

Improvements to the maize (*Zea mays* L.) *in vivo* maternal doubled haploid system

by

Gerald N. De La Fuente

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Plant Breeding

Program of Study Committee:
Thomas Lübberstedt, Major Professor
Dan Nettleton
Michael Blanco
Paul Scott
Erik Vollbrecht

Iowa State University

Ames, Iowa

2015

Copyright © Gerald N. De La Fuente, 2015. All rights reserved.

DEDICATION

This dissertation is dedicated to Mr. Mark J. Lamon. My passion for agriculture was found working on the Lamon family farm, and it continues to this day. Many were the hours spent discussing farming and agriculture with Mr. Lamon, and the lessons I learned were put to use in the development of this dissertation. Thank you, Mark.

TABLE OF CONTENTS

DEDICATION	ii
ABSTRACT	vi
CHAPTER ONE: GENERAL INTRODUCTION	1
Evolution of maize breeding	2
Doubled haploids	3
Organization of the thesis	9
References	11
CHAPTER TWO: A DIALLEL ANALYSIS OF A MAIZE DONOR POPULATION RESPONSE TO <i>IN VIVO</i> MATERNAL HAPLOID INDUCTION I: INDUCIBILITY	15
Abstract	15
Introduction	16
Materials and methods	20
Results	24
Discussion	26
References	31
Tables and figures	34
CHAPTER THREE: REGISTRATION OF BHI306 MAIZE MATERNAL HAPLOID INDUCER GERMPLSAM	38
Abstract	38
Introduction	38
Methods	43
Characteristics	45

Availability	47
References	48
Tables and figures	50
CHAPTER FOUR: DISCRIMINATION OF HAPLOID AND DIPLOID MAIZE KERNELS VIA MULTISPECTRAL IMAGING	52
Abstract	52
Introduction	53
Methods	56
Results	62
Discussion	64
References	70
Tables and figures	73
CHAPTER FIVE: A DIALLEL ANALYSIS OF A MAIZE DONOR POPULATION RESPONSE TO <i>IN VIVO</i> MATERNAL HAPLOID INDUCTION II: SPONTANEOUS CHROMOSOME DOUBLING	84
Abstract	84
Introduction	85
Materials and methods	88
Results	92
Discussion	94
References	103
Tables and figures	107
CHAPTER SIX: ACCELERATING PLANT BREEDING	114
Abstract	114
Glossary box	114

Keeping up with demand	116
The breeder's equation.....	117
Speeding up	118
The <i>in vitro</i> nursery.....	118
Challenges	123
References	126
Boxes	134
Figures	136
CHAPTER SEVEN: GENERAL CONCLUSIONS AND THE FUTURE OF MATERNAL <i>IN VIVO</i> DOUBLED HAPLOIDS	141
ACKNOWLEDGEMENTS	146

ABSTRACT

The maize (*Zea mays* L.) *in vivo* maternal doubled haploid system has been widely applied to maize breeding and genetics in recent decades and is an important part of the majority of public and private maize breeding programs today. The principal advantage of the doubled haploid system is the ability to generate completely homozygous inbred lines in as little as two seasons. Other advantages to this system include more rapid integration of loci of interest and increased usefulness over traditional lines developed through self-pollination. In this dissertation, some of the major problems in the maternal doubled haploid system are addressed. Namely, improvement of maternal inducers, improved understanding of the genetics controlling inducibility, development of an automated system to sort haploid kernels, investigation and application of spontaneous chromosome doubling, and a proposal for the acceleration of the breeding cycle beyond doubled haploids through the *in vitro* nursery. This dissertation provides some new insight into these problems, as follows. The development and release of a new improved maternal haploid inducer for use in doubled haploid programs. Improved understanding of the quantitative nature of inducibility and the effects of misclassification are discussed. Successful automated discrimination of haploid and diploid kernels using optical and fluorescence methods is described. In an effort to make the doubled haploid system more efficient and safe, a bypass of the colchicine doubling step is proposed through the application and investigation of spontaneous chromosome doubling in haploid plants. Finally, as a proposal for what could be the next step in accelerating the breeding cycle, the *in vitro* nursery and its applications is discussed.

CHAPTER ONE

GENERAL INTRODUCTION

The second highest produced crop and the number one cereal grain in the world (1.02 billion tons in 2013 valued at \$67.1 billion) (FAOSTAT, 2013), maize (*Zea mays* L.) is an important row crop used as food, feed, and fuel. From its most obvious use as an animal feed, to use in cosmetics, tires, and molded plastic, maize is a highly versatile crop providing raw materials for many industries.

Originating from Mexico and Central America from its ancestor Teosinte, maize has been bred for over 5,000 years. The maize that is commonly known today is vastly different from maize as it was 10,000 years ago. Teosinte resembles maize. However, its utility as a crop was not sufficient and over the years Native Americans selected and improved maize into the crop grown today. Today's maize is an annual, monoecious grass with imperfect flowers. Feed maize, the most commonly grown type, is typically a single stalk which produces one ear per plant. Maize is a unique grass in that its male flower (tassel) is located atop the plant and the ear (female) is found typically midway up the stalk growing between stalk and leaf. This unique architecture makes maize a highly outcrossing (allogamous) crop.

Over time, maize was slowly domesticated and adapted to increasingly higher latitudes expanding from the tropics into the temperate zones of the United States and eventually to Europe and around the world. In North America, maize production is concentrated on the plains of the Midwestern United States with most of the production

occurring in Iowa, Illinois, Nebraska, and Indiana. Temperate climate, high average rainfall, and fertile soil provide an ideal environment for production. Average maize yields in high production regions range from 10.5-12.9 metric tons/hectare with maximum yields exceeding 21 metric tons/hectare. These high yields resulted from a combination of both breeding and agronomic improvement (Duvick, 2005).

Evolution of maize breeding

Maize breeding, as we know it today has evolved from a very primitive, essentially mass selection scheme, to the multi season per year single cross hybrid programs today. Beginning with Shull (1908) and East (1908), the concept of extracting homozygous and homogeneous inbred lines from heterozygous and heterogeneous open-pollinated varieties and crossing them to create single cross hybrids began the transition to maize breeding as we know it today. However, due to poor performance of inbred lines due to inbreeding depression it was not economically viable to produce hybrid seed until the proposal of double cross (4-way) hybrids was introduced (Jones, 1918). The recognition that double cross hybrids could provide a yield advantage over open-pollinated varieties led to an increase in the research and development of inbred lines. However, it was soon realized that there is an enormous number of possible combinations between lines and that not all line combinations out yielded their open pollinated variety progenitors. This, coupled with the seven generations it takes to generate an inbred created a major problem (Hallauer, 1988). Research then turned towards the prediction of hybrid performance, and the use of testers to evaluate combining ability of inbred lines. Jenkins (1934), proposed a method for the prediction of four way hybrid performance as

a way to more easily select those four way cross combinations to be made. Slowly, double cross hybrids completely phased out open pollinated varieties and over time, the double cross hybrids were replaced by single cross hybrids due to increased performance and understanding and exploitation of heterotic groups. In 1933 no double cross hybrids were being grown in the U.S., but by 1945 over 50% of the area was planted in double cross hybrids and quickly rose to 100% by 1960 (Hallauer, 1988). Subsequently, in 1960 only a very small percentage of the area was planted in single cross hybrids, but by the year 2000, nearly 100% of all maize grown in the U.S. was single cross hybrids. Despite all of these advancements, development of inbred lines remained an expensive and time consuming endeavor, requiring at least 4 years even with off season nurseries.

Doubled Haploids

First described by Randolph (1932) haploid plants in maize were found by Chase (1947) to be a natural occurrence in typical maize populations in the United States. It was soon understood that these sterile haploid maize plants could be useful for genetic and breeding efforts. With the discovery that the chemical colchicine can be used to artificially double the genome present in haploid plants, fertile diploid lines could be artificially produced and are known today as doubled haploids (DHs) (Blakeslee and Avery, 1937; Gayen et al., 1994). There are two methods that are commonly used to produce DHs: in vitro anther/spore culture (Germanà, 2010) and in vivo induction of haploidy using specialized inducer lines (Chalyk, 1994; Coe, 1959; Rober et al., 2005). The advantage of DHs is that they are instantly 100% homozygous. This is a substantial

time saving over traditional self-pollination, where it takes four years to develop an F_8 , which is theoretically 99.6% homozygous, assuming two seasons per year.

Maize DH lines are most commonly developed using the in vivo maternal induction method, which is simple and more genotype-independent than in vitro methods in maize (Geiger, 2009). Over time, lines that produced a significantly higher proportion of haploid kernels per ear, when used as maternal inducers, were identified (Bordes et al., 1997; Chalyk, 1994; Coe, 1959; Rober et al., 2005). Inducer parents have all of the same characteristics of other maize, but heritably produce haploid kernels when crossed with another (donor) line of interest. Inducers serve to “induce” haploid embryos which are later doubled to generate DHs. The donor population is analogous to any breeding population. However, instead of some form of selection and self-pollination the population is crossed to the inducer. In vivo inductions can be made with either maternal or paternal inducers each having a different genetic consequence. In the case of a maternal inducer, the goal is to produce maternally derived haploid kernels and the donor population serves as the female. Bulk pollen from the inducer (male) is used to pollinate the donor population, since the inducer is used as the male parent the cytoplasm of the maternal haploids will come from the donor population. For paternal inducers the goal is to produce paternally derived haploid kernels and the donor population serves as the male parent. Pollen from the donor population is used to pollinate the inducer, and thus the cytoplasm of the paternally derived haploids originates from the inducer.

Four objectives are proposed for the overall improvement and adaptation of the DH process to the Midwestern U.S. The objectives are 1) investigation of the maternal

genetic components controlling inducibility of donor parents, 2) investigation of rates, genetic components and utility of spontaneous chromosome doubling, 3) development of improved and adapted maternal haploid inducers, and 4) automation of the selection of haploid progeny.

The cytological process that is responsible for the induction of haploid embryos during *in vivo* induction is not known, but two competing hypotheses exist. The first involves a failure of fertilization. The pollen does not fertilize the embryo, but induces its development, leading to a haploid embryo (Chalyk et al., 2003). The second hypothesis involves successful fertilization and subsequent expulsion of the chromosomes from the inducer parent (Wedzony et al., 2002; Zhang et al., 2008). Though inducibility may seem qualitative, a study by Lashermes and Beckert (1988) showed that a cross between an inducer line and a non-inducer line generated quantitative variation for inducibility. This prompted a study by Prigge et al. (2012) to investigate the genetics underlying the quantitative variation for inducibility. A total of seven QTL on five chromosomes were found, with two major effect QTL on chromosomes 1 and 9, which explained 80-90% of the genetic variance for inducibility (Prigge et al., 2012). This study provides breeders, who use the DH process and/or develop haploid inducer lines, the ability to use marker assisted selection to accelerate the breeding of improved and adapted inducer lines. However, the discovery that there is a heritable genetic component for inducibility in the maternal inducer leads to the question, whether there is a heritable genetic component in the donor parent that affects inducibility. This question was initially addressed by Kebede et al. (2011) using a set of ten white seeded inbreds

developed by CIMMYT in a half diallel. The ten inbreds were crossed in all possible combinations without selfs and reciprocals. The F₁ progeny of the half diallel were induced and inducibility was scored. A significant general combining ability effect was found, while specific combining ability was not significant, suggesting that inducibility is an additive trait.

Artificial doubling of genomes in plants using colchicine has first been reported by Blakeslee and Avery (1937). Colchicine has long been known to artificially double ploidy levels by inhibition of mitotic spindle apparatus development (Borisy and Taylor, 1967). Artificial genome doubling became more efficient in maize, when it was discovered that colchicine treatment applied to exposed coleoptiles improves the doubling rate (Gayen et al., 1994; Geiger, 2009). However, this method is not 100% successful, with doubling rates ranging from 16% to 49%, depending on the method used (Eder and Chalyk, 2002). Moreover, colchicine treatment costs time and labor. Spontaneous doubling of chromosomes has been reported in many grass species and has played an important role in the formation of polyploid crops. However, when considering spontaneous doubling, rates are usually too low to be implemented in a DH program (Castillo et al., 2009).

The inducer used by our group is a F₁ hybrid between RWS and RWK-76 (Rober et al., 2005), and developed in Germany. It is poorly adapted to the temperate climate of the U.S. When grown in the Midwest U.S., the inducer hybrid is small and prone to lodging and has other poor agronomic qualities. Introgressing induction loci into elite

temperate lines will help to solve most of the agronomic problems that the inducer currently has such as lodging and adaptation to the Iowa climate. Resistance to lodging, improved adaptation and the ability to produce haploid kernels are essential qualities of any haploid inducer line developed. However, another important quality of any good inducer is the ability to shed copious amounts of pollen for an extended period of time. Elite lines of maize have been bred for reduced tassel size over time due to the negative correlation between tassel size and grain yield caused by shading of upper leaves (Duncan et al., 1967; Fischer et al., 1987; Hunter et al., 1969). Therefore, introgressing the induction loci into elite maize lines will likely result in the development of lines with reduced tassel size and pollen shed density. Another pollen trait that is beneficial is pollen heat tolerance. An inducer with pollen that can remain fertile under extreme heat, which is common in the summer and winter nurseries used for DH development, would produce more successful induction crosses and extend the pollination window.

Haploid kernels encompass only a small fraction of the total number of kernels produced in an induction cross (~8%). Haploid kernel selection is thus an important and time consuming step in the development of DH lines. The most common selection method is based on the *RI-navajo* (*RI-nj*) allele of the *rl* gene (Geiger, 2009; Nanda and Chase, 1966). The *RI-nj* allele is a dominant allele in the anthocyanin pathway that produces a visible purple/red hue in the crown of the aleurone, anthers, roots, and coleoptiles (Ludwig and Wessler, 1990). The utility of *RI-nj* is that the presence of pigment in the crown of the aleurone and embryo can be readily identified in mature kernels through visual inspection. Haploid kernels exhibit color in the crown of the

aleurone, but not in the embryo. This signifies that fertilization of the central cell was successful, but fertilization of the egg cell was not. On average 8% (Geiger, 2009) of all kernels of an induction cross are haploids, though this varies between genotypes (Prigge et al., 2011; 2012).

There are several problems with the *R1*-based selection system, the most obvious being the time required for selection of haploid kernels. Each individual kernel must be visually inspected and the majority of kernels are not easily distinguishable. The extent to which the embryo coloration is seen through the seed coat also varies within and between genotypes, complicating selection. If the donor genotype carries the *CI*^l allele (colorless aleurone), that is dominant to *CI* (colored aleurone), then the resulting kernels will all be yellow or white depending on the state of the Y1 locus (Ford, 2000). Those kernels cannot be selected based on red color, since *C*¹ is epistatic to *R1-nj*. Despite these problems, the *R1-nj* marker system continues to be the most popular system employed for haploid kernel selection. Though it is a simple process, the desire for automation is substantial.

The objectives of this thesis were to contribute to improving the efficiency of the maize DH system by 1) identification of highly inducible lines, 2) identification of spontaneous chromosome doubling lines, 3) development of new inducer lines adapted to Midwest U.S., and 4) establishment of a method for potential automated sorting of haploid kernels. In chapter 5, an outlook on further acceleration of breeding cycles for maize and other crops has been proposed.

Organization of the thesis

This thesis contains one published research article (Chapter 6), one manuscript submitted for publication (Chapter 5), and three manuscripts (Chapters 2-4) in preparation for submission for publication. The general conclusion of all the chapters is summarized in Chapter 7 as an overview of the improvements identified and future works. The general introduction provides a historical perspective to maize breeding and how this work relates. Literature for each individual experiment is provided for each chapter as a context.

Author Contributions

Chapter 2

GD designed the experiment and conducted most of the data collection, all the data analysis and summary, as well as writing of the manuscript.

UF provided input for experimental design and assisted with data collection.

BT assisted with data collection.

DN provided input for experimental design and the statistical analysis

TL provided input for the experimental design and assisted with the interpretation of results and critical commentary for the manuscript.

Chapter 3

UF generated populations for line development, collected data, made selections, and provided critical commentary for the manuscript

GD made selections in the populations, collected data, assisted with design of experiments, assisted with data analysis and wrote the manuscript.

TL assisted with design of experiments and population selection, assisted with data analysis, and provided critical commentary for the manuscript.

Chapter 4

GD designed the experiment and conducted most of the data collection, all the data analysis and summary, as well as writing of the manuscript.

UF provided input for experimental design and assisted with data collection.

BT assisted with data collection.

JR assisted with data collection.

MB conducted data collection at offsite location.

NY conducted data collection at offsite location.

AV assisted with data collection.

SM conducted data collection at offsite location.

DN provided input for the experimental design and the statistical analysis.

TL provided input for the experimental design and assisted with the interpretation of results and critical commentary for the manuscript.

Chapter 5

GD assisted with design of the experiment, developed materials for data collection, analyzed the data and wrote the manuscript.

JC assisted with design of the experiment, interpretation of results and critical commentary for the manuscript.

ME conducted all data collection

TL assisted with design of the experiment, interpretation of results and critical commentary for the manuscript.

Chapter 6

GD provided development of the concept for the manuscript as well as the writing of the manuscript.

UF provided critical commentary on the concept and the manuscript.

TL provided the initial concept and development of the concept as well as critical commentary for the manuscript.

References

- Blakeslee, A.F., and A.G. Avery. 1937. Methods of inducing doubling of chromosomes in plants by treatment with colchicine. *Journal of Heredity* 28:393-411
- Bordes, J., R.D. deVaulx, A. Lapierre, and M. Pollacsek. 1997. Haplodiploidization of maize (*Zea mays* L.) through induced gynogenesis assisted by glossy markers and its use in breeding. *Agronomie* 17:291-297
- Borisy, G.G., and E.W. Taylor. 1967. The mechanism of colchicine. *Journal of Cell Biology* 34:525-533
- Castillo, A.M., L. Cistue, M.P. Valles, and M. Soriano. 2009. Chromosome doubling in monocots. In: Touraev, A. et al., editors, *Advances in haploid production in higher plants*, Springer, Netherlands. p. 329-338
- Chalyk, S., A. Baumann, G. Daniel, and J. Eder. 2003. Aneuploidy as a possible cause of haploid-induction in maize. *Maize Genetics Cooperation Newsletter* 77:29

- Chalyk, S.T. 1994. Properties of maternal haploid maize plants and potential application to maize breeding. *Euphytica* 79:13-18
- Chase, S.S. 1947. Techniques for isolating monoploid maize plants. *Journal of Botany* 34:582
- Coe, E.H. 1959. A line of maize with high haploid frequency. *American Naturalist* 93:381-382
- Duncan, W.G., W.A. Williams, and R.S. Loomis. 1967. Tassels and the productivity of maize. *Crop Science* 7:37-39
- Duvick, D.N. 2005. The contribution of breeding to yield advances in maize (*Zea mays* L.), in : L.S. Donald (Ed.), *Advances in Agronomy*, Academic Press. pp. 83-145
- East, E.M. 1908. Inbreeding in corn. *Connecticut Agric. Exp. Stn. Rep.* 1907:419-428
- FAOSTAT. 2013. Food and agricultural commodities production/Commodities by regions. Food and Agriculture Organization of the United Nations: Statistics Division. http://faostat3.fao.org/faostat-gateway/go/to/browse/rankings/commodities_by_regions/E (accessed 28 Oct 2015)
- Fischer, K., G. Edmeades, and E. Johnson. 1987 Recurrent selection for reduced tassel branch number and reduced leaf area density above the ear in tropical maize populations. *Crop Science* 27:1150-1156
- Ford, R.H. 2000. Inheritance of kernel color in corn: Explanations and investigations. *Am. Biol. Teach.* 6:181-188
- Gayen, P., J.K. Madan, R. Kumar, and K. Sarkar. 1994. Chromosome doubling in haploids through colchicine. *Maize Genetics Cooperation Newsletter* 68:65

- Geiger, H.H. 2009. Doubled Haploids. In: Bennetzen, J.L. and S. Hake, editors, Handbook of Maize. Springer, New York. p. 641-657
- Germana, M.A. 2010. Anther culture for haploid and doubled haploid production. Plant Cell, Tissue and Organ Culture. 104:283-300
- Hallauer, A.R., M.J. Carena, and J.B. Miranda. 1988. Quantitative genetics in maize breeding. In: Prohens, J., F. Nuez, and M.J. Carena, editors, Handbook of Plant Breeding. Springer, New York, NY. p. 1-31
- Hunter, B., T. Daynard, D. Hume, J. Tanner, J. Curtis, and L. Kannenberg. 1969. Effect of tassel removal on grain yield of corn (*Zea mays* L.). Crop Science 9:405-406
- Jenkins, M.T. 1934. Methods of estimating the performance of double crosses in corn. Journal of the American Society of Agronomy 26:199-204
- Jones, D.F. 1918. The effects of inbreeding and crossbreeding upon development. Conn. Agric. Exp. Stn. Bull. 207:5-100
- Kebede, A.Z., B.S. Dhillon, W. Schipprack, J.L. Araus, M. Banziger, K. Semagn, G. Alvarado, and A.E. Melchinger. 2011. Effect of source germplasm and season on the in vivo haploid induction rate in tropical maize. Euphytica 180:219-226
- Lashermes, P. and M. Beckert. 1988. Genetic control of maternal haploidy in maize (*Zea mays* L.) and selection of haploid inducing lines. Theoretical and Applied Genetics 76:405-410
- Ludwig, S.R., and S.R. Wessler. 1990. Maize R gene family: Tissue specific helix-loop-helix proteins. Cell 62:849-851
- Nanda, D.K., and S.S. Chase. 1966. An embryo marker for detecting monoploids of maize (*Zea mays* L.). Crop Science 6:213-215

- Prigge, V., C. Sanchez, B.S. Dhillon, W. Schipprack, J.L. Araus, M. Banziger, and A.E. Melchinger. 2011. Doubled haploids in tropical maize: I. Effects of inducers and source germplasm on in vivo haploid induction rates. *Crop Science* 51:1498-1506
- Prigge, V., X. Xu, L. Li, R. Babu, S. Chen, G.N. Altin, and A.E. Melchinger. 2012. New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize. *Genetics* 190:781-793
- Randolph, L.F. 1932. Some effects of high temperature on polyploidy and other variations in maize. *PNAS* 18:222-229
- Rober, 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
- Shull, G.H. 1908. The composition of a field of maize. *Am. Breeders' Assoc. Rep.* 4:296-301
- Wedzony, M., F. Rober, and H.H. Geiger. 2002. Chromosome elimination observed in selfed progenies of maize inducer line RWS, VII International Congress on Sexual Plant Reproduction. Maria Curie-Sklodowska. University Press, Lublin, Poland. pp. 173
- Zhang, Z., F. Qui, Y. Liu, K. Ma, Z. Li, and S. Xu. 2008. Chromosome elimination and in vivo haploid production induced by Stock-6 derived inducer line in maize (*Zea mays* L.). *Plant Cell Rep.* 27:1851-1860

CHAPTER TWO**A DIALLEL ANALYSIS OF A MAIZE DONOR POPULATION RESPONSE TO
IN VIVO MATERNAL HAPLOID INDUCTION I: INDUCIBILITY**

*Gerald N. De La Fuente*¹, Ursula K. Frei¹, Benjamin Trampe¹, Dan Nettleton², and
Thomas Lübberstedt^{1,*}

Manuscript in preparation for submission to Crop Science. Abstract, structure, and references are all formatted according to journal standards.

Abstract

The maize *in vivo* maternal doubled haploid (DH) system is an important tool used by maize breeders and geneticists around the world. The ability to rapidly produce DH lines of maize for breeding allows breeders to quickly respond to new selection criteria based on the ever changing biotic and abiotic stresses that maize is subjected to across its growing area. There are two important steps in the generation of DH lines using the *in vivo* maternal DH system: 1) the production and identification of haploid progeny, and 2) the doubling of chromosomes to create fertile, diploid inbred lines that can be used for topcross/per se evaluation. For this study, the focus is the first step, the production and identification of haploid progeny. In this study, a diallel mating between six inbred lines of maize GF1, GF2, GF3, GF4, GF5 and GF6 was produced to study the genetic makeup of inducibility in temperate maize germplasm. A maximum estimated rate of inducibility was found in GF1/GF2 at 15%. Significant general combining ability (GCA) effects as well as significant effects for specific combining ability (SCA), reciprocal effects (REC), environmental effects (ENV), as well as GCA by ENV and SCA by ENV interactions were found. Misclassification rates ranged from 0-45% in the 30 hybrids

considered. This study supports the use of germplasm with improved inducibility for breeding to improve rates of inducibility in germplasm which has low induction rates.

Introduction

Maize (*Zea mays* L.) is a diverse crop, whose primary application is for food and feed. However, its reach extends far beyond this, providing raw materials for different industries. Maize has been an exemplary model for understanding the power of selection as over the past century we have seen the transition from open pollinated varieties to the highly productive single cross hybrids of today. In maize breeding, as in breeding of all crops, speed is key. The ability to rapidly, efficiently, and economically run breeding cycles is essential to both a successful breeding program, and to the overall security of a robust food, feed, and fuel supply. Though plant breeding is currently limited biologically by the necessity of floral organs for minimization of generation time (De La Fuente, et al. 2013), over time breeders have integrated tools such as winter nurseries and embryo rescue techniques to increase the number of generations possible in each year.

Despite these advantageous techniques, there still remains the complexity that it takes some time (8 generations/4 years) to develop an inbred line suitable for full scale topcross evaluation. Although, it is true that early generation testing provides a hint as to the performance of lines in development (Hallauer et al., 1988), there still remained the necessity to produce a homozygous and homogeneous inbred line which can be stably reproduced and protected which could take 7 years to complete. Doubled haploids (DHs) provided a solution to this challenge.

First recognized by Chase (1947), it is understood that haploid plants occur naturally in maize at a low frequency. Theoretically, their utility for maize breeding and genetics was immediately understood (Chase, 1947). However, it was not until further advancements in maternal inducers and chromosome doubling techniques were refined that maize DHs became commonplace in breeding programs (Coe, 1959; Geiger, 2009, Rober et al., 2005). Today, a large percentage of commercial and public maize breeding is conducted through the use of the DH system. The ability to generate a new fully homozygous and homogeneous inbred line in 12 months or less provides a speed advantage that allows breeders to quickly respond to new market demands and shifts in selection targets.

The maize DH system used today is known as the *in vivo* maternal haploid system. Though this method is most popular (due to its ease of use and less genotype dependency), two other methods exist: *in vivo* paternal haploid system, and *in vitro* anther culture system (Geiger, 2009). Herein, the focus will be on the *in vivo* maternal haploid system and the genetics which control specific steps in the process. This system involves two key biological steps: 1) production of haploid progeny, and 2) doubling of chromosomes (De La Fuente et al., in preparation).

In most inbred line development programs, the breeding cycle begins with the cross of two (or more) parents of interest. Note that it is possible to produce DH lines from any type of line, population, cross, backcross etc. It is even possible to generate DH

lines from an advanced inbred, though it would be of little benefit. Often (since speed is most critical) breeders use an F_1 breeding cross as the donor parent. The donor parent, which is the female in this system, provides all of the genetic information that is passed on to the haploid progeny. These donor F_1 are pollinated by a male that is termed a maternal haploid inducer. Though somewhat counterintuitive, these inducers are termed ‘maternal’ since the resulting haploid progeny from the induction cross are produced on the female leading to the following: 1) the genotype dependency of this system is much lower than the *in vivo* paternal haploid system, 2) this system is more economical as it allows for use of isolation nurseries to generate induction crosses, and 3) the cytoplasm of the resultant line will be from the maternal donor.

The phenomenon of haploid induction in maize, though it is extensively used, is poorly understood. Two competing hypotheses exist. One hypothesis is that some percentage of the pollen from the inducer is able to ‘induce’ the egg cell to begin development leading to a functional haploid embryo without fusion of the sperm and egg cell (Chalyk et al., 2003). The second, and more supported theory, involves the union of the sperm of the inducer and the egg of the donor which stimulates development of the embryo (Wedzony et al., 2002; Zhang et al., 2008). The genome of the inducer is subsequently eliminated from the embryo that is haploid and contains only the genome of the donor parent, which is a result of one meiosis recombination during egg cell development. The production of haploid kernels is a complex phenomenon involving genetic control by both the maternal inducer and the donor (Prigge et al., 2011; Rober et al., 2005). The genetic control of induction (the ability to induce haploids – trait carried

by the inducer) has been extensively investigated through both breeding and genetics (Rober et al., 2005; Prigge et al., 2011; Prigge et al., 2012). Since the discovery and development of ‘Stock 6’ (Coe, 1959), many new maternal inducers have been developed with increasing rates of haploid induction (Rober et al., 2005; Prigge et al., 2011). QTL studies have been conducted and two major QTL, explaining over 60% of the phenotypic variation have been identified which both trace back to ‘Stock 6’ (Prigge et al., 2012). A maternal haploid inducer with a high induction rate will provide, on average, over 30% haploid kernels. However, this is a two sided phenomenon. When inducing a large number of diverse germplasm, any given maternal haploid inducer will likely produce a normal distribution of inducibility (ability to be induced to create haploid kernels – trait of the donor) pointing to quantitative control also on the donor side due to several factors (Rober et al., 2005; Prigge et al., 2011; Kebede et al., 2011). Although the genetics of haploid induction have been well studied, to our knowledge few previous studies exist that consider the effect of the donor population and were mostly conducted in tropical germplasm (Rober et al., 2005; Prigge et al., 2011; Kebede et al., 2011).

