Modern Tools for Line Development and Predicting Hybrid Performance

By Thomas Lübberstedt, Walter Suza (ISU)

Overview

Introduction

Fig. 1 Maize seeds are shown at Victoria Seeds at Victoria Seeds production facility on Kampala,Uganda. Photo By Iowa State University

Several steps are involved in hybrid seed production, including creation of genetic variability, production of inbred lines by continuous selfing for several generations, testing lines for their combining ability and crossing the best inbred lines to create hybrids. There are two drawbacks facing the selection of the promising line combinations. Selecting the best breeding population is similar to the above mentioned usefulness problem in line breeding programs. The majority of the base populations are usually discarded after preliminary evaluation for per se and performance in an “early testing” programs. As inbred lines are typically produced in two opposite heterotic groups, the main challenge in hybrid breeding ultimately is, to identify the best inbred line combination among those two heterotic groups. Presence of 100 inbred lines in each of two heterotic groups would potentially enable production of 10,000 hybrids. Thus, prediction of hybrid performance and heterosis without having to assess thousands of single-cross hybrids in field trials would reduce the time and efforts required to identify promising inbred combinations substantially.

Objectives

- Breeding schemes for line development
- Doubled haploid technology
- Marker applications for heterotic pool formation and assignment
- Genomic tools to understand nature of heterosis
- Genomic tools for predicting hybrid performance

Fig. 2 Pearl millet seed production plots at ICRISAT (Patancheru, Hyderabad, India), the panicles covered in parchment paper bags to ensure self-pollination in this normally mainly cross-pollinating crop. Photo by Rik Schuiling / TropCrop. Licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons

Breeding Schemes for Line Development

There are two main methods by which lines are developed: pedigree method and bulk method. Both methods start with the generation of genetic variation by the hybridization of two parents (Phase I).

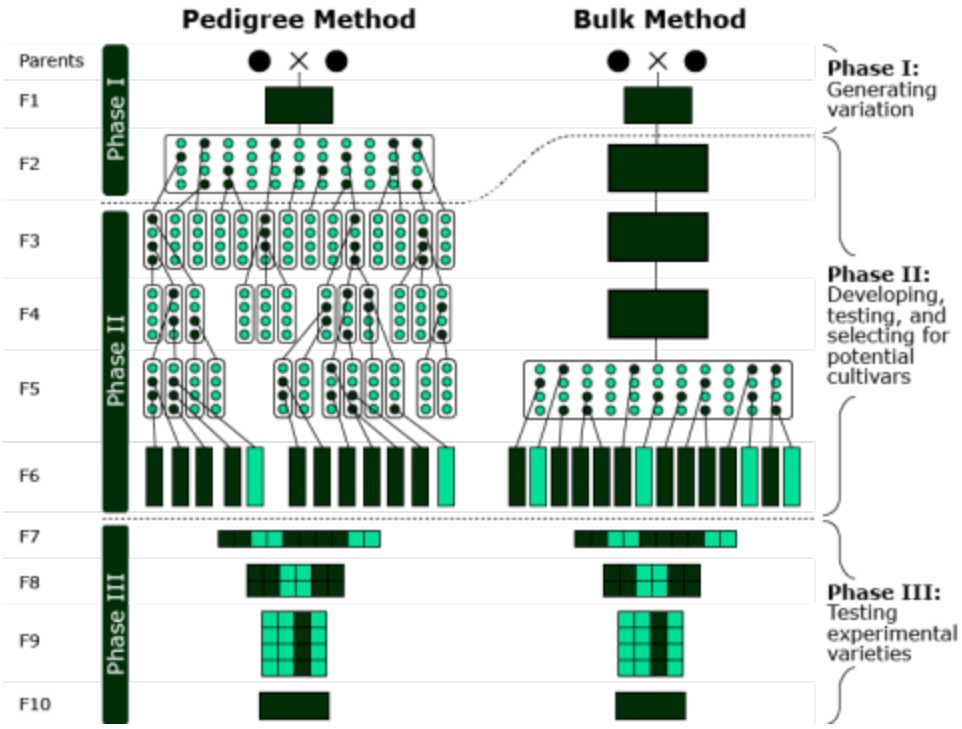


Fig. 3 The application of pedigree and bulk methods in line breeding. Inbred lines in hybrid breeding schemes can be developed similarly, but will be evaluated for their testcross performance in addition.

Doubled Haploids

Doubled Haploid - Definition

A doubled haploid (DH) cell contains the doubled chromosome number of the haploid and two identical gene sets (Fig. 4). As illustrated in Fig. 5, haploids can be induced either spontaneously or by various *in vitro* methods using female or male gametes. The methods of haploid induction are described below.

