2019

Assessment of inducibility and spontaneous haploid genome doubling in maize (Zea mays L.)

Benjamin Thomas Trampe
Iowa State University

Assessment of inducibility and spontaneous haploid genome doubling in maize (*Zea mays* L.)

by

Benjamin Thomas Trampe

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
Jianming Yu
Marna Yandeau-Nelson
Paul Scott
Jessica Barb

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

DEDICATION

To my late father who always had confidence in me and offered me encouragement and support in all my endeavors.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS

I would like to thank all the wonderful individuals who have assisted and supported me along this journey. I want to highlight a few of those individuals. First, I would like to thank Thomas Lübberstedt for the outstanding opportunity and his patience. Second, I would like to thank each of my committee members for your expertise and suggestions along the way. Next, I want to thank you, Kolin, LaNae, and Justin. All of you have been very encouraging and supportive during this adventure of my life. Finally, I couldn't do it without my family for their gracious love and support.

ABSTRACT

Maize is a staple food, fuel, and feed crop grown around the world. Doubled haploid technology allows for the quick of development of inbred lines for hybrid development. The maternal *in vivo* doubled haploid system has gained rapid adoption by the maize breeding sector within the last 10 years. There have been significant improvements in the doubled haploid technology, which made it commercially viable. Within the doubled haploid system, there is limited genetic information about the two important traits that control the ability of generating doubled haploids, which are inducibility and spontaneous haploid genome doubling. Better understanding of these two traits could drastically improve the efficiencies and reduce labor needs for producing doubled haploid lines.

In this dissertation, the genetic control of both inducibility and spontaneous haploid genome doubling were studied. A Quantitative Trait Loci (QTL) mapping study was conducted for both traits using an $F_{2:3}$ population derived from inbred A427 and CR1Ht. Inducibility QTL were identified and the improvement of inducibility is examined. A major QTL was found for spontaneous haploid genome doubling and its application to doubled haploid breeding is discussed.

CHAPTER ONE: GENERAL INTRODUCTION

Maize and Doubled Haploids

Maize is one of the essential feed, fuel, and food crops grown worldwide. Systematic hybrid breeding has increased the productivity of novel hybrid combinations[1]. To produce those hybrid combinations, elite inbreds are developed in divergent heterotic pools. Traditionally, maize has undergone subsequent generations of selfing (pedigree method) to produce inbred lines. The challenges of the pedigree method include considerable record keeping, the need to select in only one or few environments, and land and labor requirements. Advancement in technology allowed the efficient production of doubled haploid (DH) lines[2]. A doubled haploid (D0) plant is created when a haploid (n) individual undergoes a genome doubling event, to generate a DH line[3] after self-pollination. This DH line (DHL) is genetically completely homogeneous, with plants that are 100% homozygous across all loci. This technology allows faster production of inbred lines by reducing cycle time to two generations from the six to eight generations of inbreeding in traditional inbred development[4].

Production of DHLs in maize includes in vitro[5] and in vivo[6] methods. In vitro DH production is based on the use of gametophytic cells in culture[7]. In vitro maize DH production has been investigated. However, most germplasm is recalcitrant and non-responsive to haploid plant regeneration in tissue culture[8]. In vivo DH production is based on interspecific or intraspecific hybridization depending on the crop species. Maize uses an intraspecific hybridization system via haploid inducers. In vivo DH production is commonly used for the creation of DHLs due to the reliance on the donor germplasm compared to in vitro DH production. Within in vivo DH

production there are two different approaches for haploid production, paternal and maternal[9]. In paternal haploid induction, the inducer is used as the female parent. Paternal induction is controlled by the ig1 (indeterminate gametophyte 1) gene, which increases the production of haploids[10]. Paternal haploids contain the haploid genome from the pollen parent and the cytoplasm of the haploid inducer. In maternal haploid induction, the inducer is used as the male parent. Maternal haploids contain both the haploid genome and the cytoplasm from the female parent. Maternal haploid induction is the most frequently used approach to produce DHLs in maize breeding programs, because of substantially higher haploid induction rates[9].

The three main steps of in vivo maternal DH production are haploid induction, identification of haploid kernels, and genome doubling of subsequent haploids. Each of the steps has been enhanced to make DH production a viable method of creating homozygous and homogenous lines that are ready to be used for hybrid production within maize.

The first step of the in vivo maternal haploid DH system is haploid induction. Donor heterozygous germplasm (female) can be referred to as any breeding population, which is used for generation of DHLs. Haploid induction is achieved by pollinating donor germplasm with a haploid inducer. In 1959, Coe discovered a genotype that was able to induce haploids, named Stock 6[11]. Stock 6 became the progenitor of all current inducer lines. Improvements in the induction rates of inducers made the DH system more efficient by increasing percentage of haploids from donor germplasm. For example, the Procera Haploid Inducers (PHI) have induction rates of 12.1% to 14.5%[12]. A sperm-specific phospholipase, MATRILINEAL, is a major gene affecting haploid induction, which was identified by map-based gene isolation[13]. The improvements in inducers has made the maternal in vivo haploid system economically viable on

a commercial scale. Moreover, it has been shown that there is genetic variation among donors with regard to their response to haploid induction, called inducibility[4].

The second step is haploid identification, which also is essential to make DH technology commercially viable. The objective is to differentiate haploid kernels from regular fertilized diploid kernels. Early haploid identification reduces the costs and labor in later steps. Several approaches have been evaluated or used for haploid selection, including the use of color markers[14] and various automated selection techniques[15]. Kernel-based selection using R1-Navajo (*R1-nj)* [16] is the most widely used selection technique for haploid identification. This dominant gene present in maternal haploid inducers is expressed in the embryo and endosperm of regular fertilized kernels originating from the cross between donor germplasm and inducer. In contrast, kernels with a haploid embryo from the same cross show anthocyanin expression in the endosperm, but no embryo coloration[16]. Anthocyanin expression can vary based on donor background. If the expression is low, it can make identification of haploids difficult, increasing the amount of time needed for haploid selection and the number of false positives. False positives reduce the efficiency of the DH system and increase the costs to develop a new population of DH lines. Some germplasm maybe recalcitrant to *R1-nj*, due to the masking of anthocyanin expression by the *C1-I* allele [17].

The third step is genome doubling in haploid individuals. Haploid fertility was found to be very low, with rates reported as low as 0.41%[18]. Thus, increase in genome doubling capability is needed to improve haploid fertility. Genome doubling is a critical step, since it is the most limiting step in producing DH lines. Genome doubling is typically induced using artificial doubling methods. Colchicine is customarily used to double haploid genomes artificially, which increases

the rate of fertile haploid flower structures[19]. Colchicine duplicates genomes by prohibiting the formation of microtubules during meiosis[20]. Once haploids are treated with colchicine, they are transplanted into the field. To generate DH lines, both male and female germ lines need to have doubled their genomes for production of fertile flower structures and selfing seed production. An alternative to artificial genome doubling is spontaneous haploid genome doubling. An example of this is the first division restitution (*fdr1)* mutation, which increases haploid male fertility[21]. DH lines are created within two generations, which simplifies logistics and improved selection efficiency[22].

QTL Mapping

Linkage mapping depends on population developed for genetic mapping. Mapping populations for linkage mapping include F_2, recombinant inbred line (RIL), doubled haploid, and backcross (BC) populations. Mapping populations are developed from parents that differ genetically. In contrast to other types of mapping populations, DH and RIL populations consist of families that are genetically homogeneous with highly homozygous individuals within families. For this reason, DHL and RIL populations have a better accuracy compared to mapping populations consisting of genetically segregating families (e.g., BC populations), when measuring phenotypic traits that are under complex quantitative control.

Genetic markers are used to create a genetic map based on recombination events across the genome. The genetic map underpins quantitative trait locus (QTL) analysis, with the goal to establish associations between phenotypes and genotypes within the mapping population[23]. QTL mapping can be achieved by a variety of methods. The simplest method is a single marker analysis. Single marker analysis monitors the difference in phenotype between marker classes for each of the markers to search for significant marker – trait associations, without need to generate a linkage map[24]. However, single marker analysis does not allow to disentangle estimates for position and genetic effect of identified QTL. Interval mapping uses a genetic map with the use of flanking markers to identify the most likely position of QTL and to estimate the QTL effects[25]. Composite interval mapping uses in addition selected covariates to reduces the inference of linked QTL, which improves the power of QTL mapping[26], in particular for detection of linked QTL, and for determining their location in the genome. Inclusive composite interval

mapping (ICIM) utilizes stepwise regression and interval mapping first to identify the most significant marker variables in a regression approach, followed by interval mapping to detect QTL and their effects[27].

The identification of stable QTL is an important goal for breeders. It is crucial for these QTL to be stable across locations and years. A stable QTL allows for marker-assisted selection and provide consistent effects. ICIM can be used to determine whether QTL are stable across environments or whether they show QTL by environment interactions[28].

The objectives of this dissertation were assessment of inducibility and spontaneous haploid genome doubling within the maize DH system by 1) mapping of QTL controlling inducibility, 2) mapping of QTL controlling spontaneous haploid genome doubling, 3) identification of inducible lines, and 4) identification of lines with spontaneous haploid genome doubling capability.

Author Contributions

Chapter 2

B.T. experimental design, data collection, data analysis, interpretation, and writing manuscript.

U.F. and T.L. critical commentary for the manuscript.

Chapter 3

B.T. experimental design, data collection, data analysis, interpretation, and writing manuscript.

U.F, J.R., S.C., and T.L. critical commentary for the manuscript.

References

1. Moll, R. H., Lonnquist, J. H., Fortuno, J. V. & Johnson, E. C. The Relationship of Heterosis and Genetic Divergence in Maize. *Genetics* **52**, 139 (1965).

2. Prigge, V. & Melchinger, A. E. Production of Haploids and Doubled Haploids in Maize. in *Plant Cell Culture Protocols* (eds. Loyola-Vargas, V. M. & Ochoa-Alejo, N.) 161–172 (Humana Press, 2012). doi:10.1007/978-1-61779-818-4_13

3. Prasanna, B. M., Chaikam, V., & Mahuku, G. *Doubled haploid technology in maize breeding: theory and practice*. (CIMMYT, 2012).

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

5. Büter, B. In vitro haploid production in maize. in *In Vitro Haploid Production in Higher Plants: Volume 4 - Cereals* (eds. Jain, S. M., Sopory, S. K. & Veilleux, R. E.) 37–71 (Springer Netherlands, 1997). doi:10.1007/978-94-017-1862-2_2

6. Eder, J. & Chalyk, S. In vivo haploid induction in maize. *Theor Appl Genet* **104**, 703–708 (2002).

7. Tang, F., Tao, Y., Zhao, T. & Wang, G. In vitro production of haploid and doubled haploid plants from pollinated ovaries of maize (Zea mays). *Plant Cell Tiss Organ Cult* **84**, 233–237 (2006).

8. Genovesi, A. D. & Collins, G. B. In Vitro Production of Haploid Plants of Corn via Anther Culture 1. *Crop Science* **22**, 1137–1144 (1982).