The efficient and economic production of DH lines relies on the ability to produce sufficient numbers of haploid kernels. Based on experience at the Iowa State Doubled Haploid Facility, there, in all germplasm pools (even in elite adapted material) there exists some germplasm with low inducibility. This limits the potential pool of breeding material for those programs which conduct breeding primarily or exclusively with DHs. For this reason, and to prime and facilitate future mapping experiments we screened a set of diverse maize inbreds (publicly and privately developed) for inducibility and

spontaneous doubling potential (see accompanying paper). Though most produced an inducibility rate near the average (data not shown), three lines were identified which produced average induction rates in excess of 25% when induced with RWS/RWK-76 (Rober et al., 2005). These three lines, along with the three lines selected for their spontaneous doubling ability were then mated in a full diallel to study the genetic components of inducibility and spontaneous doubling ability (De la Fuente et al., in preparation). A diallel mating scheme was selected because F_1 donors are typically the generation used for induction, and it is also of interest to answer the question of whether or not poor lines can be ‘supplemented’ by lines which are superior for inducibility, if specific combinations produce superior inducibility rates, and finally if there is an effect of the direction of the cross. This is also a good system to study the interaction of genes controlling induction.

Using a six parent diallel, where three of the inbreds were identified as high inducible and three were identified as having high spontaneous chromosome doubling ability, the objectives of this study were to 1) investigate the genetics and practical use of inducibility for use in the maternal DH system, 2) evaluate the inheritance of inducibility, and 3) understand the interactions between genes controlling induction.

Materials and methods

Germplasm

Six inbred lines were selected for use in a complete diallel: three of which are highly inducible (IND), but do not have spontaneous chromosome doubling (SCD)

potential and three of which are poorly inducible, but have high SCD potential. The lines and details of their heterotic grouping, flowering time, pedigree, and origin are presented in Table 1. Seed from all three lines was acquired from the USDA North Central Regional Plant Introduction Station. The six inbred lines were crossed in a full diallel, creating a set of 30 unique F₁ hybrids. These hybrids were pollinated with the maternal haploid inducer RWS/RWK-76 in 2014 at the ISU-AEA to produce haploid seed for each of the hybrids. Seed was then visually sorted using the *RI-nj* color marker to determine the percentage of haploid seed.

Production of diallel

A diallel mating design was chosen due to the high preference of the use of F₁ donors for *in vivo* maternal haploid induction, the ability to answer various hypotheses, and to subsequently use generated haploid seed to study SCD. As previously mentioned, the six parents were specifically selected for their high trait values of either inducibility or spontaneous doubling potential. The six inbreds were mated in a full diallel (reciprocals included) producing 30 unique F₁ combinations between the six inbreds. The diallel crosses were first made in winter 2013 at Tuniche Seed Services in Graneros, Chile. Fortunately, enough seed was made of each of the crosses to perform the first experiment, however, due to nicking issues (mostly with GF4 and GF2) the diallel crosses were repeated in summer of 2014 in Ames, IA at the Iowa State University Agricultural Engineering and Agronomy Farm (AF). All ears from each specific cross were bulk harvested.

Trials and trait scoring

The diallel was grown in two separate environments over the course of two years. The first environment was in summer of 2014 (AF14) at the Iowa State University Agronomy Farm (AF), the second in summer of 2015 (AF15) at AF. The experiment was grown in a four-replication randomized complete block design with 5.4m single row plots in AF14 and 3.8m plots in AF15 both with 0.76m row spacing. All plots were manually detasseled and shoots were covered prior to silk emergence. All plots were pollinated with bulk pollen from either RWS/RWK-76 or an F_2 generated from this F_1 whose induction rate is not statistically different (data not shown). Plots were then bulk harvested, dried and shelled.

A random sample of 1000 (or as many as possible) kernels were then counted from the bulk. 1000 kernels were selected as the sample based on simulation studies conducted with data from a preliminary experiment. Data from the preliminary experiment was used to model inducibility and the associated variances and tests were conducted using different numbers of binomial trials (i.e. sorted kernels). A balance was sought between the ability to detect significant differences between small percent changes in inducibility rates as well as the number of kernels needed to be sorted (i.e. it is easier to sort 100 kernels per sample, but 100 does not provide enough power to detect differences of 5%, while sorting 10,000 would provide sufficient power, but would make the experiment larger than labor available for sorting). These kernels were then sorted into putative hybrids and haploids. Kernels were sorted based on the R1-nj seed based marker system (Nanda and Chase, 1966). Kernels that showed coloration in the aleurone

(expected for successful fusion of inducer sperm with the central cell) and in the embryo (expected for successful fusion of inducer sperm with the egg cell) were scored as hybrid. Those kernels which showed coloration in the aleurone, but not in the embryo (expected for successful induction of haploid embryo) were scored as putative haploids. A base inducibility rate (IR) was calculated as number of putative haploids divided by total number of kernels. However, as noted in other studies (Kebede et al., 2011; Prigge et al., 2011) the R1-nj system is not 100% accurate. Haploid seed were grown and based on this a misclassification rate (following Kebede et al., 2011) was calculated to adjust for misclassified haploid kernels. After approximately three weeks of growth, the hybrid plants distinguish themselves from the haploids due to increased vigor. The hybrid-appearing plants were counted as misclassified haploids and then divided by the total number of planted seed (non-germinated seed were assumed to be haploids). The corrected induction rate was then calculated as follows. Corrected induction rate = $\text{\#haploids} \times (1 - \text{misclassification rate}) / \text{total number of planted seed}$.

Statistical Analyses

The combining ability analysis was conducted using DIALLEL-SAS05 (Zhang et al., 2005) considering all F_1 s and reciprocals, also known as method 3 (Hallauer, 1988). IRs were adjusted for the misclassification rate, and then transformed for analysis using the angular transformation to normalize the distribution (Prigge et al., 2011). In our experiments, we did not sample germplasm, but characterized defined lines. Thus, a fixed effect model was considered. Estimates for general combining ability (GCA), specific combining ability (SCA), $GCA \times \text{Environment}$, $SCA \times \text{Environment}$, reciprocal (REC),

RECxEnvironment, were all computed. All calculations were done using DIALLEL-SAS05 and respective estimations as described in Bolboaca et al. (2011).

Preplanned contrasts of interest for the application of the trait were computed. SAS PROC MIXED (version 9.4, SAS Institute, 2013) was implemented using the angular transformed corrected induction rates. The model considered here was $Y_{ijk} = Env_i + Rep(Env)_{i(j)} + Entry_k + Env*Entry_{ik} + e_{ijk}$. Where Y_{ijk} is the mean angular transformed induction rate across the whole experiment, Env_i is the random effect of the i th environment, $Rep(Env)_{i(j)}$ is random effect of the j th replication nested in the i th environment, $Entry_k$ is the fixed effect of the k th entry (F_1 s from diallel), $Env*Entry_{ik}$ is the random interaction between the i th environment and the k th entry, and e_{ijk} is the residual error. Contrast statements were used to test the effect of using the inducible lines as males and females when crossed to other inducible lines and when crossed to other non-inducible lines to test if inducible lines can be used as parents in a cross to produce higher inducibility in a non-inducible background.

Results

For this experiment, across both environments, a total of 233,665 (120,000 in AF14, and 113,665 in AF15) seed were sorted. A total of 27,174 putative haploids were identified with visual sorting giving an uncorrected IR of 11.6% with uncorrected values ranging from 2.6%-32.5%. For AF14, the uncorrected IR was 13.1% and for AF15 the uncorrected IR was 10.2% (Table 2). Misclassification rates ranged from 0%-45.2% with averages for each parental line of: 5.8% (GF1), 4.1% (GF2), 4.5% (GF3), 11.8% (GF4),

16.19% (GF5), and 9.17% (GF6). GF4/GF5 and GF6/GF5 had the highest misclassification rates at 45.2% and 32.1% respectively. After accounting for misclassifications, the average IR was 10.5% with values ranging from 2.4%-30.5%.

In the combining ability analysis significant sources of variation were estimated for GCA (p-value= <0.001), SCA (p-value=0.0019), REC (p-value=0.0028), ENV (p-value= <0.001), GCA by ENV interaction (p-value=0.0012), and SCA by ENV interaction (p-value=0.016) (Table 3). Along the diagonal, GCA estimates for the six inbreds are presented. Above the diagonal, SCA effects are presented, and below the diagonal, REC effects are presented. In general, the estimated effects are low. For GCA, GF2, GF3, GF4, GF5 and GF6 had significant GCA estimates at: 0.05%, 0.063%, -0.01%, 0.003%, and -0.14%, respectively. For SCA, two specific crosses had significant estimates: GF2/GF6 (-0.05%) and GF3/GF6 (0.06%). For reciprocal estimates, only GF1/GF4 had a significant reciprocal effect of 0.06%.

Least square means of IR across both environments are presented in Figure 1. The highest estimated IR, corrected for misclassification, is GF1/GF2 (15%) and the lowest estimated IR is GF6/GF4 (7%). GF3 and GF5 predominate the crosses in the top ten estimates for IR (10/20 possible parents). GF6 and GF1 predominate the crosses in the lowest ten estimates for IR (11/20 possible parents). As seen in Figure 1, GF4/GF5, GF5/GF4, GF4/GF6 and GF6/GF5 have high rates of misclassification (also seen in Table 2). For inducibility the narrow sense heritability estimate was 0.6, while the broad sense heritability estimate was 0.1. For the preplanned contrasts, when crossed as a

female GF2 provided a significant difference (p-value=0.046) when crossed to inducible lines and non-inducible lines. The same is true for GF4 when used as a female (p-value=0.049), and as a male (0.021), and for GF6 when used as a female (p-value=0.033). Finally, in conjunction with another study on the same 30 hybrids, but considering rates of SCD in the haploid plants, the correlation between the two traits is -0.04. Overall there is not much similarity between the hybrids with high inducibility and high SCD. However, interestingly, GF1/GF2 has the highest rates of both SCD and inducibility. Also of interest, GF2/GF5 has the second highest rate of inducibility, but has the worst rate of SCD.

Discussion

The effective and efficient use of the maize *in vivo* maternal DH system relies on the ability to efficiently produce and sort, and effectively identify haploid progeny. The most ideal situation would be that the IR is 100% and all seed produced on the ear are haploids. This is of course theoretically impossible, because if the IR for a given inducer is 100%, then when you self-pollinate the inducer 100% of the progeny will be haploid and thus sterile and there would be no way to maintain the inducer for future use. A balance between IR and number of haploids the breeder is willing to have in the male rows must be sought. There are multiple avenues by which to increase the efficiency of this system 1) the development and use of improved inducer lines (Frei et al., in preparation; Prigge et al., 2011; Coe, 1959; Geiger, 2009; Rober et al., 2005), 2) improved methods of haploid selection through new traits and/or automation (De La Fuente et al., submitted; Melchinger et al. 2015; Smelser et al., 2015; Jones et al., 2012),

and 3) the method considered in this and other studies (Prigge et al, 2011; Kebede et al., 2011), improvement of rates of induction in the donor material. It is possible that there are interactions between 2 and 1, but conflicting reports show that in some cases there is interaction between donor and inducer (Prigge et al., 2011), and in some cases not (Kebede et al., 2011).

Induction Rates

Induction rates in this study average 10% across both environments and all 30 hybrids, which is fairly consistent with averages seen at the ISU-DHF (Frei, personal communication). When compared to other studies done, IRs in this study are slightly higher than those reported: 6.74% (Kebede et al., 2011), and 7.63% (Prigge et al., 2011). The higher induction rates seen in this study are likely due to the more adapted nature of the germplasm in consideration. Both studies (Kebede et al., 2011; Prigge et al., 2011) consider tropical germplasm both pollinated with RWS (Rober et al., 2005). It is possible that more elite germplasm from temperate adapted environments provide better overall IR.

Environmental effects on induction rates

Induction rates presented for AF14 (13.1%) were higher than those for AF15 (10.2%). Significant effect of environment in the combining ability analysis also supports this difference. An effect of environment was also reported by Kebede et al. (2011) and Prigge et al. (2012), however, in this study a significant effect of GCA by ENV was seen unlike that reported in Kebede et al. (2011). The significant GCA by ENV interaction could have come from the abnormally cool season that was experienced at AF15 during pollination. Unseasonably cool temperatures and high amounts of rainfall during the peak

of pollinations in this experiment may have had some effect on the IR. In addition to the odd weather pattern, a significant amount of biotic stress was experienced due to the cold weather which may have reduced the IR, this has been previously reported in maize (Geiger, 2009).

Combining ability

GCA and SCA estimates in this study were, in general, very low. This is not surprising when the range in average corrected IR for each of the inbred lines in the diallel was only 10%-12%. Ranges reported for a similar study in tropical maize (Kebede et al., 2011) were 2.9-9.66% across environments. Their inclusion of more contrasting germplasm provided with higher estimates for GCA. The highest estimate for GCA in this study was 0.06% for GF3 while the highest reported in Kebede et al., (2011) was 1.06%. However, similar low percentage effects were also reported for other lines in their study. In contrast to results from Kebede et al. (2011), however, this study reports significant SCA and REC effects. Significant SCA effects were reported for GF2/GF6 (0.05%) and GF3/GF6 (-0.06%). It may be that SCA effects were not found in Kebede et al. (2011) due to the fact that their germplasm was all sourced from breeding lines and breeding material from CIMMYT (International Center for Maize and Wheat Improvement), while germplasm for this study was sourced from both private and public breeding programs from different times and locations. More diversity in the six parents in this study may have lead to the SCA seen. No other study has reported on reciprocal effects for IR. In this study a significant effect of REC was estimated and GF5/GF4 had a significant REC effect (-0.06%). REC effect would be an important consideration as it could allow breeders to increase their IR by simply planning the direction of their cross.

Misclassification rates

Misclassification rates of haploid seed can become a significant problem if they are too high. For example, if a breeder needs to plant a total of 1000 haploid seed to produce enough DH lines, and the misclassification rate is 50%-75% then $\frac{1}{2}$ to $\frac{3}{4}$ of the plants in the field will be hybrids and will need to be removed. This is a waste of time, materials, and field space all of which add cost to the breeding program. The objective of DHs is speed and efficiency, and wasting time and money is counterproductive. In this study, misclassification rates ranged from 0%-45%. For the specific lines included in the study, GF1, GF2 and GF3 had misclassification rates that were within acceptable limits 4.5-5.8% while GF5 had an unacceptable rate of misclassified haploids (16%) averaged across all crosses. Misclassification rates were, on average, lower than those reported in other studies (Kebede et al., 2011; Prigge et al., 2011; Rober et al., 2005). Both Prigge et al. (2011) and Rober et al. (2005) report that the inclusion of unadapted landraces and populations increase the misclassification rates. This could be an explanation of why misclassification rates in this study are lower. Alleles that modify the R locus may still be segregating in unadapted material. Based on the results presented here and the genetic contributions of GF4, GF5, and GF6 and their parents to the dent germplasm it may be desirable for breeders to obtain a small sample (i.e. 100-250 seed) and germinate or cut the seed (cutting the seed allows for better visualization of the pigmentation) to come up with an estimate for the misclassification rate so that it can be accounted for when estimating the number of haploid seed that need to be planted.

Breeding for increased IR

This study is in agreement with others (Kebede et al., 2011; Prigge et al., 2011; Rober et al., 2005) that there is significant variation in IR and that the environment has an effect on the average induction rate, and that selection could be possible for improvement of IR. As noted above, heritabilities of IR are low, so it will be more difficult of a trait to select for. This study in addition concludes, that there are significant interactions between the germplasm and the environment and that specific combinations and the direction of the cross have an effect on IR. As is reported by others (Kebede et al., 2011; Prigge et al., 2011; Rober et al., 2005) when unadapted germplasm is used the rates of misclassification go up and the rates of induction go down. It may then be possible for breeders to utilize more adapted lines with high induction rates and low misclassification rates to cross into unadapted material to raise the rates of induction and lower the rates of misclassification through selection. When considering using the lines and/or hybrids in this study for breeding, it follows that GF1/GF2 would be the superior hybrid to breed with. Not only does GF1/GF2 have the best rates of inducibility, but it also has been estimated to have the highest rates of SCD (De La Fuente, et al., in preparation). Combining both of these traits into the germplasm would be beneficial. Overall, however, it seems this is the only combination that could provide this advantage as the correlation between SCD and inducibility is low ($r^2=-0.04$) and it seems that in general there is no relationship between the two. It is evident that IR is a quantitative trait, so combining both traits would take some time, but if it could be coupled through cycles of selection utilizing the DH system with some form of SCD (De La Fuente et al., in preparation) then

it could be a cheap and quick method to develop germplasm that could be rapidly incorporated into breeding programs for evaluation of other useful traits.

Acknowledgments: Authors 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 funding this work.

References

- Bolboaca, S.D., L. Jantschi, A.D. Baci, and R.E. Sestras. 2011. Griffing's experimental method II: step by step descriptive and inferential analysis of variances. *Journal of Biostatistics* 6:31-52
- Chalyk, S., A. Baumann, G. Daniel, and J. Eder. 2003. Aneuploidy as a possible cause of haploid-induction in maize. *Maize Genetics Cooperation Newsletter* 77:29
- Chase, S.S. 1947. Techniques for isolating monoploid maize plants. *Journal of Botany* 34:582
- Coe, E.H. 1959. A line of maize with high haploid frequency. *American Naturalist* 93:381-382
- De La Fuente, G.N., U.K. Frei, and T. Lubberstedt. 2013. Accelerating plant breeding. *Trends in Plant Science* 18:667-672
- Geiger, H.H. 2009. Doubled Haploids. In: Bennetzen, J.L. and S. Hake, editors, *Handbook of Maize*. Springer, New York. p. 641-657

- Hallauer, A.R., M.J. Carena, and J.B. Miranda. 1988. Quantitative genetics in maize breeding. In: Prohens, J., F. Nuez, and M.J. Carena, editors, Handbook of Plant Breeding. Springer, New York, NY. p. 1-31
- Jones, R.W., T. Reinot, U.K. Frei, Y. Tseng, T. Lubberstedt, and J.F. McClelland. 2012. Selection of haploid maize kernels from hybrid kernels for plant breeding using near-infrared spectroscopy and SIMCA analysis. *Applied Spectroscopy* 66:1-4
- 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 Science* 55:188-195
- Nanda, D.K., and S.S. Chase. 1966. An embryo marker for detecting monoploids of maize (*Zea mays* L.). *Crop Science* 6:213-215
- Prigge, V., C. Sanchez, B.S. Dhillon, W. Schipprack, J.L. Araus, M. Banziger, and A.E. Melchinger. 2011. Doubled haploids in tropical maize: I. Effects of inducers and source germplasm on in vivo haploid induction rates. *Crop Science* 51:1498-1506
- Prigge, V., X. Xu, L. Li, R. Babu, S. Chen, G.N. Altin, and A.E. Melchinger. 2012. New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize. *Genetics* 190:781-793
- Rober, 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
- SAS Institute. 2013. SAS version 9.4. SAS Institute, Cary, NC

- Smelser, A., M. Blanco, T. Lubberstedt, A. Schechert, A. Vanous, and C. Gardner. 2015. Weighing in on a method to discriminate maize haploid from hybrid seed. *Plant Breeding* 134:283-285
- Wedzony, M., F. Rober, and H.H. Geiger. 2002. Chromosome elimination observed in selfed progenies of maize inducer line RWS, VII International Congress on Sexual Plant Reproduction. Maria Curie-Sklodowska. University Press, Lublin, Poland. pp. 173
- Kebede, A.Z., B.S. Dhillon, W. Schipprack, J.L. Araus, M. Banziger, K. Semagn, G. Alvarado, A.E. Melchinger. 2011. Effect of source germplasm and season on the in vivo haploid induction rate in tropical maize. *Euphytica* 180:219-226
- Zhang, Y., M.S. Kang, and K.R. Lamkey. 2005. DIALLEL-SAS05: a comprehensive program for Griffing's and Gardner-Eberhart analyses. *Agronomy Journal* 97:1097-1106
- Zhang, Z., F. Qui, Y. Liu, K. Ma, Z. Li, and S. Xu. 2008. Chromosome elimination and in vivo haploid production induced by Stock-6 derived inducer line in maize (*Zea mays* L.). *Plant Cell Rep.* 27:1851-1860

Table 1. Pedigree, origin, and general information for six parents included in diallel.

Line Name	Heterotic Group	Origin	Pedigree	GDD to Silk	Selection Reason
GF1	non-stiff stalk	Minnesota	CC36 x A405	1318	spontaneous chromosome doubling
GF2	stiff stalk	Minnesota	CO106 x A321	1522	spontaneous chromosome doubling
GF3	non-stiff stalk	Nebraska	W117Ht x Mo17Ht	1178	high inducibility rate
GF4	stiff stalk	Hawaii	Oh40B, L317, GF5	1640	high inducibility rate
GF5	stiff stalk	Indiana	Indiana strain of Stiff Stalk Synthetic	1522	high inducibility rate
GF6	stiff stalk	Minnesota	W117 x B37Ht	1400	spontaneous chromosome doubling

Table 2. Misclassification rates of the 30 hybrids from this diallel. Rates presented here as percentages.

Entry	Misclassification Rate (%)	Std. Dev.
GF1/GF2	6.0	0.02
GF1/GF3	4.8	0.02
GF1/GF4	6.0	0.04
GF1/GF5	6.0	0.02
GF1/GF6	4.8	0.04
GF2/GF1	4.8	0.02
GF2/GF3	1.2	0.00
GF2/GF4	3.6	0.02
GF2/GF5	3.6	0.02
GF2/GF6	1.2	0.02
GF3/GF1	2.4	0.02
GF3/GF2	4.8	0.02
GF3/GF4	3.6	0.02
GF3/GF5	6.0	0.04
GF3/GF6	6.0	0.04
GF4/GF1	11.9	0.05
GF4/GF2	3.6	0.28
GF4/GF3	4.8	0.28
GF4/GF5	45.2	0.29
GF4/GF6	4.8	0.04
GF5/GF1	11.9	0.04
GF5/GF2	8.3	0.09
GF5/GF3	7.1	0.06
GF5/GF4	21.4	0.12
GF5/GF6	20.2	0.11
GF6/GF1	0.0	0.04
GF6/GF2	4.8	0.04
GF6/GF3	4.8	0.23
GF6/GF4	13.1	0.22
GF6/GF5	32.1	0.22

Table 3. Combining ability analysis. Reported as percent corrected IR. Diagonal values are GCA, above the diagonal are SCA, and below the diagonal are reciprocal effects.

*denotes significance at the 0.05 level.

Inbred	GF1	GF2	GF3	GF4	GF5	GF6
GF1	-0.0003	-0.0333	-0.0210	0.0014	0.0032	-0.0033
GF2	0.0835*	0.0500*	-0.0339	0.0256	0.0035	-0.0478*
GF3	-0.0044	-0.0050	0.0630*	0.0046	0.0005	-0.0569*
GF4	0.0058	-0.0067	0.0029	-0.0129*	-0.0339	0.0000
GF5	0.0258	0.0702	-0.0041	-0.0645*	0.0027*	0.0019
GF6	-0.0025	0.0161	-0.0324	0.0092	0.0700	-0.1357*

Table 4. Preplanned contrasts for combinations of inducible (IND) and non inducible (non-IND) lines when used as males and females. *=significance at the 0.05 level.

Contrast	F Value	Pr > F
GF1 as female: crossed to IND vs non-IND	0.02	0.8809
GF1 as male: crossed to IND vs non-IND	1.54	0.2248
GF1 as female vs GF1 as male	0.43	0.5192
GF2 as female: crossed to IND vs non-IND	4.32	0.0467*
GF2 as male: crossed to IND vs non-IND	1.98	0.1704
GF2 as female vs GF2 as male	0.05	0.8254
GF3 as female: crossed to IND vs non-IND	0.37	0.5504
GF3 as male: crossed to IND vs non-IND	0.66	0.4235
GF3 as female vs GF3 as male	0.01	0.912
GF4 as female: crossed to IND vs non-IND	4.2	0.0494*
GF4 as male: crossed to IND vs non-IND	5.92	0.0214*
GF4 as female vs GF4 as male	0.03	0.8567
GF5 as female: crossed to IND vs non-IND	0.91	0.347
GF5 as male: crossed to IND vs non-IND	0.36	0.5509
GF5 as female vs GF5 as male	0.01	0.9304
GF6 as female: crossed to IND vs non-IND	4.99	0.0334*
GF6 as male: crossed to IND vs non-IND	4.22	0.0491*
GF6 as female vs GF6 as male	0.4	0.5317

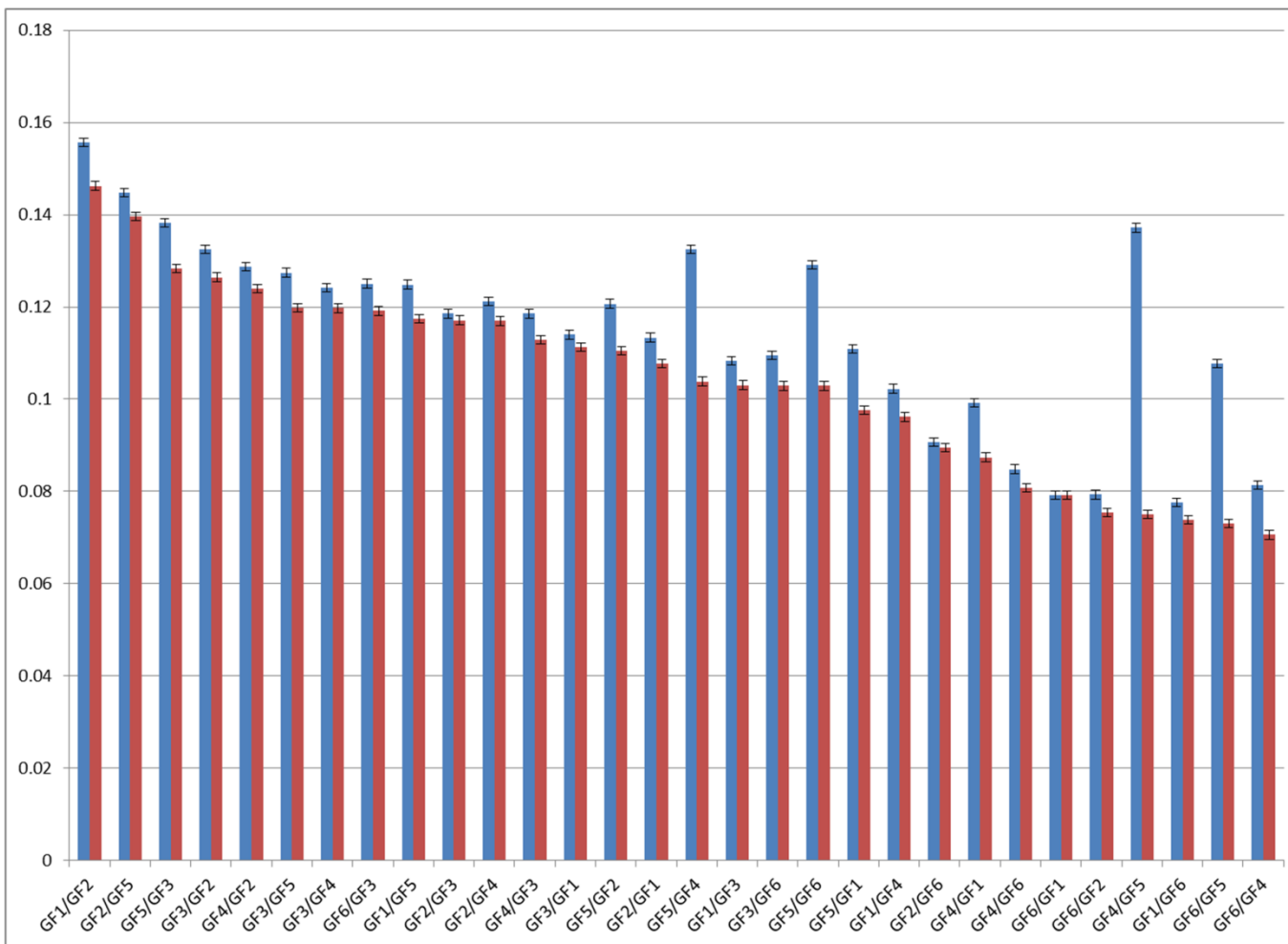


Figure 1. Least square means of 30 hybrids for inducibility with (red) and without (blue) correction for misclassification. Least square means are presented as ratios.

CHAPTER THREE

REGISTRATION OF BHI306 MAIZE MATERNAL HAPLOID INDUCER GERMPLASM

Ursula K. Frei^{1,*}, Gerald N. De La Fuente¹, and Thomas Lübberstedt¹

Manuscript in preparation for submission to Journal of Plant Registrations. Abstract, structure, and references are all formatted according to journal standards.

Abstract

BHI306 maize germplasm was developed by the Doubled Haploid Facility, Department of Agronomy, Iowa State University. This germplasm line is a maternal haploid inducer that can be used to generate haploid progeny for the development of doubled haploid lines in maize using the maternal *in vivo* haploid induction system. This germplasm is unique as it combines multiple traits that make it a versatile inducer which can be used in various programs and it is adapted to the Midwest U.S. BHI306 has an 11% induction rate which is as good as the popular maternal haploid inducers RWS and RWK-76, it also has a germination rate of 78% which is significantly higher than RWS (36%) and RWK-76 (21%). BHI306 carries the *P11* root marker, which allows selection on red root color when the popular *RI-nj* marker does not work. Finally, BHI306 carries the *Gal* allele which makes it compatible for pollination onto popcorn donor populations, which increases its versatility of use over other publicly available inducers.