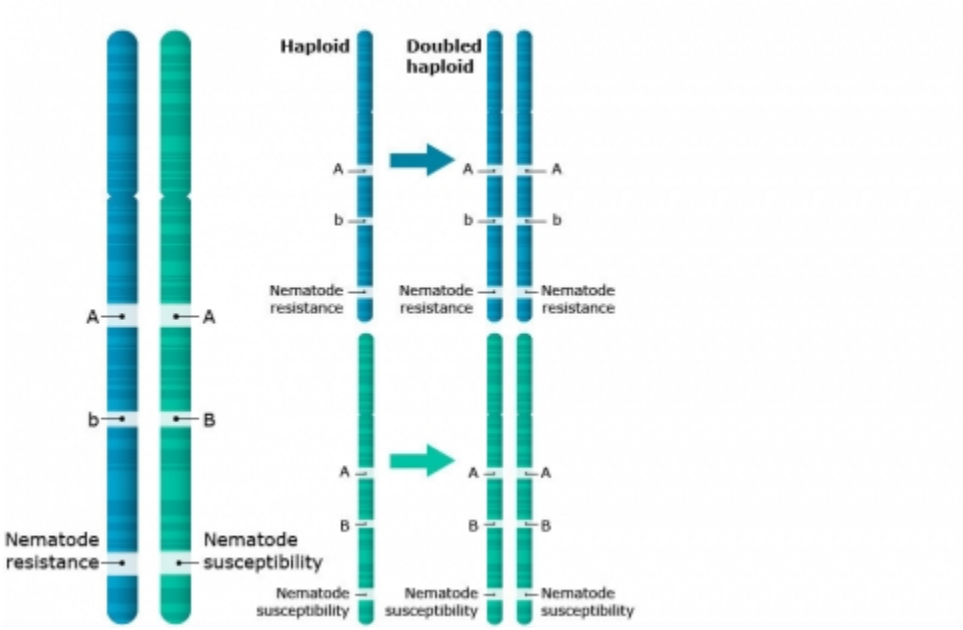


Fig. 4 A plant cell containing two sets of chromosomes which are not identical (A). Pollen has only one set of chromosomes (i.e., it is haploid). The genome can be duplicated by various methods to produce a doubled haploid (B). For this example, doubled haploid plants resistant to nematodes will survive nematode inoculation, but those that are susceptible, would be eliminated. In addition, molecular markers linked to nematode resistance can be used to pre-screen desirable individuals before field trials.

Methods of Producing Haploid Plants

An extensive discussion of development of haploids and doubled haploids in plant breeding was recently published (Murovec and Bohanec, 2012), and Fig. 5 illustrates various methods for plant haploid production.

Androgenesis is defined as male parthenogenesis in which the embryo contains only paternal chromosomes owing to the failure of the egg nucleus to participate in fertilization or the regeneration of whole plants from sexual male cell culture: anthers or isolated immature pollen, at extremely low frequencies. Gynogenesis refers to spontaneous or induced female parthenogenesis in which the embryo contains only maternal chromosomes owing to the failure of the sperm cell to fuse with the egg nucleus.

Interspecific crossing is used to develop a haploid embryo by fertilizing an ovule with pollen of another species and the subsequent elimination of the chromosomes of the pollen.

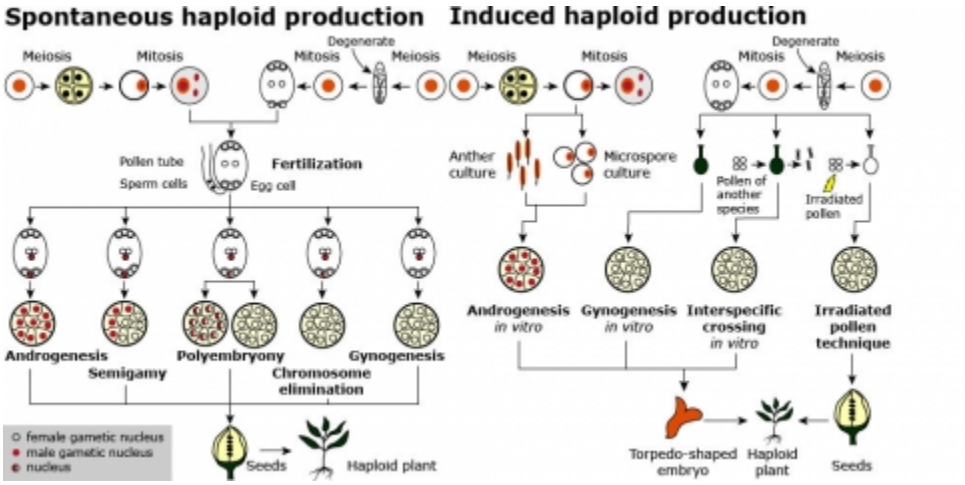


Fig. 5 Methods of plant haploid production. Spontaneous haploids can be observed via semigamy, polyembryony, chromosome elimination, gynogenesis and androgenesis.

Use of Inducers

The in vivo haploid induction can result in either paternal or maternal haploidy. For maternal haploid induction the target germplasm is pollinated with pollen from a haploid inducer genotype. For paternal haploidy, specific inducer genotypes are used as female parent. An example of haploid induction in maize is illustrated in Fig. 6.

