New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize

Vanessa Prigge*†, 1, Xiaowei Xu‡, 1, Liang Li‡, Raman Babu†, Shaojiang Chen‡, Gary N. Atlin†, and Albrecht E. Melchinger*

1 These authors contributed equally.

*Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany,
† International Maize and Wheat Improvement Center (CIMMYT), 06600 Mexico DF, Mexico,
‡ National Maize Improvement Center of China, China Agricultural University, 100193 Beijing, China
Genetics of \textit{in vivo} Haploid Induction in Maize

Keywords: \textit{in vivo} haploid induction, maize, QTL mapping, doubled haploid, haploid inducer development

Corresponding author:
Albrecht E. Melchinger,
Institute of Plant Breeding, Seed Science, and Population Genetics,
University of Hohenheim,
70593 Stuttgart,
Germany.
E-mail: melchinger@uni-hohenheim.de
ABSTRACT

Haploids and doubled haploid (DH) inbred lines have become an invaluable tool for maize genetic research and hybrid breeding, but the genetic basis of in vivo induction of maternal haploids is still unknown. This is the first study reporting comparative quantitative trait locus (QTL) analyses of this trait in maize. We determined haploid induction rates (HIR) in testcrosses of a total of 1061 progenies of four segregating populations involving two temperate haploid inducers, UH400 (HIR=8%) and CAUHOI (HIR=2%), one temperate and two tropical inbreds with HIR=0%, and up to three generations per population. Mean HIR of the populations ranged from 0.6 to 5.2% and strongly deviated from the mid-parent values. One QTL (qhir1) explaining up to $\hat{p} = 66\%$ of the genetic variance was detected in bin 1.04 in the three populations involving a non-inducer parent and the HIR-enhancing allele was contributed by UH400. Segregation ratios of loci in bin 1.04 were highly distorted against the UH400 allele in these three populations, suggesting that transmission failure of the inducer gamete and haploid induction ability are related phenomena. In the CAUHOI×UH400 population, seven QTL were identified on five chromosomes, with qhir8 on chromosome 9 having $\hat{p} > 20\%$ in three generations of this cross. The large-effect QTL qhir1 and qhir8 will likely become fixed quickly during inducer development due to strong selection pressure applied for high HIR. Hence, marker-based pyramiding of small-effect and/or modifier QTL influencing qhir1 and qhir8 may help to further increase HIR in maize.

We propose a conceptual genetic framework for inheritance of haploid induction ability, which is also applicable to other dichotomous traits requiring progeny testing, and discuss the implications of our results for haploid inducer development.
In vivo induction of maternal haploids in maize has paved the way for large-scale production of doubled haploid (DH) inbred lines which nowadays form the backbone of the global hybrid maize industry. Traditionally, the maize plants’ cross-breeding nature required recurrent self-pollinations for 6-10 generations to obtain sufficiently homozygous inbred lines (HALLAUER et al. 2010). Application of DH technology reduces the time for inbred development by more than half compared to the traditional method and additionally provides several quantitative genetic, operational, logistical, and economic advantages (NEI 1963; SCHMIDT 2003; MELCHINGER et al. 2005; SEITZ 2005; SMITH et al. 2008; CHANG and COE 2009; GEIGER 2009). By using haploid inducer genotypes as pollinators in crosses with source germplasm, ears obtained carry a proportion of seeds containing haploid embryos of maternal origin. Subsequent treatment of haploids with mitotic inhibitors facilitates chromosome duplication resulting in diploid and completely homozygous inbred lines (see PRIGGE and MELCHINGER (2011) for a detailed description of DH production).

Modern maize inducers have haploid induction rates (HIR) of about 8% on average (e.g., RÖBER et al. 2005; PRIGGE et al. 2011). The genetic mechanisms underlying in vivo induction of maternal haploids in maize are not yet fully understood. The two major hypotheses for possible mechanisms are: (i) failure of fertilization of the egg cell and subsequent parthenogenetic development of the reduced egg into a haploid embryo (SARKAR and COE 1966; CHALYK et al. 2003; BARRET et al. 2008), and (ii) normal fertilization followed by elimination of inducer chromosomes (FISCHER 2004; ZHANG et al. 2008; LI et al. 2009). Further, KATO (1997) suggested that the aberrant fertilization mechanisms leading to haploidy
may be related to mechanisms leading to hetero-fertilization.

Continuous variation of segregating populations developed from inducer by
non-inducer crosses suggests that HIR is a quantitative trait (Lashermes and
Beckert 1988). The first exploratory quantitative trait locus (QTL) mapping
study, conducted with RFLP markers in an F$_3$ population involving Stock6 (HIR
$= 2.3\%$; Coe 1959) as inducer parent, provided evidence for two QTL for HIR
on chromosomes 1 and 2, together explaining 17.9% of the phenotypic variance
(Deimling et al. 1997). Barret et al. (2008) used marker segregation ratio
distortion analyses in small samples from the two phenotypic extremes of a segre-
gating population developed from a cross between a non-inducer and an inducer
line and also identified a locus on chromosome 1 associated with HIR.

None of the most recently developed haploid inducers with high HIR have been
subjected to genome-wide QTL analysis yet. Further, QTL must be examined
for stable expression in different germplasm because high congruency of QTL in
different genetic backgrounds is desirable to facilitate marker-assisted introgression
approaches. Therefore, we conducted comparative QTL analyses for HIR in four
populations involving two inducers, UH400 and CAUHOI. Our objectives were to
(1) study the inheritance of in vivo haploid induction ability and its association
with segregation distortion and embryo abortion rate (EAR), (2) estimate the
number, genomic positions, and genetic effects of QTL associated with HIR and
EAR, and (3) discuss possible mechanisms underlying in vivo haploid induction
in monocots as well as implications of the results for fine-mapping and improving
HIR in maize.
MATERIALS AND METHODS

**Genetic Materials:** Four mapping populations were generated involving haploid inducer inbred UH400 as pollinator with two temperate (CAUHOI, 1680) and two tropical (CML395, CML495) inbreds as female parents (Table 1). UH400 was developed at the University of Hohenheim, Germany, and has a HIR of approximately 8% averaged across several tropical (Prigge *et al.* 2011) and temperate (W. Schipprack 2011, pers. comm.) environments. It carries the dominantly inherited marker gene *R1-nj* conferring a purple coloration of the scutellum and the aleurone of seeds (Nanda and Chase 1966; Neuffer *et al.* 1997), which can be used as embryo and endosperm marker, respectively, to identify putative haploid seeds. CAUHOI is a haploid inducer developed by China Agricultural University with high kernel oil content (78 g kg$^{-1}$) and a HIR of approximately 2% (Li *et al.* 2009). Inbreds 1680 (developed by China Agricultural University) as well as CML395 and CML495 (both developed by the International Maize and Wheat Improvement Center, CIMMYT) have no induction ability but excellent agronomic characteristics.