Introduction

The advent of the maize (*Zea mays* L.) doubled haploid (DH) system has accelerated the cycle of maize breeding and genetics in a significant way. Over the past

two decades, an increasing number of breeding programs (both public and private) have used doubled haploids with great success to accelerate both line development and population development for genetic studies. Under normal conditions maize is a diploid organism with $2n=20$. However, Chase (1947) found that at a low frequency (1/1000) haploid individuals are produced naturally under normal conditions. This prompted continued studies and experiments, whereby the mechanism of haploid induction was investigated, as well as its utility for maize improvement.

Production of DH lines in maize can be accomplished through both *in vivo*, and *in vitro* methods. *In vitro* methods (microspore or anther culture) are highly genotype dependent as only a limited number of genetic backgrounds respond well to tissue culture techniques. For this reason, *in vivo* haploid induction is the preferred method of haploid production. There are two methods for production of *in vivo* haploids. In the paternal haploid induction system, the inducer carries the indeterminate gametophyte (*ig*) gene and is used as the female in the cross. The population of interest pollinates the inducer and provides the genetic information present in the haploid seed that is produced at a 0-8% frequency (Kindiger and Hamann, 1992). The paternal system is also dependent on the donor genotype (Geiger, 2009), with only some genetic backgrounds able to successfully produce haploids at a high enough percentage to be feasible. Paternal haploid induction is widely used to produce cytoplasmic male sterile analogs to breeding lines. The most commonly used method for haploid production is *in vivo* maternal haploid induction where the inducer is the male parent in the cross and haploids are produced on and from the female donor population. Haploid induction rates for this method vary as well depending on the inducer, the donor population, and the

environment. Though the method is highly successful, it is still unclear what the mechanism for haploid induction involves. Two competing hypotheses exist (Geiger, 2009). The first assumes, that during double fertilization, one pollen nucleus fuses successfully with the embryo sac cells to form a viable endosperm, while the other fails to fuse with the egg cell. The egg cell develops into a haploid embryo supported by a normal triploid endosperm. The second, is that the inducer pollen successfully fuses with the egg cell stimulating division, and the haploid genome from the inducer is subsequently expelled from the embryo early in development (Geiger, 2009).

Despite the unknown mechanism for maternal haploid induction, this method is used with great success in private breeding companies as well as public institutions such as the Doubled Haploid Facility at Iowa State University (ISU-DHF). Recent QTL mapping studies (Prigge et al., 2012; Lashermes and Beckert, 1988) have identified loci which are required for haploid induction in the maternal inducer with two major QTL located on chromosomes 1 and 9 and the QTL on chromosome 1 explaining up to 66% of the variation in their population. The first true maternal haploid inducer was identified by Coe (1959) in a source obtained from Charles R. Burnham in 1950 and was designated 'Stock 6'. Stock 6 carries the QTL on chromosome 1 and on average will provide a haploid induction rate of 1-2%. After this discovery, the interest in haploids declined slightly, but was subsequently revitalized when it was found that through selective breeding the induction rate of inducers could be increased and several popular inducer lines have been developed to date including 'WS14', 'MHI', 'RWS' (Rober et al., 2005), and 'UH400' (Prigge et al., 2011).

The basic outline for maternal haploid induction and production of DH lines using this method is as follows. Season 1: production of breeding crosses for induction, typically a biparental cross F_1 is used, but any generation (including random mating populations) can be used as the donor population for induction. The F_1 is most popular because a major objective of the DH system is speed. Season 2: the donor population (female) is pollinated with bulk pollen from the maternal haploid inducer. A percentage of the seed will have haploid embryos which have undergone one recombination during gamete formation. At harvest, the seed can visually be sorted (in most cases) using the dominant *RI-nj* (Nanda and Chase, 1966) marker. Hybrid seed (which is discarded) has coloration on both the cap of the aleurone and the embryo indicating that there has been successful fertilization in both the egg and central cell. Haploid seed has coloration on the cap of the aleurone, but not on the embryo indicating that there was successful fertilization in the central cell and failed fertilization in the egg which has developed into a haploid embryo. The sorted haploid seed is collected for each donor population and stored. Season 3: haploid seed is germinated and treated with colchicine at the seedling stage using either immersion or injection (Geiger, 2009). Colchicine is a spindle poison and causes a failure of chromosomes to be pulled to opposite poles during mitosis (Borisy and Taylor, 1967). Thus, if successful, a ‘copy and paste’ of the haploid genome is created. Treated plants are subsequently transplanted into the field. Haploid plants develop normally, but are generally shorter, weaker, lighter in color, and have more erect leaves. These plants will be chimeras, having haploid and diploid sectors. If genomes of cell lines developing into reproductive cells have successfully doubled, the plants will

shed small amounts of pollen and have fertile ears. These plants are self-pollinated. Seed of each successfully pollinated plant is a new DH line.

The maternal inducer is used in season 2 to produce the haploid seed and is a vital part of a successful DH production system. Obviously, the maternal inducer must carry two non-negotiable traits: haploid induction ability and a dominant marker gene for selection. However, the inducer must also carry traits of a good male parent. Specifically, the inducer must have a strong tassel which sheds copious amounts of pollen over the course of at least 5-7 days. This translates into a thick rachis with high glume per cm number, many branches, and tolerance to heat. The inducer must also carry other favorable agronomic qualities desirable in any elite line of maize.

Though many maternal inducers are publicly available, it was clear that none were well suited to the temperate environment of the Corn Belt. Many of the most popular inducers are extremely weak plants with poor agronomic characteristics. Specifically, the F₁ hybrid RWS/RWK-76 and its parents, though they have excellent induction rates and a good tassel, are poorly adapted to the U.S. Corn Belt. RWS and RWK-76 are adapted to cooler temperatures and environments of central Europe. In the U.S., this inducer is extremely susceptible to lodging during the high winds experienced during the frequent storms which pass through the Midwest and also to the more extreme heat which is common in the middle of summer. RWS/RWK-76 also only uses the *Rlnj* marker for haploid selection making its use in colored corn impossible. The utility of RWS/RWK-76 in popcorn DH line development is limited due to incompatibility between dent and popcorn types of maize (Kermicle et al., 2006). For these reasons a line

development program was instituted within the DH facility at ISU to develop new, improved, and adapted inducer lines which can be used by researchers in temperate U.S. environments.

Methods

BHI306 was developed using a pedigree breeding scheme utilizing marker assisted selection for loci known to control induction rate. The first cross was made in the summer of 2010 at the Iowa State University Agricultural Engineering and Agronomy Farm (AF) in Boone, IA between A632.75 *A1 A2 C1 R1-r B1 P11*/(RWS/RWK-76). A632.75 is a B14 derived stiff stalk line that has been converted to contain alleles that confer purple coloration in all parts of the plant. Its inclusion in the cross was for donation of B14 adaptation and for the *P11* allele which provides the ability to select for root color as haploid identification criteria. RWS/RWK-76 is an F₁ maternal haploid inducer (Rober et al., 2005) that is widely used, and used by the ISU-DHF for inductions. RWS/RWK-76 provides high induction rates and good tassel traits such as degree of pollen shed and duration of pollen shed, but is poorly adapted to the U.S. Midwest. The F₁ was chosen to incorporate induction traits from both RWS and RWK-76. The seed from the three-way cross was then grown in the greenhouse at ISU in the winter of 2010. A single plant was selected and pollinated with bulk pollen from an F₂ population of the RWS/RWK-76 cross to increase the contribution of loci conferring induction rate. The progeny of this cross were grown in summer 2011 at AF. A single plant was selected and pollinated with bulk pollen from a RWS/RWK-76 F₂ population. Additionally, in summer 2011, a cross was made between (RWS/RWK-76)/'B15-dent sterile'. B15-dent

sterile was included in the cross as the donor of the *Gal* allele which provides fertility when crossed to popcorn which typically carries the *Gal* allele to make it non-receptive to dent maize pollen. In winter 2011 the progeny of the second A632.75 cross were grown alongside the progeny of the B15-dent sterile F₁. A single plant was selected and crossed with pollen from the B14-dent sterile F₁. The resultant population was given the designation PCPOP and was the population used for subsequent selections. In summer of 2012, seed from PCPOP was grown at Agronomy Farm (AF) and the progeny were screened for the major QTL on chromosome 1 in bin 1.03 known to control induction ability (Prigge et al., 2012). Plant 562 was selected and self-pollinated. The resultant F₂ was seed was designated PCPOP562. In winter 2012, PCPOP562 F₂ plants were grown at the winter nursery location for the ISU-DHF in Graneros, Chile with Tuniche Seed Services (CL). The F₂ was grown in three rows for test inductions and in a single row for single plant selection. Single plants were selected and self-pollinated based on favorable plant phenotypes, with special consideration given to sturdy, large tassels which shed pollen over several days. F_{2;3} lines were then grown at AF in summer 2013 in three row plots and single plants within the row were test induced on dent and popcorn hybrids. After three test inductions were made, the plant was self-pollinated. Based on induction rates of the test pollinations as well as the ability to pollinate popcorn, plant 460 was selected (PCPOP562-1-460). The resultant F_{3;4} line was grown in CL in a single row for test inductions on dent and popcorn hybrids along with other lines in development. Six individual F₄ plants were in PCPOP562-1-460 based on plant phenotype as well as tassel phenotype and self-pollinated. Based on the results of induction in dent and popcorn, the six selfs of PCPOP562-1-460 were advanced and grown in F_{4;5} progeny rows in summer

2014 at AF. The lines were evaluated for uniformity, plant type, tassel phenotype, lodging resistance, and most importantly induction rate in dent and popcorn. Based on phenotype, uniformity, and high induction rate as well as good seed set in self-pollinations, single F₅ plant selfs were advanced from PCPOP562-1-460-01 and grown in Chile as F_{5:6} progeny rows. The F_{5:6} progeny rows were test induced on three popcorn hybrids as well as one dent maize hybrid. Based on uniformity, tassel and plant phenotype, and induction rates, six of the F_{5:6} progeny rows were bulk harvested and entered into the inducer line evaluation trial (ILET). The ILET was grown in summer 2015. Each inducer entry, including BHI306, was grown in single row plots replicated four times in a randomized complete block design. Each plot was scored for GDDs to flowering, plant height, haploids in the row, germination, and was test pollinated onto a dent corn hybrid for induction rate comparisons.

Characteristics

BHI306 is classified as a maternal haploid inducer, meaning that it should be used as the pollen (male) parent to be crossed onto donor populations (female) to generate haploid progeny for the development of DH lines. Table 1 summarizes the values for the traits of consideration when evaluation of new inducer lines is conducted. BHI306 has an average induction rate of 11% which is not significantly different than RWS (10.7%) and RWK-76 (10.5%), but better than that of the other maternal inducer included in the comparison: MHI (Eder and Chalyk, 2002) which induced at 2.9%. The F₁ between RWS/RWK-76 induces at 15.4%, but this is not significantly higher than BHI306. BHI 306 reaches 50% anthesis (measured as the time when 50% of plants in plot have anthers

shedding pollen) at 1184 GDDs which is slightly later than RWS and RWK-76, but earlier than MHI. One of the most important traits that was considered is germination which was evaluated at each generation by counting the number of plants that emerged from the direct planting. When using the RWS/RWK-76 inducer, the ISU-DHF incurred a significant increase in labor cost and time because these plants had to be established via transplanting to get enough plants in the field to be used for pollinations. This made planting very laborious, as this transplanting was added onto the additional transplanting of haploid plants for DH line production. This also increased winter nursery costs due to the increase in plot cost for transplanting. As seen in Table 1, BHI306 germinates at a significantly higher rate (78%) than RWS (36%) and RWK-76 (21%). The ability to plant BHI306 with a standard plot planter and get good germination rates will increase the efficiency of inductions, and likely make induction isolation fields possible. BHI306 also had good seed set when self-pollinated making it easier to maintain as an inbred line. One aspect for potential users to consider is that when a maternal inducer is self-pollinated, it creates haploid progeny on the self-pollinated ear at rates similar to induction rates of the inducer. This needs to be considered when calculating planting rates, as the haploids (if they germinate) will typically be sterile.

In addition to high induction and germination rates, as well as good plant phenotype and pollen shed characteristics, BHI306 also carries two additional alleles that increase its utility for DH line production. BHI306 carries the *P11* (Figure 1) gene which is a dominantly inherited locus that creates red coloration in the seedling roots which can

be used for haploid selection similar to the *RI-nj* system. If the *RI-nj* marker does not work (because of the presence of C1 allele in colored maize, and sometimes popcorn and sweetcorn), the seed can be germinated and selection can be made based on the root coloration. Haploid progeny will have white roots and hybrid progeny will have red roots (Figure 1). This increases the utility of the inducer to types of maize that produce problems with the *RI-nj* marker system. In addition to this marker, BHI306 also carries the *Gal* allele which allows it to pollinate popcorn. Most popcorn produced in the U.S. carries the *Gal* allele which makes it incompatible with pollen from dent maize. Dent maize, on the other hand, does not usually carry *Gal* so it cannot pollinate popcorn, or any maize with *Gal*. Since BHI 306 carries the *Gal* allele, it can be used to pollinate both dent maize and popcorn.

Availability

Seed of BHI306 will be maintained by personnel at the Doubled Haploid Facility which is housed in the Department of Agronomy, Iowa State University, Ames, IA 50011. BHI306 is available for licensing through material transfer agreement for those interested in acquiring a maternal haploid inducer line. Requests for BHI306 should be directed to Iowa State University Research Foundation, Inc. Office of Intellectual Property and Technology Transfer, Iowa State University, Ames, IA 50011.

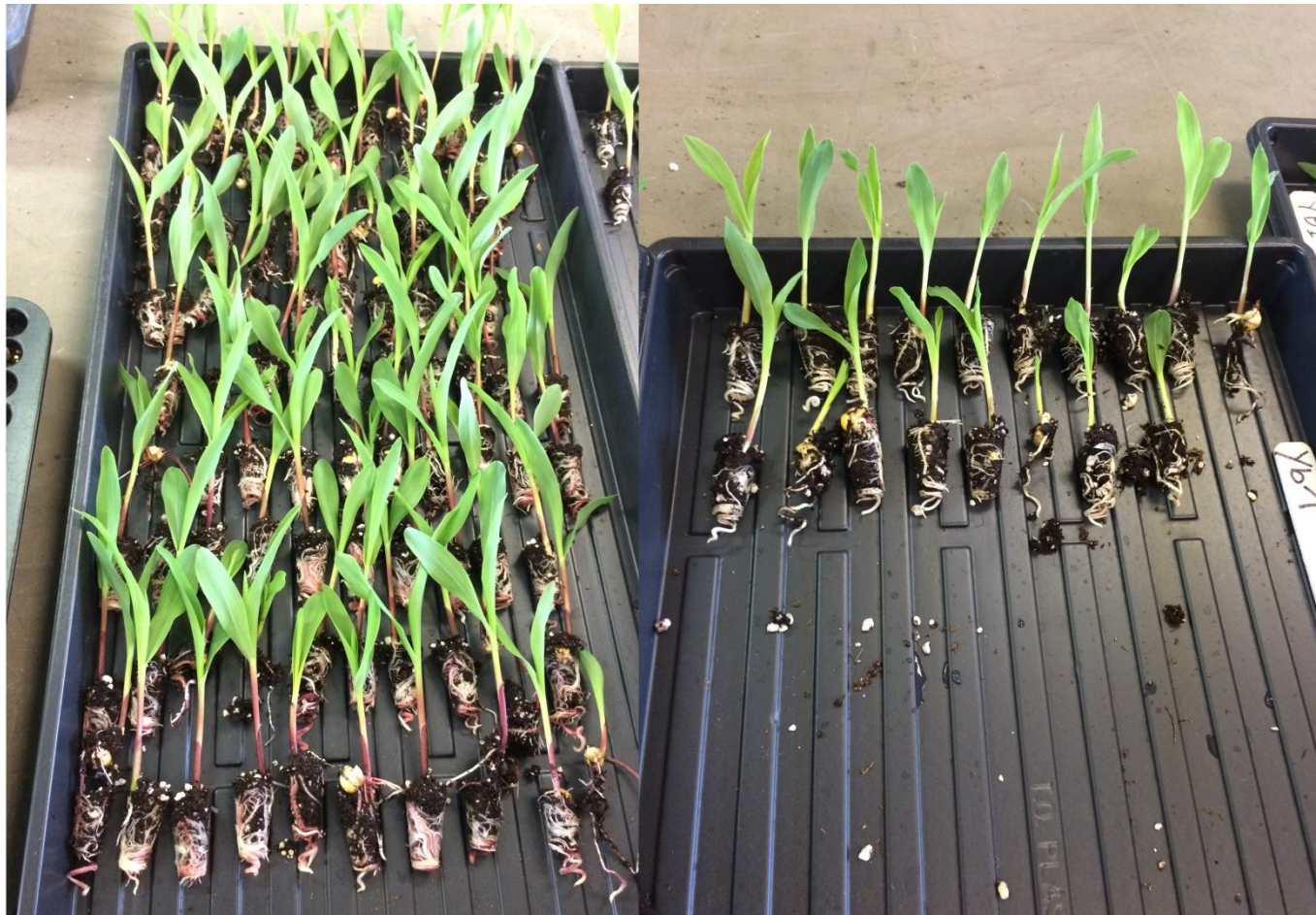
References

- Borisy, G.G., and E.W. Taylor. 1967. The mechanism of colchicine. *Journal of Cell Biology* 34:525-533
- Chase, S.S. 1947. Techniques for isolating monoploid maize plants. *Journal of Botany* 34:582
- Coe, E.H. 1959. A line of maize with high haploid frequency. *American Naturalist* 93:381-382
- Eder, J. and S. Chalyk. 2002. In vivo haploid induction in maize. *Theoretical and Applied Genetics* 104:703-708
- Geiger, H.H. 2009. Doubled Haploids. In: Bennetzen, J.L. and S. Hake, editors, *Handbook of Maize*. Springer, New York. p. 641-657
- Kermicle, J.L., S. Taba, and M.M.S. Evans. 2006. The gametophyte-1 locus and reproductive isolation among *Zea mays* subspecies. *Maydica* 51:219-225
- Kindiger, B., and S. Hamann. 1992. Generation of haploids in maize: a modification of the indeterminate gametophyte (ig) system. *Crop Science* 33:342-344
- Lashermes, P., and M. Beckert. 1988. Genetic control of maternal haploidy in maize (*Zea mays* L.) and selection of haploid inducing lines. *Theoretical and Applied Genetics* 76:405-410
- Nanda, D.K., and S.S. Chase. 1966. An embryo marker for detecting monoploids of maize (*Zea mays* L.). *Crop Science* 6:213-215

- Prigge, V., C. Sanchez, B.S. Dhillon, W. Schipprack, J. L. Araus, M. Banziger, and A.E. Melchinger. 2011. Doubled haploids in tropical maize: effects of inducers and source germplasm on in vivo haploid induction rates. *Crop Science* 51:1498-1506
- Prigge, V., X. Xu, L. Li, R. Babu, S. Chen, G.N. Altin, and A.E. Melchinger. 2012. New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize. *Genetics* 190:781-793
- Rober, 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

Table 1. Summary of inducer line evaluation trial between BHI306 and popular maternal inducers. HIR=haploid induction rate, STAND=stand count, GDD=growing degree days to anthesis?, PLHT=plant height(cm). HIR and STAND data were transformed using the logit transformation prior to analysis. Logit and untransformed data are presented. Groupings are based on LSD at the 0.05 level. Entries with the same group letter are not significantly different.

Trait	Value	BHI306	MHI	RWK-76	RWS	RWS/RWK-76	lsd 0.05
HIR	logit	-2.08	-3.3297	-2.1019	-2.0729	-1.6701	0.51
	untransformed	0.11	0.029	0.105	0.107	0.154	-
	Group	A	B	A	A	A	
STAND	logit	1.26	2.021	-1.264	-0.5654	-0.4976	1.21
	untransformed	0.78	0.886	0.217	0.36	0.376	-
	Group	AB	A	CD	CD	CD	
GDD	-	1184	1206.25	1153.75	1171.25	1114.25	34.5
PLHT	-	198	188.75	161.25	153.75	176.25	



HYBRIDS – RED ROOTS

HAPLOIDS – WHITE ROOTS

Figure 1. *Pll* ‘red root’ marker system. As seen on the left, the haploid seedlings that are hybrids between the inducer and the donor produce red seedling roots. On the right, haploids selected out of the population produce white roots due to the lack of the *Pll* allele. This allows for selection of haploid progeny even if the *R1-nj* seed marker does not work.

CHAPTER FOUR

DISCRIMINATION OF HAPLOID AND DIPLOID MAIZE KERNELS VIA MULTISPECTRAL IMAGING

*Gerald N. De La Fuente*¹, J. Michael Carstensen^{2,3}, Michael A. Edberg², Thomas Lübberstedt¹

Manuscript submitted for publication to Plant Breeding. Abstract, structure, and references are all formatted according to journal standards.

Abstract

The use of doubled haploids (DHs) in maize has become ubiquitous in maize breeding programs as it allows breeders to go from cross to evaluation in as little as two years. Two important aspects of the in vivo DH system used in maize are: 1) the identification of haploid progeny and 2) doubling of the haploid genome to produce fertile inbred lines. This study is focused on the first step. Currently, identification of maize haploid progeny is done manually using the *R1-nj* seed color marker. This is a labor intensive and time consuming process, a method for automated sorting of haploids would increase the efficiency of DH line development. In this study, six inbred lines were crossed with the maternal haploid inducer 'RWS/RWK-76' and a sample of seed was sorted manually for each line. Using the VideometerLab 3 system, spectral imaging techniques were applied to discriminate between haploids and hybrids. Using DNA markers to confirm the haploid/diploid state of the tested seed, for the majority of genotypes haploid identification was possible with over 50% accuracy when comparing the DNA analysis to the visual results.

Keywords maize, haploid, diploid, fluorescence, automated sorting

Introduction

In plant breeding, breeding cycle speed is key, as expressed in the genetic gain equation (De la Fuente et al., 2013). Per cycle gains are limited by the denominator which generally contains time and cost. Over time, breeders found various ways to accelerate the timeline by using tools such as winter nurseries and early generation testing. However, it still takes time to generate the final inbred line with a level of homozygosity and homogeneity which is acceptable. Rapid development of 100% homozygous and homogeneous lines is accomplished by development of doubled haploid (DH) lines (Geiger, 2009).

DHs are used with great success in other crops besides maize (*Zea mays* L.), and their use and acceptance continues to increase in maize breeding. Development of DH lines is more technically demanding compared to inbred line development by continued self-pollination. First discovered in the 1940s by Chase (1949), haploid plants in maize are naturally occurring at a low frequency. Their utility for genetics and breeding was recognized, but use of DHs was not immediately accepted due to the low frequency of haploid kernels and inability to efficiently produce fertile haploid plants. Subsequent development of the maternal haploid inducer 'Stock 6' (Coe, 1959) and other improved inducer lines, and development of economic and applicable protocols for the production of DH lines led to a dramatic increase in line development using the *in vivo* maternal haploid system in maize during the past two decades (Geiger, 2009).

For successful *in vivo* maternal haploid induction, a few key steps must be met. First, haploids must be generated on the maternal donor plant. Second, kernels with haploid embryo (“haploids”) must be distinguishable and separated from undesirable hybrid kernels. Third, haploid plants are treated with colchicine to double their genome number and self-pollinated to generate the final DH line. Herein, we focus on the second step: successful identification of haploids out of a mixture with undesirable hybrid kernels. On average, we expect that approximately 10% of the total number of kernels in a given lot of induced kernels will be haploid. As a consequence, 90% of the kernels are undesirable as their embryo contains 50% each donor and inducer genomes. Although alternative selectable markers are under investigation, the most widely and successfully used selectable marker is *RI-nj* (personal communication with various breeding companies). *RI-nj*, is successful due to its dominant inheritance, and ability to distinguish between its transmission to the triploid endosperm and the diploid embryo. *RI-nj* produces a red coloration in the cap of the aleurone (endosperm transmission) and in the embryo (embryo transmission). By observing this coloration, it is possible to distinguish haploids (color in the cap of the aleurone, but none in the embryo) reliably from hybrids (color in the cap of the aleurone and in the embryo). Although several other dominant inherited phenotypic markers exist in maize, *RI-nj* has so far been superior due to xenia expression and ability to select at the seed level before planting. Thus, only haploid kernels are colchicine treated and planted, reducing costs and effort compared to marker systems expressed at seedling or a later stage.

Despite of the various advantages of *RI-nj*, selection of haploid kernels is labor intensive and does not work equally well for all donors. The shape of the kernel (flat vs. round) affects the ability to see embryo coloration, as does the level of transparency of the seed coat, which is overlaying the embryo. Currently, sorting of haploids is exclusively executed by skilled labor. The challenge for commercial breeders is to sort through large numbers of kernels within a short harvest and planting window between seasons, which may lead to suboptimal outcomes in the sorting process: 1) this task is extremely repetitive which leads to fatigue and mistakes, 2) a large workforce is required during a brief period, and 3) variation in kernel shape and expression of *RI-nj* between donor populations may lead to varying false positive and false negative rates in haploid kernel detection.

The human eye is only able to detect wavelengths of light between 380 – 780 nm which limits the ability to detect subtle coloration differences. The speed at which a person can sort massive amounts of kernels is limited, prompting desire for automation of the haploid selection process. Though no fully implemented system is being commercially used, several other pilot studies have been published using other markers to discriminate between the haploid and diploid fractions using instrumentation. Traits such as the difference between the embryo weights of the haploid and diploid seed (Smelser et al., 2015), spectral differences using NIRS and SIMCA (Jones et al., 2012), fluorescence imaging (Boote et al, 2015), and oil content (Melchinger et al., 2015) are all being tested for their utility as automated selection criteria. Each method has its strengths and drawbacks. The weight, NIRS, and fluorescence methods all utilize existing markers,

while the oil content method requires the development of new high oil haploid inducers. The development of a high oil inducer is not a trivial matter as oil content is a quantitative inherited trait and can be affected by environmental conditions as well as context dependency of the germplasm. Herein, we describe an approach based on the VideometerLab 3 spectral imaging system, which has shown great success in other seed based assays. The ability to automate haploid – diploid kernel discrimination would allow for a substantial decrease in costs and increase in efficiency of the maize DH system and any other DH system in which a seed color based selectable marker is used. It is important to note that this process does not necessarily need to be 100% accurate. The ability to enrich haploid kernels to >80% would still be a desirable outcome saving both money and time.

The objectives of this study were, to (i) evaluate the utility of the VideometerLab 3 system to discriminate between haploid and diploid seed, (ii) test the system on several genotypes that display varying difficulty of manual sorting, and (iii) employ DNA marker assays for confirmation of haploid-diploid discrimination.

Methods

Germplasm

For this experiment, induced kernels were produced in the summer of 2012 at the Iowa State University Agronomy and Agricultural Engineering Research Farm in Boone, IA. As part of a larger experiment 120 inbred lines were planted in single row, 5.48 meter plots on 45.72 cm row spacing at a density of 60,000 plants per hectare. All plants had

immature ears covered before silk emergence and were detasseled to reduce foreign pollen contamination. When all plants in the plot reached approximately 50% silking, bulk pollen from the maternal haploid inducer F₁ 'RWS/RWK-76' (Rober et al., 2005) was used to pollinate all plants in the row. At maturity, all ears in the row were bulk harvested and shelled. As part of a separate experiment, each plot was visually sorted for haploid and hybrid kernels to determine induction rate. Six inbred lines were selected to be used in this experiment to test the ability of the Videometer system to select haploid kernels using the *R1-nj* marker in the kernel in a variety of genetic backgrounds.

The six inbred lines selected are as follows: 'PHR36', a DuPont Pioneer expired PVP white semi dent inbred from the non-stiff stalk heterotic group; 'PHT77', a DuPont Pioneer expired PVP yellow dent inbred from the non-stiff stalk heterotic group; 'PHK35', a DuPont Pioneer expired PVP yellow dent inbred from the stiff stalk heterotic group; 'B47', a DuPont Pioneer expired PVP yellow dent inbred from the stiff stalk heterotic group, developed from the Iowa State public inbred line 'B37'. 'NK792' is an inbred line developed by Northrop, King and Company. A PVP certificate was applied for but withdrawn for 'NK792'. 'MS198' is a yellow semi dent public inbred developed by Michigan State University. Detailed information for each of these lines can be found in the PVP certificates which can be accessed through the USDA's Germplasm Research Information System (GRIN) (<http://www.ars-grin.gov/npgs/>). From each of the six inbred lines, 100 random haploid, and 100 random diploid kernels were selected manually and used for analysis.

Sample preparation and image acquisition

From each inbred, 50 kernels were randomly selected from the hand sorted seed and placed in a 9 cm diameter petri dish with the embryo facing up. In some cases, it was difficult to orient the kernel with the embryo up due to the shape of the kernel. However, for most of the kernels it was possible to place them correctly with the embryo facing up.

For image acquisition, the VideometerLab 3 system was utilized. This instrument acquires multispectral images of the reflectance from the surface of maize kernels. Using strobed LED technology the VideometerLab 3 combines measurements at 19 different wavelengths into a single high resolution spectral image. Each pixel in the image is a reflectance spectrum which includes wavelengths ranging from ultraviolet, to visual, into the near infrared spectrum (thus outside the range visible to the human eye). In addition to the illuminated wavelength, 4 filters were included in the analysis for measuring fluorescence from the kernel surface (cutoff at 400, 500, 600, and 700 nm).

In the first step, the instrument was calibrated, and the light setup prepared to match the samples in such a way that the captured images contain an as wide a dynamic range as possible with a minimum of saturated pixels. Next, petri-dishes with kernels were placed in the VideometerLab 3 instrument and images acquired for each petri-dish. Images were saved for further processing. Images were taken with 2056x2056 pixels and 79 bands (regular reflectance, plus fluorescence measured with the above mentioned four cutoff wavelengths). Each image was 1.24 Gb in size.

