after cytogenetic verification

	Before cytogenetic verification							
	B73		Hya x B73		Mean			Mean
Investigated trait	Diploid	Haploid	Diploid	Haploid	Diploid	Haploid	Hya	Hya x B73
Seed width	6.89b	6.64b	8.08a	7.82a	7.48a	7.23b	6.76b	7.98a
Seed length	8.24b	8.20b	9.72a	9.49a	8.98a	8.84b	8.22b	9.63a
Seed circumference	25.27c	24.48c	29.48a	28.41b	27.37a	26.44b	24.9b	29.1a
Single seed weight	0.17c	0.16c	0.30a	0.27b	0.23a	0.21b	0.16b	0.28a
Root length	3.35b	3.31b	5.12a	3.69b	4.23a	3.50b	3.33b	4.57a
Root number	4.33	3.81	4.08	3.68	4.20a	3.74b	4.06	3.93
Shoot length	4.56	4.29	4.53	4.18	4.54	4.23	4.42	4.39
Plant height	21.50	19.35	21.94	19.89	21.72a	19.62b	20.4	21.2
Leaf count	6.22	6.30	6.38	6.03	6.30	6.16	6.26	6.25
SPAD value	39.02	38.71	36.85	38.44	37.93	38.57	38.9	37.5
Stomata size	37.31a	38.17a	33.83b	32.50b	35.57	35.33	37.8a	33.3b
Stomata count	62.13b	76.88ab	86.52a	71.73ab	74.32	74.30	69.8	80.9
	After cytogenetic verification							
	B73		Hya x B73		Mean			Mean
Investigated trait	Diploid	Haploid	Diploid	Haploid	Diploid	Haploid	Hya	Hya x B73
Seed width	6.80b	6.71b	8.03	7.88a	7.41a	7.29b	6.76b	7.98a
Seed length	8.21b	8.23b	9.66a	9.12a	8.93	8.67	8.22b	9.63a
Seed circumference	25.06b	24.59b	29.27a	27.78a	27.16a	26.03b	24.9b	29.1a
Single seed weight	0.17c	0.16c	0.30a	0.27b	0.23a	0.21b	0.16b	0.28a
Root length	3.36b	3.29b	4.97a	3.32b	4.16a	3.30b	3.33b	4.57a
Root number	4.14	3.95	4.00	3.40	4.07	3.67	4.06	3.93
Shoot length	4.44	4.39	4.51	4.73	4.47	4.56	4.42	4.39
Plant height	21.13	19.35	21.72	22.36	21.42a	20.85b	20.4	21.2
Leaf count	6.32	6.18	6.33	6.52	6.32	6.35	6.26	6.25
SPAD value	39.11a	38.52ab	37.16b	40.21ab	38.13	39.36	38.9	37.5
Stomata size	37.91a	37.55a	33.90b	30.24c	35.90	33.89	37.8a	33.3b
Stomata count	64.34b	77.33ab	87.15a	74.01b	75.74	75.67	69.8	80.9

Haploid and diploid seeds have different features in terms of morphometric and some other features. Studies revealed that there may be differences in haploid and diploid seed classes in terms of commonly used measurements like size and weight, especially. Research dealing with the difference in haploid and diploid samples in terms of single seed weight found similar single seed weight for haploid-diploid seeds separated according to the R1-nj marker, with lower weight reported for seeds with embryo abortion (Qu et al., 2021). Haploid seeds have smaller embryos compared to diploid seeds and seed weights are known to be lower linked to this (Melchinger et al., 2015). For all seed features investigated in our study, haploid samples had lower values both before and after cytogenetic verification.

When the mean values related to root features are considered, diploid samples had higher means compared to haploid samples. In terms of root length, the mean value for diploid samples (4.23 cm) was statistically significantly different ($p<0.05$) than the value for haploid samples (3.50 cm). A similar situation is valid for root count. According to the genotype averages, B73 line (3.33 cm) had shorter root length than the Hya x B73 hybrid. For root length, the haploid sample mean value for the Hya x B73 genotype was in the same statistical class as the mean for diploid samples from the B73 line (Table 5).