In all four populations, random F$_2$ plants were selfed to produce F$_{2:3}$ lines. In addition, for population CAUHOI×UH400, randomly chosen seeds from self-pollinated ears of F$_{2:3}$ lines were bulked and grown as F$_{2:4}$ lines. In summary, F$_2$, F$_{2:3}$, and F$_{2:4}$ generations were evaluated for HIR in CAUHOI×UH400 (designated CAU-F$_2$, CAU-F$_3$, and CAU-F$_4$, respectively), F$_2$ and F$_{2:3}$ generations in 1680×UH400 (1680-F$_2$ and 1680-F$_3$), and the F$_{2:3}$ generation was evaluated in CML395×UH400 (CML395-F$_3$) and in CML495×UH400 (CML495-F$_3$) (Table 1).
Marker Assays and Linkage Map Construction: Genotyping was performed differently for the two temperate (CAU-F$_2$, 1680-F$_2$) and the two tropical (CML395-F$_2$, CML495-F$_2$) populations following the protocols used at China Agricultural University and CIMMYT, respectively. For CAU-F$_2$ and 1680-F$_2$, leaf tissue from the parental lines and the F$_2$ plants was harvested at the three-leaf stage. Genomic DNA was extracted following the procedures of Porebski et al. (1997). One-hundred eighty-six F$_2$ individuals of populations CAU-F$_2$ and 1680-F$_2$ were genotyped with 90 and 113 polymorphic simple sequence repeat (SSR) markers, respectively, following standard protocols (Ninamango-Cárdenas et al. 2003). The SSRs were selected from the MaizeGDB database (http://www.maizegdb.org, Lawrence et al. 2004) to provide good coverage of the entire genome.

For CML395-F$_2$ and CML495-F$_2$, leaf tissue from the parental lines and the F$_2$ plants was harvested at the 10-leaf stage, placed in glassine bags, and stored at -80°C until lyophilization at CIMMYT’s biotechnology facility. Genomic DNA was extracted following standard protocols (CIMMYT 2005). Parents of both populations were screened for polymorphism with the custom GoldenGate assay containing 1536 single nucleotide polymorphism (SNP) markers (Yan et al. 2009). SNP genotyping for the parental screening was performed using the Illumina BeadStation 500 G (Illumina, San Diego, CA) at the Cornell University Life Sciences Core Laboratories Center according to the protocol of Fan et al. (2006) and data were analyzed using the Illumina BeadStudio genotyping software. The genotyping of 220 (CML395-F$_2$) and 233 (CML495-F$_2$) F$_2$ individuals with 324 and 297 polymorphic SNPs, respectively, was conducted by KBioSciences (UK) using simple-plex, automated high throughput KasPar chemistry assays. Genotypes for
all populations are available as Suppl. Material 1.

The four linkage maps were constructed by Haldane’s mapping function with software JoinMap 3.0 (VAN OOLJEN and VOORRIPS 2001). Individuals and loci with > 20% missing data were excluded from the analysis. A LOD (log_{10} of the likelihood odds ratio) value of 3.0 was used as critical threshold to declare linkage between two markers. The order of the mapped SSR markers on each linkage group was confirmed according to the IBM2 2008 Neighbors Map (http://www.maizegdb.org). SNP marker order was confirmed by their absolute physical location on the B73 reference maize sequence (http://www.maizesequence.org) and bin positions were identified from the sequence. As some markers were very close (<1.01 cM) to each other, they were combined into a “synthetic” marker resulting in a reduced number of loci used for QTL analysis (Table 1). Further, we tested expected allele frequencies of p = 0.5 of the UH400 allele with a standard χ^2 goodness-of-fit test and declared segregation distortion when loci significantly (P < 0.05) deviated from the expectation.

Assessment of Haploid Induction and Embryo Abortion Rates: The mapping populations were grown in completely randomized designs at the experimental stations of China Agricultural University near Beijing (China) in summer 2008 (CAU-F_2 and 1680-F_2) and 2009 (CAU-F_4 and 1680-F_3), and in Hainan province (China) in winter 2008/2009 (CAU-F_3), as well as at CIMMYT’s experimental station in Tlaltizapan (Mexico) in summer 2009 (CML395-F_3 and CML495-F_3). Field plots consisted of 11 plants in each experiment.

Determination of HIR was conducted based on two haploid identification systems. For the CAU and 1680 mapping populations, the R1-nj color marker in-
Inherited from the inducer parent UH400 was scored on testcross seeds harvested from the non-pigmented Chinese hybrid ZD958 after pollination with the mapping populations. Three to five plants of ZD958 were pollinated with either pollen of individual F_2 plants (CAU-F_2 and 1680-F_2) or bulked pollen of a minimum of five plants per family (CAU-F_3, 1680-F_3, and CAU-F_4). Testcross seed with purple endosperm (indicating a regular triploid endosperm resulting from successful fertilization with inducer pollen) but colorless embryo (indicating a haploid embryo of ZD958 origin) were designated as putative haploid seed following Prigge and Melchinger (2011). To verify the effectiveness of the R1-nj color marker and to account for putative misclassifications (Röber et al. 2005; Prigge et al. 2011), all haploids detected by the above method were grown near Beijing in summer 2010 and visually scored for plant type before flowering; haploid plants had shorter statures, thin stalks, and erect and narrow leaves. For 1680-F_2 and 1680-F_3, we additionally determined the embryo abortion rate (EAR) as the proportion of completely collapsed testcross seeds exhibiting embryoless pericarps.

For the CML mapping populations, a single cross obtained by crossing two temperate inbred lines carrying the recessive mutation liguleless was used as tester (hereafter referred to as lglg tester). The specific genetic source (lg1 vs. lg2) is unknown for this germplasm. The lglg tester allows clear distinction between haploid and diploid seedlings because haploids are characterized by absence of the ligule and auricle combined with an upright-positioned leaf enveloping the shoot (Neuffer et al. 1997). All lglg tester plants wereemasculated to avoid self-contamination. Bulked pollen of a minimum of five plants per F_3 family was used to pollinate a minimum of eight lglg tester plants. All testcross seeds were grown.
in the greenhouse until the four-leaf stage and were visually rated for the *liguleless* phenotype, an indication of haploidy.

For all populations, each testcross ear was shelled separately. Three testcross ears per family were selected based on the highest total number of testcross seeds to determine HIR (in all populations) and EAR (in 1680-F$_2$ and 1680-F$_3$). Repeatabilities ($w^2$) for HIR and EAR were calculated with the statistical software PLABSTAT (Utz 2004) taking the values of the three individual ears as three replicates in a completely randomized design. Combining the three ears, HIR and EAR were computed for each family as the proportion of putative haploid and embryoless testcross seeds, respectively, detected in the total testcross progeny monitored. Parent-offspring regression was performed with CAU-F$_2$, CAU-F$_3$, CAU-F$_4$, 1680-F$_2$, and 1680-F$_3$ and the regression coefficient was used as an estimator of the narrow-sense heritability ($h^2$) for HIR and EAR of the corresponding populations.

**QTL Analyses:** Analyses of QTL for HIR and EAR were performed with software PlabQTL (Utz and Melchinger 1996) using composite interval mapping (CIM, Jansen and Stam 1994; Zeng 1994) and a multiple regression procedure (Haley and Knott 1992). The appropriate numbers of marker cofactors and the appropriate genetic models were chosen based on the smallest values of the modified Bayesian Information Criterion (mBIC, Baierl et al. 2006). Using the mBIC as significance criterion, we did not detect significant first-order epistatic interactions in any of the populations. Minimized mBIC were obtained for a model including additive and dominance effects for all populations except CAU-F$_2$ and CML495-F$_3$ for HIR as well as 1680-F$_2$ and 1680-F$_3$ for EAR, where no significant
dominance effects were detected. Critical LOD thresholds were determined empirically with 1000 random permutations (Churchill and Doerge 1994) for each generation, population, and trait. LOD thresholds corresponding to genome-wide error rates of \( \alpha = 0.25 \) were employed for CIM, which is common for explorative QTL analyses (Beavis 1998). The support intervals with a LOD fall-off of 1.0 were expressed as position on the chromosome. The proportion of the genetic variance (\( \sigma^2_g \)) explained by the model was estimated as \( \hat{p} = \frac{R^2_{\text{adj}}}{w^2} \), where \( R^2_{\text{adj}} \) is the adjusted proportion of phenotypic variance explained by the model and \( w^2 \) is the repeatability of the trait. The proportion of \( \sigma^2_g \) explained by individual QTL was estimated as \( \hat{p}_{\text{part}} = \frac{R^2_{\text{part}}}{w^2} \), where \( R^2_{\text{part}} \) is the proportion of phenotypic variance explained by that QTL. All values of \( \hat{p}_{\text{part}} \) were normalized to sum to \( \hat{p} \). For those loci showing significant dominance effects, we calculated the dominance ratio \((d/|a|)\) of the estimated dominance effect over the absolute value of the estimated additive effect.