Detection of maize kernels

In order to properly identify separate kernels, each must be segmented in the image into so-called BLOBs (Binary Labelled Objects). This automated process is diagrammed in Figure 1 where the image (1.1) is divided into either foreground or background (1.2). Next, the multispectral information available in each pixel in the labelled areas is used to create a linear model (CDA – canonical discriminant analysis) (Olesen et al., 2011), that ensures the canonical discrimination function will “score” high when pixels look the same as a kernel and “score” low when pixels look like the background (1.3). Finally, this score is used to do the final segmentation of the kernels (1.4) into segregated objects (1.5).

Haploid vs diploid score

Once a list of BLOBs is created, models are constructed of the diploid embryo. Due to aforementioned differences in expression of the *RI-nj* marker, colored embryos differ between inbreds. Thus models need to be generated for each of the inbreds. This is done in a similar way to segmentation of kernels described in the previous section (Figure 2). First, regions are labelled inside the kernels (2.1 and 2.2), such that the diploid embryos have one label, and the endosperm and haploid embryos have separate labels. As before, a model is constructed (2.3) that yields a high discriminant score when the pixels “look” like a diploid embryo and low when not (2.4 and 2.5).

Extracting the haploid vs diploid score and haploid identification

Since the discriminating feature between haploid and diploid kernels is solely contained within the embryo of the kernel, the extracted measure was restricted to the embryo of the kernel. In order to do this, the “surface” (or region of interest) of the

embryo is detected and isolated. Once this region is isolated, the diploid score is extracted from only this region for each kernel (Figure 3). For each set of 100 induced kernels from each of the six inbreds, the above procedures were followed and for each of the genotypes the identification rate was estimated for three different approaches: 1) for each kernel the diploid score was used to evaluate, whether it is more similar to other diploid or haploid kernels. 2) For each kernel a score is created by combining all the diploid scores (for all genotypes) using CDA. 3) In addition to the aforementioned criteria, addition of more features related to shape and texture of the kernels was used in the model.

Testing of haploid vs diploid scores on seed genotypes

To test the system, a random sample of 20 kernels from the haploid and diploid fraction was placed on two separate petri dishes (See Figure 4). Ten of the kernels in each of the petri dishes were used to train the model for the specific genotype both for fluorescence and for visual light. Once the model was trained by selection of the optimally discriminant wavelength for that genotype, it was used to generate a haploid vs diploid score for the remaining ten kernels on the plate. Individual kernels received unique scores. These individual kernels were then subjected to marker analysis for a validation of the haploid vs diploid state.

Marker analysis for confirmation of haploid vs diploid identification

As previously mentioned, the *R1-nj* color marker is not perfect. Expression of this marker is variable in both embryo and aleurone. Other issues such as kernel size and shape, time between pollination and harvest and disease pressure create more variability in the visibility of the coloration to the human eye. Thus, to definitively confirm the

haploid vs diploid identity of the kernels, all kernels were planted in a greenhouse tray for DNA extraction from leaf tissue. It should be noted that a ‘seed chipping’ approach is not possible since the endosperm is the product of a successful fertilization between the donor plant and the inducer in all haploid kernels. The triploid endosperm, would therefore, always be a hybrid. At the 2-3 leaf stage, tissue samples were collected for DNA extraction. Prior to this analysis, the six lines and the inducer used for pollination were used to identify polymorphic markers that could be used to positively identify hybrids between the inducer and the line and also identify the line itself (Supplemental Table 1). DNA was extracted using a CTAB protocol (Stewart et al., 1993) with plant tissue which was flash frozen with liquid nitrogen. DNA was then separated using isomyl-alcohol:Chloroform solution and dissolved into ethanol. DNA was diluted to 20ng/ μ L for PCR.

Statistical Analysis

All tests were conducted using a Welch two-sample T-test (Welch, 1938) as implemented in the R function ‘t.test’. The Welch two-sample T-test is appropriate as the sample sizes are small and the variances of the two fractions are not equal. Using Satterthwaite’s approximation, percentage points of the t distribution were modified using an estimation of degrees of freedom based on the separate individual sample variances instead of a pooled variance estimates.

Results

Identification of haploids vs hybrids using fluorescence

For each individual genotype, a specific wavelength produced an optimally discriminating value for that specific genotype. In Figure 7, boxplots for each of the six genotypes were generated to show the distribution of kernel scores for the haploid and diploid fractions of each genotype. In this figure, the optimal score was used based on the model generated for each genotype, thus producing the most discriminating values possible. As seen in both the boxplots of Figure 5, and Table 1, significant differences were obtained in all genotypes except for NK792 at the 0.05 significance level. More variation within genotypes was found for genotypes PHT77 and NK792. These genotypes also have the least significant differences between the haploid and diploid fractions. This is not surprising, as these genotypes were the most difficult to sort by hand.

The effect of using the correct model for a specific genotype can be seen in Figure 6, and Table 2. For this set of boxplots, PHT77 was sorted using wavelengths optimized for all six genotypes. For example, PHT77.1 uses the wavelength that is optimized for PHR35. In the boxplots, it can be seen that the variance within genotypes increases when the non-optimal wavelength is used. In this case, the wavelength that is optimized for PHT77 (PHT77.2) and PHT77.3 and PHT77.4 produce significant differences. A similar effect can be seen for all genotypes when the non-optimal wavelength is used (data not shown).

Identification of haploids vs hybrids using visual light

In addition to using non-visible light wavelengths, the kernels were analyzed using wavelengths in the visible light spectrum. Again, each genotype had an optimal wavelength which produced the most discriminating score for differentiation of haploids

vs hybrids. In Figure 7 and Table 3, the results of this analysis are shown in boxplots and significance values for t tests of each genotype. Differences between haploids and hybrids were significant for all genotypes except for NK792. With visual light, similar results were obtained, when non-optimal wavelengths were used (Figure 8, Table 4) with significant differences for PHT77 only when using this genotype's specific wavelength (PHT77.2) or the wavelength for PHR35 (PHT77.1). Similar results were obtained with the other genotypes (data not shown).

Breeders evaluated many haploid seed from many diverse genotypes. There would be great utility in having a 'global wavelength' which can be used across all germplasm in the breeding pool. As mentioned, this study produced a unique optimized wavelength for each genotype. To evaluate the possibility of a 'global wavelength', all comparisons were made between haploids and diploids within each genotype for each fluorescence wavelength (Table 5). In PHR35, PHK35, PHB47, and MS198 there are significant differences between the haploid and diploid fractions for every wavelength. However, PHT77, PHK35 and NK792 do not show significant differences for all wavelengths. Based on this information, there would be the possibility to use a global wavelength, however, each genotype would need to be visually checked to ensure that it is being properly sorted.

Marker analysis verification of haploid vs diploid scores

For this analysis, it was assumed that all non-germinating seed were of the haploid fraction. Haploid seed produce lower germination rates on average, as these embryos are weaker since they lack half of their genetic information (Prigge et al., 2011).

Accuracies between the VideometerLab 3 score and molecular marker identification for the six genotypes ranged from 40% to 100% (Table 6), with the lowest accuracies for those genotypes which were most difficult to sort by hand (see Supplemental Table 1 for a full list of scores). Manual sorting for genotypes PHR35, PHB47, and MS198 was 100% accurate based on the haploid vs hybrid score compared to marker analysis. Haploids were detected for PHT77 with 40%, for PHK35 with 70%, and for NK792 with 50% accuracy.

Discussion

Success of identification

The overall objective of this study was, to determine whether the Videometer system can be used to accurately sort haploid and hybrid seed in maize. Sorting of haploids from a maize DH program is a very time consuming and tedious process that can take thousands of hours of labor to complete for a single average sized breeding program. Many of the sorting efforts are out sourced to countries where winter nurseries are located as labor is usually cheaper and the seed can then be readily available for planting. This study shows that there is a significant potential to utilize the Videometer system for the automation of sorting of haploid seed for maize DH programs utilizing the *RI-nj* marker system in their *in vivo* induction program. As is seen in the results, some genotypes proved difficult to sort. The six genotypes were classified into two different groups 1) easy to visually sort, and 2) difficult to visually sort. The genotypes that were easy to visually sort are PHR35, PHK35, PHB47, and MS198. The genotypes difficult to sort were PHT77, and NK792. This is a common problem seen in

all maize DH programs as the expression of *R1-nj* is dependent on the genetic background of the donor population as well as environmental factors. In the case of this study the background within genotypes is uniform as all genotypes were produced from inbred lines. A suitable next step will be to consider segregating donor populations to see what effect this has on the sort. Based on the results, it is clear that the use of the optimal wavelength for a specific genotype is important. When considering a typical maize breeding program, it is likely that the majority (at least in the U.S.) of the germplasm will be yellow dent corn. Most yellow dent corn which is properly pollinated will provide good expression of *R1-nj*. However, it remains possible that certain combinations could provide modification of the expression of *R1-nj* since this is not a trait which is selected for so any modifier loci should segregate in the germplasm. Modifications to the expression could make visual and automated sorting more difficult. It is therefore most likely that breeding programs would need to classify their induced seed into two groups as was done in this paper: those which are easy to visually sort, and those which are difficult to visually sort. As shown in this study, those which are easy to visually sort can be accurately sorted using the VideometerLab 3 system. Those which are difficult to sort could be run through the system multiple times to, at least, enrich the fraction of haploid seed in the mix which would still provide an advantage. It is also crucial to train the model with accurate visual sorts. If the visual sorts used to train the model are poor, then the model will poorly discriminate the fractions for that genotype as was the case for NK792. Visual scores (Supplemental Table 1) for NK 792 identified the haploid and hybrid fractions as they were sorted. However, marker analysis showed that only four of the haploids were correctly identified. In previous sections, it was noted that NK792

produced the least significant values between the two fractions. The poor sort for NK792 could have come from the poor expression of *RI-nj* on the kernels making visual sorting very difficult. Based on visual and marker scores, the haploid fraction contained five hybrids. It is thus likely, that the training kernels also contained a similar fraction of hybrids. Thus when the model was trained, half of the kernels were incorrectly classified producing a poor model to sort from. This explains the poor discrimination found for NK792. This will continue to pose a challenge, but it was promising to see that the automated scores corrected for a few visual sorting mistakes. The fitting of a global model would eliminate the need for manual sorting for each different genotype prior to sorting. However, it is shown in Table 5 that only the genotypes in the easy to sort group (PHR35, PHK35, PHB47, and MS198) are sortable using the optimized wavelengths from other genotypes. One aspect to consider, however, is that global models may work better within the germplasm used in specific breeding programs. For example, a breeder may be able to generate a global model for their stiff stalk dent germplasm, and a separate model for their non-stiff stalk germplasm.

Challenges and automation

A major challenge of working with maize kernels for imaging is that the kernels are of an irregular shape and do not always lay flat depending of many factors including the number of kernels on the ear, and the location of the specific kernel on the ear with flatter kernels occurring with complete pollinations in the middle of the ear and rounded kernels occurring in incomplete pollinations and on the ear base and tip. In order to use the system as it currently works, the kernels must be positioned with the embryo facing upward which proved challenging. A system will have to be devised where the kernel can

be displayed to the optics in a consistent and accurate position for imaging.

Discrimination of the embryo from the rest of the kernel is a useful aid to discrimination of the two fractions as this is the only part of the kernel which has a visible difference between the haploid/diploid fractions.

Currently, the positioning of the kernels on a petri dish is not a high-throughput method which would allow for the sorting of the millions of kernels which would be needed. Now that the system has been shown to be discriminatory between hybrids and haploids, it will be important to design an automated system which can feed a large quantity of seed to the imaging system (e.g., through a conveyor or channel system) which would then subsequently sort the seeds into two fractions using robotic picking or some kind of pneumatic or mechanical gate system. It is important to note, that a 100% accuracy rate is not necessary. A rate of 10% hybrids in the haploid fraction would be acceptable, as these can be cut out of the field easily. It is most important to identify all haploids in the seed genotype, even if this means that some hybrids are misclassified due to the setting of the sort threshold. Losing haploids into the hybrid fraction would be undesirable as haploids only occur at, on average, a 10% rate in an induction cross and each seed can be critical to the production of enough DH lines for breeders to evaluate per cross.

Comparison to other methods

As mentioned in the introduction, other pilot studies have been conducted which evaluate the ability to use an automated system to discriminate between haploid and hybrid seed in an induction cross. The method proposed herein, as mentioned, uses the

Vidometer Lab 3 system which has been documented as a useful tool for the nondestructive and automated analysis of seed phenotypes (Liu et al., 2014; Shetty et al., 2012; Olesen et al., 2011). This method is able to capitalize on the already existing marker system (*RI-nj*) and would also not require the development of a new set of optical sensors/software to handle the data. This in comparison to, for example, the previously mentioned Jones et al. (2012) which required the development of a new system for the analysis of the seeds as well as software to be adapted for its use. Though the method in Jones et al. (2012) provided a discriminative sort, there was no validation of the true identity of the haploid/hybrid kernels. The method described by Boote et al. (In press), provides the most similar method, using NIRS and fluorescence imaging. However, this method only considers one kernel at a time and while discussion of automation was provided, no current method exists for that system, while for the VideometerLab 3 system a method of automated seed feeding exists (shown here:

<ftp://videometerlab:multispec@www.videometer.com>

[/Videos/2014_July_VideometerLab_AutoFeeder.AVI](#)). The method described in Smelser et al. (2015) uses the weight of the haploid and hybrid embryos as a marker to discriminate between the two seed fractions. In this study visually selected kernels were weighed and total kernel weights were compared as a discriminative marker. Only two of the six genotypes produced significant differences. Automation of such a system could prove difficult since it would rely on single kernel weights. Also, it is unclear what the effect of kernels of different size would have. Kernels on the tip and base of the ear are sized differently than those in the center which will affect the overall weight of the kernel. There may also be loci which control the size of the embryo which segregate in

the germplasm affecting the accuracy of the sort. Again, no validation of the identity of the visual sort was provided. Finally, the method proposed by Melchinger et al. (2015; 2014; 2013) uses oil content as a marker for discrimination between haploid and hybrid seed. While detection of oil content has the potential to be automated on a single kernel basis, and the ability to discriminate the two fractions was shown, there is a dependence on the genetic background of the material as is the case in all studies discussed. Modifier loci for oil content may segregate within the germplasm causing confounding effects of the oil content expression. The most difficult aspect of the high oil marker, is the development of new high oil inducer lines. While the VideometerLab 3 system relies on existing marker technology, using oil content would require the development of new inducers with both high oil content and high induction rate. Both of these traits are quantitative and it would not be a trivial task to increase the mean induction rate and the mean oil content simultaneously, though it can be done as they describe. This would however, make the system more expensive as it would not only require the instrumentation, but also either the development or purchase of a new high oil inducer.

Considering these methods, it seems that the VideometerLab 3 system would be the easiest to implement due to the fact that 1) the instrumentation is already designed and well tested showing success in other seed phenotypes, 2) software is already developed for this system, 3) this study shows that discrimination of haploids and hybrids is possible, and 4) it utilized the existing inducers and marker system.

Acknowledgement

This project is supported by the SEVENTH FRAMEWORK PROGRAMME of EU, Industry – Academia Partnerships and Pathways (IAPP) - Marie Curie Actions (grant no.: 324433). This project was also supported by the Iowa State University Raymond F. Baker Center for Plant Breeding.

References

- Boote, B., D. Freppon, G. De La Fuente, T. Lubberstedt, B. Nikolau, and E. Smith., In
Press: Haploid Differentiation in Maize Kernels Based on Fluorescence Imaging.
Plant Breeding
- Chase, S.S, 1949: Monoploid frequencies in a commercial double cross hybrid maize,
and in its component single cross hybrids and inbred lines *Genetics* **34**, 328
- Coe, E.H., 1959: A line of maize with high haploid frequency. *American Naturalist* **93**,
381-382
- De La Fuente, G.N., U.K. Frei, and T. Lubberstedt., 2013: Accelerating Plant Breeding.
Trends in Plant Science **18**:667-672
- Geiger, H.H., 2009: Doubled Haploids. In: *Handbook of Maize: Genetics and Genomics*,
ed. Bennetzen, J.L., and S. Hake, Springer, New York, p. 641-657
- Jones, R.W., T. Reinot, U.K. Frei, Y. Tseng, T. Lubberstedt, and J.F. McClelland., 2012:
Selection of Haploid Maize Kernels from Hybrid Kernels for Plant Breeding
Using Near-Infrared Spectroscopy and SIMCA Analysis. *Applied Spectroscopy*
66,1-4
- Liu, C., W. Liu, X. Lu, W. Chen, F. Chen, J. Yang, and L. Zheng., 2014: Non-destructive
discrimination of conventional and glyphosate-resistant soybean seeds and their

hybrid descendants using multispectral imaging and chemometric methods.

Journal of Agricultural Science, doi:10.1017/S0021859614001142

Melchinger, A.E., W. Schipprack, T. Wurschum, S. Chen, and F. Technow., 2013: Rapid and accurate identification of *in vivo*-induced haploid seeds based on oil content in maize. Scientific Reports **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 Science **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 Science **55**, 188-195

Olesen, M.H., J.M. Carstensen, and B. Boelt., 2011: Multispectral imaging as a potential tool for seed health testing of spinach (*Spinacia oleracea* L.). Seed Science and Technology **39**, 140-150.

Prigge, V., C. Sanchez, B.S. Dhillon, W. Schipprack, J.L. Araus, M. Banziger, and A.E. Melchinger., 2011: Doubled Haploids in Tropical Maize: I. Effects of Inducers and Source Germplasm on *in vivo* Haploid Induction Rates. Crop Science **51**, 1498-1506

Rober, 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

Shetty, N., M.H. Olesen, R. Gislum, L.C. Deleuran, and B. Boelt., 2012: Use of partial least squares discriminant analysis on visible-near infrared multispectral image

data to examine germination ability and germ length in spinach seeds. *Journal of Chemometrics* **26**, 426-466

Smelser, A., M. Blanco, T. Lubberstedt, A. Schechert, A. Vanous, and C. Gardner., 2015:
Weighing in on a method to discriminate maize haploid from hybrid seed. *Plant Breeding* **134**, 283-285

Table 1. Comparison between haploid and diploid fractions using fluorescence for each of the six genotypes. Comparisons were made using the Welch two sample T-test. The optimal wavelength for each genotype was used in this analysis.

Genotype	Fraction	Fluorescence Mean	Std. Dev.	P-Value
PHR35	Hybrid	-0.19	0.36	< 0.001*
	Haploid	-1.06	0.1	
PHT77	Hybrid	-0.48	0.35	0.0012*
	Haploid	-1.15	0.41	
PHK35	Hybrid	-0.33	0.18	< 0.001*
	Haploid	-1.15	0.17	
PHB47	Hybrid	0.03	0.29	< 0.001*
	Haploid	-0.91	0.12	
NK792	Hybrid	-0.65	0.38	0.051 ^{NS}
	Haploid	-0.94	0.19	
MS198	Hybrid	-0.26	0.18	< 0.001*
	Haploid	-0.92	0.26	

Table 2. Comparison between haploid and diploid fractions using fluorescence for PHT77. Comparisons were made using the Welch two sample T-test. In this table, optimal wavelengths for each of the six genotypes were used on PHT77.

Genotype	Fraction	Fluorescence Mean	Std. Dev.	P-Value
PHT77.1	Hybrid	-0.73	0.18	0.09 ^{NS}
	Haploid	-0.86	0.16	
PHT77.2	Hybrid	-0.48	0.35	0.0012*
	Haploid	-1.15	0.41	
PHT77.3	Hybrid	-0.19	0.28	0.03*
	Haploid	-0.49	0.3	
PHT77.4	Hybrid	-0.31	0.33	0.03*
	Haploid	-0.61	0.22	
PHT77.5	Hybrid	-1.69	0.2	0.33 ^{NS}
	Haploid	-1.77	0.19	
PHT77.6	Hybrid	-0.73	0.18	0.17 ^{NS}
	Haploid	-0.84	0.17	

Table 3. Comparison between haploid and diploid fractions using visible light for each of the six genotypes. Comparisons were made using the Welch two sample T-test. The optimal wavelength for each genotype was used in this analysis.

Genotype	Fraction	Visible Light Mean	Std. Dev.	P-Value
PHR35	Hybrid	-0.1	0.93	<0.001*
	Haploid	-1.09	0.13	
PHT77	Hybrid	-0.72	0.24	0.002*
	Haploid	-1.09	0.2	
PHK35	Hybrid	-0.51	0.19	<0.001*
	Haploid	-1.05	0.1	
PHB47	Hybrid	-0.2	0.26	<0.001*
	Haploid	-0.92	0.13	
NK792	Hybrid	-0.81	0.24	0.051 ^{NS}
	Haploid	-1.01	0.14	
MS198	Hybrid	-0.45	0.18	<0.001*
	Haploid	-0.99	0.09	

Table 4. Comparison between haploid and diploid fractions using visible light for PHT77. Comparisons were made using the Welch two sample T-test. In this table, optimal wavelengths for each of the six genotypes were used on PHT77.

Genotype	Fraction	Visible Light Mean	Std. Dev.	P-Value
PHT77.1	Hybrid	0.003	0.19	0.03*
	Haploid	-0.17	0.14	
PHT77.2	Hybrid	-0.72	0.24	0.002*
	Haploid	-1.09	0.2	
PHT77.3	Hybrid	-0.53	0.17	0.08 ^{NS}
	Haploid	-0.69	0.21	
PHT77.4	Hybrid	-0.13	0.2	0.12 ^{NS}
	Haploid	-0.26	0.15	
PHT77.5	Hybrid	-1.02	0.2	0.32 ^{NS}
	Haploid	-1.11	0.17	
PHT77.6	Hybrid	-0.72	0.12	0.17 ^{NS}
	Haploid	-0.79	0.1	

Table 5. P-values for Welch two sample t-test for all possible combinations using fluorescence. Each genotype was tested at all six wavelengths to observe if a global model could be applied instead of producing an optimal wavelength for each genotype.

Genotype	Wavelength					
	1	2	3	4	5	6
PHR35	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
PHT77	0.09	0.001	0.03	0.03	0.33	0.17
PHK35	<0.001	<0.001	<0.001	0.16	<0.001	<0.001
PHB47	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
NK792	0.68	0.97	0.81	0.51	0.051	0.28
MS198	<0.001	0.001	<0.001	<0.001	0.002	<0.001

Table 6. Using molecular marker information that allows for the classification of a hybrid between the inbred line and the inducer and a haploid progeny of the inbred line, the accuracy of the VideometerLab 3 sort was checked.

Genotype	Fraction	# correct	# incorrect	Accuracy
PHR35	Hybrid	10	0	100%
	Haploid	10	0	100%
PHT77	Hybrid	8	2	80%
	Haploid	4	6	40%
PHK35	Hybrid	9	1	90%
	Haploid	7	3	70%
PHB47	Hybrid	10	0	100%
	Haploid	10	0	100%
NK792	Hybrid	10	0	100%
	Haploid	5	5	50%
MS198	Hybrid	10	0	100%
	Haploid	10	0	100%

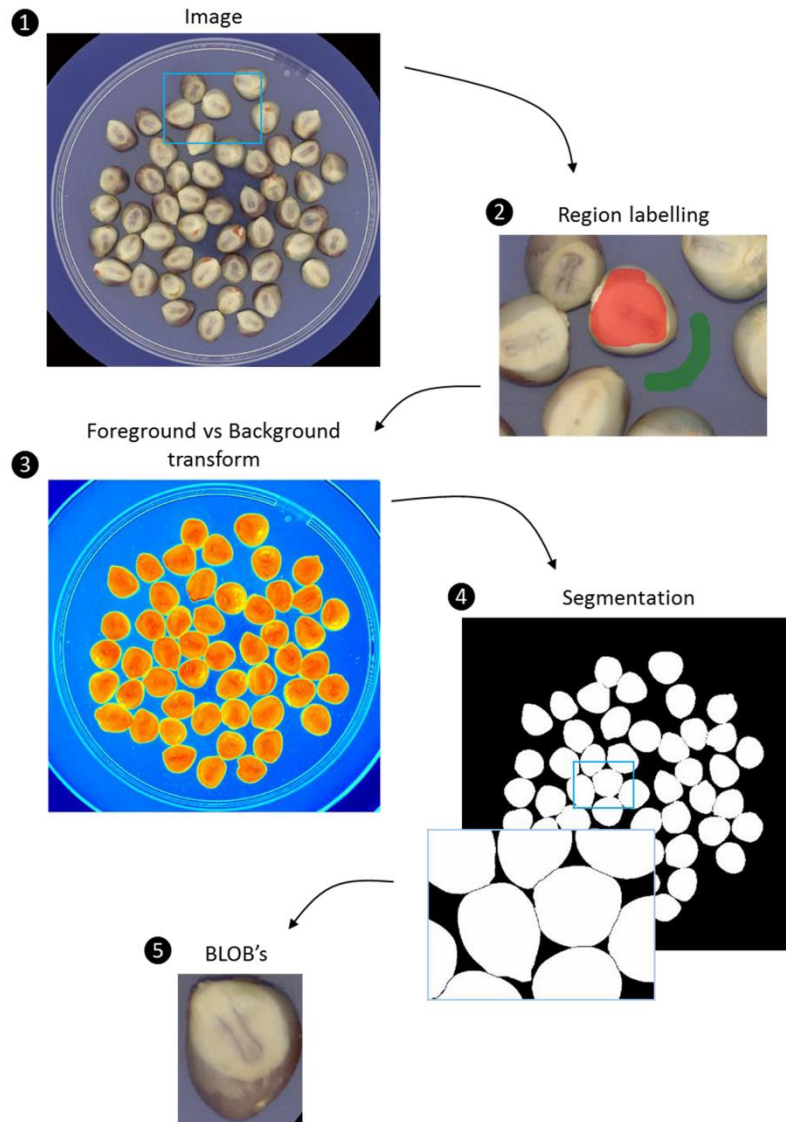


Figure 1. The segmentation process. Regions/pixels in the image (1) are labelled in “foreground” and “background” categories (2). The labelling is used to create a score (3) that can be used to segment the image into individual regions, each containing one kernel (4). Each region is extracted from the image (5) into so-called BLOBS (Binary Labelled Objects).

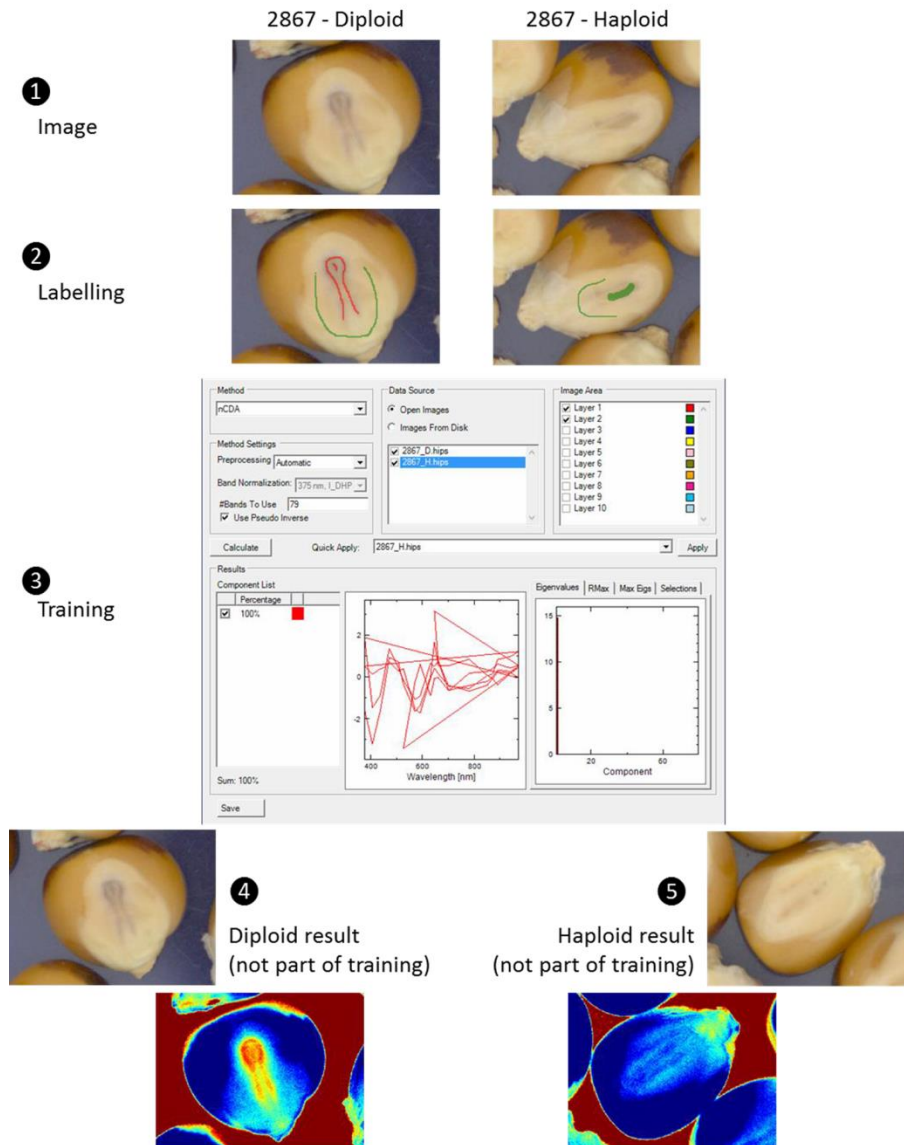


Figure 2. Identification of the diploid embryo. The diploid and haploid embryo in each kernel (1) is labelled with two different labels (2). In addition to the haploid embryo region, the surroundings, different from the diploid embryo, is included in order to create a model (3) that will score high on only diploid embryos. (4) show the resulting “diploid

embryo-score” generated from the model. (5 and left side) illustrate the resulting score from a haploid embryo.