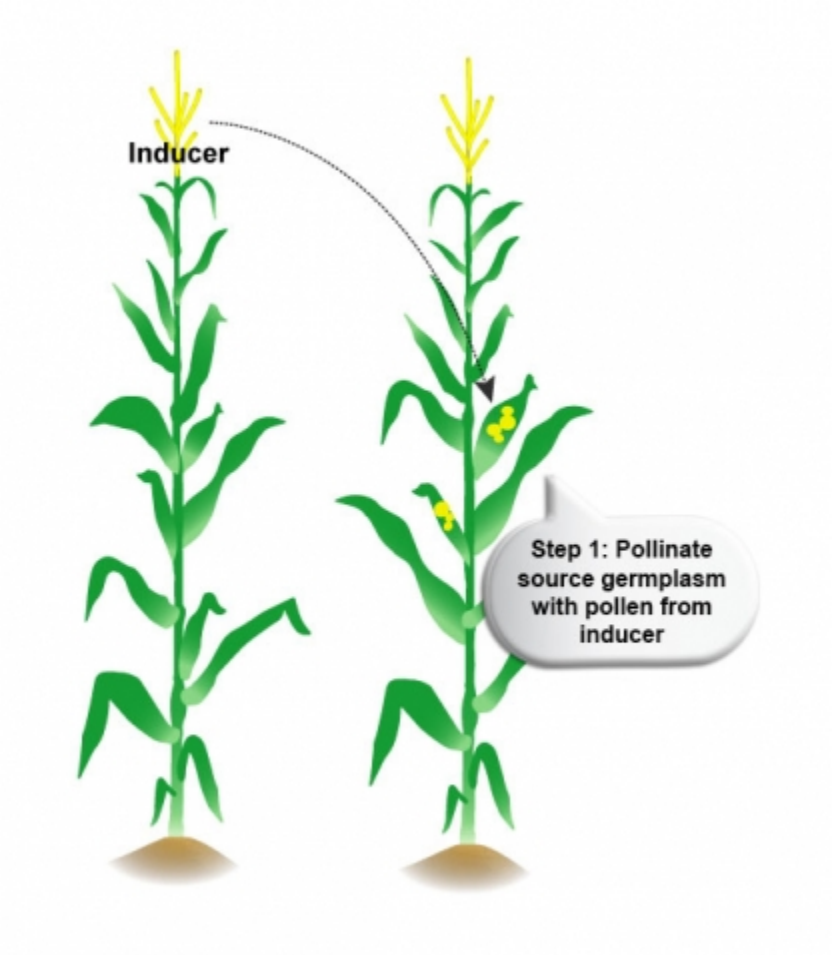
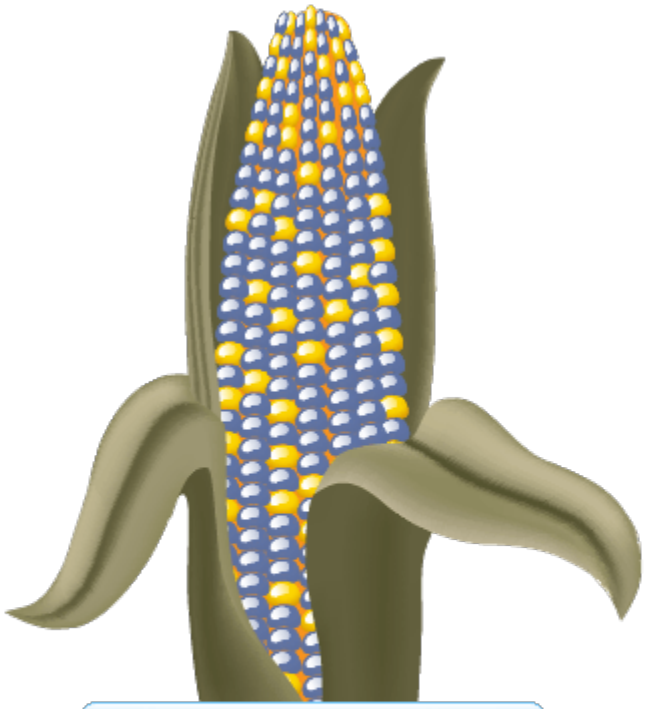
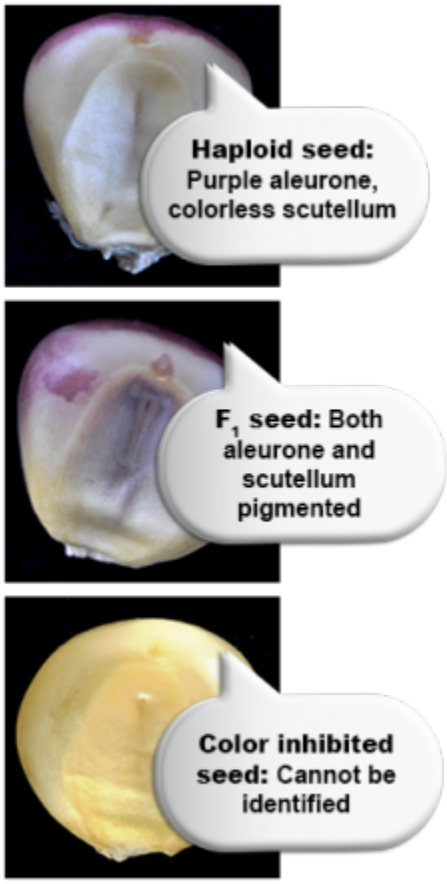


Fig. 6 Schematic description of doubled haploid line development with the in vivo haploid induction method. Adapted from Prigge and Melchinger, 2012.

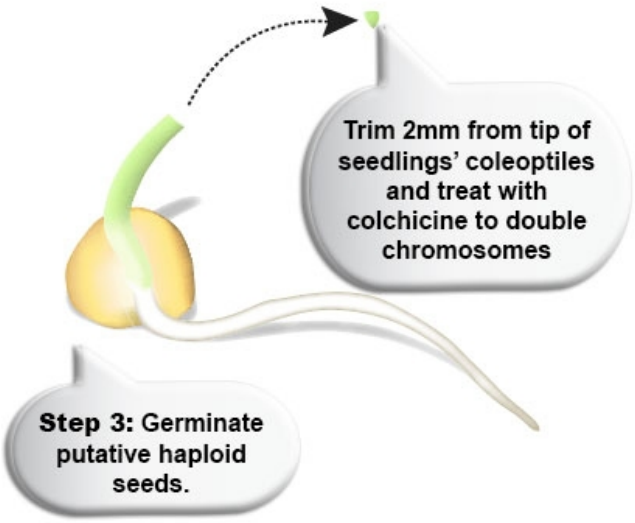
Use of Inducers (Step 2)