Further, to account for the special features of the phenotypic distributions of HIR in QTL mapping, we used the two-part model developed by Broman (2003) implemented in the software R/QTL (Broman and Sen 2009) in addition to CIM. This model combines the analysis of the binary trait (i.e., \( z_i = 1 \) vs. \( z_i = 0 \), corresponding to \( \text{HIR} > 0\% \) vs. \( \text{HIR} = 0\% \), respectively, of inducer F\(_2\) plant or F\(_{2.3}\)/F\(_{2.4}\) family \( i \)) with standard interval mapping for those having \( z_i = 1 \). For plants/families with haploid induction ability (i.e., \( z_i = 1 \)) and genotype \( g \) (AA, AB, or BB) at marker locus \( m \) in the parental F\(_2\) plant, the square root HIR is assumed to follow a normal distribution with mean \( \mu_g \) and standard deviation \( \sigma \) (Broman 2003). As HIR is determined based on the testcross progeny of \( i \),

\( \sigma_g \) (Broman 2003). As HIR is determined based on the testcross progeny of \( i \),
the phenotype of each testcross seed $ij$ can be coded by an indicator variable $z_{ij}$, with $z_{ij} = 1$ if seed $ij$ is haploid or $z_{ij} = 0$ otherwise. Further, $\pi_g = Pr(z_i = 1|m_i = g)$ denotes the conditional probability of obtaining $z_i = 1$, if the genotype of the parental F$_2$ plant $i$ at marker locus $m$ is $g$, whereas $\mu_g = E(z_{ij}|m_i = g, z_i = 1)$ denotes the conditional expectation for testcross seed $ij$ within the progeny of inducer $i$ with HIR $> 0\%$, if the genotype of the parental F$_2$ plant $i$ at marker locus $m$ is $g$. Separate LOD scores were calculated to test whether the three genotype classes $g$ have the same probability $\pi_g$ of inducing haploidy (i.e., $z_i = 1$) and/or whether those families with $z_i = 1$ have the same sample means $\mu_g$ (corresponding to HIR) according to three hypotheses (Broman 2003): (1) LOD($\pi$) to test whether $\pi_{AA} = \pi_{AB} = \pi_{BB}$; (2) LOD($\mu$) to test whether $\mu_{AA} = \mu_{AB} = \mu_{BB}$; and (3) LOD($\pi\mu$) to test the combined hypothesis that both the $\pi_g$ and $\mu_g$ were equal. The corresponding critical LOD thresholds for the two-part model were determined empirically with 1000 random permutations.
RESULTS

Haploid Induction and Embryo Abortion Rates: The average number of testcross seeds used for determination of HIR was lowest in CML395-F_3 (153) and highest in CAU-F_4 (317; Table 1) and ranged from 50 to 640 for individual families. Genotypic variances for HIR were significant \((P < 0.01)\) in all populations. Repeatabilities \((w^2)\) were intermediate for CML395-F_3 and CAU-F_3 and high for all other populations. Mean HIR were highest in CAU-F_2 and lowest in the two CML populations and HIR ranged from 0 to 15.6% for individual families (Table 1, Figure 1). Distributions of HIR were strongly skewed towards low values in populations where non-inducers (1680, CML395, CML495) were used as parents. Population means deviated significantly \((P < 0.001, \text{ Wilcoxon signed rank test})\) from the mid-parent values towards lower values in all but the CAU-F_2 population. CML395-F_3 and CML495-F_3 harbored a larger number of families with HIR = 0% than the other populations. Heritability estimates for HIR obtained from parent-offspring regression (Figure 2) were lower for the CAU population \((h^2 = 0.32 \text{ and } 0.41)\) than for the 1680 population \((h^2 = 0.80)\). Spearman’s rank correlations were higher for HIR between F_2 and F_3 generations in both the CAU \((r_s = 0.51^{**})\) as well as 1680 \((r_s = 0.71^{**})\) populations than for HIR between F_3 and F_4 generations of the CAU population \((r_s = 0.26^{**})\).

Significant \(\sigma^2_g\) and high \(w^2\) were estimated for EAR in 1680-F_2 and 1680-F_3 (Table 3). Mean EAR was higher in 1680-F_3 than in 1680-F_2 and the range of EAR was also wider in 1680-F_3. Heritability of EAR was 0.77 based on parent-offspring regression (Suppl. Fig. 1) and Spearman’s rank correlation between EAR in F_2 and F_3 generations was 0.37 \((P < 0.001)\). Moderate correlations existed.
between EAR and HIR of 1680-F\(_2\) \((r_s = 0.63^{***})\) and 1680-F\(_3\) \((r_s = 0.53^{***})\).

**Linkage Maps and Segregation Distortion:** The length of the four genetic linkage maps ranged from 1353 to 1630 cM and the average marker intervals were 4.8 cM in CML395-F\(_2\), 5.2 cM in CML495-F\(_2\), 13.1 cM in 1680-F\(_2\), and 20.4 cM in CAU-F\(_2\) (data not shown). Significant deviations from expected allele frequencies were observed in all populations (Figure 3). Most notably, strong \((P < 0.001)\) distortion occurred against the UH400 allele in bin 1.04 in 1680-F\(_2\), CML395-F\(_2\), and CML495-F\(_2\), and in a varying number of adjacent bins depending on the population. Chromosome 4 was affected in three of the four populations (although distortion happened against different alleles), while chromosome 7 was affected by segregation distortion only in CAU-F\(_2\) and 1680-F\(_2\) and chromosomes 5 and 10 were only affected in CML395-F\(_2\) and CML495-F\(_2\). Across the four populations, distortion was observed against the UH400 allele in 17 bins and against the non-UH400 allele in 11 bins.

**Quantitative Trait Locus Analyses:** Five QTL for HIR were detected with CIM in CAU-F\(_2\) on chromosomes 3, 5, and 9 explaining together \(\hat{p} = 71\%\) of the genetic variance (Table 2, Figure 3). In CAU-F\(_3\) three QTL were detected on chromosomes 4, 7, and 9 with \(\hat{p} = 64\%\), while in CAU-F\(_4\) one QTL was detected on chromosome 9 (\(qhir8; \hat{p} = 20\%\)). The HIR-enhancing allele was always contributed by UH400 except for \(qhir5\) in bin 5.01 of CAU-F\(_2\).