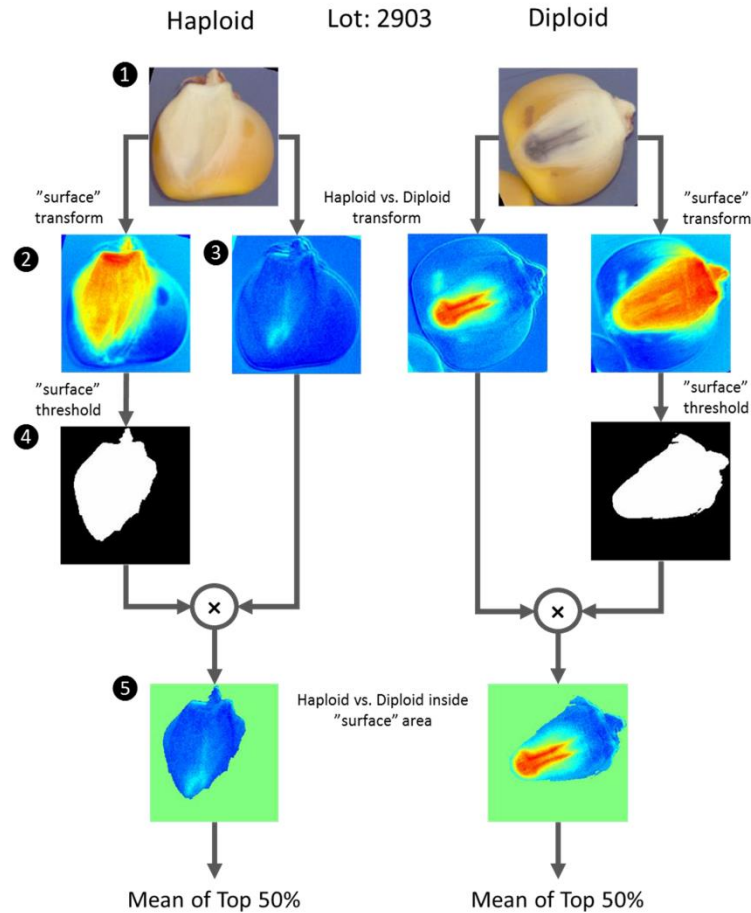


Figure 3. Before the diploid-embryo score is extracted from each kernel, we ensure that the score is extracted from the endosperm region. Hence, as for the foreground/background segmentation in Figure 6, we label the endosperm region and the rest in two labels, and create a model that we use for doing the segmentation of the embryo region.

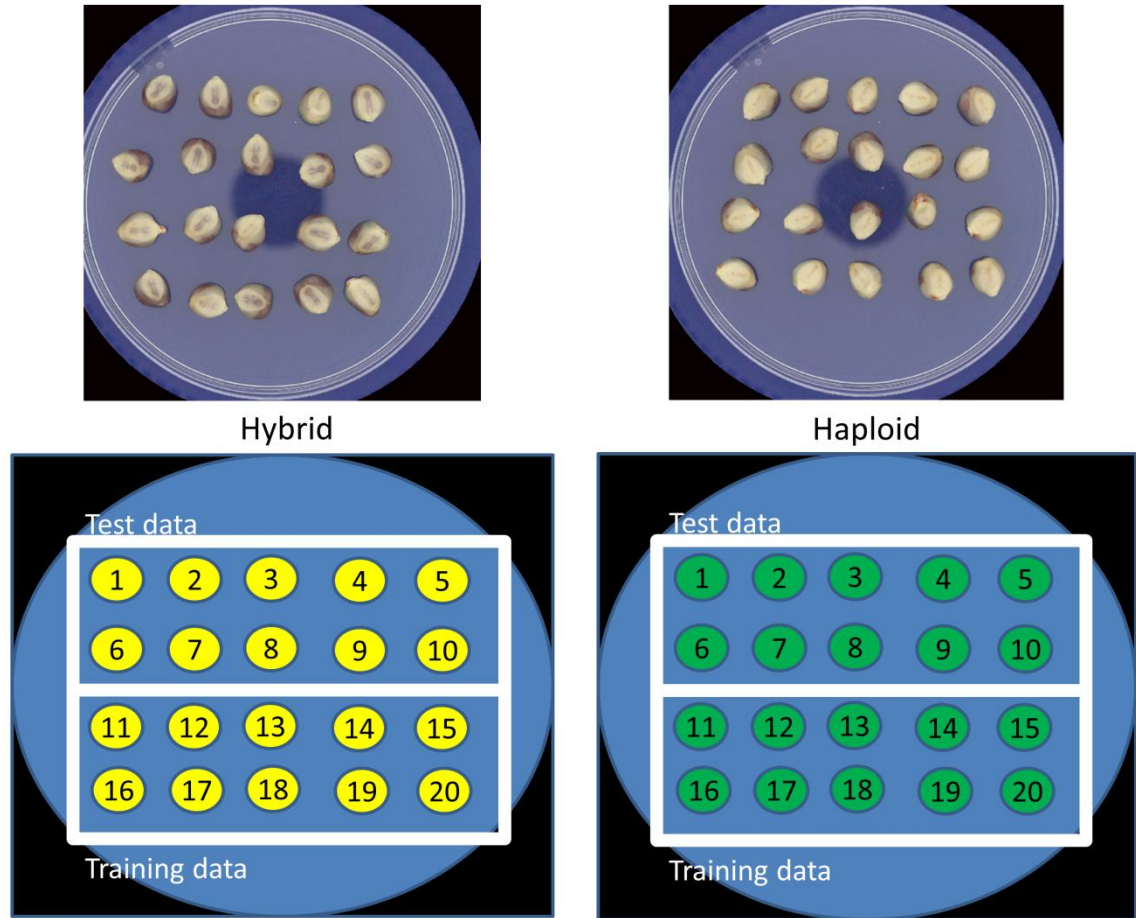


Figure 4. Experimental layout of kernels. 20 individual randomly selected kernels from the haploid and diploid fractions of each genotype were placed on a petri dish, embryo side up, for imaging. The bottom ten kernels were used to train the model, and the model was then tested on the remaining 10 kernels for each fraction.

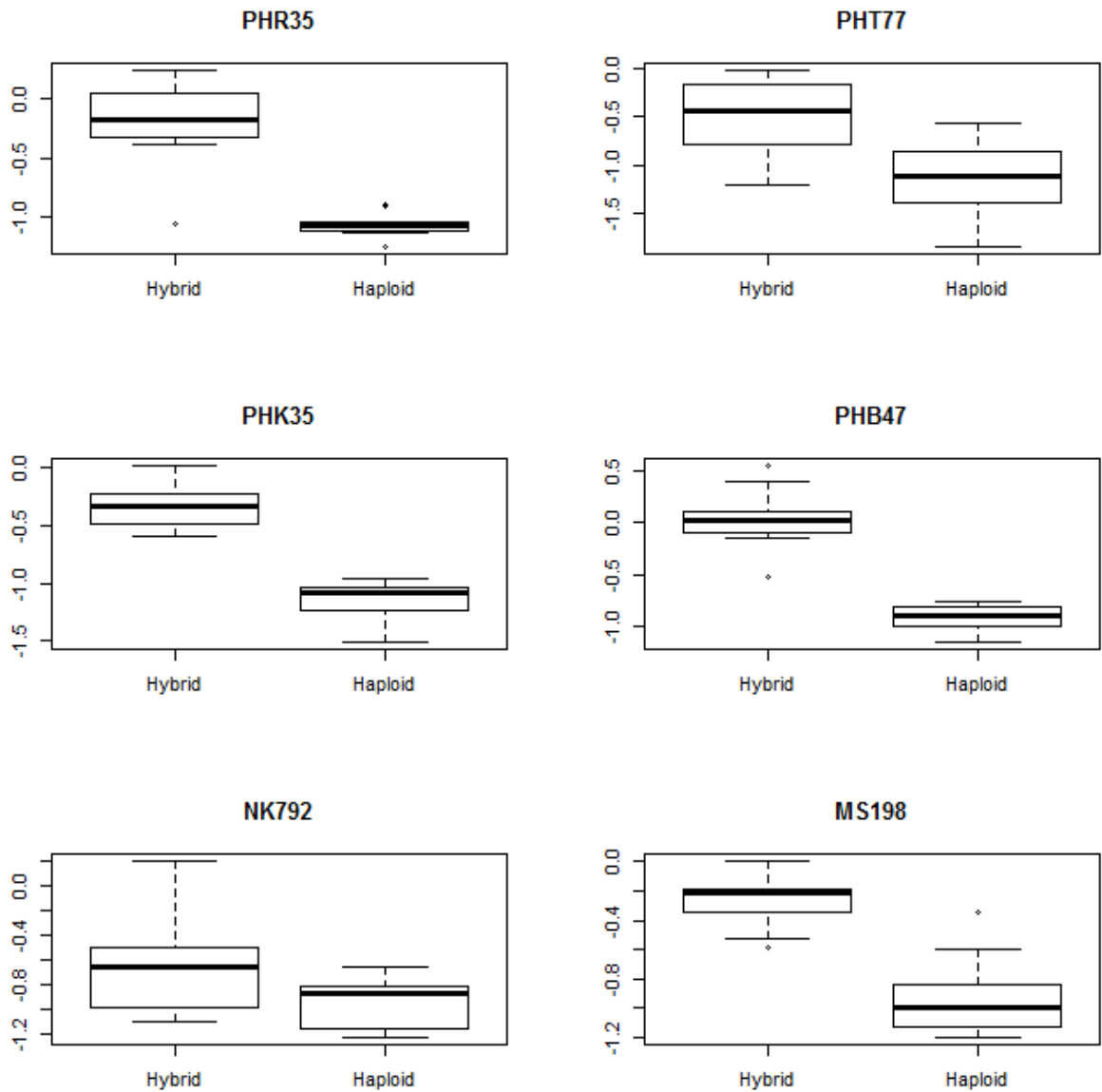


Figure 5. Boxplots showing the distribution of values for each of the six genotypes using fluorescence. As seen, some of the genotypes have small variance within each group and good separation between fractions while some have higher variance and less separation.

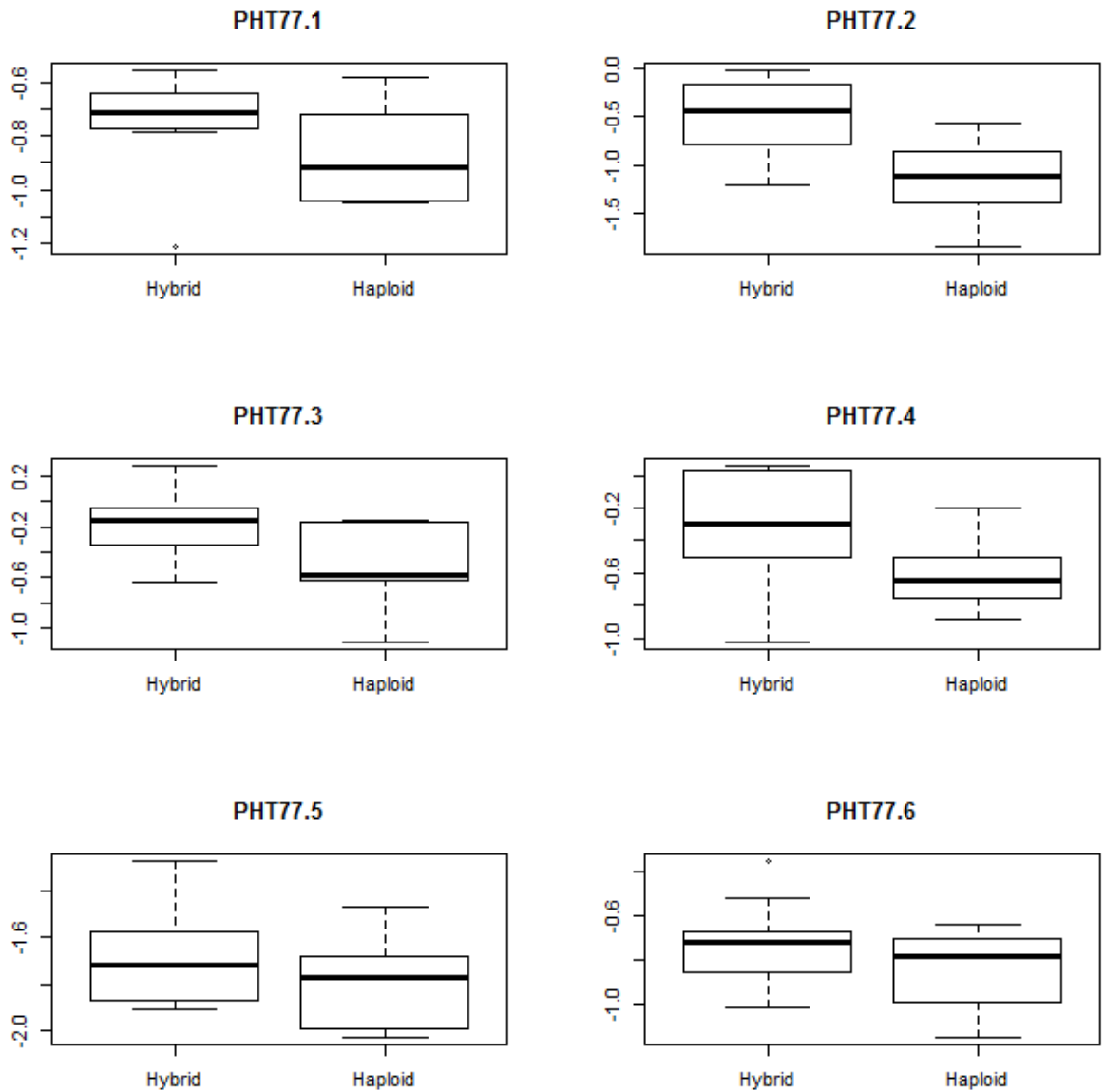


Figure 6. Boxplots showing the distribution of values for PHT77 using optimal fluorescence wavelengths for all six genotypes. As seen, the differentiation between haploid and diploid fractions is not as pronounced as seen in Figure 7.

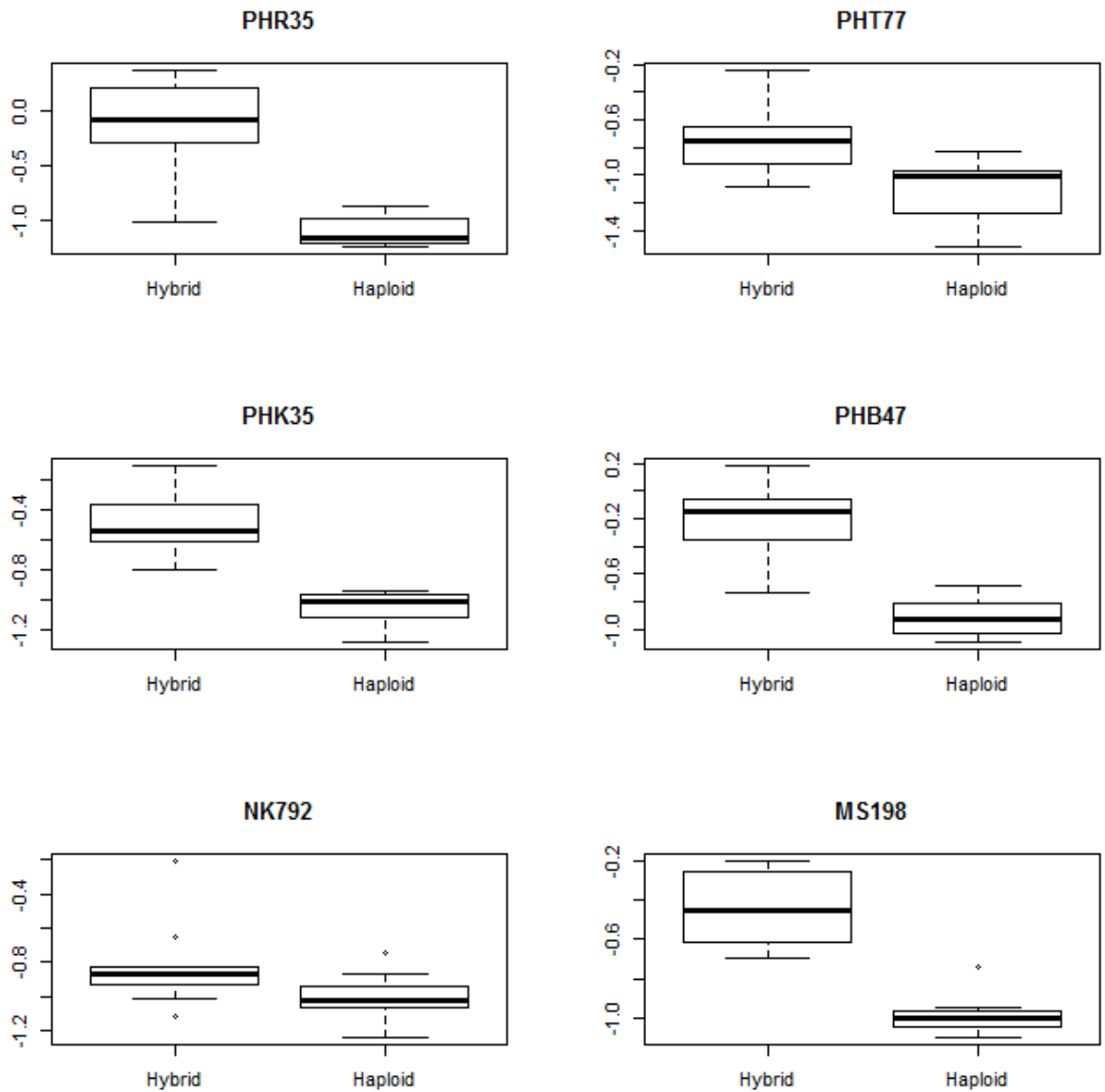


Figure 7. Boxplots showing the distribution of values for each of the six genotypes using visible light. As seen, some of the genotypes have small variance within each group and good separation between fractions while some have higher variance and less separation.

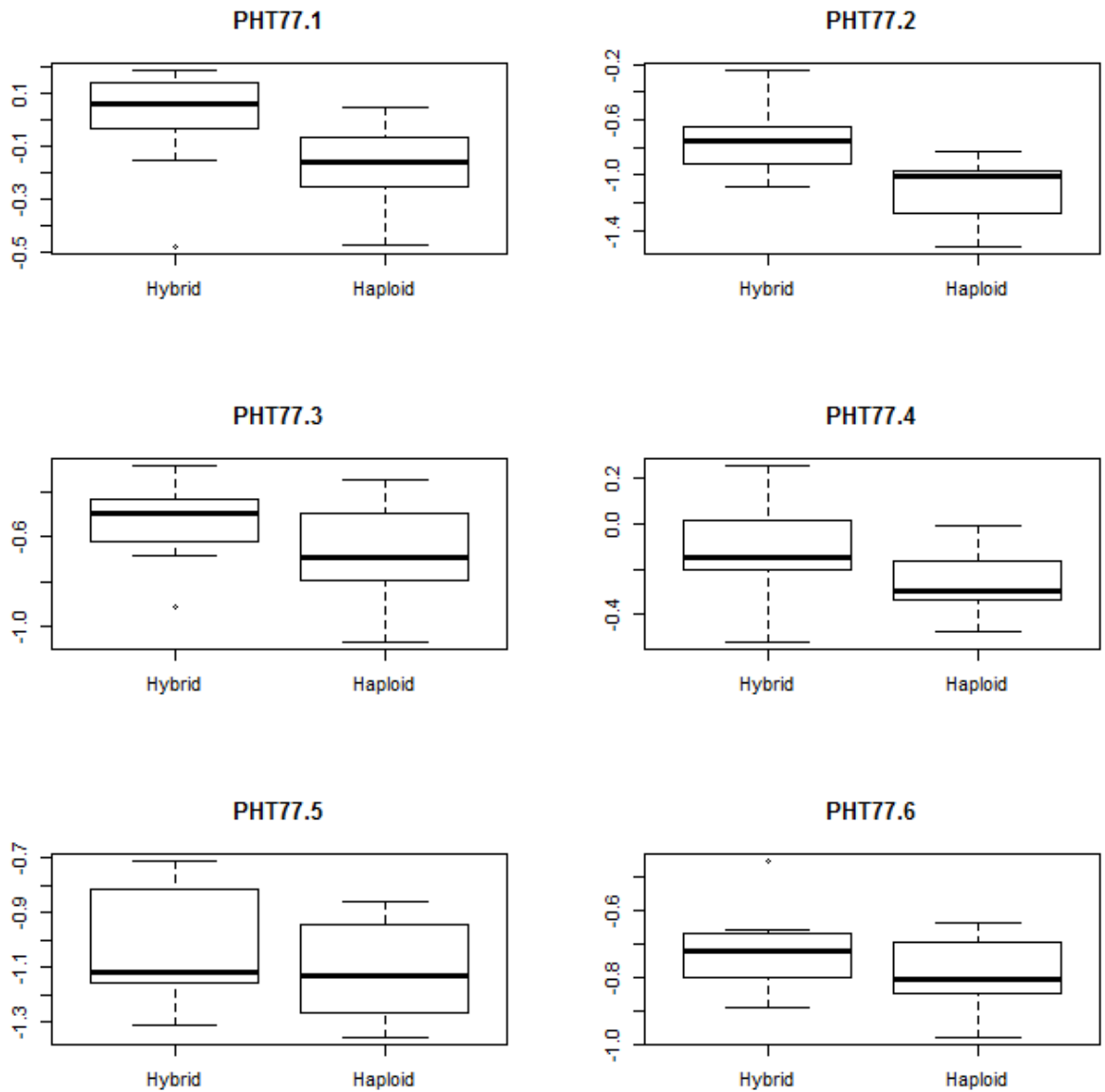


Figure 8. Boxplots showing the distribution of values for PHT77 using optimal visible light wavelengths for all six genotypes. As seen, the differentiation between haploid and diploid fractions is not as pronounced as seen in Figure 7.

CHAPTER FIVE**A DIALLEL ANALYSIS OF A MAIZE DONOR POPULATION RESPONSE TO
IN VIVO MATERNAL HAPLOID INDUCTION II:
SPONTANEOUS CHROMOSOME DOUBLING**

Gerald N. De La Fuente¹, Ursula K. Frei¹, Benjamin Trampe¹, Jiaojiao Ren¹, Martin Bohn², Nicole Yana², Anderson Verzegnazzi¹, Seth C. Murray³, Dan Nettleton⁴ and Thomas Lübberstedt^{1,*}

Manuscript in preparation for submission to Crop Science. Abstract, structure, and references are all formatted according to journal standards.

Abstract

Doubled haploid (DH) lines are used in maize breeding to accelerate the breeding cycle and create inbred lines in as little as two seasons. This allows breeders to quickly evaluate new cross combinations without wasting time inbreeding. There are two important steps in creating DH lines: 1) generation and selection of haploid progeny, and 2) doubling of chromosomes to create a fertile, diploid inbred. The second step is the focus herein. Normally colchicine is used to double chromosomes in haploid plants which is expensive and time consuming. In this study three public inbred lines GF1, GF2, and GF6 were found to have spontaneous chromosome doubling (SCD) ability as haploids. In conjunction with another study, a 6 parent full diallel between these three SCD lines and three non-SCD lines was created and male fertility in haploids was scored. Diallel analysis shows that significant GCA estimates of up to 17% exist for SCD as well as significant SCA effects of up to 25%. No reciprocal effects were found significant and broad and narrow sense heritabilities for SCD were estimated at 0.62 and 0.31, respectively. The potential to use SCD in breeding programs for the improvement of the

DH system as well as to expand the DH system into other types of maize such as tropical or popcorn exists as well as the reduction in time and monetary resources.

Introduction

Production of doubled haploid (DH) lines in maize (*Zea mays* L.) has provided breeders and geneticists a powerful tool for the rapid production of new inbred lines for testing and evaluation. The applications of DH lines and their advantages in both breeding and genetics have been discussed at length (Bernardo, 2009; Geiger, 2009; Prigge et al., 2011; Longin et al., 2011). Maize was not the first, nor is it the only crop to realize the potential applications and benefits of DHs. For many crops such as wheat (*Triticum* spp.), barley (*Hordeum vulgare* L.), rye (*Secale cereale*), rapeseed (*Brassica napus*), broccoli (*Brassica oleracea*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), sugar beet (*Beta vulgaris* L.), onion (*Allium cepa*), apple (*Malus domestica*), poplar (*Populus tremula*), oak (*Quercus* spp.), and some citrus species protocols for production of DH lines are available.

Though multiple methods exist for the production of maize DH lines, the most popular due to its ease of use, success rate, and flexibility is the *in vivo* maternal haploid induction system. With this system there are two key biological steps that must occur for the successful production of DH lines: 1) the production and identification of haploid progeny (see accompanying paper) and 2) successful genome doubling in haploid plants. In this manuscript, the focus will be on spontaneous genome doubling.

DH systems provide 100% (theoretically) inbred lines rapidly, economically, and reliably. Although widely accepted, the current maize DH system is not without shortcomings. Quantitative genetic control of the rates of haploid induction limits the number of haploid progeny which can be produced (Prigge et al., 2011, De La Fuente et al., in preparation). Without doubling treatment, the rate of haploid fertility is typically very low: 0.41% were reported by Kleiber et al. (2012). However, with doubling treatment, across a diverse set of germplasm, rates are typically near 8% (Kleiber et al., 2012). In order for the haploid plant to be fertile, genomes of cells which lead to the formation of reproductive organs must successfully be doubled. This genome doubling produces a diploid genome which is expected to be 100% homozygous and thus, results in a completely inbred line. After genome doubling, reductional division is possible and meiosis leads to formation of fertile pollen and egg cells.

The current DH system relies on use of artificial doubling treatments. The most widely used method is the use of the chemical colchicine. The first report of the use of colchicine as a doubling agent goes back to 1937 (Blakeslee and Avery, 1937). Colchicine acts to artificially double ploidy through the inhibition of the mitotic spindle apparatus development (Borisov and Taylor, 1967). This method was, at first, not very efficient, but refined practices have led to reported doubling rates of between 16-49% depending on the method of application (Eder and Chalyk, 2002). Colchicine is a toxic chemical to both user and plant. The use of colchicine requires specialized training and supervision as well as appropriate disposal protocols. Perhaps, the most problematic issue with this system is the need for germination, treatment, and subsequent transplanting,

which requires a substantial labor input compared to sowing maize seed. Dependence on colchicine limits application of DH technology in developing countries (Kleiber et al., 2012). Observations made at the Iowa State University Doubled Haploid Facility (ISU-DHF) over multiple seasons show that even with the use of colchicine treatment, skilled labor, and appropriate facilities some populations used for DH production are recalcitrant and double at low rates (<1%, data not shown). This is commonly occurring in specialty maize such as popcorn and sweet corn. However, it is also found in elite temperate adapted maize populations, leading to potential reduction in selectable variation in a breeding program when applying DH technology.

Spontaneous chromosome doubling (SCD) has been reported in maize (Wu et al., 2014; Sugihara et al., 2013; Kleiber et al., 2012) and also in other grass species and has likely been an important factor in the formation of some of our polyploid crops (Castillo et al., 2009). However, for economic production of DH lines the rate of SCD must be at least that observed when artificial chromosome doubling is used (~8-10%). SCD has been reported for tropical and elite temperate maize with European and North American origin (Kleiber et al., 2012). Reports on SCD tend to concentrate on female fertility (Chalyk, 1994; Geiger et al., 2006), though one does consider male fertility (Kleiber et al., 2012). It is important to consider male fertility, since usually female fertility is not the limiting factor (Chalyk, 1994). So far, no studies on SCD were performed for maize germplasm adapted to Midwest U.S.

Based on preliminary experiments with inbred line derived haploid lines, a six parent diallel was produced between these three SCD inbred lines and three inbreds which have low SCD but high inducibility rates. The objectives of this study were to 1) investigate the genetics and potential practical use of SCD in DH line production, 2) evaluate the inheritance of SCD, and 3) evaluate, if a cross between and non-SCD line and an SCD line produces SCD haploids.

Materials and Methods

Preliminary Trial

As part of a study on inducibility a total of 102 public and expired Plant Variety Protection (ExPVP) inbred lines were pollinated with the maternal haploid inducer ‘RWS/RWK-76’ (Rober et al., 2005). Haploid kernels from these inbred lines were grown at the Iowa State University Agricultural Engineering and Agronomy Farm (ISU-AEA) in Boone, IA in the summer of 2013 in a preliminary screening experiment. Twenty-five putative haploid kernels from each entry were directly planted with a plot planter in a two replication randomized complete block design. Observations of fertile male tassels with healthy anthers dehiscing at ISU-DHF along with discussions with users of ISU-DHF led us to the conclusion that male fertility limits successful self-fertilization of haploids. As consequence, pollen-shedding haploids were scored. For control, self-pollinations and crosses were made in lines with high SCD. All crosses produced intact seed, which was subsequently grown for other experiments and confirmed as derived from a cross. Three inbreds were identified with superior (>50%) male fertility. Each of these inbreds produced fertile anthers, successful self-pollinations

and crosses onto other haploid lines (data not shown). The three identified inbreds and their corresponding haploids were grown side-by-side for confirmation in summer 2014 and 2015 (Figure 1) at ISU-AEA and again the haploids produced fertile anthers and were phenotypically distinct from their inbred counterpart. Additionally, a cross made between an SCD haploid from the highest doubling line and another non-doubling line was grown in the greenhouse in the winter of 2013 at ISU producing fertile F₁ plants which were induced with 'RWS/RWK-76'. These haploids were directly planted in the field and segregated for SCD male fertility at a near 1:1 ratio, confirming our observations that this is a heritable, stable, and selectable trait.