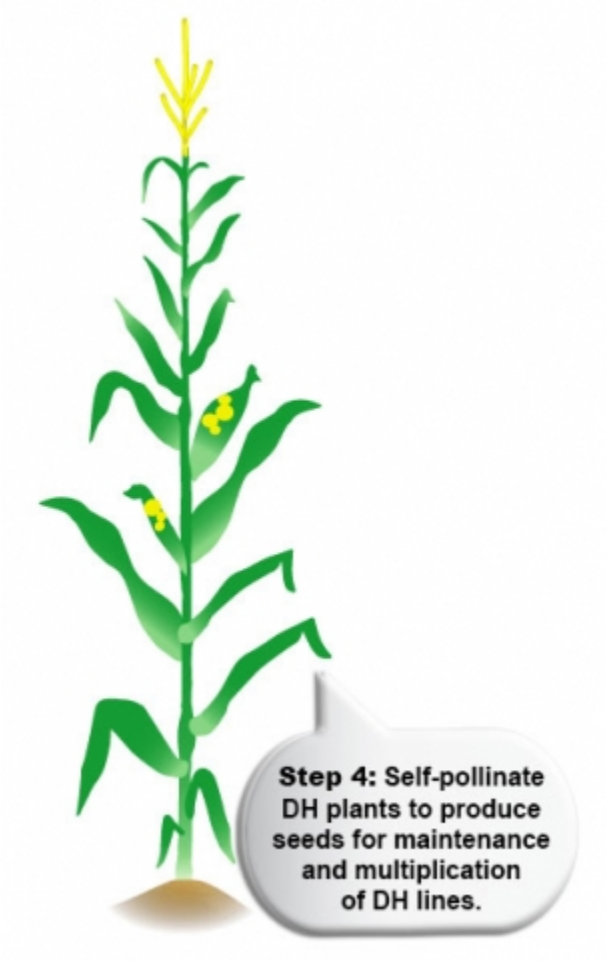


Step 2: Score pollinated ears using a marker system to identify haploid kernels.



Use of Inducers (Steps 3 and 4)





Inducer Effects

Use Of Inducers

- **Pollination with irradiated pollen** may be used to develop a haploid embryo by fertilizing an ovule with irradiated (inactive) pollen that nevertheless is capable of introducing cellular divisions in the ovule and in the development of the embryo.
- **Semigamy** refers to an abnormal type of fertilization whereby either reduced or unreduced male and female gametes participate in embryo formation but fertilization does not occur.
- **Polyembryony** is the production of two or more embryos in one seed, owing either to the existence and fertilization of more than one embryonic sac or to the origination of embryos outside of the embryonic sac.

Application of DH Technology

DH lines are usually produced from F_1 or F_2 plants. DH lines are comparable to lines obtained by the bulk method (**Fig. 3**), only in shorter time. DH technology allows development of completely homozygous plants, from which breeding lines or cultivars are derived, within two generations.

To identify best genotypes, breeders perform a multi-stage selection by first testing many genotypes with low precision/efforts and subsequently testing fewer and fewer genotypes with high precision and effort (with respect to locations, replications, etc.).

The advantages of DH technology are:

- Rapid generation of homozygous genotypes (**Fig. 8**)
- No masking of undesirable genes in the heterozygotes
- Maximum genetic variance from the first generation
- Perfect compliance with DUS criteria
- Short time to market
- Simplified logistics
- Reduced expenses for selfing and maintenance breeding

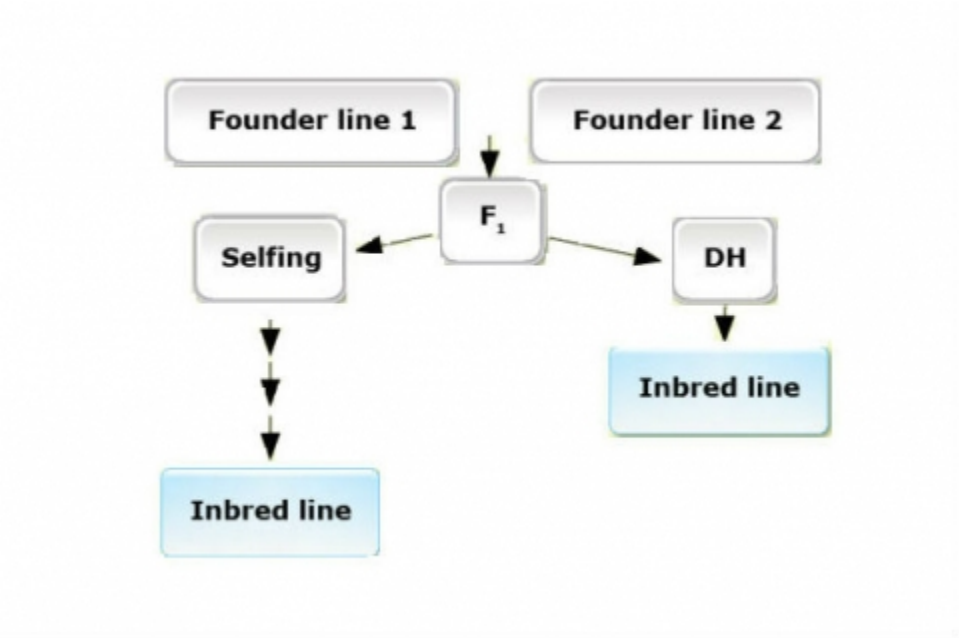


Fig. 7 DH technology helps speed up line development process. Plants selected from conventional breeding population do not breed true resulting in increased generations of inbreeding and selecting desirable lines.

Phenotypic Markers

The key is to have an early expressed marker, which enables discrimination of seed with a haploid versus diploid embryos. Only kernels with a haploid embryo are useful for DH line production. The R1-nj marker provides easy and fast visual assessment of DH and hybrid grain (Fig. 8A). Also, other dominant color marker genes expressed in other organs can be used, for example, the PI1 gene that is expressed in primary roots (Fig. 8B).