9. Geiger, H. H. Doubled Haploids. in *Handbook of Maize: Genetics and Genomics* (eds. Bennetzen, J. L. & Hake, S.) 641–657 (Springer New York, 2009). doi:10.1007/978-0-387-77863-1_32

10. Kermicle, J. L. Indeterminate Gametophyte (ig): Biology and Use. in *The Maize Handbook* (eds. Freeling, M. & Walbot, V.) 388–393 (Springer New York, 1994). doi:10.1007/978-1-4612-2694-9_58

11. Coe, E. H. A Line of Maize with High Haploid Frequency. *The American Naturalist* **93**, 381–382 (1959).

12. Rotarenco, V., Dicu, G., State, D. & Fuia, S. New inducers of maternal haploids in maize. *Maize Genetics Cooperation Newsletter* 21–22 (2010).

13. Kelliher, T. *et al.* MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* **542**, 105–109 (2017).

14. Chaikam, V. *et al.* Analysis of effectiveness of R1-nj anthocyanin marker for in vivo haploid identification in maize and molecular markers for predicting the inhibition of R1-nj expression. *Theor Appl Genet* **128**, 159–171 (2015).

15. Wang, H. *et al.* Fully-Automated High-Throughput NMR System for Screening of Haploid Kernels of Maize (Corn) by Measurement of Oil Content. *PLOS ONE* **11**, e0159444 (2016).

16. Nanda, D. K. & Chase, S. S. An Embryo Marker for Detecting Monoploids Of Maize (Zea Mays L.) 1. *Crop Science* **6**, 213–215 (1966).

17. Coe, E. H. Spontaneous Mutation of the Aleurone Color Inhibitor in Maize. *Genetics* **47**, 779–783 (1962).

18. Kleiber, D. *et al.* Haploid Fertility in Temperate and Tropical Maize Germplasm. *Crop Science* **52**, 623–630 (2012).

19. Borisy, G. G. & Taylor, E. W. THE MECHANISM OF ACTION OF COLCHICINE: Binding of Colchincine-3H to Cellular Protein. *The Journal of Cell Biology* **34**, 525–533 (1967).

20. Wan, Y., Petolino, J. F. & Widholm, J. M. Efficient production of doubled haploid plants through colchicine treatment of anther-derived maize callus. *Theoret. Appl. Genetics* **77**, 889–892 (1989).

21. Sugihara, N., Higashigawa, T., Aramoto, D. & Kato, A. Haploid plants carrying a sodium azide-induced mutation (fdr1) produce fertile pollen grains due to first division restitution (FDR) in maize (Zea mays L.). *Theor Appl Genet* **126**, 2931–2941 (2013).

22. Mujeeb-Kazi, A. & Sitch, L. A. *2nd International Symposium on Genetic Manipulation in Crops*. (Int. Rice Res. Inst., 1989).

23. Myles, S. *et al.* Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. *The Plant Cell* **21**, 2194–2202 (2009).

24. Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. & Pang, E. C. K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **142**, 169–196 (2005).

25. Lander, E. S. & Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199 (1989).

26. Zeng, Z. B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *PNAS* **90**, 10972–10976 (1993).

27. Li, H., Ye, G. & Wang, J. A Modified Algorithm for the Improvement of Composite Interval Mapping. *Genetics* **175**, 361–374 (2007).

28. Meng, L., Li, H., Zhang, L. & Wang, J. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *The Crop Journal* **3**, 269–283 (2015).

CHAPTER TWO: QTL MAPPING OF MATERNAL INDUCIBILITY OF MAIZE (*ZEA MAYS* L.)

Benjamin Trampe, Ursula Karoline Frei, and Thomas Lübberstedt

Modified from a manuscript to be submitted Theoretical and Applied Genetics.

Abstract

The *in vivo* maternal doubled haploid system is used within maize breeding programs. Inducibility of maternal germplasm is an important trait for the production of haploids. To identify QTL controlling inducibility, an $F_{2:3}$ mapping population was developed from A427 and CR1Ht. $F_{2:3}$ families were planted in three locations in a completely random design. $F_{2:3}$ families averaged 8.3% inducibility rate across all three environments. A linkage map was constructed using 4,791 markers and 247 $F_{2:3}$ families covering a genetic map length of 2,090 cM were used for QTL mapping. A total of four QTL were found for inducibility evaluated on chromosomes 2, 4, 5, and 8. No stable QTL were found across environments. A QTL was found on chromosome 8 showing a high QTL by environmental interaction.

Introduction

Maize has traditionally been selfed recurrently to produce inbred lines. *In vivo* haploid production allows for production of doubled haploid (DH) lines, which accounts for a substantial time savings in inbred line development within a breeding program (Geiger 2009). In vivo haploid production relies on inducer and source germplasm. The inducer requires a relatively high haploid induction rate (HIR) to make the in vivo haploid system economically viable. Stock 6 is the progenitor of many haploid inducers, with an induction rate of 2.3% (Coe 1959). There have been improvements to inducers to increase haploid induction rate above 10% (Liu et al. 2016).

The *R1-nj* color marker can be used for identification of haploids (Nanda and Chase 1966). This marker produces anthocyanin, which causes a purple coloration in the endosperm and embryo of diploid seed. Haploids display coloration in the endosperm, but no coloration in embryo of the seed. Prigge et al. (2011) showed that there was variation in expression of the *R1-nj* marker across source germplasm. *R1-nj* can be masked by the *C1-I* allele, impairing haploid identification in certain backgrounds of germplasm (Coe 1962).

Inducers produce haploids by fertilization and selective chromosome elimination may occur during haploid formation (Zhang et al. 2008; Zhao et al. 2013). This results in the elimination of the inducer genome from embryo cells, leaving only the source germplasm haploid genome within these embryo cells. A QTL study comparing four populations of haploid inducers identified two major QTL (qhir1 and qhir8), explaining large percentages (>50%) of the genetic variance for haploid induction (Prigge et al. 2012). A sperm-specific phospholipase, MATRILINEAL, has meanwhile been isolated, and is a major gene underlying qhir1 affecting haploid induction (Kelliher et al. 2017). Haploid induction varies significantly depending on the background of the source germplasm (Eder and Chalyk 2002; Prigge et al. 2011). Inducibility (IND) is the ability of source germplasm to produce haploids. A breeder may be reluctant to make crosses and develop lines from germplasm with low induction rates. Improvements in inducibility would facilitate the production of DH lines within recalcitrant materials. No study has been conducted with temperate U.S. germplasm for IND thus far.

In this study, we used a mapping population of $F_{2:3}$ families developed from a cross between A427 and CR1Ht. The selection of parents was based on a preliminary experiment (data not shown) where A427 showed moderate HIR and CR1Ht showed high HIR. The objectives of

this study were to i) construct a high-density linkage map based on genotyping-by-sequencing (GBS) single nucleotide polymorphisms (SNPs), and (ii) detect QTL and QTL by environment interactions affecting IND.

Materials and Methods

Genetic Materials and Population Development

A biparental population of 247 $F_{2:3}$ families was formed from a cross between A427 and CR1Ht. A427 is a public non-stiff stalk inbred line developed at University of Minnesota (GRIN). CR1Ht is an exPVP non-stiff stalk inbred line developed by the J.C. Robinson Seed Company. Based on preliminary data, CR1Ht had a high inducibility rate of 16.8% and A427 had an average inducibility rate of 11.2% (data not shown). Iowa State University Haploid Inducers (BHI305, BHI306, BHI307, and BHI310) were bulked together and used as maternal haploid inducer (BHI Bulk). The 247 $F_{2:3}$ families were planted as donor in three separate isolations and induced with BHI bulk.

Experimental design

Field trials were conducted during the 2016 and 2017 growing seasons in isolation fields in Ames, IA and Boone, IA. We used three environments: two in 2016 in Ames, IA (N 42°00'38.1", W 93°37'45.9") and one in 2017 in Boone, IA (N 41°59'19.1", W 93°41'06.0"). The environments were Ames 1 2016 (AM1), Ames 2 2016 (AM2), and Boone 1 2017 (BO1). AM1 and AM2 were planted on 5/17/2016. BO1 was planted on 5/15/2017. The experiment was planted in a 4-3 isolation design with four female rows ($F_{2:3}$ families) for every three male rows (BHI Bulk). The three separate plantings of BHI Bulk were planted 0, 100, and 200 growing degree days after

planting donors in each environment. The experiment was grown in a completely random design with 3.81 m plots and 0.76 m row spacing. All trials were grown under rainfed field conditions using standard agronomic practices. All $F_{2:3}$ families were detasseled and open pollinated by BHI Bulk.

Phenotypic Evaluation

Inducibility was evaluated on a plot basis using the *R1-nj* seed-based marker system. All of the kernels from each plot were pooled and then sorted into putative haploid and hybrid groupings and counted. Inducibility rate was calculated as:

$$IND = \frac{number\ of\ haploids}{total\ number\ of\ kernels}\ X\ 100\% \quad [1]$$

Statistical Analysis

IND was angular transformed to normalize the distribution of data. Best linear unbiased predictions (BLUPs) were calculated from the proportion and angular-transformed data for the QTL analysis. The angular transformed data were analyzed using the following model:

$$Y_{ij} = \mu + E_i + G_j + \varepsilon_{ij}$$

where Y_{ij} is the angular transformed IND rate, μ is the overall mean, E_i is the random effect of the *i*th environment, G_j is the random effect of the *j*th $F_{2:3}$ family, and ε_{ij} is the residual error.

To estimate variance components and entry mean based heritability for IND, a random model was calculated using SAS PROC MIXED version 9.4 (SAS Institute 2013). Entry mean based heritability was calculated using the formula:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_r^2}{e}}$$

where, σ^2_g is the variance component for genotypes, σ^2_r is the variance component for the residual; e is the number of environments.

Genotyping

Genotyping of $F_{2:3}$ families was completed using GBS (Elshire et al. 2011). Plant tissues were collected at V2 growth stage from 10 maize plants per $F_{2:3}$ family and pooled to represent the parental F_2 individual. DNA extraction and genotyping were conducted by Cornell University Genomic Diversity Facility. For GBS, DNA was digested with the *Ape*KI restriction enzyme and DNA fragments were pooled for sequencing. The Buckler Laboratory for Maize Genetics and Diversity conducted the alignment to the Maize B73 RefGen_v2 genome assembly (https://www.maizegdb.org) and the calling of SNPs using the Tassel 5.0 GBS Production Pipeline (Glaubitz et al. 2014).

GBS Correction

Incorrect genotype calls can cause problems in generation of a genetic map due to inflated genetic map lengths as a result of improper map orders (Buetow 1991). GBS with low sequencing coverage may cause incorrect genotyping calls for SNPs that are notably heterozygous in nature. TASSEL software version 5 was used to filter GBS data (Glaubitz et al. 2014). The filtering was performed to remove single nucleotide polymorphisms (SNPs) with > 25% missing data and minor allele frequencies below 5%. A custom R script was created to eliminate SNPs with more than two alleles to avoid potential genotyping errors. TASSEL plugin GenostoABHPlugin converted the SNP nucleotides to a parent-based format for genotype correction (Bradbury et al. 2007). Genotype-Corrector (Miao et al. 2018) was used for correction of genotyping errors of primarily heterozygous SNP calls from GBS data. A sliding window

approach using a window size of 25 SNPs was used to correct genotyping errors and to impute missing data across the genome.