Germplasm

Six inbred lines were selected for use in a complete diallel: three of which are highly inducible, but do not have SCD potential and three of which are poorly inducible, but have high SCD potential. The lines and details of their heterotic grouping, flowering time, pedigree, and origin are detailed in Table 1. Seed from all three lines was acquired from the USDA North Central Regional Plant Introduction Station. The six inbred lines were crossed in a full diallel, creating a set of 30 unique F₁ hybrids. These hybrids were pollinated with the maternal haploid inducer RWS/RWK-76 in 2014 at the ISU-AEA to produce haploid seed for each of the hybrids. Seed was then visually sorted using the *RInj* color marker.

Field Trials

Haploids of each of the 30 F₁s were grown at a total for four locations: Iowa State University's Agricultural Engineering and Agronomy Farm (AF) in Boone, IA, USDA's North Central Regional Plant Introduction Station (PI) in Ames, IA, University of Illinois

Research Farm (IL), Urbana, IL, and Texas A&M University's Field Research Station (TX) near College Station, TX. In addition, the haploids of the six inbreds were grown in a smaller two replication trial at AF in paired rows with their corresponding inbred line for comparison. The trial at Texas A&M was planted in two row 9.14m plots on 0.76 m spacing with a total of 90 seed planted in each plot. The TX trial was delayed in planting due to heavy rains that fell in the early spring in TX. Once the plots were planted and established, heavy rains came again in May causing the nearby Brazos River to flood the nursery. However, data could still be taken as the plants were at flowering already. The remaining three trials were planted on 5.48 m plots with 0.76 m spacing with 28 kernels planted per plot in 2015. Weather in Iowa (PI and AF) was cooler and wetter than average, especially around flowering time. The trial at IL experienced a rainy and cold spring, but a normal summer season. All plots were planted using untreated haploid seed, not transplants. Normally, haploids are germinated, treated with colchicine, and then transplanted into the field. All plots were maintained with standard agronomic practices. Once plants reach ~V5 (Elmore et al., 2011) misclassified hybrids were visually identified and removed from the plot so that only haploid plants remained. A stand count was taken of the final number of plants in the row.

Traits Scored

Plots were walked daily to score flowering haploid plants for male fertility (Figure 2). Each day, individual plants were evaluated for the presence of healthy fertile anthers. Haploids are most commonly sterile. Most sterile plants do not even exert anthers out of the glumes. Plants were scored male fertile when at least one healthy anther was extruded from a glume on the tassel. If scored fertile, the plant was tagged for

counts after flowering was complete. After flowering, the number of fertile (tagged) plants was counted and divided by the total number of haploid plants in the row to compute the fraction of male fertile plants. For confirmation, one replication at AF was hand pollinated to show that fertile anthers contained indeed viable pollen.

Statistical Analyses

Two types of statistical analyses were conducted. The first considered the mating design of the diallel and broke down the variance components of the SCD trait to investigate the genetic inheritance and nature of the trait. The second analysis was based on preplanned contrasts that are of interest in the practical use of the trait. For the first analysis the SCD = fertile plant/total plant ratio was logit transformed: $SCD_{\text{logit}} = \log[(SCD+0.005)/(1-SCD+0.005)]$ as was reported in Kleiber et al. (2012) to normalize the data. The combining ability analysis was conducted using DIALLEL-SAS05 (Zhang et al., 2005) considering all F_1 s and reciprocals, also known as method 3 (Hallauer, 1988). In our experiments, we did not sample germplasm, but characterized defined lines. Thus, a fixed effect model was considered. Estimates for general combining ability (GCA), specific combining ability (SCA), $GCA \times \text{Environment}$, $SCA \times \text{Environment}$, reciprocal (REC), $REC \times \text{Environment}$, were all computed, as well as genetic variance (σ^2_g), additive variance (σ^2_A), and dominance variance (σ^2_D) determined for calculation of broad and narrow sense heritabilities. All calculations were done using DIALLEL-SAS05 and respective estimations as described in Bolboaca et al. (2011).

For the second analysis, preplanned contrasts of interest for the application of the trait were computed. SAS PROC GLIMMIX (version 9.4, SAS Institute, 2013) was

implemented using the binomial count data of number of fertile plants as successes, and total number of plants as trials. The model considered here was $Y_{ijk} = \text{Env}_i + \text{Rep}(\text{Loc})_{i(j)} + \text{Entry}_k + \text{Env} * \text{Entry}_{ik} + e_{ijk}$. Where Y_{ijk} is the mean across the experiment of the logit transformed SCDs, Env_i is the random effect of the i th environment, $\text{Rep}(\text{Loc})_{i(j)}$ is random effect of the j th replication nested in the i th environment, Entry_k is the fixed effect of the k th entry (F_1 s from diallel), $\text{Env} * \text{Entry}_{ik}$ is the random interaction between the i th environment and the k th entry, and e_{ijk} is the residual error. Contrast statements were used to test the effect of using the SCD lines as males and females when crossed to other SCD lines and when crossed to other non-SCD lines to test if SCD lines can be used as parents in a cross to produce SCD in non-SCD background.

Results

Figure 3 summarizes the least square means across environments for the 30 hybrids ordered from highest to lowest SCD. GF1/GF2 has the highest estimated SCD across environments at 46% followed by its reciprocal cross at 38%. The top five SCD hybrids are all SCDxSCD crosses. The lowest estimated SCD is GF2/GF5 at 9%. There are no SCDxSCD crosses in the worst five hybrids with two of the five being non-SCD x non-SCD crosses. The average value for SCD across all hybrids and locations is 23%. As part of the separate two replication experiment at AF, the estimates for the inbred SCD rates were as follows: GF1 (94%), GF2 (65%), GF3 (71%), GF4 (71%), GF5 (0%), GF6 (71%). Figure 4 summarizes the distribution of DH lines generated from the pollination of the single replication at AF. Summarized is both the DH lines made per attempted pollination (left bar) and the DH lines made per total haploid plants in the row. The

highest percentage (75%) for DH lines made per attempted pollination was GF1/GF4 and the highest percentage for DH lines produced per number of haploid plants in the row was GF4/GF5 with (33%). In general the lowest estimates (zero successful pollinations) were combinations with GF4 and especially GF5. Also included in Figure 4 are totals across all combinations for each line. The values for DH lines per attempted pollination are as follows: GF1 (23%), GF2 (32%), GF3 (32%), GF4 (32%), GF5 (14%), and GF6 (25%).

Diallel Analysis

In the combining ability analysis (Table 2) the effect of environment, GCA, SCA, and GCAxEnv interaction are all significant at the 0.05 level. There was no significant effect for the direction of the cross (reciprocal effect) (Table 2). Positive general combining abilities were estimated for GF1, GF2, and GF6 (Table 3) though only GCA for GF1 was significant. GF3, GF4, and GF5 produced negative GCA estimates with significant GCA estimates for GF4 and GF5. The highest positive GCA estimate (17.1%) was obtained for GF1. GF4 produced the lowest negative GCA with a -8.3% reduction in SCD when combined with the other lines in the diallel. Significant SCA estimates were found for GF1/GF2 (12.4%), GF3/GF4(-13.9%), GF2/GF5(-13.4%), GF3/GF5(23.3%) and GF4/GF5(25.5%). Heritability estimates of SCD from the diallel analysis were 0.62 for broad sense heritability and 0.31 for narrow sense heritability.

Contrast Analysis

In Table 4 for SCD lines (GF1 p-value=0.45, GF2 p-value=0.84, GF6 p-value=0.73) it does not matter whether they are used as male or female. There is a significant difference between the SCD and non-SCD group when crossed to them. For

the following, the difference in consideration is whether or not for the SCD and non-SCD line groups (GF1,GF2, GF6, or GF3, GF4,GF5, respectively) there is a significant difference between when the line in question is used as the female or the male in the cross. For GF1, it matters whether it is used as a female (p-value=0.0004), or a male (p-value=0.0147) when crossing to SCD and non-SCD lines. The same is true for GF2, when used as a female (p-value=0.001) or a male (p-value=<0.0001). This is also true for GF6 when used as a female (p-value=0.02) or a male (p-value=0.04). For GF4 there are no significant differences. For GF5 and GF3 it only matters when they are used as the male in the cross (p-values=0.01 and 0.05, respectively).

Discussion

Rates of SCD

The efficient and economical production of DH lines requires the ability to produce fertile haploid plants at a high enough percentage that it is not necessary to plant out too many haploids per population. Based on experience with standard colchicine doubled DH line production at the ISU-DHF, an average of about 20-25% of colchicine treated plants will produce fertile anthers for pollinations (Frei, 2015, personal communication). These 20-25% of plants are the result of thousands of seeds being germinated in trays, subsequently individually treated with colchicine and then transplanted into the field. Values of 24% were obtained in a replicated experiment conducted on haploids produced from B73 with different colchicine application methods (Vanous, 2011). However, across the 30 hybrids in this experiment as seen in Figure 3, an average male fertility rate of 23% was achieved without any transplanting and

colchicine treatment. This is a substantial improvement and might lead to substantial savings of both time and money when introduced into breeding programs. Our inferences are limited to the six lines of our diallel. However, breeding with SCD lines was started in 2014 and haploids grown in 2015 as part of a line development program showed fertility ranging from 5-30% without colchicine from crosses between the SCD lines and non-SCD elite germplasm (data not shown). Preliminary experiments identified the SCD lines as GF1, GF2, and GF6 based on male fertility in self and cross combinations. However, as was seen across all environments, GF3 and GF4 expressed some SCD male fertility based on our criteria. This is especially evident when significant SCA estimates are found for GF4/GF5 (25.5%) and GF3/GF5 (23.3%). This is further supported by the percentage of successful pollinations per plants in the row in GF4/GF5 (33%) and GF3/GF5 (11%). It seems that GF3 and GF4 complement well with GF5 and express SCD. This was not expected, however, it shows that this is not a simply inherited trait, and there are more complex genetic factors that must be considered.

Utilization of SCD in breeding populations

Though it is clear that there is potential for the use of SCD in breeding programs that utilize the maize *in vivo* maternal DH system, there are challenges that must be addressed before utilization of SCD in breeding populations can be fully adopted. Unlike maternal haploid inducer development, where the objective is to create a single inducer that works well with all germplasm, the incorporation of SCD will require the trait be moved into all the active breeding populations being used for DH line development. The ideal situation would be that SCD is a single gene trait which is simply inherited and not affected by the environment or the genetic background it is moved into. However, based

on the results of this study, it is evident that SCD will be more challenging to work with. Effects of environment and genotype by environment interactions will complicate the usage of the trait across the germplasm. The more complex nature of its inheritance, with both significant GCA and SCA complicates the way it is utilized in the germplasm. The effect of the genetic background which it is in also will affect the nature in which it is utilized in crosses and breeding.

Effects of environment complicate the use of many traits when incorporated into a breeding program. In this experiment, the effect of environment was significant, which is not surprising as we used very different environments both in Texas and the Midwest. Estimates for rates of male fertility were 65% (max=78%, min=5%) at AF, 43% (max=51%, min=0%) in CS, 41% (max=67%, min=0%) in IL, and 47% (max=71%, min=0%) at PI. Though the lowest average rates of male fertility were in CS, the location with the most plots with zero plants fertile was IL with a total of 19, ten of which were SCD line combinations. The highest observed rate was 78% at AF. Of the few reports considering SCD in the literature, only one (Kleiber et al., 2012) used multiple environments for their trials. Kleiber et al. (2012) reported no significant effect of environment. It will be important, moving forward, to consider the effect of the breeding environment when utilizing SCD germplasm.

A simple effect of environment would mean that certain environments are more conducive to the expression of SCD, however, with significant genotype by environment interaction, not only are environments different, but specific lines are performing

differently in specific locations. For example, for the GF1 crosses, when crossed by GF2 it performed, on average, 19% worse at AF as compared to PI, IL, and CS. In contrast, GF1/GF3 performed 25% better at PI than the average of AF, IL, and CS, and over 30% better and AF alone. For AF, the best performer was GF6/GF5 (68%), and the worst performer was GF1/GF3 (34%). For CS, the best performer was GF4/GF6 (61%) and the worst performer was GF5/GF2 (36%). For IL, the best performer was GF2/GF1 (59%) and the worst performer was GF2/GF4 (34%). For PI the best performer was GF1/GF3 (69%) and the worst performer was GF2/GF5 (40%). The only consistency was GF1/GF3 performing best at AF and PI. These two locations (AF and PI) are located in adjacent counties in Iowa. It seems that the adaptation of the inbred lines could be playing a role in the expression of the trait. The only other study to consider multiple environments (Klieber et al. 2012) did not report significant GxE interactions, but their methods used adapted elite materials from a private company in one experiment, and completely unadapted material from tropical germplasm in another. Consideration will need to be given to the SCD line used for specific environments. Alternatively, the SCD trait can be moved into the germplasm that is adapted to that specific environment.

Simply moving the trait into new germplasm, however, may not be the entire answer. The combining ability analysis summarized in Table 3 shows that there are both additive and epistatic effects that must be considered when working with SCD. Significant GCA indicates additive genetic effects that are passed on to progeny making line conversion, and/or development of new elite germplasm straightforward. This supports earlier findings where recurrent selection for SCD was practiced on a small

scale, leading to an increase in the rate of SCD in the population (Zabirova et al., 1993), as well as reports that variation for SCD exists (Geiger and Schonleben, 2011). A larger broad sense heritability as compared to the narrow sense heritability shows that there are dominance and/or epistatic genetic components involved. This is supported by significant SCA effects for some hybrid combinations: 12.4% for GF1/GF2, -13.85% for GF3/GF4, -13.41 for GF2/GF5, 23.28% for GF3/GF5, and 25.46% for GF4/GF5. GF1/GF2 was far superior to all other combinations, and could thus be used to make three way crosses to increase the rates of SCD. However, it is important here to consider what is giving this significant SCA effect. Though the nature of SCD is unknown, it is likely being expressed in the haploid plant since if the genome doubled very early in development (i.e. in the embryo formation), the plants would look exactly like their inbred parent which is not the case. Since the haploid plant only has a single copy of the genome and even if doubled it is a 100% homozygous inbred there is no possibility for dominance effects. Thus, what remains to explain this is epistatic effects in the haploid genome. Epistatic effects can be exploited when making breeding crosses in DH breeding programs. If there are two sets of germplasm that combine to create good SCA for SCD such as the GF1/GF2 combination shown in Table 3, then crossing them for production of donor populations will result in high SCD potential. Also, the ability to produce SCD in hybrids where a SCD line and a non-SCD line are crossed makes the selection process even simpler. This ability to recover SCD is seen in the overall averages for the inbreds in Figure 4 where both SCD and non-SCD inbreds produced DH lines. Haploids will select for the breeder, as only those haploids which double will produce progeny and consequently all non-SCD progeny will be eliminated. There could, however, be potential

drawbacks to this. For example, if any unwanted loci are linked to or are in gametic phase disequilibrium with loci controlling SCD, allowing only fertile haploids through the selection bottleneck could create undesirable phenotypes in the breeding populations. The production of SCD haploids in GF4 and GF3 crosses, as evidenced by the significant SCA (Table 4) suggest that perhaps there is some complementation in this trait that is segregating in this population. Since there has been no prior selection for SCD, the loci segregating in the germplasm may have complementary loci which when combined produce SCD. Ongoing mapping studies will further elucidate this question. A study done by Laude and Carena (2014) looked into combining abilities among 16 maize populations adapted to the temperate U.S. Corn Belt for grain and grain quality traits. They found that for grain yield, the predominant factor was non-additive genetic effects (SCA), i.e. specific combinations of populations had more of an effect than did populations overall. While for grain quality traits, the predominant factor was additive genetic effects (GCA), i.e. specific populations carried more favorable additive alleles that increased values when combined with any other population. However, other, contrasting conclusions are also cited (Laude and Carena, 2014), leading to the conclusion that the conclusions are specific to the population being used. As seen for SCD, lines such as GF1 provide positive and significant GCA estimates, suggesting that it carries alleles that are additive in nature and works well with other germplasm. In contrast, there is GF4 and GF5 which produce significant negative GCA estimates, but positive SCA estimates (Table 4) suggesting that they carry alleles that are important for SCA, but only work in combination with other alleles. It seems that the inheritance of the trait is not as straightforward as originally hypothesized.

How then, should the genetic architecture of this trait be characterized for use in breeding crosses and selection? Most of the published reports of haploid fertility in maize generally conclude that there is significant genetic variation in SCD and that selection for the trait is possible (Kleiber et al., 2012; Geiger and Schonleben, 2011; Chalyk, 1994; Geiger et al., 2006). These studies concluded that in their particular germplasm, varying from proprietary breeding germplasm to unadapted tropical germplasm, that there is significant variation for SCD. When comparing an untreated control to nitrous oxide doubling, Kato (2002) showed that untreated maize haploids were doubling due to SCD at 11%, similar to the rates (12%) observed first by Chase (1952). Wu et al. (2014) also reports identification of SCD in Zhengdan958 at a maximum rate of 3.52% and rates in other germplasm ranging from 3.85%-1.06%. Finally, Sugihara et al. (2013) report on a single locus *fdr1* which was mutated with sodium azide which produced fertile haploid plants at a rate of 48%. Though these previous studies provided motivation and a basis for investigating SCD, what they lack is planned crosses in a designed mating, which this study provides. Based on the combining analysis, new hypotheses can be developed about the genetic nature of SCD. At first, it was thought that SCD was a single (or very few) gene trait that was simply inherited based on some of the preliminary data. Crosses made with SCD lines and non-SCD lines yielded fertile haploids. However, as is seen in Table 4, there are SCA effects that are significant for GF3 and GF4. This suggests that there are multiple loci controlling SCA, and that complementation is occurring leading to epistatic effects with specific combinations. GF1 may potentially carry several of these alleles, or a major allele that allows it to have a positive GCA across all the lines.

However, GF3 and GF4, while having negative GCA estimates (Table 4), have high SCA estimates for specific crosses (Table 4). It is possible that the alleles that are affecting SCD may be segregating in the germplasm, unknown since this is not a trait that is selected on and prior to DH was of no consideration. When moving this trait into breeding germplasm, the genetic background will need to be taken into consideration. It may involve moving it into different germplasm and evaluating the rates of SCD and then using an adapted and new background to move SCD into the remaining germplasm pools.

The most challenging aspect of working with SCD will be the complex nature of its expression and inheritance. Both additive and epistatic gene action, likely with some form of complementation in the germplasm is occurring. Further studies need to be conducted for mapping loci controlling SCD, investigating negative effects of the trait, and finally to study performance of SCD in diverse genetic backgrounds.

Potential Applications

The ability to directly plant haploid seed without the tedious process of treating with colchicine and transplanting not only makes the DH system cheaper and more efficient, but also makes it safer. Haploid seed could be packaged like all other nursery seed, or bulk planted if a large enough quantity mitigating the risk of weather events as well as allowing for better delay control for labor management. In general, most DH line development begins with the crossing of two parents for the development of an F_1 which is induced. However, there is some discussion about the use of F_2 populations for induction. In a simulation study, Bernardo (2009) suggested that F_2 be used for inductions instead of F_1 . This could be more possible with the use of SCD, however, a

major objective of the DH system is rapid line development. Using F_2 populations would add a season to the cycle and there is no guarantee that any given F_2 plant will produce a DH line. However, if there is a non-negotiable trait that can be easily selected for among the F_2 then this may be beneficial through F_2 enrichment. However, with the rates of SCD seen and the ease of its use, then one could simply select among the haploids which are essentially samples of F_2 gametes. Doubling rates with colchicine are generally 30% in adapted and elite material, especially if it has been through a cycle of DH. However, when tropical material, or older germplasm as well as sweet corn and popcorn are considered the rates of doubling drop down to near zero making DHs difficult if not impossible. The SCD trait could be moved into some of the tropical/sweet corn/popcorn germplasm and would allow rapid adaptation of tropical germplasm as well as give sweet corn and popcorn breeders a new tool for line development. Consider for example the cost of producing a DH line using colchicine doubling. The current rate offered by the ISU-DHF is \$38.25 per DH line for germination, colchicine treatment, transplanting, and pollination. So, for 200 lines it would cost \$7,650. Compare this to using SCD, once it is moved into the breeding population. A success rate of 10% will be used (Figure 4) for this example. Here this is no need for germination, transplanting, or colchicine treatment. So, it would simply be 2000 haploid seed planted in plots with a plot planter. The cost of space for this planted at 74,000 plants/hectare would be \$70 with a rental cost of \$2,475 per hectare. The cost of supplies and time for planting would be approximately \$45. The cost of pollinating (1 person, 20\$/hour, 2 hours per day, for 14 days) would be \$560. This would add up to \$675 to create 200 new DH lines using SCD. Thus, the cost per line would be \$3.38. This is a 91% savings over colchicine doubled DH line production.

Finally, the ability to produce large number of fertile haploids cheaply could allow for the development of new breeding strategies where selection can be conducted at the haploid level. Currently no selection, other than natural selection, is conducted at the haploid level and all DH lines are advanced to testing. It could be possible to sample DNA from haploids and use genomic selection or marker assisted selection strategies to select among the haploids and advance only those lines which are of interest based on marker information. Consideration to this possibility was given by Wu et al. (2014) where the efficiency of selection using genomic selection was contrasted with selection pressure applied at the haploid versus the DH level. It was concluded that in order to make genomic selection (GS) at the haploid level more effective a success rate for generation of DH lines from haploids without colchicine of 17% would be needed (Wu et al., 2014). Based on the results presented here, it would be possible to use the haploid selection scheme presented by Wu et al. (2014) to make GS more effective at the haploid level.

Acknowledgments: Authors 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 funding this work.

References

Bernardo, R. 2009. Should maize doubled haploids be induced among F1 or F2 plants?.

Theoretical and Applied Genetics. 119:255-262

- Blakeslee, A.F. and A.G. Avery. 1937. Methods of inducing doubling of chromosomes in plants by treatment with colchicine. *Journal of Heredity* 28:393-411
- Bolboaca, S.D., L. Jantschi, A.D. Baci, and R.E. Sestras. 2011. Griffing's experimental method II: step by step descriptive and inferential analysis of variances. *Journal of Biostatistics* 6:31-52
- Borisy, G.G., and E.W. Taylor. 1967. The mechanism and action of colchicine binding of colchicine-3H to cellular protein. *The Journal of Biology* 34:525-533
- Castillo A.M., Cistué L., Vallés M.P., Soriano M. 2009. Chromosome doubling in monocots, in: A. Touraev, et al. (Eds.), *Advances in haploid production in higher plants*, Springer Netherlands. pp. 329-338.
- Chalyk, S.T. 1994. Properties of maternal haploid maize plants and potential application to maize breeding. *Euphytica* 79:13-18.
- Chase, S.S. 1952. Monoploids in maize, in: Gowen, J.W. (ed.), *Heterosis*. Iowa State College Press, Ames. P. 390-399
- Eder, J., and S. Chalyk. 2002. In vivo haploid induction in maize. *Theoretical and Applied Genetics*. 104:703-708
- Elmore, R.W., M.J. Boyer, and S.K. Marlay. 2011. Corn growth and development. *PMR* 1009. Iowa State University Extension, Ames, IA. p. 13-27
- Geiger, H.H., and G.A. Gordillo. 2009. Doubled Haploids in Maize Breeding. *Maydica* 54:485
- Geiger, H.H., M.D. Braun, G.A. Gordillo, S. Koch, J. Jesse, and B.A.E. Krutzfeld. 2006. Variation for female fertility among haploid maize lines. *Maize Genetics Newsletter* 80:28-29

- Geiger, H.H., and M. Schonleben. 2011. Incidence of male fertility in haploid elite dent maize germplasm. *Maize Genetics Newsletter* 85:22-32.
- Hallauer, A.R., M.J. Carena, and J.B. Miranda. 1988. Quantitative genetics in maize breeding. In: Prohens, J., F. Nuez, and M.J. Carena, editors, *Handbook of Plant Breeding*. Springer, New York, NY. p. 1-31
- Kleiber, D., V. Prigge, A.E. Melchinger, F. Burkard, F. San Vicente, G. Palomino, and G.A. Gordillo. 2012. Haploid fertility in temperate and tropical maize germplasm. *Crop Science* 52:623-630
- Longin, C.F.H., H. Utz, J.C. Reif, W. Schipprack, and A.E. Melchinger. 2006. Hybrid maize breeding with doubled haploids: I. One-stage versus two-stage selection for testcross performance. *Theoretical and Applied Genetics* 112:903-912
- Prigge, V. and A.E. Melchinger. 2012. Production of haploids and doubled haploids in maize. *Plant Cell Culture Protocols* pp.161-172
- Prigge, V., C. Sanchez, B.S. Dhillon, W. Schipprack, J.L. Araus, M. Banziger, and A.E. Melchinger. 2011. Doubled haploids in tropical maize: I. Effects of inducers and source germplasm on in vivo haploid induction rates. *Crop Science* 51:1498-1506
- Rober, 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.
- SAS Institute. 2013. SAS version 9.4. SAS Institute, Cary, NC
- Sugihara, N., T. Higashigawa, D. Aramoto, and A. Kato. 2013. Haploid plants carrying a sodium azide-induced mutation (*fdr1*) produce fertile pollen grains due to first

division restitution (FDR) in maize (*Zea mays* L.). Theoretical and Applied Genetics 126:2931-2941

Vanous, A.E. 2011. Optimization of doubled haploid production in maize (*Zea mays*L.).

M.S. thesis, Iowa State University, Ames, IA

Wu, P., J. Ren, L. Li, and S. Chen. 2014. Early spontaneous diploidization of maternal maize haploids generated by in vivo haploid induction. Euphytica 200:127-138

Wu, Y., U.K. Frei, H. Liu, G. De La Fuente, K. Huang, Y. Wei, and T. Lubberstedt.

2014. Combining genomic selection and doubled haploid technology increases efficiency of maize breeding, In: Govil, J.N., editor, Recent Developments in Biotechnology. Studium Press. p. 215-237

Zabirova, E., O.A. Shatskaya, and V.S. Shcherbak. 1993. Line 613/2 as a source of a high frequency of spontaneous diploidization in corn. Maize Genetics Newsletter 67:67

Zhang, Y., M.S. Kang, and K.R. Lamkey. 2005. DIALLEL-SAS05: a comprehensive program for Griffing's and Gardner-Eberhart analyses. Agronomy Journal 97:1097-1106

Table 1. Pedigree, origin, and general information for six parents included in diallel.

Line Name	Heterotic Group	Origin	Pedigree	GDD to Silk	Selection Reason
GF1	non-stiff stalk	Minnesota	CC36 x A405	1318	spontaneous chromosome doubling
GF2	stiff stalk	Minnesota	CO106 x A321	1522	spontaneous chromosome doubling
GF3	non-stiff stalk	Nebraska	W117Ht x Mo17Ht	1178	high inducibility rate
GF4	stiff stalk	Hawaii	Oh40B, L317, GF5	1640	high inducibility rate
GF5	stiff stalk	Indiana	Indiana strain of Stiff Stalk Synthetic	1522	high inducibility rate
GF6	stiff stalk	Minnesota	W117 x B37Ht	1400	spontaneous chromosome doubling

Table 2. Analysis of variance table for Griffing's method 3 fixed model diallel analysis of SCD. GCA=general combining ability, SCA=specific combining ability, REC = reciprocal, Env = environment. *=significant at the 0.05 level

	df	SS	MS	F-value	P-value
Environments	3	20.75	6.9181	29.81	<0.001*
Reps(Environments)	8	1.7213	0.2139	0.92	0.4991
Hybrids	29	26.694	0.9204	3.97	<0.001*
GCA	5	13.5815	2.716	11.704	<0.001*
SCA	9	16.7224	1.858	8.006	<0.001*
REC	15	3.4807	0.232	0.999	0.456
Hybrid x Env	87	30.1002	0.3459	1.49	0.0098*
GCAxEnv	15	11.07855	0.7385	3.1825	<0.001*
SCAxEnv	27	8.6581	0.32067	1.3818	0.10646
REcxEnv	45	11.8316	0.26292	1.1329	0.274
Error	29	232	53.8411	0.232	

Table 3. Combining ability table for the six parent diallel for SCD. GCA estimates are on the diagonal and SCA estimates are on the off diagonal. Only the top half of the matrix was filled in as there was not significant reciprocal or maternal effects. Data is presented in percentages which were obtained from using the inverse logit function to convert the SCD data back into a ratio. *=significance at the 0.05 level.

	GF1	GF2	GF3	GF4	GF5	GF6
GF1	17.13%*	12.40%*	-3.27%	1.56%	-6.55%	-0.54%
GF2		1.57%	5.28%	1.70%	-13.41%*	3.95%
GF3			-2.15%	-13.85%*	23.28%*	3.64%
GF4				-8.28%*	25.46%*	0.36%
GF5					-6.10%*	-6.01%
GF6						5.09%

Table 4. Preplanned contrasts for combinations of SCD and non-SCD lines when used as males and females. *=significance at the 0.05 level.