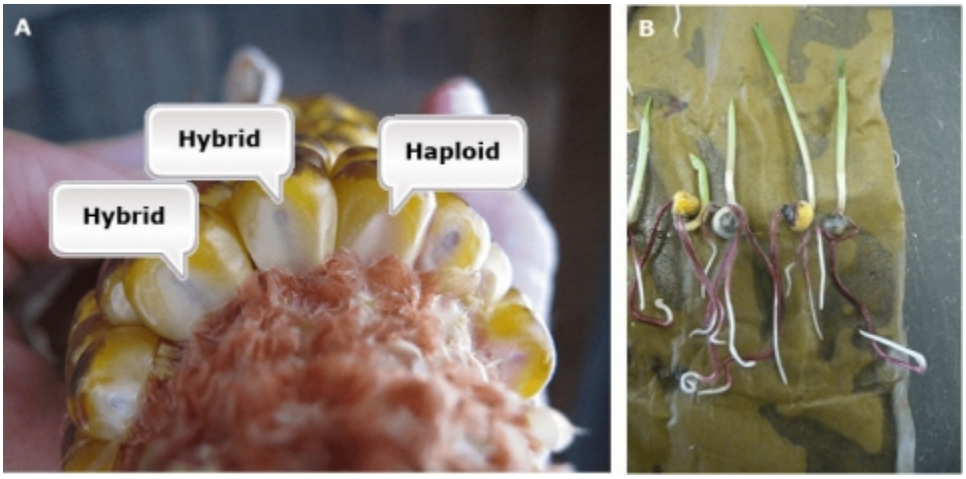


Fig. 8 The R1-nj marker gene produces diploid hybrid seed with a purple embryo. The haploid seed has a colorless embryo (A). Alternative markers, such as the PI1 gene that produces purple color in primary roots may also be used (B).

Metabolite Markers

Near infrared reflectance spectroscopy (NIRS) enables both early and automated discrimination of kernels with haploid versus diploid embryos. Thus, 10,000s of kernels can be sorted with minimal human interference.

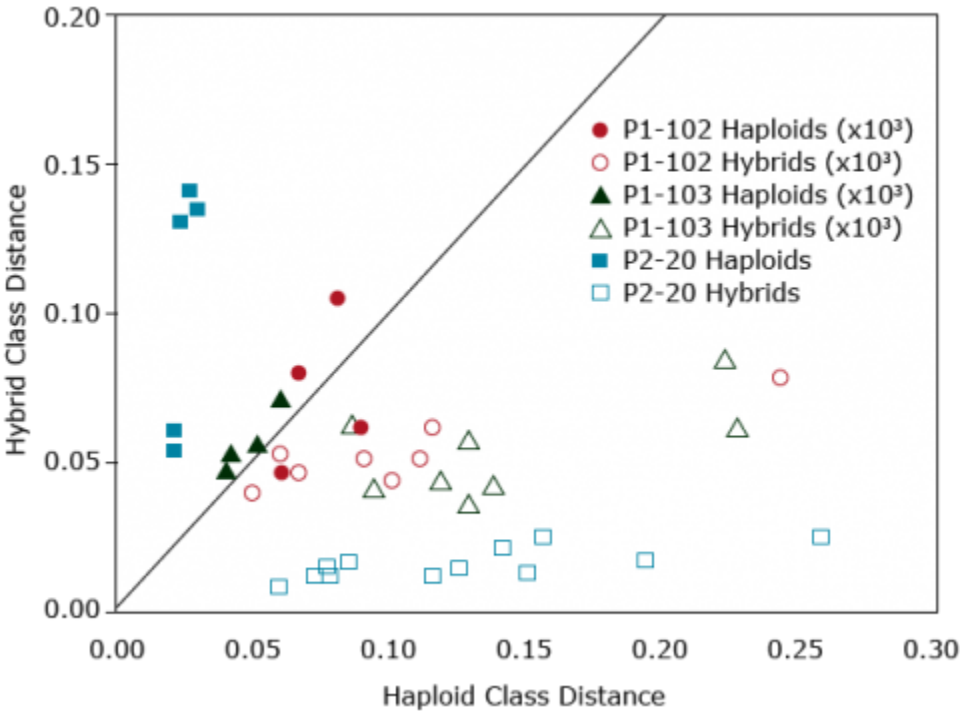


Fig. 9 Biochemical differences between haploids and hybrids of maize. In this example, the oil contents of haploids and hybrid seed is analyzed by Near-Infrared Spectroscopy (NIRS). NIRS is a spectroscopic method that uses the near-infrared region of the electromagnetic spectrum (from about 800 nm to 2500 nm). Adapted from Jones et al., 2012.

DNA-Based Markers

DNA markers are useful in gene pyramiding schemes for resistance when phenotypic selection cannot be achieved due to lack of differentiating pathogen strains, for example, Barley Yellow Mosaic Virus (Werner et al., 2005). In such gene pyramiding schemes, DH techniques are valuable because the frequency of homozygous recessive genotypes is higher in DH populations than in segregating F_2 populations.

Fig. 10 Barley Yellow Mosaic Virus symptoms. Photo by Mike Adams Rothamsted. Licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons.