Linkage Map Construction

The linkage map was constructed based on 247 F_2 plants. Binning of markers was conducted by using the BIN function in QTL IciMapping V4.1 (Meng et al. 2015). Chi square tests were used to identify SNPs with significant segregation distortion for genotypes. Markers with a value of $p < 0.001$ were removed. Linkage map construction was conducted using the MAP function in QTL IciMapping V4.1 (Meng et al. 2015). The Kosambi mapping function was used to create the linkage map. Grouping of markers was completed by chromosome. The traveling salesman algorithm (nnTwoOpt) was used to order the markers. Markers were rippled using the criterion of Sum of Adjacent Recombination Frequencies (SARF) and a window size of 5 markers. The rippling allows fine tuning of the marker order to minimize the linkage map length.

QTL Mapping

QTL mapping was conducted using QTL IciMapping v4.1 (Meng et al. 2015). Inclusive composite interval mapping was utilized for all analyses. BLUPs were used as the phenotype for all analyses. The QTL mapping in biparental mapping (BIP) functionality was used to evaluate QTL in each single environment. The QTL by environment interaction in biparental populations (MET) functionality was used to evaluate QTL by environment interactions across environments. The LOD threshold was set based on 1,000 permutation tests using a Type I error rate of P = 0 .05. The single environment LOD score threshold was 4.0 and the QTL by Environment Interaction LOD score threshold was 6.1.

QTL Analysis for Inducibility

Four separate QTL analyses were completed, based on (i) QTL by location using percentage data (Table 4), (ii) QTL across locations using percentage data (Table 5), (iii) QTL by location using angular-transformed data (Table 6), and (iv) QTL across locations using angular-transformed data (Table 7).

Results

Descriptive Statistics and Linkage Map

A total of 1,039,498 kernels were evaluated in this experiment across all environments. A total of 85,027 putative haploids and 954,471 hybrids were identified based on *R1-nj* seed-based marker system. An experiment-wise induction rate of 8.2% was observed, with 8.0% IR for AM1, 7.5% IR for AM2, and 9.3% for BB1 (Table 1). Induction rates ranged from 0.6% to 23.8% for the $F_{2:3}$ families (Table 1). A427 had an average IR of 6.7% and CR1Ht had an average of 7.5%. Locations and lines were highly significant. The entry mean based heritability for IND was 0.60. Significant environment and family effects were found (Table 2).

The linkage map was constructed using 4,791 markers and 247 $F_{2:3}$ families. The total length of the genetic map was 2090 cM with marker density of 2.3 markers / cM across the whole genome. The linkage map length was consistent with other maize population linkage maps (e.g., Beavis et al. 1991).

Single Environment QTL Analysis

A total of four QTL were detected on three chromosomes (2, 4, and 5), individually explaining between 8.1% to 10.7% of the phenotypic variance for a given environment (Table 4). Across both the percentage and angular transformed data, 71% of loci had favorable alleles that originated from CR1Ht and 29% of loci had favorable alleles that originated from A427. This indicates that both CR1Ht and A427 both carry favorable alleles for IND. There was no QTL that was found in all three environments. When comparing percentage and angular transformed data, results were very similar except for qIND2a, qIND2b, and qIND4a. For percentage data, qIND2a was detected on chromosome 2 at position 180 (Table 4). For angular-transformed data, qIND2b was detected on chromosome 2 at position 202 (Table 6). These two QTL were found 22 cM apart. The QTL qIND4a was found in the percentage analysis, but not found in the analysis of transformed data.

QTL x Environment Interaction Analysis

A total of four QTL were detected on four chromosomes (2, 4, 5, and 8). These small-effect QTL individually explained between 5.1% to 9.9% of the phenotypic variance. The QTL found on chromosomes 5 and 8 were detected in both analyses, but QTL qIND2a and qIND4b were found with percentage data only (Table 5). qIND8 was found in QTL across environments but was not identified in the single environment analysis. Only two QTL (qIND4b and qIND8) showed strong QTL x environment interactions. This is indicated by the higher phenotypic variance explained by additive x environment effects (PVE(AbyE)) than by the phenotypic variance explained by additive effects (PVE(A)).

Discussion

Source germplasm has significant influence on the production of haploids (Eder and Chalyk 2002). There was considerable variation for IND within this study. The largest variation was found between families. The HIR varied from 0.6% to 23.8% when comparing families across all environments (Table 1). The environments also had a significant effect on IND (Table 2). The average HIR was 8.3% for $F_{2:3}$ families of A427/CR1Ht, which is lower than the HIR of 10.3% by F_1 between A427/CR1Ht reported earlier (Fuente et al. 2018). Interestingly, the rates of IND of the parents were substantially lower compared to the preliminary evaluation (data not shown). There was a considerable reduction of HIR of 4.5% for A427 and 9.3% for CR1Ht when comparing the preliminary HIR data to the results found in this study. Environments in which HIR is evaluated can produce drastically different results.

Misclassification Rates

Haploid misclassification rates can influence the accurate calculation of HIR and correction of HIR may be needed to improve accuracy. This study didn't adjust the haploid induction rates for false positives and false negatives during haploid selection, because of the cost of evaluation and the acceptable misclassification rates based on Fuente et al. (Fuente et al. 2018). Correcting for false positives and false negatives would increase the accuracy and precision of the HIR and, therefore, potentially improve the QTL analysis. The difficultly in mapping and evaluating HIR is the large number of kernels that needs to be evaluated. A total of 1,000 kernels would need to be selected to detect a difference of 5% and smaller differences would require much larger selection screens (Fuente et al. 2018). Visual scoring can be a viable

selection method, if properly trained labor is acquired, nevertheless issues of misclassification rates are still present. An automated and high throughput method is needed to select haploids to achieve better accuracy and precision. Alternatively, nuclear magnetic resonance is a type of automated high throughput technology that can evaluate oil content within the seed to selected haploids when paired with a high oil inducer for the production of haploids (Wang et al. 2016).

Environment Effects on Inducibility

The QTL identified all had small genetic effects. The QTL by environment results were similar between both datasets. qIND4b and qIND5 were found in both data analyses and the effect sizes were very comparable between the two. The single environment analysis did find two different QTL, qIND4a and qIND2b. All the identified QTL were found in different environments and displays the instability of IND (Tables 4 and 6). The instability of IND shows the highly quantitative character of this trait. Significant QTL by environment interactions were found for two of the QTL, qIND8 and qIND4b (Table 5). For qIND8 and qIND4b, a significant amount of the phenotypic variance can be explained by additive x environment interactions. The use of these two QTL would be problematic to use within a breeding program. The breeder would be limited as to when or where the induction of source germplasm could be performed.

Breeding for Improved Induction Rate

Improvements in HIR within the source germplasm can make the in vivo DH system more cost effective and reduce the amount of haploid selection. HIR is highly quantitative in nature, but the heritability in this study was moderate at 0.49, which is similar to 0.45 found by Kebede et al. (Kebede et al. 2011). The highly quantitative nature and additive effects of IND indicates that a genomic selection strategy could be used to improve IND, because of the small genetic

effects. The breeding germplasm of maize programs is genetically unique, which makes it a complex problem in improving IND by introgression of IND QTL. Introgression of QTL from this study could be too laborious, because of the small effects and the influence of QTL by environment interactions.

Author Contributions

B.T. – design the experiment, phenotyping, data, construction of linkage map, statistical and QTL analysis, interpretation, and manuscript writing. U.F. – Critical commentary. T.L. – project leader and corresponding author who oversaw project conceptualization and implementation, and manuscript development. All authors and revised and approved the manuscript.

Acknowledgements

The authors are thankful for financial support from USDA NIFA project IOW04314 and IOW01018. BT was supported by the ISU Plant Sciences Institute, the ISU R.F. Baker Center for Plant Breeding, and the ISU K.J. Frey Chair in Agronomy.

References

Beavis WD, Grant D, Albertsen M, Fincher R (1991) Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. Theoret Appl Genetics 83:141–145. doi: 10.1007/BF00226242

Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi: 10.1093/bioinformatics/btm308

Buetow KH (1991) Influence of aberrant observations on high-resolution linkage analysis outcomes. Am J Hum Genet 49:985–994

Coe EH (1959) A Line of Maize with High Haploid Frequency. The American Naturalist 93:381–382

Coe EH (1962) Spontaneous Mutation of the Aleurone Color Inhibitor in Maize. Genetics 47:779–783

Eder J, Chalyk S (2002) In vivo haploid induction in maize. Theor Appl Genet 104:703–708. doi: 10.1007/s00122-001-0773-4

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379. doi: 10.1371/journal.pone.0019379

Fuente DL, N G, Frei UK, et al (2018) A Diallel Analysis of a Maize Donor Population Response to In Vivo Maternal Haploid Induction: I. Inducibility. Crop Science 58:1830–1837. doi: 10.2135/cropsci2017.05.0285

Geiger HH (2009) Doubled Haploids. In: Bennetzen JL, Hake S (eds) Handbook of Maize: Genetics and Genomics. Springer New York, New York, NY, pp 641–657

Glaubitz JC, Casstevens TM, Lu F, et al (2014) TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLOS ONE 9:e90346. doi: 10.1371/journal.pone.0090346

Kebede AZ, Dhillon BS, Schipprack W, et al (2011) Effect of source germplasm and season on the in vivo haploid induction rate in tropical maize. Euphytica 180:219–226. doi: 10.1007/s10681-011-0376-3

Kelliher T, Starr D, Richbourg L, et al (2017) MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature 542:105–109. doi: 10.1038/nature20827

Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. The Crop Journal 3:269–283. doi: 10.1016/j.cj.2015.01.001

Miao C, Fang J, Li D, et al (2018) Genotype-Corrector: improved genotype calls for genetic mapping in F 2 and RIL populations. Scientific Reports 8:10088. doi: 10.1038/s41598-018-28294-0

Nanda DK, Chase SS (1966) An Embryo Marker for Detecting Monoploids Of Maize (Zea Mays L.) 1. Crop Science 6:213–215. doi: 10.2135/cropsci1966.0011183X000600020036x

Prigge V, Sánchez C, Dhillon BS, et al (2011) Doubled Haploids in Tropical Maize: I. Effects of Inducers and Source Germplasm on in vivo Haploid Induction Rates. Crop Science 51:1498–1506. doi: 10.2135/cropsci2010.10.0568

Prigge V, Xu X, Li L, 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. doi: 10.1534/genetics.111.133066

SAS Institute (2013) SAS version 9.4. SAS Institute, Cary, NC

Wang H, Liu J, Xu X, et al (2016) Fully-Automated High-Throughput NMR System for Screening of Haploid Kernels of Maize (Corn) by Measurement of Oil Content. PLOS ONE 11:e0159444. doi: 10.1371/journal.pone.0159444

Zhang Z, Qiu F, Liu Y, et al (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. doi: 10.1007/s00299-008-0601-2

Zhao X, Xu X, Xie H, et al (2013) Fertilization and Uniparental Chromosome Elimination during Crosses with Maize Haploid Inducers. Plant Physiology 163:721–731. doi: 10.1104/pp.113.223982

Tables and Figures

Table 1. Phenotypic summary of inducibility (IND) evaluated in three environments. Traits means are shown for both parents and $F_{2:3}$ families.