Contrast	F Value	Pr > F
GF1 as female: crossed to SCD vs non-SCD	13.74	0.0004*
GF1 as male: crossed to SCD vs non-SCD	6.2	0.0147*
GF1 as female vs GF1 as male	0.59	0.4455
GF2 as female: crossed to SCD vs non-SCD	11.6	0.001*
GF2 as male: crossed to SCD vs non-SCD	25.01	<.0001*
GF2 as female vs GF2 as male	0.04	0.8356
GF3 as female: crossed to SCD vs non-SCD	2.79	0.0987
GF3 as male: crossed to SCD vs non-SCD	4.11	0.0457*
GF3 as female vs GF3 as male	1.1	0.2963
GF4 as female: crossed to SCD vs non-SCD	1.67	0.1995
GF4 as male: crossed to SCD vs non-SCD	3.18	0.0781
GF4 as female vs GF4 as male	0.28	0.595
GF5 as female: crossed to SCD vs non-SCD	3.43	0.0673
GF5 as male: crossed to SCD vs non-SCD	5.98	0.0165*
GF5 as female vs GF5 as male	0.1	0.7487
GF6 as female: crossed to SCD vs non-SCD	5.26	0.0242*
GF6 as male: crossed to SCD vs non-SCD	4.3	0.041*
GF6 as female vs GF6 as male	0.12	0.728

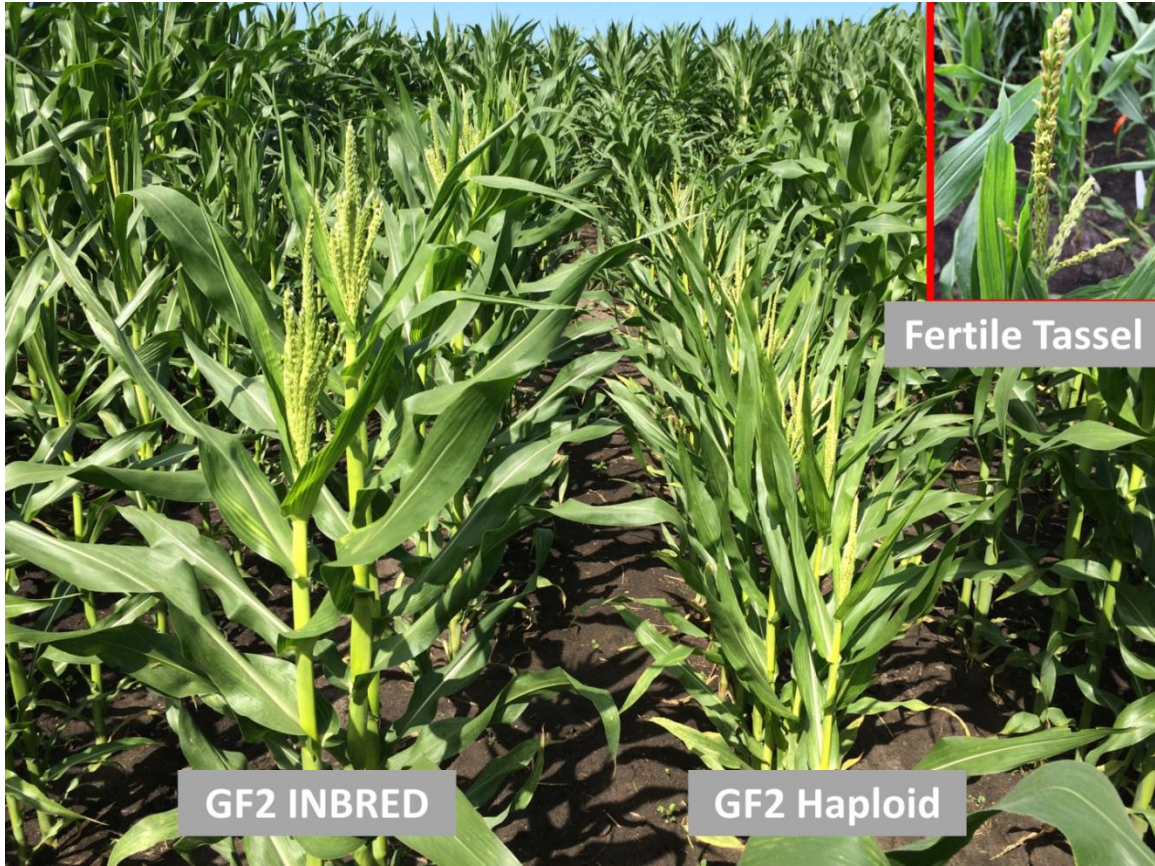


Figure 1. Side by side comparison of GF2 inbred and GF2 haploid. General phenotype of plants is identical. However, the haploid plants are always shorter and less vigorous, they typically have smaller leaves, ears, and tassels. Inset is a picture of a fertile tassel on a GF2 haploid.



Figure 2. Fertile tassels in haploid field trials. Outer four pictures show examples of the tassel phenotypes seed when healthy anthers are shedding pollen on an SCD haploid plant. The center picture gives a close-up of the anthers and pollen.

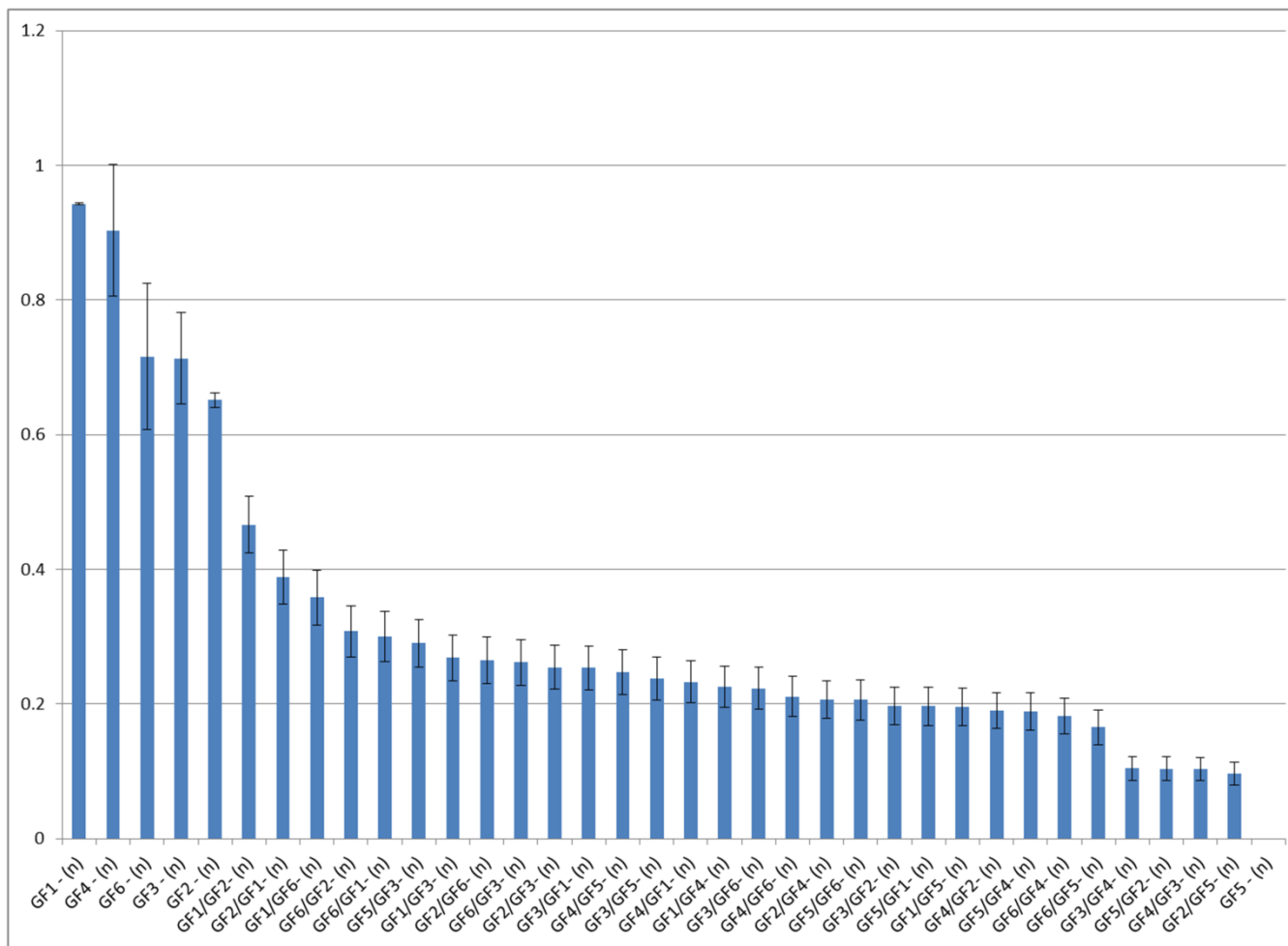


Figure 3. Least square means of SCD across all environments from the contrast analysis. SCD is presented in a ratio of fertile plants / total plants.

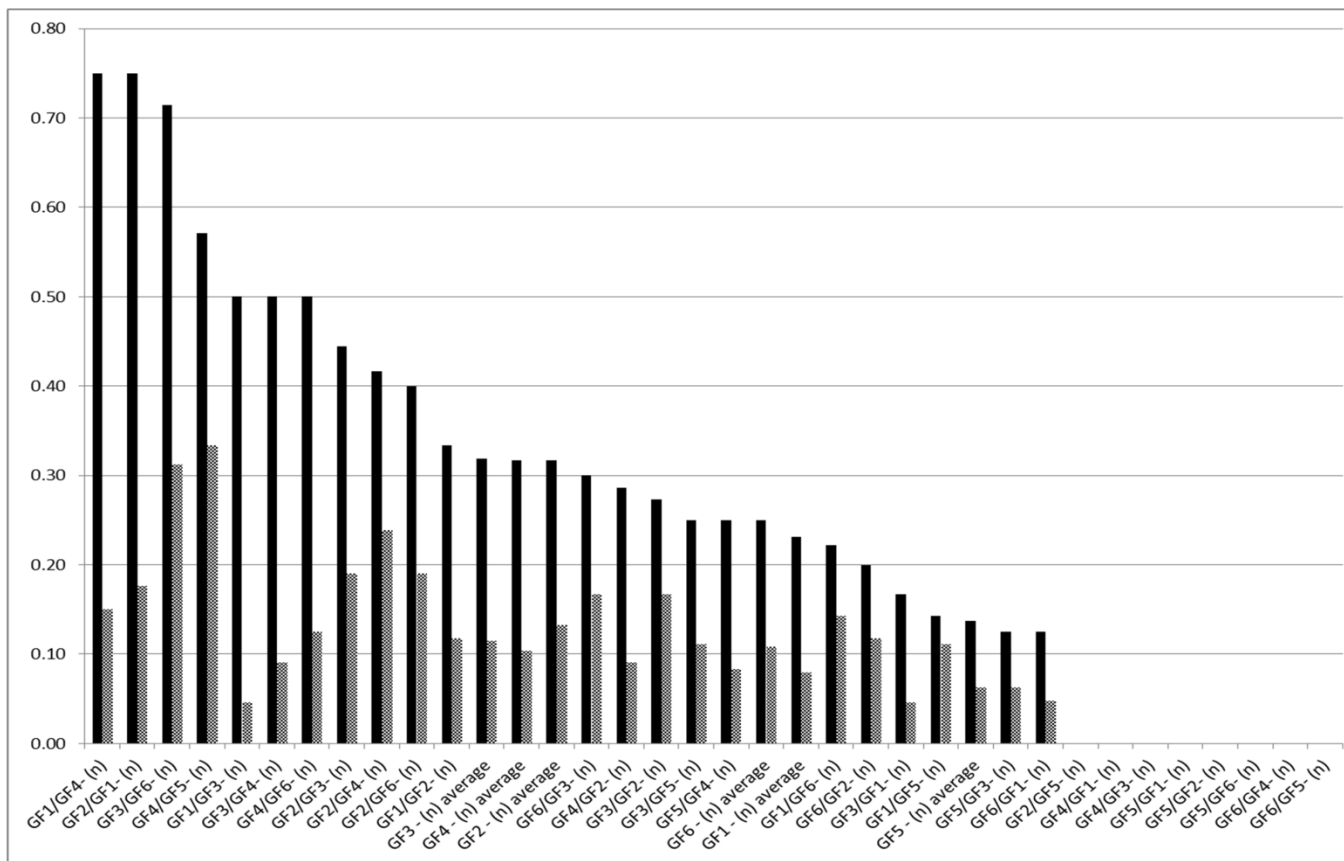


Figure 4. Summary of pollination success for pollinations made in single rep at Iowa State University Agronomy Farm. Solid bar denotes ratio of successful pollinations (DH lines) made per fertile plants (attempted pollinations). Dotted bar denotes ratio of successful pollinations made per total haploid plants in the row. Also included are averages of all combinations for each of the inbred lines in the diallel.

CHAPTER SIX**ACCELERATING PLANT BREEDING**

*Gerald N. De La Fuente*¹, Ursula K. Frei¹, Thomas Lübberstedt¹

Manuscript published in Trends in Plant Science. Abstract, structure, and references are all formatted according to journal standards.

Abstract

The growing demand for food with limited arable land available, necessitates that the yield of major food crops continues to increase over time. Advances in marker technology, predictive statistics, and breeding methodology have allowed for continued increases in crop performance through genetic improvement. However, one major bottleneck is the generation time of plants, which is biologically limited and has not been improved since the introduction of doubled haploid technology. In this opinion article we propose to implement *in vitro* nurseries, which could substantially shorten generation time through rapid cycles of meiosis and mitosis. This could prove a useful tool for speeding up future breeding programs with the aim of sustainable food production.

Glossary box

backcross: a breeding methodology where a gene or few genes (for example, resistance to a disease) usually contained within a wild or less than acceptable line are transferred to high performing lines by crossing the two lines and then repeatedly crossing the progeny back to the high performing parent while selecting for the gene or few genes of interest. The objective is to produce progeny that are as genetically similar to the high performing

parent as possible while containing the gene or few genes desired from the less than acceptable parent.

BC4 line: backcross 4 line. Lines which are derived after four generations of backcrossing.

full-sib recurrent selection: a method of genotypic recurrent selection where individuals are evaluated for performance by paired plant cross pollinations which generates a set of full-sib (i.e. two shared parents) families which are tested in replicated trials to generate data for selection. Requires 2 seasons per cycle.

half-sib recurrent selection: a method of genotypic recurrent selection where individuals are evaluated for performance by cross pollination with a tester which generates a set of half-sib (i.e. one shared parent) families which are tested in replicated trials to generate data for selection. Requires 1-3 seasons per cycle depending on the specific method used.

genetic improvement/gain: the change in mean performance of a population that occurs as the result of the selection and recombination of superior performing individuals in a population.

introgression: a relatively small portion of the genome of an unadapted individual which is transferred through conventional crossing to adapted germplasm for evaluation of its utility for genetic improvement.

linkage drag: the undesirable transfer of unwanted genes along with the gene/locus of interest due to physical linkage causing a decrease in performance of the progeny.

MABC: marker assisted backcross. A variation of the backcross breeding methodology where molecular markers are used to select for the trait of interest, and if desired for maximum recovery of the desired parent genome.

self incompatibility: the inability of a plant with functional male and female gametes to produce a zygote through self fertilization.

selfed progeny recurrent selection: a method of genotypic recurrent selection where individuals are evaluated for performance by development of selfed families (i.e. F2:3, F3:4, F4:5, etc.) which are tested in replicated trials to generate data for selection. Requires 3+ seasons per cycle depending on how advanced the generation of self pollination is (i.e. more time is required for F4:5 than F2:3).

Keeping up with demand

Crop production has steadily increased over time and it has been suggested that 50% of the progress is attributable to advances in crop management and breeding [1, 2]. For example, the three major crops in the US, maize (*Zea mays*), wheat (*Triticum spp.*), and soybean (*Glycine max*), show positive linear increases in average yield from 1930 to 2012 [3] (Figure 1). However, changes in climatic patterns, land and water availability now provide additional challenges for plant breeders and geneticists to ensure yield stability in varying environments [4]. In order to meet the projected increase of global demand for food, feed, and fibre (100% by 2050 [5]), the linear progress seen in Figure 1 will need to be increased. In order to increase the rate of genetic improvement (see Glossary) the efficiency, reliability, and speed of genetic improvement must be increased.

In this opinion article we propose an idea benefitting the speed of genetic improvement through the implementation of rapid generation cycling by the use of the *in vitro* nursery. Through rapid cycles of meiosis and mitosis conducted in tissue culture, generation times of crop species can be decreased allowing more opportunities for recombination and selection in a given unit of time.

The breeder's equation

Five modifiable components are used to estimate genetic gain (Box 1): additive genetic and phenotypic variance (which can be combined as narrow sense heritability), selection intensity, parental control, and time [6-9]. Choice of germplasm for formation of segregating populations affects additive variation (genetic variation that can be transmitted to the next generation), while choice and management of selection environments affects phenotypic variance. A combination of these components affects selection efficiency. Selection intensity, corresponding to percentage of individuals advanced after a cycle of selection, can be modified easily. The aforementioned factors can be optimized through knowledge of the germplasm and the use of predictive tools. The most critical remaining factor to maximize genetic gain is time. The number of generations per year is biologically limited. The most extreme cases are short generation times (six/year) in *Arabidopsis* (*Arabidopsis thaliana*) versus long generation times in tree species (multiple years/generation). Advances in cycle time have been limited, except for the use of off season nurseries, and doubled haploid technology.

Speeding up

Off season nurseries, popularized by the pioneering plant breeder Norman Borlaug among others, can help to reduce the time needed to release new cultivars, e.g., the time for producing a new wheat cultivar was shortened from 10-12 to 5-6 years [10]. For pure line and hybrid crop breeding, the ability to generate homozygous and homogeneous lines is another time constraint. However, by using doubled haploids (DHs) in different crop species, homozygous and homogeneous lines have been produced in two rather than five or more generations, and was the last major breakthrough to reduce cycle time [11-13]. The most popular being the maize DH system using the *R1-nj* color marker [14]. But the different steps of the DH process (Figure 2) have biological and genotypic limitations. The success rates for haploid induction [11, 15-17], adaptation to tissue culture (in the case of anther culture) [18], and doubling [19] have all been shown to be genotype dependent in different crop species. Breeders using DHs will unintentionally practice recurrent selection for loci increasing success rates of the DH process [20], which might constrain genetic variation in breeding populations, at least for respective genome regions.

The *in vitro* nursery

Currently, the most efficient way to produce homozygous and homogeneous lines is through a combination of off season nurseries (generations per year) and DH technology (homozygosity per generation). We propose the concept of an *in vitro* nursery, where new genotypes are formed by *in vitro* production of gametes and their

subsequent fusion. Here, generation time is limited by how quickly somatic cells can form new gametes and how quickly these gametes can be fused.

The general progression of the *in vitro* nursery is outlined in Figure 3. Tissue is extracted from the basal leaf section of selected genotypes and converted into an *in vitro* cell culture and induced to mitotically divide through application of growth regulators such as 2-4D [21], which can be maintained in minimal space requirements in a laboratory setting with each cell callus occupying about 4 mm² [22]. Genotypes of interest are subsequently isolated and single somatic cells are induced to undergo meiosis for generation of new gametes. These gametes are subsequently fused to generate new genotypes in a similar way to the *in vivo* unification of pollen and egg cells. However, in contrast to the *in vivo* system, where the breeder would need to wait until seed maturity and the flowering of progeny to produce the next generation, fused diploid cells could immediately be induced to undergo meiosis within the *in vitro* system, and produce gametes for new crosses, or for artificial genome doubling to produce a new homogeneous/homozygous cell line. [23]. Several techniques exist for fusion of plant gametes *in vitro*: electrically induced fusion, chemically induced fusion, and calcium induced fusion [24, 25]. Successful fusion of plant gametes *in vitro* has been reported for maize [23, 26], wheat [27], rice (*Oryza sativa* L.) [28], and tobacco (*Nicotiana tabacum* L.) [29]. The main biological bottlenecks are now limited to the induction of meiosis and the rate of cell division, whose estimation is critical to successful tissue culture [22]. It is estimated that plant cell division rates can range from 22 to 48 hours [21].

This entire process would need to be coupled with marker-based and/or genomic selection. Evaluation and selection within the *in vitro* nursery would be accomplished by running marker analyses on new cell lines and/or gametes. Time can be saved by using single cells for whole genome amplification and subsequent marker analysis [30, 31]. Selection efficiency can be increased by selecting gametes versus zygotes. In traditional breeding practices, selection is limited to the diploid (or polyploid) plant in most cases. A notable exception would be selection on haploid plants in a DH system. In the *in vitro* system, specific and targeted matings could be achieved through mitotic division of gametes and subsequent marker analysis for genomic gamete selection (GGS). Though no examples exist of the mitotic division and callus formation of artificially induced gametes, other biological examples such as yeast, the ability to grow haploid callus in anther culture, and the normal (though weak) functionality of haploid maize plants provide evidence that this is possible. These haploid mitotic divisions allow for the selection of gametes without their destruction. This could also be coupled with optimization procedures for generating optimal genotypes with minimal numbers of resources and time [32] increasing selection efficiency. Selected cell lines could then be converted to mature plants, which can be used for phenotypic evaluation. In maize, converting cell lines into mature plants will be the most time-demanding step, currently requiring 148 to 215 days from gamete fusion to the harvest of mature seed. Plant regeneration is not 100% efficient, and varies in different species with percentages reported as 37-73% in tobacco, 25-48% in rice, 41-59% in maize, 5-33% in cotton (*Gosypium hirsutum*), and 93-100% in soybean[33]. This step is likely also genotype dependent and warrants more research into the regeneration of plants from tissue culture.

The obvious advantage of this system is the reduction in time for line development. With a conservative estimate for a division rate of 48 hours per cycle, a new cycle could be generated every week, provided that marker analyses could be conducted at a similar pace. For comparison, a DH line can be produced in one year with only a single recombination event. Alternatively, in the same time period, a line produced from the *in vitro* nursery could result from 12 cycles of recombination and selection (at 1 week intervals), assuming that meiotic induction and division takes 48 hours similar to the division rate for mitosis, before plant regeneration is limiting seed production. The utility of *in vitro* nurseries is obvious for both mapping and marker-assisted backcrossing (MABC). Mapping experiments require the development of large (i.e., >200 families) populations, which can be used for genotyping and phenotyping. Development and maintenance of large populations require significant resources including both labor force and field space. This is particularly true for species with large generation time and space requirements. The *in vitro* nursery system could allow for the quick and efficient development of cell lines that can be subsequently stored and/or converted into plants to be used for phenotyping and/or production.

In MABC the ultimate goal is to transfer a gene of interest into an existing cultivar/line. One main challenge of MABC is, to remove unfavorable alleles of closely linked genes, i.e., to eliminate linkage drag, particularly in the case of exotic introgressions. Thus, multiple individuals need to be evaluated, which is costly and requires a significant amount of resources. MABC programs could alternatively be conducted within *in vitro* nurseries. Large numbers of individuals could be generated

within a controlled laboratory setting and evaluated using markers. This would allow rapid and efficient introgression of genes of interest. The utility of this system becomes increasingly superior, as the number of loci to be introgressed increases [13].

Another application of an *in vitro* nursery would be to overcome self incompatibility (SI), which is present in many cultivated species [34-37]. In order to successfully produce single cross hybrids in SI crops, breeders must be able to generate homogeneous and homozygous parental inbred lines to produce the hybrid. The generation of these inbred lines is impossible in case of SI. This process, however, occurs through the interaction of pollen tubes with stigma [38]. In the *in vitro* nursery, this pollination stage can be bypassed and gametes can be fused directly, thus overcoming the issue of SI. We envisage a system, where somatic cells of these species are used to generate gametes which could be subsequently fused with gametes from the same cell (simulating self-pollination) or artificially doubled simulating the DH process to generate homozygous and homogeneous lines that can be used subsequently to generate hybrids. This idea can be taken one step further. Gametes from selected homozygous and homogeneous lines could be fused *in vitro* to generate zygotes which are the desired hybrid combination. This process could be combined with the development of synthetic seeds where somatic embryos are encapsulated to form artificial seed, which can be packaged and distributed to growers similar to a normal seed. Successful germination of artificial seed generated from somatic tissue has been demonstrated in species such as alfalfa (*Medicago sativa*) [39], cyclamen (*Cyclamen persicum* Mill.) [40], and salparni (*Desmodium gangeticum* L.) [41].

The utility of this system is more beneficial for plant species with long generation times such as those of the genus *Leucaena*, which can take up to 2 years to flower [42], pecan (*Carya illinoensis*) which flower at 6-7 years of age [43], and other woody species. However, its utility could also extend to species which normally require vernalization or a chilling cycle to induce flowering such as peaches (*Prunus persica*) [44], and wheat [45] as a method to overcome these requirements and produce new sexual progeny at any time in the year. Apomictic species for which recovery of sexually generated populations to be used as variation for selection is difficult may also benefit such as those of citrus species [46]. Finally annual crops, such as maize and soybean, could also benefit through rapid generation of new populations for selection and line conversion. For example, consider the time and expenses used to convert new breeding lines of maize and soybean into those which contain desirable genes for resistance to a pathogen or transgenes. This process which works in tandem with line development can require up to six seasons to produce a suitable BC4 line and assuming 3 seasons per year would take two years to complete. Using the proposed *in vitro* nursery, this process could be shortened to 257 days assuming one week per cycle and 215 days to regenerate a fertile plant. The savings will not only be in time, but also in cost of land, seed shipment to off season nurseries, labor, and a smaller number of lines converted.

Challenges

The purpose of this manuscript is to combine recent advances in different fields of biology and conceptualize a technique that could substantially advance efficiency of plant breeding, once becoming available. The idea of an *in vitro* nursery presented in the

previous sections, while new and innovative, does have obvious problems and gaps at current. The first, and most important, is the ability to stimulate meiosis and to generate gametes *in vitro*. Recent advances in both plant and animal models provide insight into gamete formation *in vitro*. For animals, the production of egg cells *in vitro* has been reported [47], as well as the successful production of artificial gametes in mice [48]. The first study required the use of stem cells, whereas the second used testicular tissue and thus, did not induce gametes from purely somatic cells. A recent review [49] outlined current advances in development of artificial gametes in animals and the significant obstacles that remain. The authors note, that the knowledge needed to generate functional germ cells *in vitro* exists, but the methodology is in its infancy [49]. In contrast to animals, whose germ lines are established early in development, plants specify germ lines later in development and can have multiple germ lines [50]. For example, a hypoxic environment causes any cell in an early maize anther to convert to a germ cell [51]. It is currently unclear if recombination is occurring, though it is likely since meiosis is induced, and more research is needed to confirm. The genetic mechanisms which underlie the control of plant meiosis are being elucidated and research is ongoing with practical applications, including the *in vitro* nursery, across the plant sciences. The complexity and breadth of the research in this field is beyond the scope of this manuscript, but the reader is referred to [52] for an up to date description of the latest breakthroughs. This provides an initial framework for producing gametes *in vitro* from somatic cells. Like most techniques in biology, it is likely that this process will not be 100% efficient. There would therefore be a need to distinguish between haploid and diploid cells, which may not be trivial.

The use of the *in vitro* nursery will also require the continued advancement of predictive tools that can be used in genomic selection schemes. This research is not specific to applications for the *in vitro* nursery as it would also assist current breeding programs.

An array of issues still remains with this proposed idea. Growing cells in tissue culture can generate genotype dependencies [53-55] and the use of the *in vitro* nursery will cause unintended selection for loci, which control success of cell culture. Genotype-dependency of regeneration is the major challenge in tissue culture techniques [56-58]. However, genes or QTL for regeneration in tissue culture have been identified [59, 60] and can help to overcome this bottleneck. Recent reports show that targeting young zygotes or isolating cells during the early callus phase for plant regeneration has less genotype dependency than those which are allowed to go through a callus growth phase and are then regenerated [61-63]. Another issue is the phenomenon of somaclonal variation. When plants are grown *in vitro*, stress induces changes in regenerated plants. Somaclonal variation can provide useful variation [64]. In the *in vitro* nursery changes due to somaclonal variation, such as activation of transposable elements, can counteract generation of homogeneous and homozygous lines.

Despite these challenges, a major benefit would be a larger number of generations per year with the potential to increase the rate of genetic gain which in turn may increase the rate at which the mean yield of crops improves (Figure 1).