In 1680-F\(_2\) and 1680-F\(_3\), a major QTL (\(qhir1; \hat{p}_{\text{part}} = 63–66\%)\) for HIR was detected on chromosome 1 (bin 1.04) and the favorable allele was contributed by UH400 in both generations. Furthermore, a minor QTL on chromosome 3 (\(qhir2\)) was detected in 1680-F\(_2\) resulting in \(\hat{p} = 71\%\) for 1680-F\(_2\) and the favorable allele at
qhir2 was contributed by the non-inducer parent 1680. One QTL for EAR (qear1) was detected on chromosome 1 (bin 1.04) in 1680-F2 (\( \hat{p} = 38\% \)) and 1680-F3 (\( \hat{p} = 14\% \); Table 3). qear1 shared one flanking marker each with qhir1 in 1680-F2 (umc1917) and 1680-F3 (umc1811). The EAR-enhancing allele was from UH400.

In CML395-F3 and CML495-F3, qhir1 was detected in bin 1.04 explaining \( \hat{p} = 20 \) and 9\%, respectively. In both CML populations, SNP PHM5306 flanked qhir1 and the favorable allele was contributed by UH400. Based on the physical positions of flanking SNPs (Table 2), the delimited physical interval for qhir1 was about 40Mbp (data not shown).

Significant dominance effects were revealed for qhir8 in CAU-F4 as well as for qhir1 in 1680-F2, 1680-F3, and CML395-F3 (Table 2). At these loci, the dominance ratio ranged from -0.37 to -1.71 indicating weak to strong recessive gene action adopting the nomenclature suggested by Stuber et al. (1987). Individuals homozygous for the UH400 allele at the nearest flanking marker (see Table 2) to qhir1 always showed HIR > 0\% in 1680-F2, while about one third of the individuals of this genotype class showed HIR = 0\% in CML495-F3 (Figure 4).

Using the two-part model approach on HIR data, we confirmed qhir2, qhir6, and qhir8 for CAU-F2; qhir4, qhir7, and qhir8 for CAU-F3; and qhir8 for CAU-F4 (Table 2). Furthermore, qhir1 of 1680-F2, 1680-F3 and CML495-F3 was confirmed with this method. All QTL identified with the two-part model only affected the mean HIR of the non-null phenotypes, except for qhir1 in 1680-F2 (Figure 5), which additionally affected the probability of occurrence of non-null phenotypes.
DISCUSSION

Conceptual genetic framework for inheritance of haploid induction ability: Results of our study and others (Lashermes and Beckert 1988; Barret et al. 2008) showed that (1) the phenotypic distributions of HIR obtained with segregating populations are mostly skewed and (2) a large proportion of null phenotypes is observed. This suggests that the properties of haploid induction ability differ from those of most agronomic traits, which display continuous phenotypic variation that usually approximates a Gaussian distribution.

Haploid induction ability of maize inducer genotypes is considered to be a pollen characteristic, because pollen of inducers under evaluation is used to pollinate a tester and expression of haploidy is assessed on the testcross seed. Under this premise, the genotype of the inducer as such does not matter, but only its gametic array transmitted through the pollen. Hence, dominance effects should be absent for HIR because testcross genotypic values follow a simple additive model in the absence of epistasis (Bernardo 2002). However, segregation distortion as detected in our study at four QTL in three populations can potentially mimic dominance, because if gene frequencies in the gametic arrays produced by heterozygous F$_2$ individuals and even more so of their selfing progenies in generations F$_3$ and F$_4$ deviate from 0.5, this causes a deviation from the additive genetic model for testcross progeny. Nevertheless, it cannot be ruled out entirely that haploid induction ability is not only a property of the pollen itself but also influenced by its parental genotype, similar to factors causing sporophytic self-incompatibility in plants (e.g., Hiscock and Tabah 2003), in which case our observation would reflect dominance between HIR-affecting alleles.
Genetic analysis of haploid induction ability is complicated further by the dichotomous nature of the trait. In the testcross progeny of inducer genotype $i$ the phenotype of each testcross seed $ij$ is coded by the indicator variable $z_{ij}$, with $z_{ij} = 1$ if $ij$ is haploid or $z_{ij} = 0$ otherwise. For our conceptual framework, we assume that the phenotype $z_{ij}$ is determined by the concentration $y_{ij}$ of the un-observable genetic factor(s) $X$ (e.g., expression level of gene(s) or concentration of metabolite(s)) in the pollen grain triggering the development of testcross seed $ij$. We suppose that $y_{ij}$ displays a quantitative distribution and decompose it according to the model commonly employed in quantitative genetics: $y_{ij} = G_i + \varepsilon_{ij}$, where $G_i$ refers to the genotypic value of genotype $i$ (i.e., $G_i = E(y_{ij}|i)$ for factor $X$ and $\varepsilon_{ij}$ reflects the experimental noise as well as genetic segregation in case of a heterozygous inducer genotype. This situation is depicted schematically in Figure 6. If $y_{ij}$ exceeds a certain threshold $T$, the phenotype of seed $ij$ will be haploid. Consequently, $z_{ij} = I[y_{ij} > T]$, which takes $z_{ij} = 1$ if $y_{ij} > T$ and $z_{ij} = 0$ if $y_{ij} \leq T$, links the distribution of the unobservable variable $y_{ij}$ with the phenotype $z_{ij}$. Thus, genotype $i$ induces haploidy in testcross seed $ij$ with a true probability of $p_i$ such that $Pr(z_{ij} = 1) = p_i$, and it follows that $Pr(z_{ij} = 0) = 1 - p_i$. Hence, the expectation of $z_{ij}$ is $E(z_{ij}) = p_i$ and its variance is $var(z_{ij}) = p_i(1 - p_i)$. The genetic factor(s) $X$ affecting $p_i$ are unknown, but once identified (e.g., through gene expression studies with near-isogenic lines) they could potentially be measured on a continuous scale and the inheritance could be studied directly.

In summary, we propose that the trait haploid induction ability is a threshold character under polygenic control, yet leading to dichotomous expression (haploid vs. diploid) on the level of the individual testcross seed. Prominent examples
of threshold characters include disease susceptibility (affected vs. not affected) in plants or litter size in large mammals (single vs. twin births; Falconer and Mackay 1996). Further, the two common assumptions for genetic analysis, i.e., normality of residuals and homoscedasticity, are not fully met for haploid induction ability. Composite interval mapping (Jansen and Stam 1994; Zeng 1994) and alternative approaches such as non-parametric tests (Kruglyak and Lander 1995) or two-part model tests (Broman 2003) are expected to work well for non-normally distributed data as long as critical LOD scores are determined via permutation tests from the original data set. However, heteroscedasticity of error variances remains an unsolved problem and warrants further research.

**Heritability of HIR:** Between-family heritability estimates for HIR (represented by $h^2$ estimates from parent-offspring regression) were lower than within-family heritability estimates (represented by repeatabilities $w^2$ estimated from HIR of three ears per family). Since parents and offspring were not evaluated in the same environments, genotype by environment interactions for HIR may have reduced $h^2$ estimates. Further, the discrepancy between $h^2$ and $w^2$ can be explained by natural selection disfavoring the haploidy-inducing gametes during selfing as suggested by the observed segregation distortion. From an evolutionary point of view, a higher proportion of haploids in the progeny caused by high HIR of the pollinator would result in fitness disadvantages, because haploid maize plants are less vigorous, often male sterile, and, therefore, generally less likely to produce offspring than diploid maize plants (Chase 1952; Chalyk 1994), which explains the counteracting force of natural selection. The latter may also explain the difficulties encountered during maintenance of inducers reported by maize breeders.
Surprisingly, $h^2$ for HIR estimated from regression of the 1680 generations was
twice as high as those estimated from regression of the CAU generations or the
estimate DEIMLING et al. (1997; $h^2 = 0.44$) reported for HIR. Since $h^2$ is a pop-
ulation parameter (FALCONER and MACKAY 1996), the different germplasm base
explains this discrepancy. Further, from our conceptual genetic framework it fol-
low that $h^2$ for HIR estimated from a segregating population such as F2 or F3
depends on the genotypic variance for the probability parameter $p_i$ as well as on
the sample size $N_i$ of the testcross progeny used for determining HIR of inducer
genotype $i$. Consequently, estimation of $h^2$ remains difficult because (1) phenotyp-
ing of HIR in multi-location experiments on various testers is extremely laborious,
and (2) heteroscedasticity as well as the associated segregation distortion compli-
cate quantitative genetic analyses including estimation of $h^2$, so novel biometric
approaches are urgently needed.