Application Example

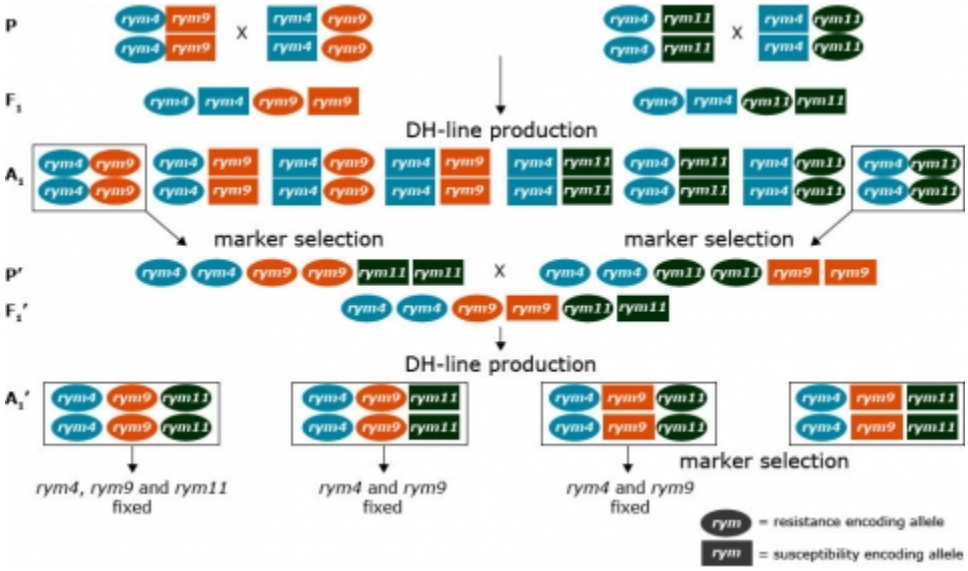


Fig. 11 Scheme of pyramiding Barley Yellow Mosaic Virus resistance genes using marker selection in combination with the doubled haploid method. Adapted from Werner et al., 2005.

Challenges in Application

Like any other technology, the DH technology has its own strengths and weaknesses. The strengths and weaknesses of DH technology as applied to maize breeding are summarized in Table 1 below.

Table 1 Comparison of DH methods in maize.

Approach	Strengths	Weaknesses
In vitro	• No need of inducer	• Low induction rate • Genotype dependency • Need of tissue culture
In vivo — paternal	• Simple inheritance • cms conversion	• Low induction rate • Genotype dependency • Need of tissue culture
In vivo — maternal	• Limited genotype dependency • Induction rate (10%)	• Backgrond effects • Complex inheritance

Other Concerns: Adapted Inducers

Need for developing adapted inducers:

For large scale haploid seed production, it is important to use inducer genotypes that are adapted to the haploid seed production environment (Fig. 12).

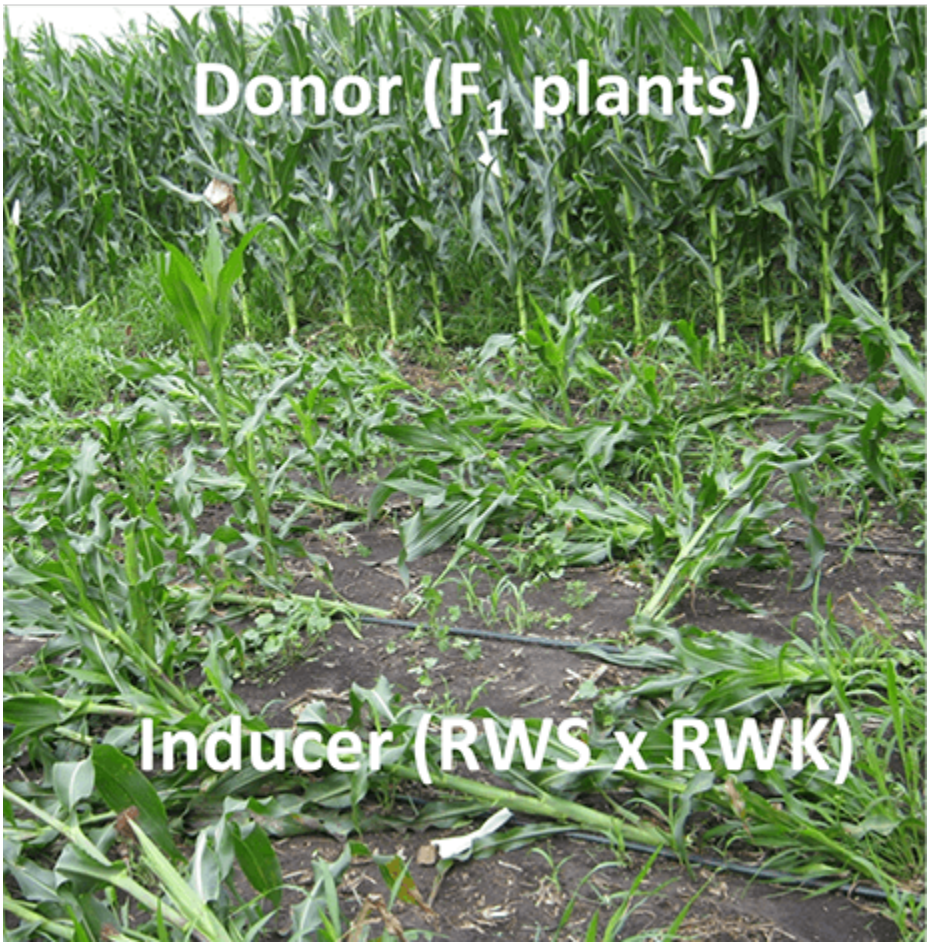


Fig. 12 Storm damage of European inducer grown in the Midwest US. Photo by Iowa State University.

Other Concerns: Alternative Markers

Need to apply alternative markers: The *R1-nj* marker works in a in a wide range of donor genotypes since the majority of commercial corn is unpigmented. However, the marker may be suppressed by inhibitor genes (e.g. C1-I), that are carried by the female parent (Fig. 13B).