		IND %			
		AM1	AM2	BO1	Average
Parents	A427	N/A	N/A	6.74	6.74
	CR1Ht	7.52	7.13	7.75	7.50
$F_{2:3}$ Families	Mean	8.04	7.47	9.31	8.26
	Min	0.80	0.75	0.63	0.63
	Max	18.01	16.61	23.76	23.76
	SD	2.91	2.85	3.74	3.27
	CV	36.15	38.08	40.17	39.64

Table 2. Analysis of variance for inducibility across three environments.

Source	DF	SS	MS	F Value	Pr > F
Environment	2	0.044	0.022	35.59	<.0001**
Families	246	0.416	0.001	2.51	<.0001**
Residual	490	0.331	0.0006		

DF = Degrees of freedom, SS = Sum of Squares, MS = Mean Square, * Significant difference at the 5% level, ** Significant difference at the 1% level, ns Non-significant

Table 3. Marker statistics of linkage map from $F_{2:3}$ families derived from A427 and CR1Ht.

Chromosome	No. of Markers	% of Markers	Total Length (cM)	Bins	Average Marker/Bin Interval	Marker per cM
1	737	15.38%	286.73	504	1.46	2.57
2	616	12.86%	238.31	396	1.56	2.58
3	654	13.65%	245.65	419	1.56	2.66
4	503	10.50%	240.86	291	1.73	2.09
5	508	10.60%	231.89	327	1.55	2.19
6	433	9.04%	188.2	277	1.56	2.30
7	402	8.39%	188.72	235	1.71	2.13
8	261	5.45%	147.42	163	1.60	1.77
9	262	5.47%	157.88	161	1.63	1.66
10	415	8.66%	164.11	234	1.77	2.53
Total	4791		2089.77	3007	1.59	2.29

Table 4. List of significant QTL identified in three environments for inducibility (IND) using percentage data.

Env[1]	QTL	Chr[2]	Pos[3]	Marker Interval	LOD	PVE[4]	Add[5]
AM1	qIND5	5	106	S5.163229787-S5.163889760	5.11	9.07	0.01
AM2	qIND2a	2	180	S2.213848716-S2.213842789	4.83	9.5	-0.01
	qIND4a	4	119	S4.173795662-S4.174915619	4.11	8.1	-0.01
BO1	qIND4b	4	97	S4.158130766-S4.158136040	6.21	10.7	-0.02

[1]Environment. [2]Chromosome of identified QTL. [3]Position of the QTL in cM. [4]Phenotypic variance explained. [5]Additive Effect (Positive values signify that alleles came from A427 and negative values signify that alleles came from CR1Ht).

Table 5. List of significant QTL x environment interactions identified in three environments for inducibility (IND) using percentage data.

QTL	Chr[1]	Pos[2]	Marker Interval	LOD	PVE[3]	PVE(A)[4]	PVE(AbyE)[5]	Add[6]
qIND2a	2	180	S2.213848716-S2.213842789	7.51	6.24	5.74	0.5	-0.0067
qIND4b	4	97	S4.158130766-S4.158136040	6.83	9.91	3.55	6.37	-0.0053
qIND5	5	105	S5.160081320-S5.163229787	7.23	6.5	5.78	0.72	0.0042
qIND8	8	27	S8.7675588-S8.7748928	6.57	5.97	1.9	4.07	0.003

[1]Chromosome of identified QTL. [2]Position of the QTL in cM. [3]Phenotypic variance explained. [4]Phenotypic variance explained by additive effects. [5]Phenotypic variance explained by additive x environment effects. [6]Additive Effect (positive values signify that alleles came from A427 and negative values signify that alleles came from CR1Ht).

Table 6. List of significant QTL identified in three environments for inducibility (IND) using angular-transformed data.

Env[1]	QTL	Chr[2]	Pos[3]	Marker Interval	LOD	PVE[4]	Add[5]
AM1	qIND5	5	106	S5.163229787-S5.163889760	4.95	8.82	0.01
AM2	qIND2b	2	202	S2.222648035-S2.222649363	5.19	8.03	-0.02
BO1	qIND4b	4	97	S4.158130766-S4.158136040	5.48	9.54	-0.03

[1]Environment. [2]Chromosome of identified QTL. [3]Position of the QTL in cM. [4]Phenotypic variance explained. [5]Additive Effect (Positive values signify that alleles came from A427 and negative values signify that alleles came from CR1Ht).

Table 7. List of significant QTL x environment interactions identified in three environments for inducibility (IND) using angular-transformed data.

QTL	Chr[1]	Pos[2]	Marker Interval	LOD	PVE[3]	PVE(A)[4]	PVE(AbyE)[5]	Add[6]
qIND5	5	105	S5.160081320-S5.163229787	8.01	6.48	6.06	0.42	0.0089
qIND8	8	27	S8.7675588-S8.7748928	6.12	5.11	1.65	3.46	0.0055

[1]Chromosome of identified QTL. [2]Position of the QTL in cM. [3]Phenotypic variance explained. [4]Phenotypic variance explained by additive effects. [5]Phenotypic variance explained by additive x environment effects. [6]Additive Effect (Positive values signify that alleles came from A427 and negative values signify that alleles came from CR1Ht).

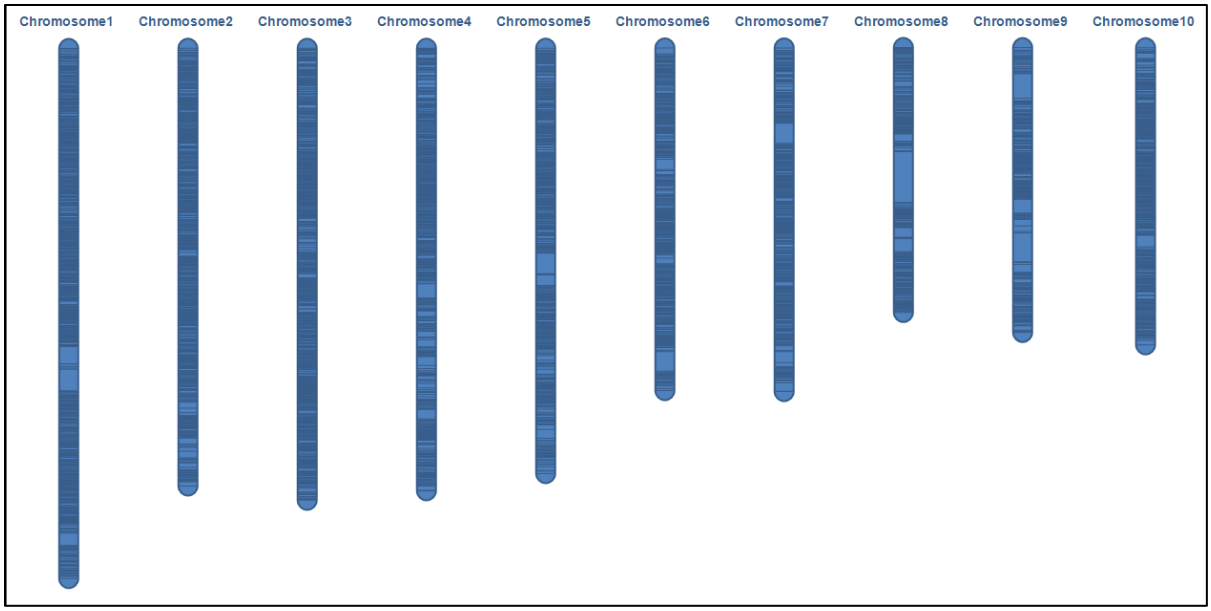

Figure 1. Linkage map constructed from $F_{2:3}$ families derived from A427 and CR1Ht

CHAPTER THREE: QTL MAPPING OF SPONTANOUS HAPLOID GENOME DOUBLING USING GENOTYPING BY SEQUENCING IN MAIZE (*ZEA MAYS* L.)

Benjamin Trampe, Ursula Karoline Frei, Jiaojiao Ren, Shaojiang Chen, and Thomas Lübberstedt

Modified from a manuscript to be submitted to Scientific Reports.

Abstract

Low efficiency of haploid genome doubling is a bottleneck in producing doubled haploid lines in maize. To identify QTL controlling spontaneous haploid genome doubling (SHGD), a $F_{2:3}$ mapping population was developed from the inbred lines A427 and CR1Ht. Haploids from $F_{2:3}$ families were planted in three environments and scored for anther emergence, pollen emergence, and tassel size. A linkage map of 4,171 markers covering a genetic map length of 2,141 cM was used for QTL mapping. A total of 15 QTL were found for anther emergence, pollen production, tassel size, and haploid male fertility traits on chromosomes 1,5,6,7, and 10. A major QTL was detected on chromosome 5, which showed pleiotropic effects for all four traits. It explained 51.3% of the variation for anther emergence, 55.9% for pollen production, 48.5% for tassel size, and 45.7% for haploid male fertility. A marker-assisted backcrossing program could be an effective way to incorporate this major QTL conferring SHGD into other elite germplasm.

Introduction

Maize is primarily grown as a hybrid crop across the world. An essential step in producing maize hybrids is the development of parental inbred lines. Traditionally, maize inbreds have been created using ear-to-row selections, which includes several generations of self-pollination to create homogeneous and homozygous lines[1]. The adoption and use of doubled haploid (DH)

technology increased speed and efficiency of producing inbred lines[2]. The production of DH lines in maize is primarily done using maternal in vivo haploid induction, which includes three steps: (i) induction, (ii) identification, and (iii) genome doubling of haploids[3].

The first step of the induction of haploids is accomplished by crossing a donor genotype, from which DH lines are developed, and an inducer line used as male. Stock 6 was the first line identified that could induce maternal maize haploids with an induction rate of 2.3%[4]. Stock 6 is the progenitor of improved inducer lines. Current inducers have induction rates ranging from 2% to 15%[5,6]. The efficiency of haploid identification depends on the developmental stage and type of selection. It is preferable that reliable identification and selection happen as early as possible in the process to reduce the costs and labor associated with DH line development. Haploids can either be identified at the kernel or a later stage of development. Several approaches have been evaluated or used for haploid selection, including the use of color markers[7] and various automated haploid selection techniques[8]. Kernel-based selection is based on double fertilization. In a regular fertilization event, two genetically identical male sperm cells within a pollen tube[9] fuse with the egg cell and the two central cells, resulting in the embryo (2n) and endosperm (3n) formation, respectively. In crosses with maternal haploid inducers, hybrids contain normal embryo (2n) and endosperm (3n). Haploids contain an abnormal haploid embryo and normal 3n endosperm. The biological mechanism for the production of haploids is not yet understood. Two alternative hypotheses are either single fertilization or selective chromosome elimination during the formation of the embryo[10]. The most widely used approach for haploid selection to date is the use of *R1-Navajo* [11], which codes for a transcription factor in the anthocyanin pathway[12]. This dominant allele present in maternal haploid inducers is expressed in the embryo and endosperm

of diploid kernels originating from the cross between donor and inducer. A haploid kernel from the same cross shows anthocyanin expression within the endosperm, but no expression in the embryo[11].