**Agreement between conceptual genetic framework and QTL results:**
To the best of our knowledge, this is the first study that used multiple populations
and generations to identify QTL responsible for HIR. A common approach for
comparison of QTL locations across populations is using joint linkage analysis as
proposed by BLANC et al. (2006), but this was not possible due to a lack of com-
mon markers between the four populations. Yet, simple addition of LOD curves
from CIM runs of individual populations revealed that only qhir1 was significant
across both CML populations (see Suppl. Figures 3 and 4), while no joint QTL was
significant for the two Chinese populations (data not shown). Further, a variable
degree of segregation distortion in different populations complicates joint analysis.
Therefore, we refrained from joint linkage analysis and compared QTL across pop-
ulations using the "bin concept", which is well established in the literature (e.g., 
Gardiner et al. 1993; Schön et al. 2010).

The number of QTL and the magnitude of QTL effects obtained for HIR in the present study varied depending on populations and generations. There was a clear trend, however, for $qhir1$ in bin 1.04 having a major effect on HIR in populations involving non-inducer genotypes as females (1680, CML395, and CML495). Previous studies have also identified this chromosomal region to segregate in populations involving only one inducer parent (Deimling et al. 1997; Barret et al. 2008). In contrast, no QTL was detected in bin 1.04 in CAU-F$_2$, CAU-F$_3$, and CAU-F$_4$, which were derived from a cross of the two haploid inducers CAUHOI and UH400. Most likely, the CAU generations are not segregating for $qhir1$ because the favorable allele is fixed in the parents, which provides a simple biological explanation for the incongruency of QTL. In fact, further analysis of the bin 1.04 region with 16 markers that were all polymorphic between inducer line UH400 and non-inducer line 1680, revealed monomorphism between inducers UH400 and CAUHOI (Suppl. Fig. 2). As large-effect QTL are subjected to strong selection pressure, $qhir1$ has likely become fixed quickly when selection for high HIR occurred during the development of UH400 and CAUHOI. Further, the two inducers may be identical by descent in this chromosomal region due to derivation from the same original inducer source, Stock6 (Coe 1959).

QTL detection frequencies obtained with 200 runs of five-fold cross-validation (Utz et al. 2000) were 95–100% for the large effect QTL in bins 1.04 ($qhir1$) for each of the populations except CAU, and 9.01 ($qhir8$) for the CAU populations in the present study. This explains why these QTL were detected in more than one
population and/or all generations of the respective population and underlines their significance for HIR. In contrast, QTL detection frequencies ranged from 23 to 80% for QTL with smaller effects and, accordingly, none of the minor QTL except \textit{qhir2} was detected in more than one population. The power of simultaneous detection of QTL in different experiments is obtained by multiplication and, therefore, low congruency of minor QTL across experiments meets expectations and has also been reported previously (Melchinger \textit{et al.} 1998; Mihaljevic \textit{et al.} 2004).

The use of cofactors in CIM has been shown to increase power of QTL detection (Jansen and Stam 1994; Zeng 1994). Analyses with the two-part model did not confirm all of the QTL identified with CIM. Higher critical LOD thresholds were employed for this model and the minor QTL identified with CIM (\textit{qhir3} and \textit{qhir5}) fell short of the critical LOD during two-part analyses due to lower power for QTL detection (Broman 2003).

The two-part analyses also revealed that \textit{qhir1} in 1680-F\textsubscript{2} significantly affected the chance whether genotypes have HIR $> 0\%$, while the remaining QTL only affected the genotype means once HIR $> 0\%$. Hence, we speculate that \textit{qhir1} acts as a key modulator for HIR, i.e., genes located in bin 1.04 trigger genetic processes that enable passing of the threshold outlined in our conceptual genetic framework. In this case, introgression of \textit{qhir1} would equip any germplasm with haploid induction ability. However, in the two CML populations not all individuals homozygous for the inducer allele at the flanking markers of \textit{qhir1} showed HIR $> 0\%$, a phenomenon known as incomplete penetrance (e.g., Barret \textit{et al.} 2008; Eldar \textit{et al.} 2009; Gaudet \textit{et al.} 2010), while complete penetrance was observed in 1680-F\textsubscript{2} and 1680-F\textsubscript{3}. Yet, genotypes at the actual QTL position are unknown.
and denser marker coverage would be necessary to more conclusively infer about penetrance.

The QTL *qhir1* explained a much larger proportion of the genetic variance in 1680-F_2 and 1680-F_3 than in CML395-F_2 and CML495-F_2, possibly due to incomplete penetrance. This discrepancy can be attributed to different testers and the associated haploid identification systems employed for the different mapping populations (Kebede *et al.* 2011; Prigge *et al.* 2011). Further, the presence of different modifier genes (Allard 1999) acting as enhancers or suppressors of major genes in different germplasm sources as well as QTL by genetic background interactions (Pumphreys *et al.* 2007) may be responsible. Future studies should employ different testers (such as liguleless, glossy, *R1-nj*) and near-isogenic lines developed from crosses of inbreds with contrasting HIR to focus on modifier QTL regions in greater detail and investigate putative epistatic interactions. Our results suggest that populations developed from crossing two inducers will be particularly helpful in revealing QTL other than *qhir1*.

**Effects of Segregation Distortion:** Distorted allele frequencies were observed in all populations indicating the presence of segregation distortion loci (SDL) that cause linked markers to deviate from expected Mendelian segregation patterns due to gametic or zygotic selection (Xu 2008). It is expected that recombination frequencies and, consequently, marker positions during linkage map construction will not be affected by segregation distortion if only one SDL is present in a distorted region and codominant markers such as SSRs or SNPs are used (Lorieux *et al.* 1995; Lu *et al.* 2002). The total map lengths estimated for our F_2 populations were comparable to those reported in previous studies with F_2
populations (e.g., Melchinger et al. 1998; Mihaljevic et al. 2004). Further, both marker order as well as bin positions have been confirmed for all markers based on the published IBM2 2008 Neighbors Map or physical locations. Thus, our QTL results should be robust and reliable as far as the flanking markers and chromosomal bins are concerned.

In fact, had SDL been excluded from analysis, \textit{qhir1} would have been missed in our study. This confirms previous results that genomic regions exhibiting strong segregation distortion are equally or even more likely to contain QTL (e.g., Wang et al. 2005). The strong distortion against the UH400 marker allele observed in the region of \textit{qhir1} in three populations of the present study supports the putative association between HIR and transmission failure of the inducer gamete suggested by Barret et al. (2008). Perhaps the same or tightly linked loci are responsible for causing both segregation distortion as well as triggering induction of haploidy.

Under the assumption that \textit{qhir1} is responsible for the observed segregation distortion in bin 1.04, this relationship could potentially be exploited for fine-mapping of \textit{qhir1}. By saturating this QTL region with SNPs and identifying those SNPs with maximum segregation distortion, fine-mapping of the QTL may be achieved without the need of elaborate designs in a simple large-size F\textsubscript{2} population. Schneeberger and Weigel (2011) proposed a similar approach using next-generation sequencing technology for mutation mapping in \textit{Arabidopsis}.