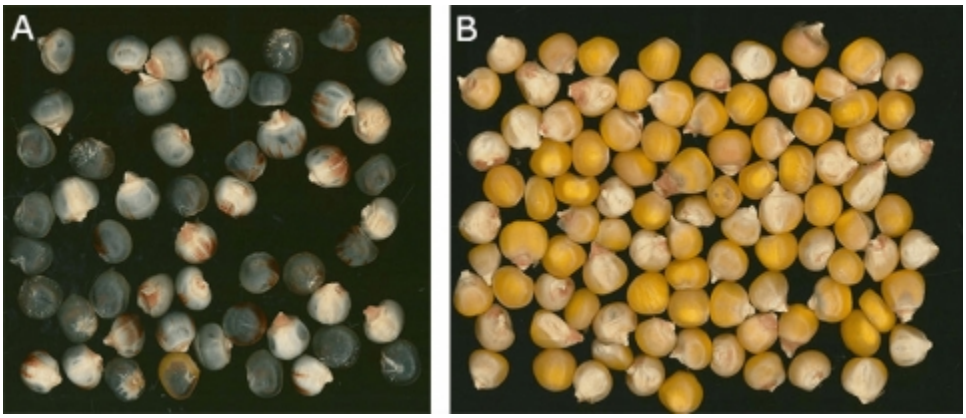


Fig. 13 Phenotypic evaluation of haploid seed may not work all the time. For example, due to coloration (left) or inhibition of R1-nj expression (right). Photos by Iowa State University.

Other Concerns: Toxicity

Toxicity of chemical inducers: Colchicine inhibits **microtubule** polymerization during meiosis by binding to tubulin, one of the main constituents of microtubules. However, colchicine is also very toxic. Less toxic inhibitors of mitosis than colchicine are presently under evaluation or already in use for large-scale chromosome doubling programs. These include (a) herbicides, e.g., Pronamid, Trifluralin, and Oryzalin; (b) caffeine; and (c) nitrous oxide.

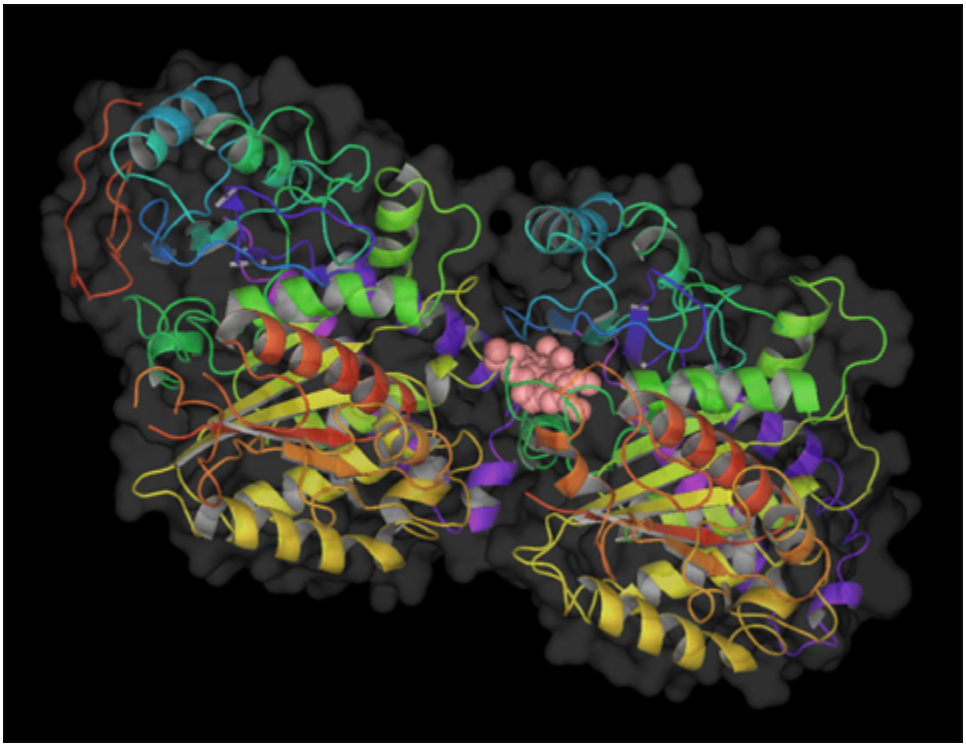


Fig. 14 Colchicine binds to tubulin, one of the main constituents of microtubulues. Image by Group6-3. Licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons.

Genomic Tools for Hybrid Breeding

Description

The seed of a hybrid variety used for a commercial planting is produced by crossing two inbred parent lines from different heterotic groups. Individuals within a F_1 hybrid variety are genetically heterozygous and homogeneous.

The two main goals of hybrid breeding are to maximize the agronomic performance (hybrid performance) and to identify the best performing genotype, while being able to reproduce this one genotype from its homozygous parents.

Part of the superiority of hybrids compared to inbred lines comes from heterosis. Parental lines have to perform sufficiently well, in particular the “seed parent”, on which hybrid seed will be produced. More important for selecting inbred lines in the breeding process is their general and specific combining ability.

Per se performance of inbred lines is a poor predictor for their combining ability, i.e., the yield potential of respective hybrids produced with those inbred lines. Thus testcrosses to determine general and later specific combining ability are crucial to identify best inbred line combinations.

Fig. 15 Hybrid seed from a production company in Uganda. Photo by Iowa State University

Breeding Scheme

As only 100 inbred lines in each of two heterotic groups result in 100 x 100 = 10,000 potential hybrids (Fig. 16), any procedures that identify the most promising combinations contribute substantially to the efficiency of hybrid breeding programs. Molecular and biotechnological tools contribute to more efficient hybrid breeding schemes (see bullets in Fig. 16).