The third step is the genome doubling of haploid individuals to produce DH lines. Tassels of most haploid plants are completely sterile, which prevents self-pollination of individual plants to produce DH lines[13]. Artificial doubling is usually used in the creation of DH lines. Colchicine is the "gold standard" to double haploid genomes, substantially increasing the rate of fertile haploid flower structures[14]. Challenges of treating putative haploids with colchicine are the significant labor requirement and toxicity of colchicine. Treated haploid seedlings need to be transplanted into the field, which is a considerable labor requirement compared to direct sowing of seed. Permits and trained labor are necessary for the application of colchicine. Colchicine has high toxicity for humans and can be fatal when an overdose occurs[15]. Alternatives to colchicine include herbicides like oryzalin or prinomide, and nitrous oxide[16–18]. The alternatives still require a substantial labor investment.

Spontaneous doubling of haploid genomes has been found in different genetic backgrounds of maize and may be an alternative to chemical genome doubling[19]. Spontaneous haploid genome doubling (SHGD) requires the duplication of genomes in cells leading to both female and male inflorescences [20]. Haploid female fertility seems not to be a limiting factor in haploid fertility with reports of > 90% of haploid ears setting seed [21]. In Chinese germplasm, Jiang et al. found minimal levels of haploid male fertility (HMF) from 4% to 11% with an average of 8.28% [22]. In a previous mapping study using Chinese germplasm, four QTL were identified for HMF: *qhmf1*, *qhmf2*, *qhmf3, qhmf4*, located in chromosome bins 1.11, 3.06, 4.02/4.03, and

6.07[23]. QTL *qhmf4* showed the strongest segregation distortion and was fine mapped to a ~800 kb region on chromosome 6, which includes the candidate gene *absence of first division 1* (*adf1)*[23]. *Adf1* is important for the development of axial element elongation and homologous pairing. Ma et al. reported 14 QTL for HMF using GWAS, which was located in bins 2.05, 2.06, 3.07, 5.05, 6.01, 7.05, 9.01, and 10.04 and collectively explained 22.5% of the total phenotypic variance[24]. No study has been conducted with temperate U.S. germplasm for HMF thus far.

In this study, we used a mapping population of $F_{2:3}$ families developed from a cross between inbred A427 and CR1Ht. The selection of parents was based on a preliminary experiment (data not shown) where A427 showed the highest HMF and CR1Ht showed moderate rates of HMF. The objectives of this study were to (i) construct a high-density linkage map based on genotyping by sequencing (GBS) SNPs, (ii) detect QTL and QTL by environment interactions (QXE) affecting anther emergence (AE), pollen production (PP), tassel size (TS) of fertile haploids, and haploid male fertility (HMF), and (iii) identify the best trait for mapping and selection of HMF.

Materials and Methods

Genetic Materials and Population Development

A biparental population of 218 $F_{2:3}$ families was developed from a cross between A427 and CR1Ht. A427 is a public non-stiff stalk inbred line developed at the University of Minnesota[25]. CR1Ht is an expired proprietary non-stiff stalk inbred line developed by J.C. Robinson Seed Company in Nebraska. Both A427 and CR1Ht were part of a preliminary experiment, which consisted of a larger panel of 102 inbred lines that were screened for HMF in 2013[26]. A427 showed a high rate of HMF at 78% (data not shown). In contrast, CR1Ht showed a moderate rate

of HMF of 22% (data not shown). Iowa State University haploid inducers (BHI305, BHI306, BHI307, and BHI310)[27] were bulked and used as maternal haploid inducer (BHI bulk). The 218 $F_{2:3}$ families were planted as donors in an isolation field and pollinated by the BHI bulk. A representative sample of haploids from each donor plant within families was bulked to maximize genetic variation within each family. The representative sample was produced by counting the number of haploids from each ear and then selecting haploids from all ears until the ear with the fewest number of haploids had been adequately represented, followed by randomly selecting from the remaining ears to produce a representative sample of 120 haploids for each $F_{2:3}$ family.

Experimental Design

Field trials were conducted during the 2017 growing season at the Agricultural Engineering and Agronomy Farm (AEA) in Boone, IA (N 42°01'14.4" W 93°46'36.1") and the Plant Introduction Station (PI) in Ames, Iowa (N 42°00'38.5" W 93°39'32.5"). We used three environments. AEA Early (AEAE) was planted 5/26/2017, AEA 2017 Late (AEAL) was planted 6/13/2017, and PI 2017 (PI) was planted 5/16/2017. The AEAE and AEAL were both rainfed locations, while PI was irrigated with surface drip irrigation. The haploid families derived from the 218 $F_{2:3}$ families, including haploids from parents A427 and CR1Ht as checks, were evaluated in a randomized complete block design with two replications in each of the three environments. The trials were planted in 3.81 m plots with 0.76 m row spacing. All plots were directly sown into the field without application of a chemical doubling agent. Planting density was 69,000 putative haploids per hectare. Hybrid plants (misclassified haploids) were removed from the field at the V4 growth stage based on visual appearance of increased vigor, plant size, and leaf number.

Standard agronomic practices of fertilization, weed control, and pest management were used for all field trial locations.

Phenotypic and Statistical Analysis

Phenotyping was conducted for anther emergence, pollen production, and tassel size on fertile haploids. Anther emergence (AE) was evaluated using a rating scale of 0-5. The rating scale for anther emergence follows rating scale from Wu et al.[28], as follows: (0) sterile tassel with no anthers; scores 1-5 are tassels with the following percentage of fertile anthers (1) below 5%, (2) 5-20%, (3) 20-50%, (4) 50-75%, (5) 75-100%. Pollen production (PP) was evaluated using a rating scale of 0-5. The quantity of pollen produced was scored with a 0-5 rating scale: (0) no pollen produced, (1) little pollen, only obtained by pressing anthers, (2) small amount of pollen by shaking the tassel, (3) modest amount of pollen by shaking the tassel, (4) ample pollen by touching the tassel, (5) abundant pollen released by moving the tassel. Pollen viability was not evaluated within this study. However, self-pollination of male fertile haploids was attempted and was successful (data not shown). Tassel size (TS) was assessed on a 1-5 scale: (1) central spike only, (2) central spike and 1 tassel branch, (3) central spike and 2 - 3 tassel branches, (4) central spike and 4 - 5 tassel branches, (5) central spike with 6 or more tassel branches. Haploids were rated every day throughout the pollination season. The highest daily rating for each of the traits was used for subsequent analyses. In addition to AE, PP, and TS, haploids were scored as fertile or sterile based on the presence of pollen according to Ma et al.[24]. Any haploid plant extruding pollen to the visible eye was counted as fertile, regardless of the PP scores 1- 5. HMF was calculated by counting the number of the pollen shedding haploid plants divided by the total number of plants in each plot:

$$HMF = \frac{number\ of\ pollen\ shedding\ plants}{total\ number\ of\ haploid\ plants}\ X\ 100\%$$

Data from the environments were analyzed individually. An ordinal logistic regression with proportional odds assumption was used to obtain best linear unbiased predictors (BLUPs) for each $F_{2:3}$ family for AE, PP, TS. The analysis was conducted using *clmm* functionality from the ordinal package[29] in the R software version 3.5.0[30]. The model for field trials was:

$$Y_{jkl} \sim \text{Multinomial}(1, \pi_{ijk})$$

$$\text{logit}(\pi_{ijkl}) = \theta_i + R_j + G_k$$

where the effects are as follows. θ_i is the intercept for the *i*th response category (*i* = 0,1,2,3,4,5), R_j is the effect of the *j*th replication, and G_k is the effect of *k*th genotype.

HMF was logit transformed as the data were percentages and many values were close to zero. The response variable was equal to HMF logit = log[(HMF+0.005)/(1-HMF+0.005)], to normalize the data. Best linear unbiased predictors were calculated from the logit-transformed scores and used for QTL analysis. The logit-transformed data were analyzed using the model:

$$Y_{ijk} = \mu + E_i + R_{(i)j} + G_k + EG_{ik} + \varepsilon_{ijk}$$

where Y_{ijk} is logit transformed HMF rate, μ is the overall mean, E_i is the random effect of the *i*th environment, $R_{(i)j}$ is the random effect of the *j*th replication nested in *i*th environment, G_k is random effect of the *k*th $F_{2:3}$ family, EG_{ik} is the random interaction between the *i*th environment and the *k*th $F_{2:3}$ family, and ε_{ijk} is the residual error.

A random model was used to estimate variance components, and entry mean based heritabilities for AE, PP, TS, and HMF using SAS PROC MIXED version 9.4[31]. Entry mean based heritabilities were estimated using the formula:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{e} + \frac{\sigma_r^2}{re}}$$

where σ^2_g is the variance component for genotypes, σ^2_{ge} is the variance component for genotype x environment, σ^2_r is the variance component for the residual; e is the number of environments and r is the number of replications. Trait correlation analysis was completed using the average untransformed scores for each plot of haploids for AE, PP, and TS and HMF.

Genotyping

Genotyping of $F_{2:3}$ families were completed using GBS[32]. Plant tissues were collected at the V2 growth stage from 10 maize plants per $F_{2:3}$ family and pooled to represent parental F_2 individuals. DNA extraction and genotyping were conducted by Cornell University Genomic Diversity Facility. For GBS, DNA was digested with the *Ape*KI restriction enzyme, and DNA fragments were pooled for sequencing. The Laboratory for Maize Genetics and Diversity at Cornell University conducted the alignment to Maize B73 RefGen_v2 (https://www.maizegdb.org) and the calling of SNPs using the Tassel 5.0 GBS Production Pipeline [33]. A total of 955,690 SNPs was called for 218 families.

GBS Correction

Incorrect genotype calls can cause problems in the generation of a genetic map such as inflated genetic map lengths as a result of improper map orders[34]. GBS with low sequencing coverage may cause incorrect genotyping calls for SNPs. TASSEL software version 5[35] was used to filter GBS data. The filtering was performed across all single nucleotide polymorphisms (SNPs) with > 25% missing data and minor allele frequencies below 5%. A custom R script was created to eliminate SNPs with more than two alleles to avoid potential genotyping errors. TASSEL plugin

GenostoABHPlugin converted the SNP nucleotides to a parent-based format for genotype correction[35]. Genotype-Corrector[36] was used for correction of genotyping errors of primarily heterozygous SNP calls from GBS data. A sliding window approach using a window size of 25 SNPs was used to correct genotyping errors and to impute missing data across the genome. This left 23,102 SNPs after correction.

Linkage Map Construction

The linkage map was constructed based on 218 F_2 genotypes. Binning of markers was conducted by using the BIN function in QTL IciMapping V4.1[37]. Chi-square tests were used to identify SNPs with significant segregation distortion for genotypes. Markers with a value of $p < 0.001$ were removed. Linkage map construction was conducted utilizing the MAP function in QTL IciMapping V4.1[37]. The Kosambi mapping function was used to create a linkage map. Grouping of markers was completed by chromosome. The traveling salesman algorithm (nnTwoOpt) was used to order the markers. Markers were rippled using the criterion of Sum of Adjacent Recombination Frequencies (SARF) and a window size of 5 markers. The rippling allows fine-tuning of the marker order to minimize the linkage map length.