**Putative Candidate Genes for Haploid Induction Ability:** Identification of candidate genes for haploid induction ability is premature because the underlying biological mechanisms are still unknown. Considering that limited pollen production capacities and protogyny have been observed for inducers (data not
shown; W. SCHIPPRACK 2011, pers. comm.), weak or partial male sterility may be involved in haploid induction. Particularly relevant appear mutants producing pollen that does not always lead to complete fertilization of egg and central cells. Almost 50 male sterility genes (ms; NEUFFER et al. 1997) are known. For example, mutants carrying ms12, which is located on chromosome 1, show partial male sterility probably depending on the genetic background (ALBERTSEN and PHILLIPS 1981).

Further, defective kernel mutations (dek; NEUFFER et al. 1997) map to similar chromosomal regions as QTL identified in our study in bins 1.04 (dek32), 3.02 (dek5, dek24), 3.06 (dek17), 4.03 (dek11), 5.01 (dek18), 5.04 (dek9, dek26, dek27, dek33), and 9.01 (dek12). These mutations are characterized by a variation of degenerated embryo and endosperm tissues. Several mutants such as carriers of dek12, dek32, and dek33 show completely collapsed kernels appearing as embryoless pericarps without any starch deposition (SHERIDAN and NEUFFER 1980). This phenotype was employed to determine embryo abortion rates (EAR) in our study. Significant variation for EAR was detected and EAR estimates of > 10% were observed, which is more than expected in normal crosses. The joint presence of qhir1 and qear1 in bin 1.04 as well as the highly significant rank correlations between HIR and EAR in 1680-F2 and 1680-F3 suggest a possible relationship with inducer properties. However, further genetic mapping or gene expression studies will be necessary to study the involvement of dek loci in HIR. Examining the genotypes of endosperm and embryo via comparative seed and leaf DNA-based analyses (GAO et al. 2011) may be helpful to study the fertilization process(es) inducing haploidy and to better understand the biological basis of in vivo haploid...
induction in maize.

In *Arabidopsis thaliana*, Ravi and Chan (2010) report the production of haploid plants by crossing the source germplasm with a mutant carrying a genetic alteration of the centromeric histone CENH3. In our study, no QTL for HIR was detected in bin 6.06 which harbors CENH3 of maize (Zhong et al. 2002). This indicates that *in vivo* induction of haploidy in maize is not governed by the mechanism reported by Ravi and Chan (2010), although such centromere-mediated genome elimination may also work in maize to produce haploids. Nonetheless, *in vivo* induction of haploidy may be fundamentally different in mono- and dicots and maize could serve as a model species for studying this phenomenon in monocots.

**Implications for Haploid Inducer Development:** Transgressive segregants were identified in all populations except CML395-F$_3$ and also non-inducer parents contributed HIR-enhancing alleles, indicating that inducers with increased HIR can be generated through targeted parental recombination and selection. To reduce the laborious and time-consuming task of phenotypic selection for HIR, marker-based approaches such as genomic selection are suitable to increase the frequency of favorable alleles for HIR in inducer development programs. This seems particularly relevant if haploid induction ability is indeed governed by one or few major genes influenced by several modifier genes (Allard 1999), as major genes will quickly become fixed and gain from selection will then depend entirely on small effect modifier genes which will be extremely difficult to select for phenotypically.

In addition, targeted introgression of relevant QTL into adapted and agronomically superior non-inducer germplasm will enable the development of custom-made...
inducers that can be employed for efficient DH line production in various breeding programs and agroecologies. \textit{qhir1} represents an interesting chromosomal region for map-based cloning and the required fine-mapping can be achieved by developing near-isogenic lines. However, incomplete penetrance observed for this QTL in CML395-F$_2$ and CML495-F$_2$ suggests that the relationship between genotype and phenotype is not that reliable in all genetic backgrounds. Thus, phenotyping of marker-selected inducers still seems necessary as a final evaluation step.

We greatly appreciate the skilled technical assistance of Li Haochuan, Dong Xin, Guo Mingxin, Chen Liming and co-workers at China Agricultural University as well as of Ciro Sánchez, Alberto Vergara, Ana Mely Islas Montes, and Germán Bastian de León and co-workers at CIMMYT during DNA extraction, testcross seed production, and HIR assessment. We are grateful to Jianbing Yan for facilitating the parental screening with the 1536 SNP chip, and to Stefan Scholten and H. Friedrich Utz for valuable comments on an earlier version of this manuscript. This research was financially supported by National Projects nycytx-02 and 2009CB118400 at China Agricultural University, and by the Bill and Melinda Gates Foundation, Seattle, WA, USA, through the Drought Tolerant Maize for Africa (DTMA) project at CIMMYT. Vanessa Prigge is indebted to Eiselen Foundation, Ulm, Germany for financial support of her research as a joint Ph.D. student between CIMMYT and the University of Hohenheim, Germany.
LITERATURE CITED


CIMMYT, 2005 Laboratory protocols: CIMMYT Applied molecular genetics laboratory. CIMMYT, Mexico (DF), Mexico.


MIHALJEVIC, R., H. F. UTZ, and A. E. MELCHINGER, 2004 Congruency of
quantitative trait loci detected for agronomic traits in testcrosses of five maize populations of European maize. *Crop Sci.* **44:** 114–124.


**Nei, M.,** 1963 *The efficiency of haploid methods of plant breeding.* *Heredity* **18:** 95–100.


WANG, C., C. ZHU, H. ZHAI, and J. WAN, 2005 Mapping segregation distortion
Genetics of in vivo Haploid Induction in Maize

loci and quantitative trait loci for spikelet sterility in rice (Oriza sativa L.).

XU, S., 2008 Quantitative trait locus mapping can benefit from segregation
distortion. Genetics **180**: 2201–2208.


ZENG, Z. B., 1994 Precision mapping of quantitative trait loci. Genetics **136**:
1457-1468.

elimination and in vivo haploid induction by stock 6-derived inducer line in

ZHONG, C., J. B. MARSHALL, C. N. TOPP, R. J. MROCKEZ-WILLIAMSON, A.
KATO, et al. 2002 Centromeric retroelements and satellites interact with
TABLE 1

Number of families ($N_F$) and marker loci ($N_L$) used for QTL analyses in the four populations and various generations; the average number of seeds ($N_S$) used to determine the haploid induction rates (HIR); and the corresponding estimates of the genetic variance components ($\sigma_g^2$), repeatabilities ($w^2$), means, and ranges for HIR.