Germination features of plant seeds are used for different purposes like genotypic characterization, early development performance and tolerance of abiotic stress. Studies about this topic in the literature showed that haploid maize seeds had lower root length, shoot length and side root numbers compared to diploid seeds (Chaikam et al., 2017). Similarly, Baleroni et al. (2021) reported that a certain duration after haploid maize seeds germinated, they had lower root length, shoot length and side root numbers compared to diploid seeds, independent of genotype and germination temperature. According to the results obtained in our study, there were clear differences in terms of root length and root count before and after cytogenetic verification. Among these features, the difference in terms of root length was statistically significant ($p<0.05$).

When the differences between haploid and diploid samples for plant height and leaf count before and after cytogenetic verification are investigated, the Hya x B73 genotype had higher mean values compared to the B73 line in terms of plant height before and after verification. For leaf count, both before and after verification, the B73 line had higher values than the Hya x B73 genotype (Table 5).

Before and after cytogenetic verification, the differences in SPAD value, stomata size and stomata count for haploid and diploid samples are presented in Table 4. Considering the mean values for haploid and diploid seed samples, both before and after cytogenetic verification, the differences between SPAD value, stoma size and stomata count were statistically insignificant. The B73 line had higher mean values for both SPAD value and stomata size both before and after verification compared to the Hya x B73 genotype. For leaf count, the Hya x B73 genotype had higher values compared to the B73 line both before and after verification. With the increase in ploidy level, notable differences formed for some plant and physiological features. Kwon et al. (2016) reported that tetraploid Platycodon grandiflorus plants had taller height, and higher values for chlorophyll content and SPAD compared to diploid plants. Studies were performed related to some physiological features in maize and the differences between haploid and diploid plants were dealt with. In these studies, stomata cell length was the leading physiological feature used to differentiate haploid maize plants from diploid plants. In terms of this feature, haploid plants had lower values compared to hybrid and diploid plants (Molenaar et al., 2019; Da Silva et al., 2020). In our research, before cytogenetic verification, the mean values for haploid and diploid groups did not have a clear difference in terms of stomata size. After cytogenetic verification, diploid plants had higher mean values for stomata size compared to haploid plants; however, this difference was statistically insignificant. Phenotyping using Soil Plant Analyzer Development (SPAD)-502 (SPAD-502 Konica Minolta Sensing Inc., Japan) gives SPAD readings to indicate the index of chlorophyll a and chlorophyll b in the thylakoid membrane in the leaf mesophyll chloroplasts (Kandel, 2020). In terms of SPAD value, differences were observed in haploid and diploid samples of the genotypes, with haploid plants

from the hybrid genotype having lower SPAD values than diploid plants. The opposite situation was observed for the pure line. Finally, according to general averages for haploid and diploid classes, diploids had lower SPAD values than haploids before and after cytogenetic verification; however, this difference was insignificant (Table 5).

Threshold Determination by Decision Tree

To identify whether the features investigated in the study could determine the seed class or not, the decision tree graphics were separately prepared before (Figure 3) and after (Figure 4) cytogenetic verification. According to the graph for before cytogenetic verification, root length and seed circumference may be effectively used to differentiate haploid and diploid seed samples. The seed samples were classified under three nodes, with samples with root length below 3.41 cm located on the second node (Node 2). There was a total of 60 samples and the majority comprised haploid seeds. The fourth node (Node 4) classified samples with root length longer than 3.415 cm and seed circumference below 25.526 mm. This included a total of 15 samples and 60% were haploid seeds. The final node (Node 5) classified 90 samples. These were samples with root lengths longer than 3.41 cm and seed circumferences above 25.526 mm. At this node, it is notable that nearly 75% were diploid samples.