QTL Mapping

QTL mapping was conducted using inclusive composite interval mapping using QTL IciMapping v4.1 [37]. BLUPs were used as the phenotype for all analyses. The QTL mapping in bi-parental mapping (BIP) functionality was used to evaluate QTL in every single location. The QTL by environment interaction in biparental populations (MET) feature was used to assess QTL by environment interactions across environments. The LOD threshold was set based on 1,000 permutation tests using a Type I error rate of P = 0 .05[37].

Results

Phenotypic Statistics and Linkage Map

Both parents were evaluated across the three environments used in this study. A427 had a HMF rate of 73%, and CR1Ht had an HMF rate of 44% (Table 1), which differed slightly from preliminary data (78% and 22%, respectively). For the $F_{2:3}$ families, the highest scores for AE, PP, and TS, were found in PI, followed by AEAE, and lowest scores were found for AEAL (Table 1). The location averages for AE were as follows: PI with 1.23, followed by AEAE with 1.07, and AEAL with 1.00 (Table 1). The location averages for PP are as follows; PI with 1.06, followed by AEAE with 0.93, and AEAL with 0.92 (Table 1). The location averages for TS were as follows: PI with 1.60, followed by AEAE with 1.27, and AEAL with 1.15 (Table 1). The location averages for HMF were 36% at PI, 30% for AEAE, and 28% for AEAL (Table 1). The averages values across locations were 1.10 for AE, 0.97 for PP, 1.34 for TS, and the average HMF rate was 31% (Table 1). A considerable amount of variation was found for all traits and environments. The range across environments for AE was 0-4.48, 0-5 for PP, 0-5 for TS, and for HMF it was 0% - 100%. Variance components for environment and genotype by environment interactions were significant, which suggest environment is very important factor to consider (Tables S1-S4). Very high correlations were found between all traits. The closest correlation was found between TS and HMF (0.97) and the lowest was found between PP and TS, and PP and HMF at 0.88 (Table 2). The $F_{2:3}$ families had entry-mean based heritabilities of 0.47 for AE, 0.48 for PP, 0.53 for TS, and 0.85 for HMF (Table 3). The linkage map was 2141.2 cM in length and contained 4,171 SNP markers across 10 linkage groups with an average distance between adjacent markers of 0.51 cM (Table 4).

QTL analysis for AE, PP, TS, and HMF

A total of 27 QTL were detected for AE, PP, TS, and HMF in AEAE, AEAL, and PI. The identified QTL were distributed over five chromosomes (1, 5, 6, 7, and 10). Individual QTL explained between 3.5% and 55.9% of the phenotypic variance in a given environment (Table 5). For the 27 QTL identified, 81% of loci had favorable alleles that originated from A427. This indicates that both A427 and CR1Ht both carry favorable alleles for SHGD. Most QTL were small effect QTL, except for a pleiotropic QTL on chromosome 5, controlling AE, PP, TS, and HMF.

For AE, a total of seven QTL were identified in three environments. The detected QTL were dispersed over chromosomes 1, 5, 6, 7, and 10. The phenotypic variance explained by additive effects of QTL ranged from 3.5% to 42.2%. A major QTL, qAE5, was pinpointed on chromosome 5 and found in all three environments explaining a considerable percentage of phenotypic variance: 30.1% at PI, 41.6% at AEAL, and 42.2% at AEAE (Table 5). The favorable allele came from inbred A427 for this region. The QTL by environmental interaction analysis revealed that phenotypic variance explained by qAE5 was 51.3% (Table 6).

For PP, a total of six QTL were identified in three environments. The detected QTL were dispersed over chromosomes 1, 5, 6, and 7. The phenotypic variance explained by additive effects of QTL ranged from 3.5% to 41.8%. The same region on chromosome 5 identified for AE also controlled PP: qPP5. Like qAE5, qPP5 also explained considerable phenotypic variance for the three environments: 32.5% at PI, 41.8% at AEAE, and 43.3% for AEAL (Table 5). The QTL by environmental interaction analysis uncovered that the phenotypic variance explained for qPP5b was 55.9% (Table 6).

For TS, a total of eight QTL were identified in three environments. The detected QTL were distributed over chromosomes 5, 6, 7, and 10. The phenotypic variance explained by additive effects of QTL spanned 4.1% to 40.0%. The chromosome 5 region that was identified for AE and PP was shown to be associated with TS of male fertile haploids. Like both qAE5 and qPP5, qTS5 explained substantial phenotypic variance: 28.2% at PI, 38.6% at AEAL, and 40.0% for AEAE (Table 5). The QTL by environmental interaction analysis uncovered a phenotypic variance explained was 48.5% for qTS5a (Table 6). Only qTS6b showed a strong QTL by environment interaction, with 70% of variance controlled by additive by environment interaction.

For HMF, a total of six QTL were identified in three environments. The detected QTL were dispersed over chromosomes 1, 5, and 6. The phenotypic variance explained by additive effect QTL was 5.2% to 38.9%. The chromosome 5 region explained a substantial fraction of the phenotypic variance. Like qAE5, qPP5, and qTS5, qHMF5 explained substantial phenotypic variance: 23.6% at PI, 36.0% at AEAE, and 38.9% for AEAL (Table 5). The QTL by environmental interaction analysis uncovered a phenotypic variance explained was 45.7% for qHMF5b (Table 6). Only qHMF6b showed a strong QTL by environment interaction, with 71% of variance controlled by additive by environment interactions.

Discussion

The limiting factor in producing doubled haploids is haploid male fertility. Ren et al.[23] and Wu et al.[28] evaluated haploids based on anther emergence score, anther emergence rate, pollen production score, and pollen production rate. Ma et al.[24] evaluated haploids by HMF rate. Different scoring methods have been developed as a result of the complex phenotypic expression of SHGD within the male reproductive system. In this study, we evaluated AE and PP following

Wu et al.[28] and Ren et al.[23], HMF following Ma et al.[24], and scored TS as well. The entry-mean based heritabilities were moderate at 0.47 for AE, 0.48 for PP, 0.53 for TS and high for HMF at 0.85 (Table 3). Heritabilities have been reported for HMF at 0.62[26] and 0.65[24]. All traits were closely correlated with each other, with the highest correlation found between TS and HMF.

The close positive correlations should help to reduce and simplify data collection by evaluating only a single or at best two of the four traits. SHGD evaluation within the male reproduction is difficult as a result of the polygenic nature of the trait. In comparing methods of assessment, the HMF rate is the most straightforward evaluation of SHGD. It can be used to rapidly and efficiently screen a sizable amount of germplasm for SHGD and drastically reduces the subjectivity of scoring on a 1 to 5 scale. HMF rate would be most important for breeding because of the focus on plants producing pollen. Assuming simultaneous availability of a fertile ear and pollen, a few pollen grains should be sufficient for self-pollination. In contrast, scoring for AE and PP is beneficial since the phenotypic expression of SHGD is highly variable based on this study and Ma et al.[24] This intensive phenotyping would be needed for characterization and identification of genetic mechanisms underlying SHGD. It could additionally be necessary for evaluation of new SHGD donors for introgression into breeding germplasm and fine mapping, due to expression differences between lines and families. The challenges of visual scoring of AE, PP, and TS include the subjectivity of scoring these traits and the tedious nature of phenotyping. The implementation of high throughput imaging and machine learning algorithms could be combined to produce more accurate and precise phenotypes over visual observations.

Genotypic variation was found to be highly significant for all traits studied, showing that there were significant differences between families (Table S1-S4). QTL analyses were completed

using ordinal logistic regression (AE, PP, TS) and logit-transformed data (HMF). All QTL identified were novel based on current literature. A novel large effect QTL with pleiotropic effects was found on chromosome 5 between the positions of 91-93 cM. It showed pleiotropic effects for AE, PP, TS, and HMF and was found to be stable across all three environments tested.

The QTL region spans markers S5.86261290-S5.92805032 (Table 5). This QTL is located near the centromere of chromosome 5, which is located from 101.6 Mb to 104.8 Mb, based on the ZmB73v1 genome assembly[38]. This makes fine mapping of this region challenging since recombination is repressed near the centromere. A minor effect QTL with pleiotropic effects was also found on chromosome 6, which is located from 79.5 Mb to 82.5 Mb. It showed pleiotropic effects for AE, PP, TS, and HMF as well. Only two QTL, qTS6b and qHMF6b showed a high level of QTL by environment interactions (Table 6). All but these QTL were largely additive in nature, which concurs with results found by Wu et al[28], and is expected as phenotyping was done on haploids plants.

Breeding of SHGD in DH System

Across all locations, 73% of A427 haploids exerted anthers and shed pollen, 44% of CR1Ht haploids displayed anthers and pollen, and 31% of haploids from the mapping population displayed HMF (Table 1). Based on QTL results from this study and Ma et al.[24], SHGD is quantitative in character and has a complex genetic architecture. QTL were identified from both parents. This supports that different genetic backgrounds carry QTL that contribute to SHGD. Screening material for SHGD may be an effective strategy to identifying new sources of SHGD. After the identification of the new sources of SHGD, the favorable alleles could be pyramided with the goal of improving SHGD using marker-assisted backcrossing or similar approaches.

There are two challenges in using SHGD within commercial DH breeding programs. First, the derived haploid must be both male and female fertile to produce an ear with seed set as well as synchronization of anthesis silking interval. The second challenge is the introgression of SHGD within breeding germplasm. A breeding program would require resources to effectively and efficiently introgress this trait into their germplasm. Marker-assisted backcrossing, or gene editing of major QTL after a causal gene is identified could be the most promising approaches. The complex nature makes understanding of the biological control difficult, because of the highly variable expression of HMF within and among lines. Further studies examining the genetic mechanism of SHGD and its biology be would be beneficial for the advancement and development of SHGD breeding materials.

In summary, a high-density linkage map was developed using GBS SNPs to identify QTL controlling SHGD. The large effect QTL on chromosome 5 explains over 45% of the observed variance for all four traits across all three environments. The use of introgression projects for SHGD into germplasm could make SHGD derived commercial DH lines economically viable.

Author Contributions

B.T. – design the experiment, phenotyping, data, construction of linkage map, statistical and QTL analysis, interpretation, and manuscript writing. U.F. – Critical commentary. J.R. – Critical commentary. S.C. – Critical commentary. T.L. – project leader and corresponding author who oversaw project conceptualization and implementation, and manuscript development. All authors and revised and approved the manuscript.

Acknowledgments

A special thanks to Katherine Goode for assistance with statistical analysis. The authors are thankful for financial support from USDA NIFA project IOW04314 and IOW01018. BT was supported by the Plant Sciences Institute, R.F. Baker Center for Plant Breeding at ISU, and K.J. Frey Chair in Agronomy at ISU.