<table>
<thead>
<tr>
<th>Population–Generation</th>
<th>Code</th>
<th>$N_F$</th>
<th>$N_L$</th>
<th>$N_S$</th>
<th>$\sigma_g^2$</th>
<th>$w^2$</th>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAUHOI×UH400-F$_2$</td>
<td>CAU-F$_2$</td>
<td>185</td>
<td>89</td>
<td>246</td>
<td>8.25**</td>
<td>69</td>
<td>5.2</td>
<td>0.5-15.5</td>
</tr>
<tr>
<td>CAUHOI×UH400-F$_2$ 3</td>
<td>CAU-F$_3$</td>
<td>161</td>
<td>89</td>
<td>189</td>
<td>1.52**</td>
<td>46</td>
<td>2.3</td>
<td>0.0-8.0</td>
</tr>
<tr>
<td>CAUHOI×UH400-F$_2$ 4</td>
<td>CAU-F$_4$</td>
<td>124</td>
<td>89</td>
<td>317</td>
<td>3.50**</td>
<td>76</td>
<td>2.4</td>
<td>0.0-10.3</td>
</tr>
<tr>
<td>1680×UH400-F$_2$</td>
<td>1680-F$_2$</td>
<td>124</td>
<td>106</td>
<td>237</td>
<td>6.59**</td>
<td>91</td>
<td>2.8</td>
<td>0.0-15.6</td>
</tr>
<tr>
<td>1680×UH400-F$_2$ 3</td>
<td>1680-F$_3$</td>
<td>113</td>
<td>106</td>
<td>264</td>
<td>10.65**</td>
<td>85</td>
<td>2.9</td>
<td>0.0-13.4</td>
</tr>
<tr>
<td>CML395×UH400-F$_2$ 3</td>
<td>CML395-F$_3$</td>
<td>171</td>
<td>191</td>
<td>153</td>
<td>0.40**</td>
<td>40</td>
<td>0.6</td>
<td>0.0-5.0</td>
</tr>
<tr>
<td>CML495×UH400-F$_2$ 3</td>
<td>CML495-F$_3$</td>
<td>183</td>
<td>183</td>
<td>171</td>
<td>1.88**</td>
<td>88</td>
<td>0.8</td>
<td>0.0-12.9</td>
</tr>
</tbody>
</table>

** Significant at $P < 0.01$.  

November 26, 2011
TABLE 2

<table>
<thead>
<tr>
<th>Chr. Pos.</th>
<th>Flanking marker</th>
<th>LOD</th>
<th>Type</th>
<th>$p_{\text{part}}$</th>
<th>Effect</th>
<th>$d/a$</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAU-F$_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.02 30 24–32 umc1886–bng1523</td>
<td>5.83</td>
<td>a</td>
<td>9.39</td>
<td>0.93$^{**}$</td>
<td>na</td>
<td>qhir2</td>
<td></td>
</tr>
<tr>
<td>3.06 90 78–106 bng1601–umc2169</td>
<td>5.40</td>
<td>a</td>
<td>8.20</td>
<td>1.11$^{**}$</td>
<td>na</td>
<td>qhir3</td>
<td></td>
</tr>
<tr>
<td>5.01 28 18–32 umc1097–phi024</td>
<td>7.32</td>
<td>a</td>
<td>11.33</td>
<td>-1.01$^{**}$</td>
<td>na</td>
<td>qhir5</td>
<td></td>
</tr>
<tr>
<td>5.04 122 114–128 umc1332–umc1155</td>
<td>9.54</td>
<td>a</td>
<td>17.44</td>
<td>1.55$^{**}$</td>
<td>na</td>
<td>qhir6</td>
<td></td>
</tr>
<tr>
<td>9.01 14 10–26 bng1272–umc1040</td>
<td>19.76</td>
<td>a</td>
<td>24.13</td>
<td>1.82$^{**}$</td>
<td>na</td>
<td>qhir8</td>
<td></td>
</tr>
<tr>
<td>CAU-F$_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.03 34 16–50 umc2082–umc2039</td>
<td>3.23</td>
<td>a</td>
<td>15.46</td>
<td>0.72$^{**}$</td>
<td>na</td>
<td>qhir4</td>
<td></td>
</tr>
<tr>
<td>7.01 64 58–78 phi057–umc2142</td>
<td>3.66</td>
<td>a</td>
<td>17.43</td>
<td>0.63$^{**}$</td>
<td>na</td>
<td>qhir7</td>
<td></td>
</tr>
<tr>
<td>9.01 8 0–18 bng1272–umc1040</td>
<td>6.81</td>
<td>a</td>
<td>31.04</td>
<td>0.96$^{**}$</td>
<td>na</td>
<td>qhir8</td>
<td></td>
</tr>
<tr>
<td>CAU-F$_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.01 4 0–12 bng1272–umc1040</td>
<td>5.80</td>
<td>a</td>
<td>12.27</td>
<td>1.07$^{**}$</td>
<td>-1.17</td>
<td>qhir8</td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}$: 20.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1680-F$_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.04 56 54–60 umc1917–umc2390</td>
<td>28.07</td>
<td>a</td>
<td>51.61</td>
<td>4.33$^{**}$</td>
<td>-0.44</td>
<td>qhir1</td>
<td></td>
</tr>
<tr>
<td>3.02 30 22–38 bng1325–umc1886</td>
<td>2.85</td>
<td>a</td>
<td>4.86</td>
<td>-0.68$^{**}$</td>
<td>na</td>
<td>qhir2</td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}$: 71.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1680-F$_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.04 58 54–60 umc2390–umc1811</td>
<td>19.21</td>
<td>a</td>
<td>53.04</td>
<td>4.03$^{**}$</td>
<td>-0.37</td>
<td>qhir1</td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}$: 62.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CML395-F$_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.04 90 72–98 PHM5306–PHM9418</td>
<td>3.67</td>
<td>a</td>
<td>8.05</td>
<td>0.28$^{**}$</td>
<td>-1.71</td>
<td>qhir1</td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}$: 19.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CML495-F$_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.04 78 72–88 PZA02550–PHM5306</td>
<td>5.27</td>
<td>a</td>
<td>8.95</td>
<td>0.66$^{**}$</td>
<td>na</td>
<td>qhir1</td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}$: 8.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^{**}$ Significant at $P < 0.01$.

$^{a}$ Underlined bins: the QTL was identified with both composite interval mapping (CIM) as well as the two-part model; non-underlined bins: QTL was identified with CIM.

$^{b}$ The marker closest to the position of the putative QTL is italicized.

$^{c}$ a = additive, d = dominance.

$^{d}$ Portion of the genotypic variance explained by a specific QTL and summing up to the total proportion ($\hat{\rho}$) of the genotypic variance explained by the model including all QTL in a population-generation combination.

$^{e}$ Positive values indicate that UH400 contributed the favorable allele.

$^{f}$ Dominance ratio of the estimated dominance effect over the estimated additive effect. na = not applicable due to non-significant dominance effects.

$^{g}$ For codes of populations see Table 1.
TABLE 3

Summary of results from phenotypic and quantitative trait locus (QTL) analyses of embryo abortion rate (EAR) in F$_2$ and F$_3$ generations of 1680×UH400.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>1680-F$_2$</th>
<th>1680-F$_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotypic results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of families</td>
<td>No.</td>
<td>132</td>
<td>116</td>
</tr>
<tr>
<td>Number of SSR loci</td>
<td>No.</td>
<td>106</td>
<td>106</td>
</tr>
<tr>
<td>Average number of testcross seeds</td>
<td>No.</td>
<td>344</td>
<td>399</td>
</tr>
<tr>
<td>Genetic variance (σ$_g^2$)</td>
<td>%</td>
<td>0.68**</td>
<td>1.31**</td>
</tr>
<tr>
<td>Repeatability ($w^2$)</td>
<td>%</td>
<td>68</td>
<td>73</td>
</tr>
<tr>
<td>Mean</td>
<td>%</td>
<td>0.81</td>
<td>1.11</td>
</tr>
<tr>
<td>Range</td>
<td>%</td>
<td>0–6.70</td>
<td>0–11.10</td>
</tr>
<tr>
<td>QTL results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome</td>
<td>–</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Position (Support interval)</td>
<td>cM</td>
<td>52 (50–56)</td>
<td>60 (56–66)</td>
</tr>
<tr>
<td>Flanking markers</td>
<td>–</td>
<td>bnlg2180, umc1917</td>
<td>umc1811, umc1770</td>
</tr>
<tr>
<td>LOD</td>
<td>–</td>
<td>8.19</td>
<td>3.19</td>
</tr>
<tr>
<td>Proportion of genetic variance explained (ρ)</td>
<td>%</td>
<td>34.13</td>
<td>14.18</td>
</tr>
<tr>
<td>Additive effect</td>
<td>%</td>
<td>0.97**</td>
<td>0.97**</td>
</tr>
<tr>
<td>Name</td>
<td>–</td>
<td>qear1</td>
<td>qear1</td>
</tr>
</tbody>
</table>