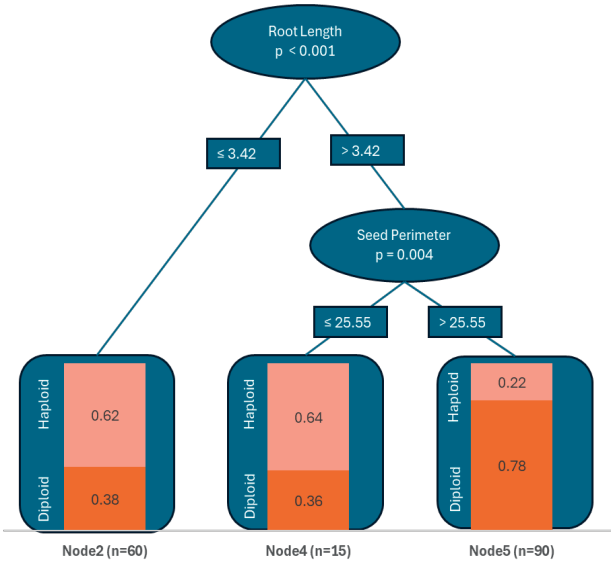

Figure 3. Decision tree used to classify maize seed samples into haploid and diploid groups based on the R1-nj marker and additional phenotypic variables prior to cytological validation

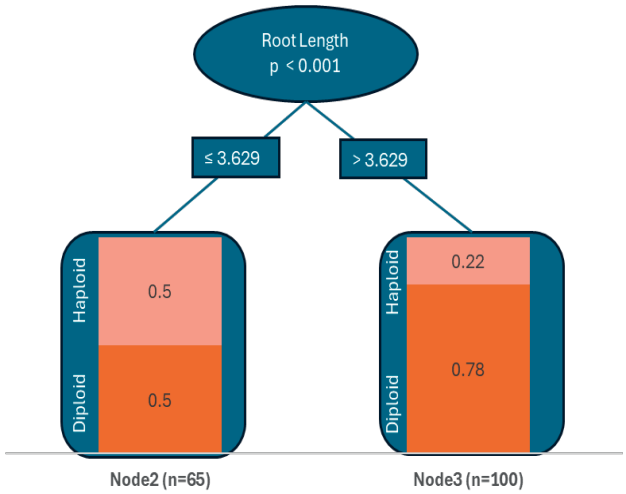

Figure 4. Decision tree used to classify maize seed samples into haploid and diploid groups based on cytological validation and additional phenotypic variables

It was determined that root length may be effectively used to separate haploid and diploid seed samples according to the graphic from after cytogenetic verification. Seed samples were classified under two nodes, with the second node (Node 2) including samples with root length less than 3.62 cm. This included a total of 65 samples and nearly half comprised haploid and diploid samples. The other node (Node 3) included a total of 100 samples and these samples had root length longer than 3.62 cm. Of these samples, more than 85% were diploid samples. According to the model after cytogenetic verification, diploid samples may be more accurately differentiated by only examining root length, while it was determined that a different classification model is required for haploid sample differentiation (Figure 4).

Different features are used for differentiation of haploid and diploid maize samples. Additionally, the number of studies stating the threshold for differentiation values for all the feature groups investigated in this study is limited. A variety of studies about verification of haploid samples using sapling features were performed in the literature about this topic.

Research by Baleroni et al. (2021) determined the threshold values for root length, coleoptile length, and seminal root count. In the study, variability was determined compared to the donor genotypes for the threshold values of the germination features. For root length, these values varied from 2.1 cm to 5.6 cm; for coleoptile length the value varied from 1.8 cm to

2.8 cm. In our study, it was determined that root length was able to be used effectively for verification of haploids and the identified threshold value is within the thresholds in the literature. In other words, more successful differentiation appears possible if seed samples differentiated by eye according to color markers are germinated and reviewed again according to root length. Though this procedure requires extra time, it may offer an easy solution in terms of increasing classification success.

In this study, haploid and diploid maize samples obtained with the in vivo doubled haploid technique, commonly used in maize breeding, were compared in terms of seed morphology, plant morphology and physiological features. According to the study results, haploid seed samples generally had lower values in terms of the measured features compared to diploid samples. Based on the statistical analysis results, differences in seed length, root length, root count, leaf count, stomata count, and SPAD values were determined to be significant between haploid and diploid samples.