References

- 1 Duvick, D.N. (2005) The contribution of breeding to yield advances in maize (*Zea mays* L.). In *Advances in Agronomy* (Donald, L.S., ed), pp. 83-145, Academic Press
- 2 Lopes, M.S., et al. (2012) Genetic yield gains and changes in associated traits of CIMMYT spring bread wheat in a “historic” set representing 30 years of breeding. *Crop Sci.* 52, 1123-1131
- 3 USDA-NASS (2013) Field crops statistics.
- 4 Brummer, E.C., et al. (2011) Plant breeding for harmony between agriculture and the environment. *Front. Ecol. Environ.* 9, 561-568
- 5 Tilman, D., et al. (2011) Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci.* 108, 20260-20264
- 6 Fehr, W.R. (1993) *Principles of Cultivar Development*. McGraw-Hill
- 7 Bernardo, R. (2010) *Breeding for Quantitative Traits in Plants*. Stemma Press
- 8 Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*. Pearson
- 9 Hallauer, A.R., et al. (2010) *Quantitative Genetics in Maize Breeding*. Springer
- 10 Borlaug, N.E. (2007) Sixty-two years of fighting hunger: Personal recollections. *Euphytica* 157, 287-297

- 11 Ferrie, A.M.R. and Möllers, C. (2010) Haploids and doubled haploids in *Brassica* spp. For genetic and genomic research. *Plant Cell Tissue Organ Cult.* 104, 375-386
- 12 Geiger, H.H. (2009) Doubled haploids In *Handbook of Maize* (Bennetzen, J.L. and Hake, S., eds), pp. 641-657, Springer New York
- 13 Lübberstedt, T. and Frei, U.K. (2012) Application of doubled haploids for target gene fixation in backcross programmes of maize. *Plant Breed.* 131, 449-452
- 14 Nanda, D.K. and Chase, S.S. (1966) An embryo marker for detecting monoploids of maize (*Zea mays* L.). *Crop Sci.* 6, 213-215
- 15 Prigge, V., et al. (2011) Doubled haploids in tropical maize: I. Effects of inducers and source germplasm on in vivo haploid induction rates. *Crop Sci.* 51, 1498-1506
- 16 Prigge, V., et al. (2012) New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize. *Genetics* 190, 781-793
- 17 Kebede, A.Z., et al. (2011) Effect of source germplasm and season on the in vivo haploid induction rate in tropical maize. *Euphytica* 180, 219-226
- 18 Germanà, M.A. (2010) Anther culture for haploid and doubled haploid production. *Plant Cell Tissue Organ Cult.* 104, 283-300
- 19 Castillo, A.M., et al. (2009) Chromosome doubling in monocots. In *Advances in Haploid Production in Higher Plants* (Touraev, A., et al., eds), pp. 329-338, Springer Netherlands

- 20 Prigge, V. and Melchinger, A. (2012) Production of haploids and doubled haploids in maize. In *Plant Cell Culture Protocols* (Loyola-Vargas, V.M. and Ochoa-Alejo, N., eds), pp. 161-172, Humana Press
- 21 Bhojwani, S. and Dantu, P. (2013) Tissue and cell culture. In *Plant Tissue Culture: An Introductory Text*, pp. 39-50, Springer India
- 22 Godoy-Hernández, G. and Vázquez-Flota, F. (2012) Growth measurements: Estimation of cell division and cell expansion. In *Plant Cell Culture Protocols* (Loyola-Vargas, V.M. and Ochoa-Alejo, N., eds), pp. 41-48, Humana Press
- 23 Kranz, E. and Lorz, H. (1993) *In vitro* fertilization with isolated, single gametes results in zygotic embryogenesis and fertile maize plants. *Plant Cell* 5, 739-746
- 24 Kranz, E. and Scholten, S. (2008) *In vitro* fertilization: Analysis of early post-fertilization development using cytological and molecular techniques. *Sexual Plant Reproduction* 21, 67-77
- 25 Bhojwani, S. and Dantu, P. (2013) *In vitro* pollination and fertilization. In *Plant Tissue Culture: An Introductory Text*, pp. 155-171, Springer India
- 26 Faure, J.-E., et al. (1994) An *in vitro* system for adhesion and fusion of maize gametes. *Science* 263, 1598-1600
- 27 Ponya, Z., et al. (2004) Structural change in the endoplasmic reticulum during the in situ development and *in vitro* fertilisation of wheat egg cells. *Sexual Plant Reproduction* 17, 177-188

- 28 Uchiumi, T., et al. (2007) Establishment of an *in vitro* fertilization system in rice (*Oryza sativa* L.). *Planta* 226, 581-589
- 29 Sun, M.-X., et al. (2001) *In vitro* double fertilization in *Nicotiana tabacum* (L.): The role of cell volume in cell fusion. *Sexual Plant Reproduction* 13, 225-229
- 30 Woyke, T., et al. (2011) Decontamination of MDA reagents for single cell whole genome amplification. *PLoS ONE* 6, e26161
- 31 Treff, N.R., et al. (2010) Accurate single cell 24 chromosome aneuploidy screening using whole genome amplification and single nucleotide polymorphism microarrays. *Fertility and Sterility* 94, 2017-2021
- 32 Xu, P., et al. (2011) An optimization approach to gene stacking. *European Journal of Operational Research* 214, 168-178
- 33 Greenway, M.B., et al. (2012) A nutrient medium for diverse applications and tissue growth of plant species *in vitro*. *In vitro Cell. Dev. Biol.-Plant* 48, 403-410
- 34 Iwano, M. and Takayama, S. (2012) Self/non-self discrimination in angiosperm self-incompatibility. *Curr. Opin. Plant Biol.* 15, 78-83
- 35 Sun, P. and Kao, T.H. (2013) Self-incompatibility in *Petunia inflata*: The relationship between a self-incompatibility locus f-box protein and its non-self s-rnases. *Plant Cell* 25, 470-485
- 36 Chantha, S.-C., et al. (2013) Secondary evolution of a self-incompatibility locus in the Brassicaceae genus *Leavenworthia*. *PLoS Biol* 11, e1001560

- 37 De Nettancourt, D. (1972) Self-incompatibility in basic and applied researches with higher plants. *Genet. Agr.* 1-2, 163-216
- 38 Kaothien-Nakayama, P., et al. (2010) Self-incompatibility systems in flowering plants. In *Plant Developmental Biology - Biotechnological Perspectives* (Pua, E.C. and Davey, M.R., eds), pp. 459-485, Springer Berlin Heidelberg
- 39 Liu, W., et al. (2013) Preservation and faithful expression of transgene via artificial seeds in alfalfa. *PLoS ONE* 8, e56699
- 40 Winkelmann, T., et al. (2004) Germination of encapsulated somatic embryos of *Cyclamen persicum*. *Hort. Sci.* 39, 1093-1097
- 41 Puhan, P. and Rath, S.P. (2012) Induction, development and germination of somatic embryos from *in vitro* grown seedling explants in *Desmodium gangeticum* L.: A medicinal plant. *Res. J. Med. Plant* 6, 346-369
- 42 Brewbaker, J.L. (2008) Registration of KX2-Hawaii, interspecific-hybrid leucaena. *J. Plant Reg.* 2, 190
- 43 Thompson, T. and Conner, P. (2012) Pecan. In *Fruit Breeding* (Badenes, M.L. and Byrne, D.H., eds), pp. 771-801, Springer US
- 44 Byrne, D., et al. (2012) Peach. In *Fruit Breeding* (Badenes, M.L. and Byrne, D.H., eds), pp. 505-569, Springer US

- 45 Chen, A. and Dubcovsky, J. (2012) Wheat tilling mutants show that the vernalization gene *vrn1* down-regulates the flowering repressor *vrn2* in leaves but is not essential for flowering. *PLoS Genetics* 8, e1003134
- 46 Ollitrault, P. and Navarro, L. (2012) Citrus. In *Fruit Breeding* (Badenes, M.L. and Byrne, D.H., eds), pp. 623-662, Springer US
- 47 Cheloufi, S. and Hochedlinger, K. (2012) Reproductive biology: Stem cells bear eggs. *Nature* 491, 535-536
- 48 Sato, T., et al. (2011) *In vitro* production of functional sperm in cultured neonatal mouse testes. *Nature* 471, 504-507
- 49 Kashir, J., et al. (2012) Viability assessment for artificial gametes: The need for biomarkers of functional competency. *Biology of Reproduction* 87, 114
- 50 Whipple, C. (2012) Defining the plant germ line-nature or nurture? *Science* 337, 301-302
- 51 Kelliher, T. and Walbot, V. (2012) Hypoxia triggers meiotic fate acquisition in maize. *Science* 337, 345-348
- 52 Wijnker, E. and Schnittger, A. (2013) Control of the meiotic cell division program in plants. *Plant Reproduction* 26, 143-158
- 53 Karp, A. (1994) Origins, causes and uses of variation in plant tissue cultures. In *Plant Cell and Tissue Culture* (Vasil, I. and Thorpe, T., eds), pp. 139-151, Springer Netherlands

- 54 Sharma, V., et al. (2004) A highly efficient plant regeneration system through multiple shoot differentiation from commercial cultivars of barley (*Hordeum vulgare* L.) using meristematic shoot segments excised from germinated mature embryos. *Plant Cell Reports* 23, 9-16
- 55 Simmonds, D.H. and Donaldson, P.A. (2000) Genotype screening for proliferative embryogenesis and biolistic transformation of short-season soybean genotypes. *Plant Cell Reports* 19, 485-490
- 56 Anami, S., et al. (2010) Somatic embryogenesis and plant regeneration of tropical maize genotypes. *Plant Cell Tissue Organ Cult.* 102, 285-295
- 57 Khan, T., et al. (2010) High-frequency regeneration via somatic embryogenesis of an elite recalcitrant cotton genotype (*Gossypium hirsutum* L.) and efficient *Agrobacterium*-mediated transformation. *Plant Cell Tissue Organ Cult.* 101, 323-330
- 58 Targońska, M., et al. (2013) The specificity and genetic background of the rye (*Secale cereal* L.) tissue culture response. *Plant Cell Reports* 32, 1-9
- 59 Krakowsky, M.D., et al. (2006) Quantitative trait loci for callus initiation and totipotency in maize (*Zea mays* L.). *Theor. Appl. Genet.* 113, 821-830
- 60 Nishimura, A., et al. (2005) Isolation of a rice regeneration quantitative trait loci gene and its application to transformation systems. *Proc. Natl. Acad. Sci. USA* 102, 11940-11944

- 61 Holme, I., et al. (2012) Transformation of barley (*Hordeum vulgare* L.) by *Agrobacterium tumefaciens* infection of *in vitro* cultured ovules. In *Transgenic Plants* (Dunwell, J.M. and Wetten, A.C., eds), pp. 151-161, Humana Press
- 62 Parmar, S., et al. (2012) Plant regeneration from mature embryo of commercial indian bread wheat (*Triticum aestivum* L.) cultivars. *Physiol. Mol. Biol. Plants* 18, 177-183
- 63 Ebrahimie, E., et al. (2006) Direct shoot regeneration from mature embryo as a rapid and genotype-independent pathway in tissue culture of heterogeneous diverse sets of cummin (*Cuminum cyminum* L.) genotypes. *In vitro Cell. Dev. Biol.-Plant* 42, 455-460
- 64 Bhojwani, S. and Dantu, P. (2013) Somaclonal variation. In *Plant Tissue Culture: An Introductory Text*, pp. 141-154, Springer India

Box 1. Genetic gain: breeders equation

The objective of plant breeding is the identification and development of superior individuals and families. The mean performance of breeding populations is increased through selection of individual plants with higher than average performance. This change in mean performance of the breeding population can be expressed as genetic gain in different forms, depending on the situation [6].

Genetic gain per cycle:

$$(I) G_c = kch^2\sigma_P$$

$$(II) G_c = \frac{kc\sigma_A^2}{\sigma_P} \quad \text{where } h^2 = \frac{\sigma_A^2}{\sigma_P^2}$$

As seen in equation I in the case of one cycle of selection, k is the selection differential expressed in standard deviation units, representing the percentage of individuals selected and advanced to the next generation. The degree of parental control (i.e., genetic control of males, females, both sexes) is quantified in c . Narrow sense heritability (h^2) is a measure of what proportion of phenotypic variance (σ_P^2) can be explained by additive genetic variance (σ_A^2). Equation II can be derived by substituting σ_A^2 / σ_P^2 for heritability. The additive genetic variance is the component of the genetic variance that is transmitted to the progeny (except in polyploids where some dominance variance is transmitted and in clonal breeding, where all genetic variance is transmitted).

Different selection schemes (e.g., half-sib, full-sib, selfed families) require different numbers of seasons to complete a full selection cycle. For comparison of alternative breeding schemes, the calculation of genetic gain per year is more informative than gain per cycle. This is achieved by dividing equation II by the number of years (y) required per cycle.

Genetic gain per year:

$$(III) G_y = \frac{kc\sigma_A^2}{y\sigma_P}$$

Equation III can be expanded further for specific situations, when different environments and replications are used and to quantify variance that is contained within and among families in the selection scheme. These expansions are beyond the scope of this article, the reader is referred to [6] for an in depth discussion of the different forms of the genetic gain equations.

By modifying the components in equation III, breeders are able to maximize genetic gain. Some components are simpler to manipulate than others. This article focuses on the management of time (expressed as y) as a method to maximize genetic gain.

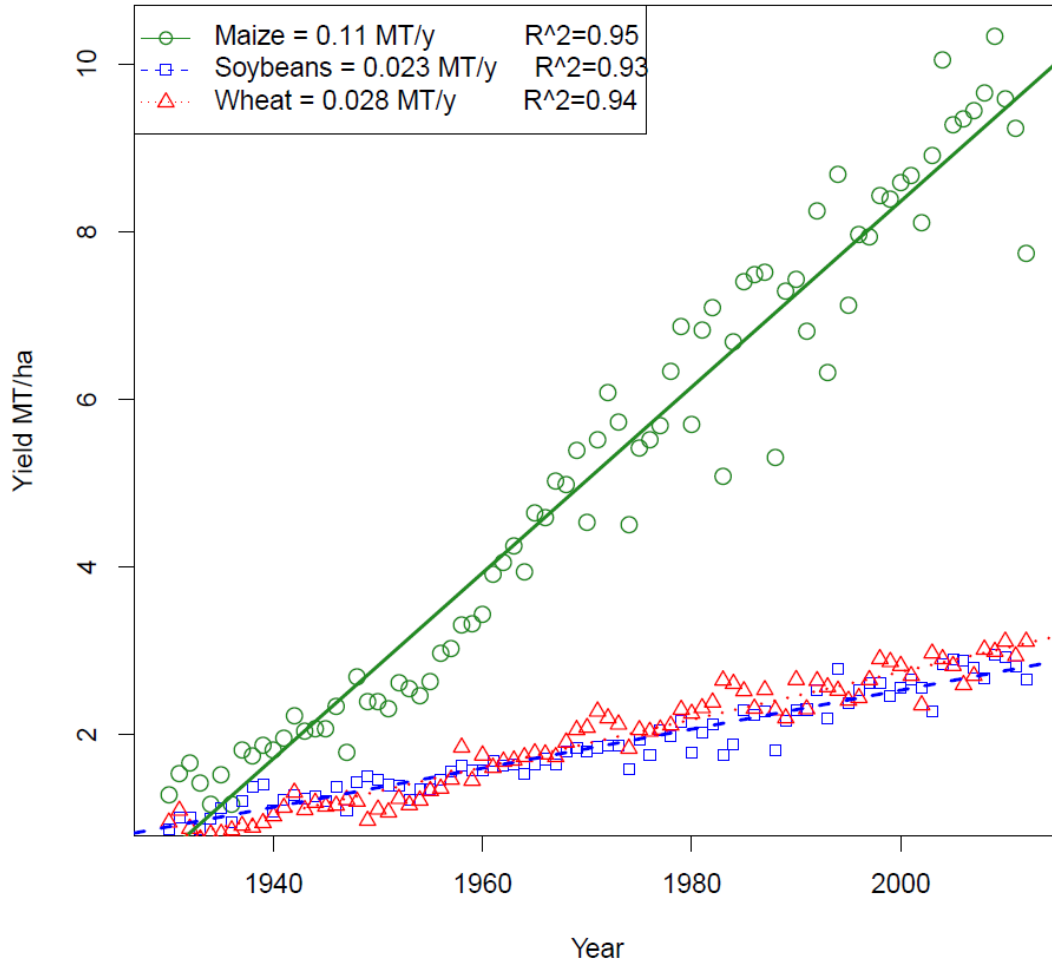


Figure 1. Yield gains of major U.S. crops. Average yield per year in metric tons/ha (MT/ha) for each of the three major U.S. crops (maize, wheat, and soybean) from 1930-2012 [3]. Each crop shows a linear increasing trend over time with maize having the highest annual gain of 0.11 MT/year followed by wheat at 0.028 MT/year and soybeans at 0.023 MT/year for average grain yield. This increase in mean yield per hectare needs to be increased in order to meet the demands of a growing human population.

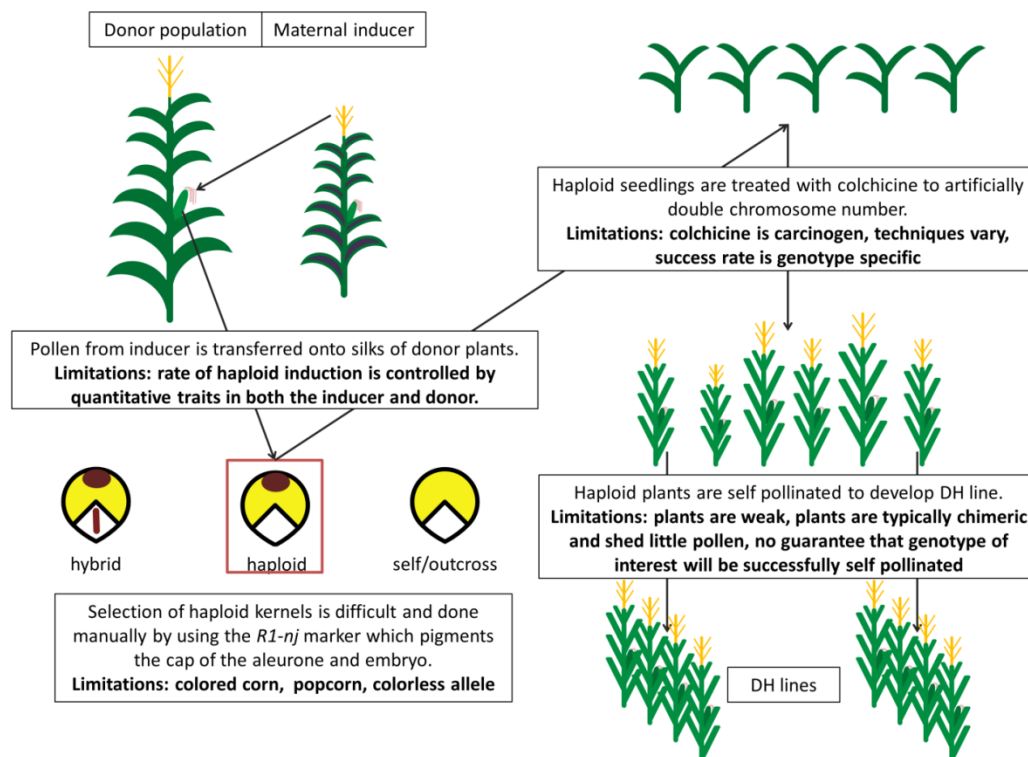


Figure 2. The maize doubled haploid system. Maize doubled haploid (DH) technology is a specific example of DH technology used with great success by public and private plant breeders to shorten the time it takes to generate a homozygous line from eight to two seasons. This is arguably the latest major breakthrough in cycle time (a reduction in years per cycle: see Box 1). Though DH technology is used with success in maize and other crop species, there are limitations as are noted in the figure. The rate of haploid induction is genetically controlled by quantitative trait loci (QTL) in both the inducer and donor population. The *R1-nj* [14] marker allele used to identify haploid kernels is useless, if the kernels are colored or if they carry the colorless allele. Rates of doubling in haploid plants are typically low and highly dependent on both technique and genotype. The doubling agent, colchicine, is a carcinogen. Those plants that successfully double their

genomes typically shed little pollen and there is no guarantee that the optimal genotype will set seed and advance to testing. For now, the benefits of time savings outweigh the drawbacks.

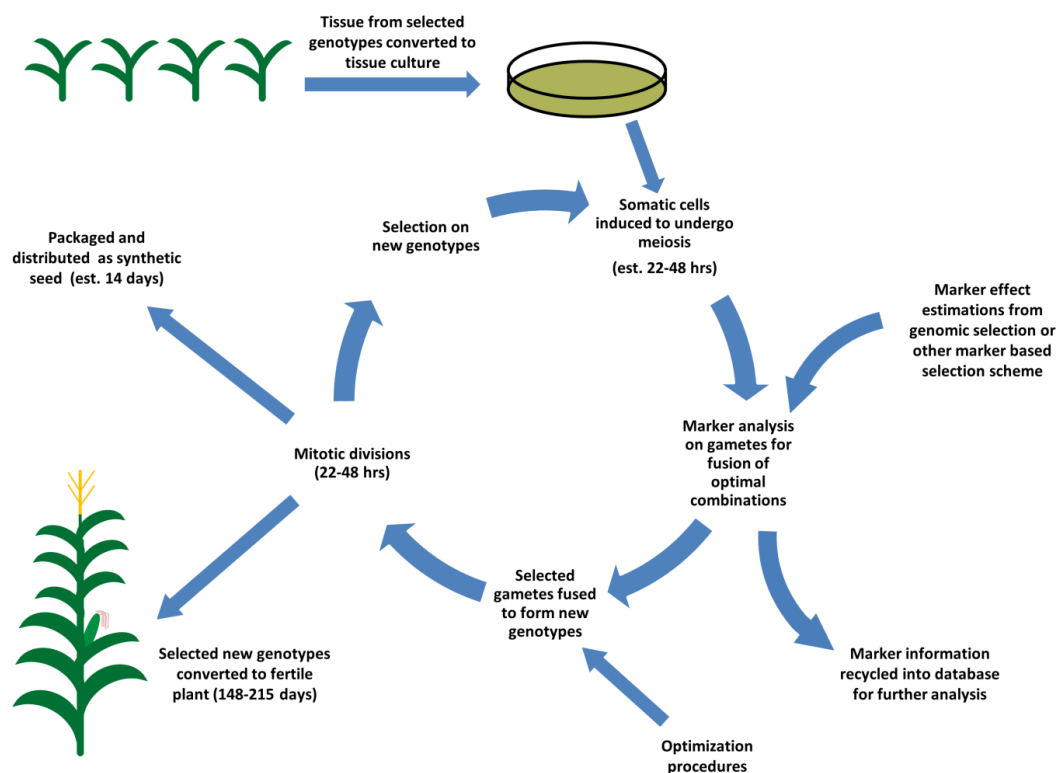


Figure 3. The *in vitro* nursery. The general scheme of the *in vitro* nursery. First, tissue from selected genotypes must be extracted and converted into a tissue culture. A genotype dependency for tissue culture conversion and success is likely. Once the somatic cells have stabilized in culture, they are induced to undergo meiosis. After gametes are formed, they are allowed mitotic cycles which lead to clonal cells, so that DNA can be extracted from some of those cells for marker analyses. Marker effect estimation based on genomic selection, marker-assisted backcrossing, or marker assisted selection are incorporated. Optimization procedures can then be incorporated to make the stacking of optimal loci as efficient as possible. Optimal gametes are then selected and fused to form a new diploid individual. Mitotic divisions are required to enable DNA extraction. At this junction, selected new genotypes can either be converted into fertile

plants or into synthetic seed for phenotypic evaluation. The cell line can then be immediately recycled in the nursery and induced to form new gametes, in order to complete the cycle.

CHAPTER SEVEN

GENERAL CONCLUSIONS AND THE FUTURE OF MATERNAL *IN VIVO* DOUBLED HAPLOIDS

In the previous chapters the reader has been given, first, an introduction to the progression of maize breeding from early research to modern commercial programs and a description of the maternal *in vivo* DH system and problems associated with it. As previously mentioned modern maize breeding programs extensively utilize the *in vivo* maternal DH system for production of new inbred lines for testing. Maize breeders heavily rely on the ability to quickly adapt to new selection targets by utilizing the speed available through this DH system. Through Chapters 2-6, the reader is taken through the normal progression of the DH system: 1) generation of a donor population which is suitable for production of new inbreds through sufficient haploid progeny (Chapter 2), 2) use of maternal inducer lines for production of haploid progeny on the donor population from step 1 (Chapter 3), 3) efficient and accurate selection of haploid progeny (Chapter 4), and 4) doubling of the haploid genome for generation of homozygous diploid lines (Chapter 5). Finally, the reader is presented with the potential for the next advancement in breeding cycle speed in Chapter 6.

Results from the studies included in this dissertation show promise for the ability to continue to improve the maize DH system to make maize breeding more efficient. Conclusions presented from the experiments on inducibility suggest that improvement of the rates of induction on the donor population is possible through selective breeding. This is in agreement with published literature in other maize germplasm as well. Utilization of

improved lines for inducibility can provide the potential to expand the germplasm base of maize breeding programs by including otherwise recalcitrant material in DH breeding programs. Considering now, the other side of the induction cross, the maternal inducers, the release of BHI306 maternal haploid inducer will expand the pool of germplasm that can be used for DH line development. The ability of BHI306 to use multiple selectable markers (*RI-nj* and *pl1*) and its ability to pollinate popcorn germplasm will allow breeders of specialty maize types to utilize DHs in their line development programs. BHI306 also provides an improvement in agronomic traits, especially germination rates, over the existing inducer lines available in the temperate U.S.

After the induction cross is made, next, the seed must be sorted to identify the haploid progeny. In Chapter 4 the description of a potential automated system for haploid/diploid discrimination is included. Results show that the Videometer Lab 3 system has the potential to discriminate between diploid and haploid seed with greater than 50% accuracy in most cases. The expression of *RI-nj* plays a role in the accuracy of the system, but as the results indicated, the optical system corrected human sorting errors, and even if the system cannot sort at 100% accuracy, enrichment of the sample for haploid progeny is a desirable outcome.

Once the haploid progeny have been identified, typically, they would undergo a very laborious and expensive process of germination, colchicine treatment, and transplanting into the field for self-pollination. Results from Chapter 5 show that the potential to utilize spontaneous chromosome doubling (SCD) for the development of DH

lines is possible. Combinations of SCD lines with non-SCD lines allowed for the production of DH lines without colchicine treatment. Rates of doubling above 20% allow for the effective use of SCD for production of haploids. The amount of labor, space, and money saved from utilizing SCD could have the potential to restructure maize breeding programs by freeing up resources from DH line development to be used for testing and evaluation thereby increasing the ability of maize breeders to find favorable genotypes with the same budget.

Finally, the concept of the *in vitro* nursery, while not directly related to DHs, is indirectly related to DHs through the concept of speed. The purpose of the DH system is the acceleration of breeding programs, which is the main objective of the concept of the *in vitro* nursery. The ability to make selections at the gamete level is similar to the ability to select at the haploid level which was discussed with the use of SCD in Chapter 5. However, in haploid plants, breeders are still limited by the life cycle of the plant which is relatively long compared to the life cycle of a cell. This concept will require a great deal of future research, but essentially the only aspect that has not been considered is the induction of meiosis. That will not be an easy hurdle to cross, but neither was the development of the maize DH system.

While the preceding chapters provided a great deal of new information and potential avenues for advancement, like all research they also raise a new set of questions. The ultimate objective of the previous studies is the practical application of the results to an applied maize breeding program. In order to do this, the traits of inducibility

and SCD must be further evaluated. Towards this, during the development of the experiments described here, additional line development and population development was conducted for future experiments and evaluation of the potential of these topics to be used in applied maize breeding. Crosses with elite germplasm have been made for the evaluation of new lines containing both SCD and inducibility potential with improved agronomic characteristics. These traits will need to be evaluated for any detrimental effects they may have on yield and other important agronomic characteristics needed in any maize hybrid. Crosses have also been made for the development of mapping populations for the further understanding of the genetic makeup of these traits so that they can be better utilized and understood in maize, and potentially found/transferred to other crop species. Automation of haploid selection would also reduce cost to breeding programs and allow for, potentially, more accurate sorting of haploid progeny. Development of automated systems for sorting is currently underway.

Taken individually, the conclusions of the chapters in this dissertation each provide avenues for improvement to the specific steps of the DH process. However, what is important to consider is that taken together the conclusions of the chapters presented above provide a whole process improvement to the DH system by contributing understanding and time/cost savings to the system as a whole. Improved inducers and rates of induction on the donor will decrease the number of kernels needed to be sorted saving time and money. The ability to use an automated system for the sorting of haploid and diploid seed will also support the previous step by saving time and money in the sorting process. Once sorted, the ability to use SCD to eliminate the steps of colchicine

application and transplanting will save a large amount of time and money which will further improve the process. Overall, the conclusions of each chapter provide the potential to increase the efficiency and applicability of the DH system in maize. Breeders are encouraged to utilize the germplasm identified in this dissertation for the integration of increased rates of induction and the ability to use SCD for DH line development.

Maize breeding has come a long way from mass selection conducted by our ancestors that domesticated it and made it the crop we know today. Continued efforts over the years progressing through open pollinated varieties to four way hybrids, all the way to the modern single cross hybrids have made maize a staple crop worldwide which continues to be grown on more area every year. A few decades ago, DHs were the new technology on the horizon and improvements to the DH system have been presented here. These improvements have the potential to change the way maize haploids and DHs are used, in a more efficient and effective manner. The question remains, how long will it take before the next breakthrough arrives, could the *in vitro* nursery be that answer.

ACKNOWLEDGEMENTS

I would like to express my thanks to all those who helped me through my time as a doctoral student, both for academic and non-academic support. First, my thanks go to my advisor Dr. Thomas Lübberstedt. His patience and dedication are remarkable. Second, though she was not officially on my committee, Dr. Uschi Frei provided a great deal of help, guidance, and support to me throughout my time as a student. Dr. Frei and Dr. Lübberstedt allowed me a great deal of freedom which allowed me to develop my own experimental goals as well as help with the development of the lab's field program. I would also like to thank my committee, Dr. Dan Nettleton, Dr. Erik Vollbrecht, Dr. Paul Scott, and Dr. Mike Blanco for their support and guidance on so many subjects, you are all outstanding scientists and I appreciate your dedication to my education. I would also like to thank Mrs. Elizabeth Bovenmyer and Mr. Paul White who supported my research in many ways, your hard work, advice, and patience is much appreciated. Thanks also to Mr. Mike Fiscus, the director of the Iowa State Agronomy Farm and his entire team, my field trials and those of the other students were outstanding thanks to your help. Also, to my fellow graduate students, I appreciate all the help both in school and out. I would also like to thank my Ames area friends especially the Aitchison family, I would not have been able to endure three years of solitude without your welcoming me as family. Finally, I must thank my family. I cannot describe in words the amount of support I received from my entire family; I would not have made it through without it. This is especially true of my loving wife Jae. She endured three years of us being 1000 miles apart during my studies and yet still had the love and dedication to support me every day. For all those who believed in me, thank you.