** Significant at $P < 0.01$.
Genetics of *in vivo* Haploid Induction in Maize

Proportion of $\text{F}_2$ Families

<table>
<thead>
<tr>
<th>Population</th>
<th>$N_T$</th>
<th>$N_Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAU-HOI</td>
<td>185</td>
<td>0</td>
</tr>
<tr>
<td>1680×UH400</td>
<td>124</td>
<td>8</td>
</tr>
<tr>
<td>CML395×UH400</td>
<td>171</td>
<td>71</td>
</tr>
<tr>
<td>CML495×UH400</td>
<td>183</td>
<td>66</td>
</tr>
</tbody>
</table>

Proportion of $\text{F}_3$ Families

<table>
<thead>
<tr>
<th>Population</th>
<th>$N_T$</th>
<th>$N_Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAU-HOI</td>
<td>161</td>
<td>7</td>
</tr>
<tr>
<td>1680×UH400</td>
<td>113</td>
<td>2</td>
</tr>
<tr>
<td>CML395×UH400</td>
<td>183</td>
<td>66</td>
</tr>
</tbody>
</table>

Proportion of $\text{F}_4$ Families

<table>
<thead>
<tr>
<th>Population</th>
<th>$N_T$</th>
<th>$N_Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAU-HOI</td>
<td>124</td>
<td>2</td>
</tr>
</tbody>
</table>

**FIGURE 1**

Distribution of haploid induction rates (HIR) assessed in different generations of four populations: $\text{F}_2$, $\text{F}_3$, and $\text{F}_4$ of CAU-HOI×UH400 (CAU-$\text{F}_2$, CAU-$\text{F}_3$, CAU-$\text{F}_4$), $\text{F}_2$ and $\text{F}_3$ of 1680×UH400 (1680-$\text{F}_2$ and 1680-$\text{F}_3$), $\text{F}_3$ of CML395×UH400 (CML395-$\text{F}_3$), and $\text{F}_3$ of CML495×UH400 (CML495-$\text{F}_3$). $N_T$ = Total number of families phenotyped. $N_Z$ = Number of families with HIR = 0%. Black arrows indicate the parental HIR. Dashed lines represent the population mean.

November 26, 2011
Genetics of *in vivo* Haploid Induction in Maize

![Figure 2](image)

**FIGURE 2**: Relationship between haploid induction rates (HIR) obtained in (A) CAU-F$_2$ and CAU-F$_3$, (B) CAU-F$_3$ and CAU-F$_4$, and (C) 1680-F$_2$ and 1680-F$_3$ generations. $h^2$ denotes the heritability estimate and $r_s$ denotes Spearman’s rank correlation coefficient. **, *** Significant at $P<0.01$ and $P<0.001$, respectively.
FIGURE 3

Chromosomal positions, direction, and magnitude of segregation distortion (column SD) in four F₂ populations as well as positions of quantitative trait loci (QTL) for haploid induction rate (HIR, ⋆) and embryo abortion rate (EAR, ⋄) detected in generation Fₙ of the populations. Bin assignment is based on the IBM2 2008 neighbors reference map (maizegdb.org). The centromeric bin is shown between bold lines.
FIGURE 4

Box-whisker plots of haploid induction rates (HIR) of the three marker genotypes at the closest markers flanking the quantitative trait locus *qhir1* detected on chromosome 1 (bin 1.04) in populations 1680-F2 (left) and CML495-F3 (right). A represents the non-inducer allele (1680 or CML495) and B the inducer allele (UH400). Individual observations are indicated as solid circles and jittered along the X-axis to avoid overplotting.
Genetics of *in vivo* Haploid Induction in Maize

![LOD profile of chromosome 1 of population 1680-F2 for interval mapping of the square root transformed haploid induction rate (HIR) data using the two-part model. Separate LOD scores were calculated to test whether the three genotype classes $g$ (AA, AB, BB) of inducer $i$ have the same probability $\pi_g$ of inducing haploidy (i.e., $z_i = 1$) and whether those inducers with $z_i = 1$ have the same sample mean $\mu_g$ according to three hypotheses (Broman 2003): LOD($\pi$) (blue) to test whether $\pi_{AA} = \pi_{AB} = \pi_{BB}$; LOD($\mu$) (red) to test whether $\mu_{AA} = \mu_{AB} = \mu_{BB}$; and LOD($\pi\mu$) (black) to test the combined hypothesis that both the $\pi_g$ and $\mu_g$ were equal. Dashed lines represent empirically determined significance thresholds ($\alpha = 0.05$) for the three LOD curves.](image-url)

November 26, 2011
Concentration of \( y_{ij} \)

\[ T = \text{Threshold} \]

\[ p_{ij} = \Pr(z_{ij} = 1) = E(I_{y_{ij} > T}) \]

\[ z_{ij} = 0 \]

\[ z_{ij} = 1 \]

\[ G_i \]

Concentration of \( y_{ij} \)

FIGURE 6

: Scheme of the quantitative distribution of concentration \( y_{ij} \) of the unobservable factor \( X \) empowering male gametes of genotype \( i \) to induce haploidy in testcross seed \( ij \). \( G_i \) represents the genotypic value of inducer \( i \) for factor \( X \). If \( y_{ij} \) passes the threshold \( T \), then haploidy is induced in \( ij \) and the indicator variable takes \( z_{ij} = 1 \), otherwise \( ij \) becomes diploid (\( z_{ij} = 0 \)).

November 26, 2011
Supplemental Figure 1 Relationship between embryo abortion rates (EAR) observed in generations 1680-F$_2$ and 1680-F$_3$. $h^2$ denotes the heritability estimate and $r_s$ denotes Spearman’s rank correlation coefficient. *** Significant at P<0.001.
Supplemental Figure 2 Representative gel profiles generated for haploid inducer inbreds CAUHOI (C) and UH400 (U) with 16 markers located in bin 1.04.
Supplemental Figure 3 $\log_{10}$ of the likelihood odds ratio (LOD) profiles of 10 chromosomes determined during quantitative trait locus (QTL) analyses with composite interval mapping (CIM) for haploid induction rate (HIR) conducted in an $F_3$ population from the cross CML395×UH400. The solid vertical line in each plot represents the empirically determined critical LOD threshold. The line below each plot represents the length of the chromosome (in cM); dashes on this line represent marker positions, triangles represent cofactor positions.
Supplemental Figure 4 \(\text{Log}_{10}\) of the likelihood odds ratio (LOD) profiles of 10 chromosomes determined during quantitative trait locus (QTL) analyses with composite interval mapping (CIM) for haploid induction rate (HIR) conducted in an F\(_3\) population from the cross CML495×UH400. The solid vertical line in each plot represents the empirically determined critical LOD threshold. The line below each plot represents the length of the chromosome (in cM); dashes on this line represent marker positions, triangles represent cofactor positions.