Additionally, these differences may not be valid for both donors, with these differences more pronounced for Hya×B73 hybrid. Based on these findings, the success of haploid/diploid sample differentiation according to the donor material used displays variability. Another result obtained in the research is whether haploid-diploid sample differentiation can be completed using the investigated features or not. According to the decision tree method, verification may be performed by examining the root length and seed circumference among features of seeds allocated as being haploid and differentiation success may increase.

CONCLUSIONS

The findings of this study demonstrated that visual classification of haploid maize samples based on pigmentation induced by the R1-nj gene can achieve a high success rate. Significant differences in plant height between haploid and diploid groups were observed both before and after the verification stage, while no statistically significant difference was found in the number of leaves. These results highlight the potential of using specific morphometric and physiological traits to distinguish between haploid and diploid samples during early developmental stages.

Furthermore, incorporating additional measurable features such as root length and seed circumference may further enhance classification accuracy. This improvement is particularly relevant for large-scale breeding programs, where early and reliable identification of haploids can contribute to more efficient use of labor, time, and cultivation space.

In light of these findings, future studies should aim to expand the range of traits measured, including both seed-based and seedling-based characteristics, to develop more robust classification models. Increasing the sample size and integrating advanced modelling techniques, such as machine learning approaches, may also significantly improve classification performance.

Overall, the study provides a promising foundation for optimizing haploid identification processes in maize breeding programs.

ACKNOWLEDGEMENTS

This work was supported by Çanakkale Onsekiz Mart University the Scientific Research Coordination Unit, Project number: FYL-2021-3823. Also this manuscript is prepared from Talha Tunç's MSc thesis.

REFERENCES

Baleroni, A. G., Ré, F., Pelozo, A., Kamphorst, S. H., Carneiro, J. W. P., Rossi, R. M., & Scapim, C. A. (2021). Identification of haploids and diploids in maize using seedling traits and flow cytometry. *Crop Breeding and Applied Biotechnology, 21*: e38422145.

Battistelli, G. M., Von Pinho, R. G., Justus, A., Couto, E. G., & Balestre, M. (2013). Production and identification of doubled haploids in tropical maize. *Genetics and Molecular Research, 12*: 4230-4242.

Breiman, L., Friedman, J., Olsen, R., & Stone, C. (1984). *Classification and Regression Trees*. Wadsworth International (California).

Cengiz, R., & Korkut, K. Z. (2016). Development of doubled haploid maize lines by using in vivo haploid technique. *Biotech Studies, 29*(1): 1-7.

Cerit, İ., Cömertpay, G., Oyucu, R., Çakır, B., Hatipoğlu, R. & Özkan, H. (2016). Melez mısır islahında in-vivo katlanmış haploid tekniğinde kullanılan farklı inducer genotiplerin haploid indirgeme oranların belirlenmesi. *Tarla Bitkileri Merkez Araştırma Enstitüsü Dergisi, 25*(1): 52-57.

Chase, S. S. (1969). Monoploids and monoploid derivatives of maize (*Zea mays* L.). *Bot. Review*, *35*: 117-167.

Chaikam, V., Nair, S. K., Babu, R., Martinez, L., Tejomurtula, J., & Boddupalli, P. M. (2015). Analysis of effectiveness of R1-nj anthocyanin marker for in vivo haploid identification in maize and molecular markers for predicting the inhibition of R1-nj expression. *Theoretical and Applied Genetics, 128*(1): 159-171.

Chaikam, V., Antonio Lopez, L., Martinez, L., Burgueno, J., & Boddupalli, P. M. (2017). Identification of in vivo induced maternal haploids in maize using seedling traits. *Euphytica, 213*: e177.

Coe, E. H. (1959). A Line of Maize with Haploid Frequency. *Am. Nat., 93*: 381-382.

Couto, E. G. O., Davide, L. M. C., Bustamante, F. O., Pinho, R. G. V., & Silva, T. N. (2013). Identification of haploid maize by flow cytometry, morphological and molecular markers. *Science and Agrotechnology, 37*(1): 25-31.

Da Silva, H. A., Scapim, C. A., Vivas, J. M. S., do Amaral Júnior, A. T., Pinto, R. J. B., Mourão, K. S. M., Rossi, R. M., & Baleroni, A. G. (2020). Effect of ploidy level on guard cell length and use of stomata to discard diploids among putative haploids in maize. *Crop Science, 60*(3): 1199-1209.