References

1. Webel, O. D. & Lonnquist, J. H. An Evaluation of Modified Ear-to-Row Selection in a Population of Corn (Zea mays L.) 1. *Crop Science* **7**, 651–655 (1967).
2. Geiger, H. H. & Gordillo, G. A. Doubled haploids in hybrid maize breeding. *Maydica* **54**, 485–499 (2009).
3. Prigge, V. & Melchinger, A. E. Production of Haploids and Doubled Haploids in Maize. in *Plant Cell Culture Protocols* (eds. Loyola-Vargas, V. M. & Ochoa-Alejo, N.) 161–172 (Humana Press, 2012). doi:10.1007/978-1-61779-818-4_13
4. Coe, E. H. A Line of Maize with High Haploid Frequency. *The American Naturalist* **93**, 381–382 (1959).
5. Rotarenco, V., Dicu, G., State, D. & Fuia, S. New inducers of maternal haploids in maize. *Maize Genetics Cooperation Newsletter* 21–22 (2010).
6. Prigge, V., Schipprack, W., Mahuku, G., Atlin, G. N. & Melchinger, A. E. Development of in vivo haploid inducers for tropical maize breeding programs. *Euphytica* **185**, 481–490 (2012).
7. Chaikam, V. *et al.* Analysis of effectiveness of R1-nj anthocyanin marker for in vivo haploid identification in maize and molecular markers for predicting the inhibition of R1-nj expression. *Theor Appl Genet* **128**, 159–171 (2015).
8. Wang, H. *et al.* Fully-Automated High-Throughput NMR System for Screening of Haploid Kernels of Maize (Corn) by Measurement of Oil Content. *PLOS ONE* **11**, e0159444 (2016).

9. Faure, J.-E. *et al.* Double fertilization in maize: the two male gametes from a pollen grain have the ability to fuse with egg cells. *The Plant Journal* **33**, 1051–1062 (2003).
10. Zhao, X., Xu, X., Xie, H., Chen, S. & Jin, W. Fertilization and Uniparental Chromosome Elimination during Crosses with Maize Haploid Inducers. *Plant Physiology* **163**, 721–731 (2013).
11. Nanda, D. K. & Chase, S. S. An Embryo Marker for Detecting Monoploids Of Maize (Zea Mays L.) 1. *Crop Science* **6**, 213–215 (1966).
12. Ford, R. H. Inheritance of Kernel Color in Corn: Explanations & Investigations. *The American Biology Teacher* **62**, 181–188 (2000).
13. Chalyk, S. T. Properties of maternal haploid maize plants and potential application to maize breeding. *Euphytica* **79**, 13–18 (1994).
14. Borisy, G. G. & Taylor, E. W. THE MECHANISM OF ACTION OF COLCHICINE: Binding of Colchincine-3H to Cellular Protein. *The Journal of Cell Biology* **34**, 525–533 (1967).
15. Finkelstein, Y. *et al.* Colchicine poisoning: the dark side of an ancient drug. *Clin Toxicol (Phila)* **48**, 407–414 (2010).
16. Häntzschel, K. R. & Weber, G. Blockage of mitosis in maize root tips using colchicine-alternatives. *Protoplasma* **241**, 99–104 (2010).
17. Molenaar, W. S., Schipprack, W. & Melchinger, A. E. Nitrous Oxide-Induced Chromosome Doubling of Maize Haploids. *Crop Science* **58**, 650–659 (2018).
18. Melchinger, A. E., Molenaar, W. S., Mirdita, V. & Schipprack, W. Colchicine Alternatives for Chromosome Doubling in Maize Haploids for Doubled-Haploid Production. *Crop Science* **56**, 559–569 (2016).
19. Kleiber, D. *et al.* Haploid Fertility in Temperate and Tropical Maize Germplasm. *Crop Science* **52**, 623–630 (2012).

20. Chase, S. Spontaneous Doubling of the Chromosome Complement in Monoploid Sporophytes of Maize. *Proceedings of the Iowa Academy of Science* **56**, 113–115 (1949).
21. Chalyk, S. T. Properties of maternal haploid maize plants and potential application to maize breeding. *Euphytica* **79**, 13–18 (1994).
22. Jiang, L. *et al.* A study on natural recovery of tassel fertilization and doubling method in maize haploids. *Genet. Mol. Res.* **16**, (2017).
23. Ren, J., Wu, P., Tian, X., Lübberstedt, T. & Chen, S. QTL mapping for haploid male fertility by a segregation distortion method and fine mapping of a key QTL qhmf4 in maize. *Theor Appl Genet* **130**, 1349–1359 (2017).
24. Ma, H. *et al.* Genome-wide Association Study of Haploid Male Fertility in Maize (Zea mays L.). *Front. Plant Sci.* **9**, (2018).
25. *Compilation of North American Maize Breeding Germplasm*. (Crop Science Society of America, 1993). doi:10.2135/1993.compilationofnorthamerican.frontmatter
26. De La Fuente, G. Improvements to the maize (Zea mays L.) in vivo maternal doubled haploid system. *Graduate Theses and Dissertations* (2015). doi:https://doi.org/10.31274/etd-180810-4353
27. Technology - Haploid Inducing Genotype for Field and Specialty Corn. Available at: http://isurftech.technologypublisher.com/technology/20090. (Accessed: 11th November 2018)
28. Wu, P. *et al.* New Insights into the Genetics of Haploid Male Fertility in Maize. *Crop Science* **57**, 637–647 (2017).
29. Christensen, R. H. B. ordinal - Regression Models for Ordinal Data. (2018).
30. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, 2018).
31. SAS Institute. *SAS version 9.4*. (SAS Institute, 2013).

32. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379 (2011).

33. Glaubitz, J. C. *et al.* TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLOS ONE* **9**, e90346 (2014).

34. Buetow, K. H. Influence of aberrant observations on high-resolution linkage analysis outcomes. *Am J Hum Genet* **49**, 985–994 (1991).

35. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).

36. Miao, C. *et al.* Genotype-Corrector: improved genotype calls for genetic mapping in F 2 and RIL populations. *Scientific Reports* **8**, 10088 (2018).

37. Meng, L., Li, H., Zhang, L. & Wang, J. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *The Crop Journal* **3**, 269–283 (2015).

38. Wolfgruber, T. K. *et al.* Maize Centromere Structure and Evolution: Sequence Analysis of Centromeres 2 and 5 Reveals Dynamic Loci Shaped Primarily by Retrotransposons. *PLoS Genetics* **5**, e1000743 (2009).

Tables and Figures

Table 1. Phenotypic summary of the scores for anther emergence (AE), pollen production (PP), tassel size (TS), and the percentage of haploid male fertility (HMF) evaluated in three environments. Traits means are shown for both parents and $F_{2:3}$ families.

		AE				PP				TS				HMF (%)			
		AEAE	AEAL	PI	Average	AEAE	AEAL	PI	Average	AEAE	AEAL	PI	Average	AEAE	AEAL	PI	Average
Parents	A427	3.23	2.84	2.94	3.00	2.79	2.49	2.89	2.73	3.85	3.55	3.45	3.62	77.6	74.4	69.0	73.7
	CR1Ht	1.00	1.23	0.90	1.04	0.69	0.66	0.45	0.60	1.60	1.70	1.60	1.63	47.6	45.8	38.2	43.9
F2:3 Families	Mean	1.07	1.00	1.23	1.10	0.93	0.92	1.06	0.97	1.27	1.15	1.60	1.34	30.0	27.7	36.1	31.3
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0
	Max	4.10	4.38	4.22	4.38	4.00	5.00	4.33	5.00	4.40	5.00	4.44	5.00	90.0	100.0	90.0	100.0
	SD	0.87	0.83	0.90	0.87	0.77	0.79	0.83	0.80	0.92	0.89	0.97	0.94	20.4	20.3	21.1	20.9
	CV	81.0	82.4	72.9	78.9	82.8	85.4	78.0	82.1	72.7	77.6	60.5	70.7	68.1	73.4	58.4	66.8

Table 2. Phenotypic correlation coefficients for anther emergence (AE), pollen production (PP), tassel size (TS), and haploid male fertility (HMF).

	Coefficient of Correlation			
	AE	PP	TS	HMF
AE	1	0.95	0.92	0.93
PP		1	0.88	0.88
TS			1	0.97
HMF				1

Table 3. Variance components and entry-based heritabilities for anther emergence (AE), pollen production (PP), tassel size (TS), and haploid male fertility (HMF).

Variance Components	AE	PP	TS	HMF
σ^2_e	0.01189**	0.01776**	0.02058**	0.004331*
σ^2_g	0.4646**	0.3927**	0.1458**	0.04984**
σ^2_{g*e}	0.03096**	0.09176**	0.02221**	0.00155
σ^2_r	3.0311	2.3667	0.7278	0.04944
H^2	0.47	0.48	0.53	0.85

σ^2_e environment variance, σ^2_g genetic variance, σ^2_{g*e} genotype by environment variance, σ^2_r residual variance, H^2 entry-mean based heritability, * Significant difference at the 5% level, ** Significant difference at the 1% level

Table 4. Linkage map statistics for $F_{2:3}$ families between the cross of A427 and CR1Ht.

Chromosome	No. of Markers	Length in cM	Bins	Average Marker/ Bin Interval	Markers per cM
1	593	294.3	410	1.45	2.01
2	477	239.7	310	1.54	1.99
3	551	253.5	353	1.56	2.17
4	463	251.6	266	1.74	1.84
5	453	229.7	297	1.53	1.97
6	402	192.9	262	1.53	2.08
7	402	198.4	238	1.69	2.03
8	254	156.3	165	1.54	1.63
9	196	150.7	124	1.58	1.30
10	380	173.8	226	1.68	2.19
Total	4171	2141.1	2651	1.57	1.95

Table 5. List of significant QTL identified in three environments for AE, PP, TS, and MHF.

Trait	Env[1]	QTL	Chr[2]	Pos[3]	Marker Interval	LOD	PVE[4]	Add[5]
AE	AEAE	qAE5a	5	91	S5.86261290-S5.87949497	30.01	42.19	0.73
	AEAL	qAE1a	1	169	S1.210741020-S1.236903080	4.75	3.93	-0.11
		qAE5b	5	93	S5.92720589-S5.92805032	36.36	41.58	0.70
		qAE6a	6	82	S6.111018551-S6.111368312	7.51	6.50	0.27
		qAE7	7	52	S7.6435275-S7.10551600	4.60	4.00	0.03
		qAE10	10	94	S10.133552715-S10.133703892	4.27	3.51	-0.19
	PI	qAE5b	5	93	S5.92720589-S5.92805032	23.04	30.09	0.57
PP	AEAE	qPP5a	5	91	S5.86261290-S5.87949497	31.65	41.80	0.73
	AEAL	qPP1a	1	169	S1.210741020-S1.236903080	4.43	3.51	-0.09
		qPP5b	5	93	S5.92720589-S5.92805032	38.80	43.25	0.74
		qPP6a	6	82	S6.111018551-S6.111368312	7.64	6.33	0.27
		qPP7	7	52	S7.6435275-S7.10551600	5.30	4.43	0.04
	PI	qPP5b	5	93	S5.92720589-S5.92805032	23.93	32.52	0.59
TS	AEAE	qTS5a	5	91	S5.86261290-S5.87949497	25.98	39.98	0.68
	AEAL	qTS5b	5	93	S5.92720589-S5.92805032	32.21	38.62	0.64
		qTS5c	5	134	S5.191768713-S5.191990245	4.52	4.10	0.18
		qTS6a	6	82	S6.111018551-S6.111368312	7.49	7.13	0.26
		qTS7	7	52	S7.6435275-S7.10551600	4.79	4.67	0.00
		qTS10	10	87	S10.129355934-S10.130907871	4.67	4.30	-0.20
	PI	qTS5a	5	93	S5.92720589-S5.92805032	21.05	28.22	0.48
		qTS6b	6	84	S6.112600900-S6.112602435	8.09	9.42	0.27
HMF	AEAE	qHMF1b	1	167	S1.199730599-S1.210741020	4.42	5.19	-0.03
		qHMF5a	5	91	S5.86261290-S5.87949497	26.5	36.00	0.17
	AEAL	qHMF5b	5	93	S5.92720589-S5.92805032	27.15	38.94	0.17
		qHMF6a	6	82	S6.111018551-S6.111368312	5.15	6.04	0.06
	P1	qHMF5b	5	93	S5.92720589-S5.92805032	15.56	23.57	0.13
		qHMF6	6	84	S6.112600900-S6.112602435	8.38	11.77	0.09

[1]Environment. [2]Chromosome of identified QTL. [3]Position of the QTL in cM. [4]Phenotypic variance explained. [5]Additive Effect (Positive values signify that alleles came from A427 and negative values signify that alleles came from CR1Ht).