Dönmez, E. (2020). Classification of haploid and diploid maize seeds based on pre-trained convolutional neural networks. *Celal Bayar University Journal of Science, 16*(3), 323-331.

Geiger, H. H., & Gordillo, G. A. (2009). Doubled haploids in hybrid maize breeding. *Maydica, 54*: 485-499.

Kahrıman, F. (2016). Misirda polen etkisi ve bu etkinin kontrolünde uygulanan yöntemler. *Lambert Academic Publishing, Saarbrükhen,* 52 pp.

Kahrıman, F., Güz, A. M., & Pehlivan, İ. (2023). Use of machine learning models-based image analysis for classification of haploid and diploid maize. *Crop Breeding and Applied Biotechnology*, *23*(4), e45322349.

Kandel, B. P. (2020). SPAD value varies with age and leaf of maize plant and its relationship with grain yield. *BMC Res Notes, 13*: 475.

Kermicle, J. L. (1969). Androgenesis conditioned by a mutation in maize. *Science, 166*: 1422-1424.

Kwon, S. J., Lee, H. D., Seo, D. Y., Moon, Y. J., Cho, G. Y., Boo, H. O., Woo, S. H., & Kim, H. H. (2016). Comparison of morpho-physiological characteristics in diploid and tetraploid *Platycodon grandiflorum*. *Korean Journal of Crop Science, 61*(2): 138-144.

Lashermes, P. & Becekert, M. (1988). Genetic control of maternal haploidy in maize (*Zea mays* L.) and selection of haploid inducing lines. *Theoretical and Applied Genetics, 76*: 405-410.

Mathur, D. S., Sachan, J. K. S., & Sarkar, K. R. (1976). Radiation induced haploid and heterofertilization in maize. *Journal of Nuclear Agriculture and Biology, 5*: 76-77.

Mathur, M. A., & Sarkar, K. R. (1980). Induction of maternal haploids in maize through heat treatment of pollen. *Current Science, 49*: 744-746.

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

Moldovan, G., & Bărbos, A. (2019). The decline in viability and vigour of the hybrid maize (*Zea mays* L.) seed under the influence of genotype, duration and the factors of the storage ecosystem. *Agrolife Scientific Journal, 8*(1): 181-191.

Molenaar, W. S., Couto, E. G. O., Piepho, H. P., & Melchinger, A. E. (2019). Early diagnosis of ploidy status in doubled haploid production of maize by stomata length and flow cytometry measurements. *Plant Breeding, 138*: 266-276.

Ribeiro, C. B., Pereira, F. D. C., Filho, L. D. N., Braz, G. T., Ruy, M. C., Silva, M. B., Cenzi, G., Techio, V. H., & Souza, J. C. D. (2018). Haploid identification using tropicalized haploid inducer progenies in maize. *Crop Breeding and Applied Biotechnology, 18*(1): 16-23.

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.

Sekiya, A., Pestana, J. K., Silva, M. G. B., Krause, M. D., Silva, C. R. M., & Ferreira, J. M. (2020). Haploid induction in tropical supersweet corn and ploidy determination at the seedling stage. *Pesquisa Agropecuária Brasileira, 55*: e00968. https://doi.org/10.1590/S1678-3921.pab2020.v55.00968.

Tanabata, T., Shibaya, T., Hori, K., Ebana, K. & Yano, M. (2012). SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiology, 160*(4): 1871-1880.

Qu, Y., Liu, Z., Zhang, Y., Yang, J., & Li, H. (2021). Improving the sorting efficiency of maize haploid kernels using an NMR-based method with oil content double thresholds. *Plant Methods, 17*: 1-15.

Williams, G. J. (2011). *Data Mining with Rattle and R: The art of excavating data for knowledge discovery*, series Use R! Springer. https://rd.springer.com/book/10.1007/978-1-4419-9890-3.

Zuoyo, Z., & Minguang, G. (1984). Production of pure lines of maize through parthenogenesis induced by chemicals. *Acta Genetica Sinica, 11*: 39-46.

***R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.