Table 6. List of significant QTL x environment interactions identified in three environments for AE, PP, TS, and HMF.

Trait	QTL	Chr[1]	Pos[2]	Marker Interval	LOD	LOD(A)[3]	LOD(AbyE)[4]	PVE[5]	PVE(A)[6]	PVE(AbyE)[7]	Add[8]
AE	qAE1b	1	168	S1.199730599-S1.210741020	8.59	7.33	1.25	5.74	5.01	0.72	-0.09
	qAE5b	5	93	S5.92720589-S5.92805032	59.81	44.56	15.25	51.32	35.91	15.41	0.44
	qAE6a	6	82	S6.111018551-S6.111368312	11.53	7.51	4.01	8.00	5.30	2.69	0.16
	qAE7	7	52	S7.6435275-S7.10551600	6.17	4.37	1.79	3.94	3.04	0.89	0.01
	qAE10	10	94	S10.133552715-S10.133703892	7.36	6.61	0.74	4.90	4.60	0.30	-0.14
PP	qPP1b	1	168	S1.199730599-S1.210741020	8.41	7.32	1.09	5.65	5.03	0.62	-0.08
	qPP5b	5	93	S5.92720589-S5.92805032	63.08	48.54	14.53	55.93	38.99	16.93	0.45
	qPP6a	6	82	S6.111018551-S6.111368312	11.16	7.10	4.06	7.53	4.92	2.61	0.15
	qPP7	7	52	S7.6435275-S7.10551600	7.38	5.50	1.88	4.79	3.80	0.98	0.02
TS	qTS5a	5	93	S5.92720589-S5.92805032	53.47	37.16	16.31	48.48	33.76	14.72	0.38
	qTS5c	5	134	S5.191768713-S5.191990245	6.32	4.02	2.29	4.85	3.18	1.67	0.09
	qTS6a	6	82	S6.111018551-S6.111368312	11.04	7.68	3.36	9.35	6.24	3.10	0.15
	qTS6b	6	84	S6.112600900-S6.112602435	8.14	2.21	5.92	5.99	1.78	4.20	0.08
	qTS7	7	52	S7.6435275-S7.10551600	7.30	5.54	1.76	5.60	4.39	1.20	0.01
	qTS10	10	87	S10.129355934-S10.130907871	6.30	3.74	2.56	4.97	2.99	1.97	-0.09
HMF	qHMF1a	1	169	S1.210741020-S1.236903080	7.8	0.00	7.8	6.4	5.29	1.11	-0.03
	qHMF1b	1	282	S1.297804454-S1.297977885	6.16	0.00	6.16	5.04	4.11	0.94	0.03
	qHMF5b	5	93	S5.92720589-S5.92805032	42.96	0.00	42.96	45.67	31.62	14.05	0.1
	qHMF6a	6	82	S6.111018551-S6.111368312	8.85	0.00	8.85	7.47	4.92	2.56	0.04
	qHMF6b	6	84	S6.112600900-S6.112602435	8.61	0.00	8.61	8.34	2.4	5.93	0.03

[1]Chromosome of identified QTL. [2]Position of the QTL in cM. [3]LOD score explained by additive effects. [4]LOD score explained by additive x environment effects. [5]Phenotypic variance explained. [6]Phenotypic variance explained by additive effects. [7]Phenotypic variance explained by additive x environment effects. [8]Additive Effect (Positive values signify that alleles came from A427 and negative values signify that alleles came from CR1Ht).

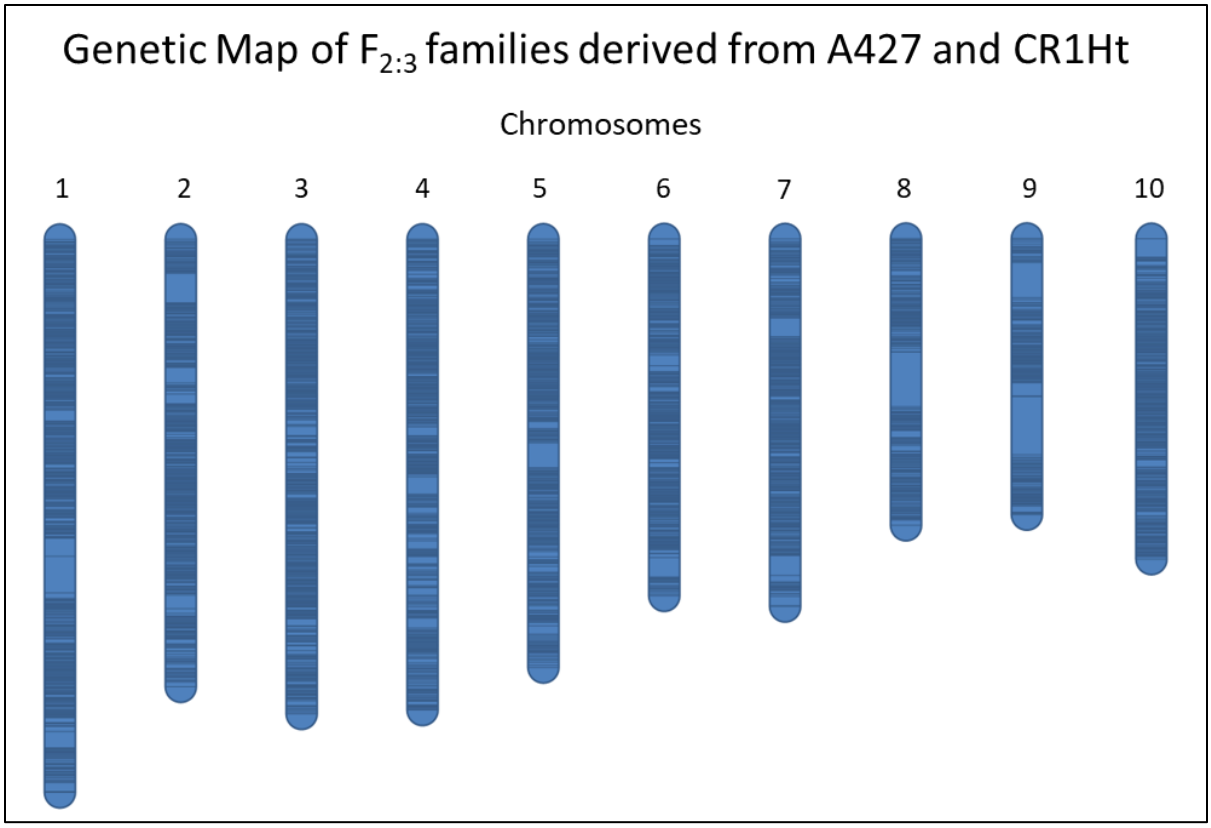

Figure 1. Linkage map constructed from $F_{2:3}$ families derived from A427 and CR1Ht.

Supplementary Tables

Table S1. Analysis of variance for anther emergence (AE) across three environments.

Source	DF	SS	MS	F Value	Pr > F
Environment	2	124.69	62.34	100.54	0.0002**
Replication (Env)	3	1.85	0.62	0.2	0.8938^{ns}
Genotype	221	8261.48	37.38	9.9	<.0001**
Environment*Genotype	442	1673.69	3.79	1.25	0.0003**
Residual	15783	47840	3.03		

DF = Degrees of freedom, SS = Sum of Squares, MS = Mean Square, * Significant difference at the 5% level, ** Significant difference at the 1% level, ^{ns} Non-significant

Table S2. Analysis of variance for pollen production (PP) across three environments.

Source	DF	SS	MS	F Value	Pr > F
Environment	2	40.58	20.29	26.01	0.0012**
Replication (Env)	3	2.35	0.78	0.33	0.8027^{ns}
Genotype	221	2533.43	11.46	3.78	<.0001**
Environment*Genotype	428	1317.24	3.08	1.3	<.0001**
Residual	4623	10941	2.37		

DF = Degrees of freedom, SS = Sum of Squares, MS = Mean Square, * Significant difference at the 5% level, ** Significant difference at the 1% level, ^{ns} Non-significant

Table S3. Analysis of variance for tassel size (TS) across three environments.

Source	DF	SS	MS	F Value	Pr > F
Environment	2	43.43	21.71	30.15	0.0012**
Replication (Env)	3	2.17	0.72	0.99	0.3941^{ns}
Genotype	220	863.54	3.93	4.44	<.0001**
Environment*Genotype	428	382.9	0.89	1.23	0.0014**
Residual	4471	3254.03	0.73		

DF = Degrees of freedom, SS = Sum of Squares, MS = Mean Square, * Significant difference at the 5% level, ** Significant difference at the 1% level, ^{ns} Non-significant

Table S4. Analysis of variance for haploid male fertility (HMF) across three environments.

Source	DF	SS	MS	F Value	Pr > F
Environment	2	3.96	1.98	40.88	0.015*
Replication (Env)	3	0.14	0.05	0.93	0.4262^{ns}
Genotype	221	78.44	0.35	6.75	<.0001**
Environment*Genotype	442	23.24	0.05	1.06	0.236^{ns}
Residual	680	33.62	0.05		

DF = Degrees of freedom, SS = Sum of Squares, MS = Mean Square, * Significant difference at the 5% level, ** Significant difference at the 1% level, ^{ns} Non-significant

CHAPTER FOUR: GENERAL CONCLUSIONS

The maize *in vivo* doubled haploid system has been a very valuable tool for plant breeders because of its easy ability to quickly produce inbred lines. A significant amount of research that has been completed and applied to made doubled haploids what they are today. There have been significant advancements in increased haploid induction rate, the selection of haploids, and doubling of haploid genomes.

In this study, two new traits of interest and their genetic control were evaluated. Inducibility was found to be a quantitative trait with strong influence by the environment. There were no stable or major QTL found for inducibility within this study. In contrast, spontaneous haploid genome doubling (SHGD) showed promise for transformative innovation in maize doubled haploid technology. A large stable QTL was found within this study. This major QTL can help to overcome the need of artificial haploid genome doubling by colchicine, which is both laborious and requires work with a toxic chemical. Incorporation of this major QTL using marker-assisted backcrossing into other elite germplasm is straightforward.

New advancements continue to make doubled haploid system easier and easier to deploy into a breeding program. With the current research on gene editing, inducers and doubled haploids may offer future avenues in incorporating new traits into corn hybrids. Doubled haploids have solidified their place in maize breeding programs and will continue to provide value to maize breeders.