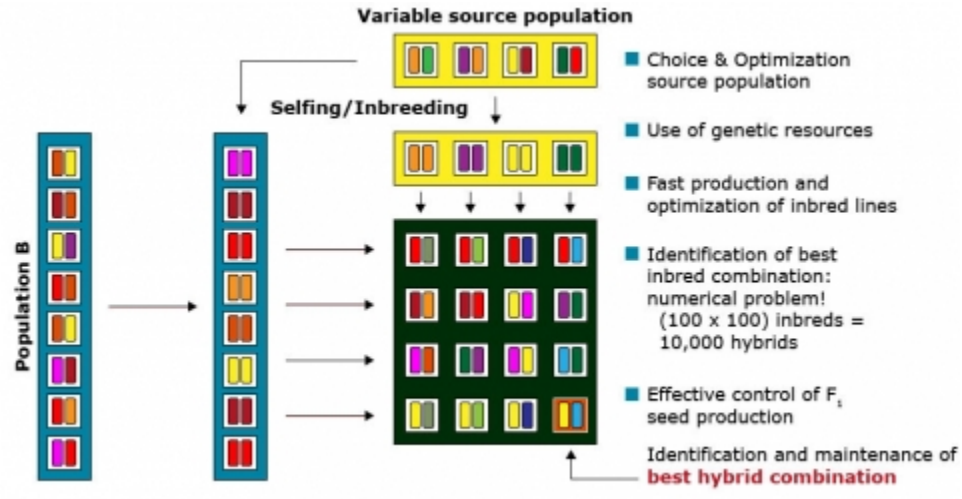


Fig. 16 Simplified hybrid breeding scheme.

Associations

Molecular markers are useful to assign inbred lines to heterotic groups based on their genetic similarity, e.g., by a principle coordinate analysis (Fig. 17).

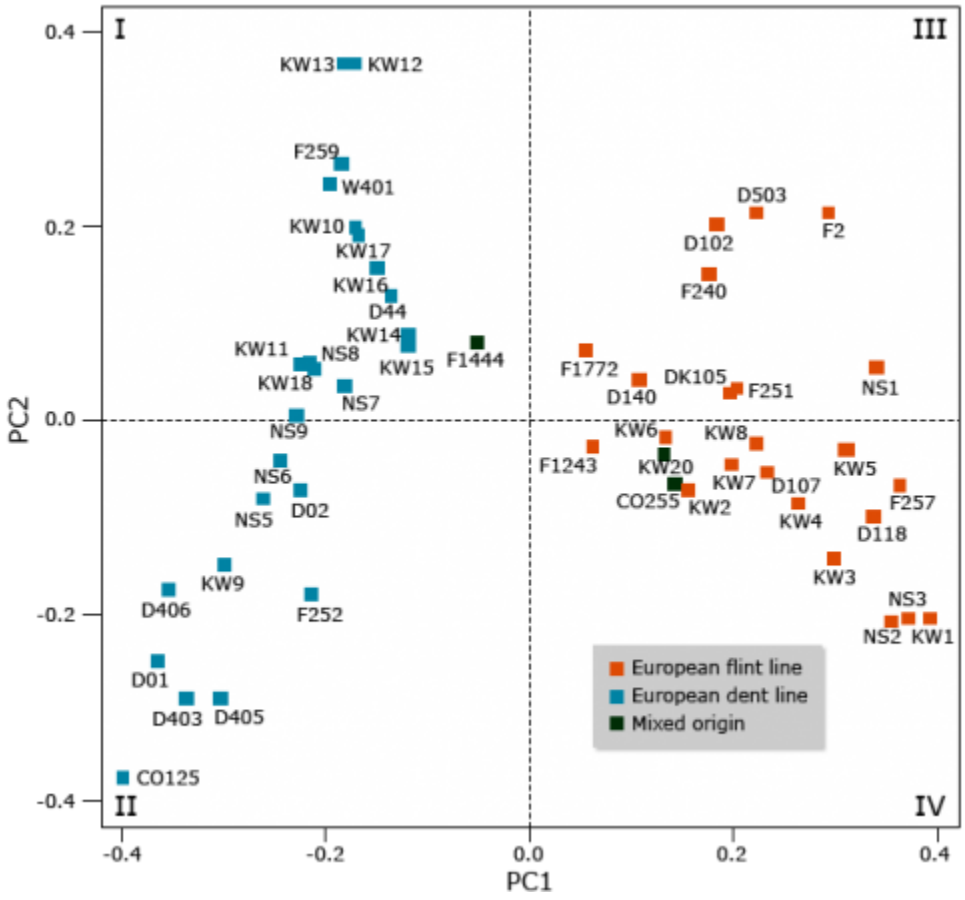


Fig. 17 Associations among maize inbred lines revealed by principal coordinate analysis performed on genetic similarity estimates calculated from AFLP data. PC1 and PC2 = first and second principal coordinates. Adapted from Lübberstedt et al., 2000.

Molecular Basis of Heterosis

Three traditional hypotheses try to explain heterosis: dominance, overdominance, and epistasis.

In the **dominance** hypothesis, superiority of hybrids is caused by total or partial dominance, due to masking of undesirable recessive alleles from one inbred parent by dominant alleles from the other inbred parent.

The **overdominance** hypothesis posits that hybrid vigor is caused by superior performance of heterozygotes due to over- dominance at loci contributing to the trait of interest.

The interaction of favorable alleles at different loci (i.e., **epistasis**) is another classical explanation of hybrid vigor.

Fig. 18 Sorghum is one commodity crop whose productivity can be enhanced by hybridization. Photo by Iowa State University.

Changes in Gene Expression

Another important factor leading to superiority of hybrids over inbred parents are changes in gene expression (Figs. 19 and 20). Gene expression describes regulation of gene activity according to the physiological demands of a particular cell type, developmental stage, or environmental condition. In the context of gene expression, DNA sequence motifs in the vicinity of the structural portion of the gene that are necessary for gene expression are referred to as cis-elements. Transcription factors that bind to cis-elements are referred to as trans-acting factors. The combination of cis- and trans- regulation in allele specific gene expression might lead to significant increase in the hybrid performance over the parental lines. However, a gene that is exclusively subjected to trans-regulation is expected to provide an equal expression of both alleles in the hybrid, whereas genes exposed to cis-regulation will exhibit unequal expression of the two alleles in the hybrid (Figure 22; Hochholdinger and Hoecker, 2007).

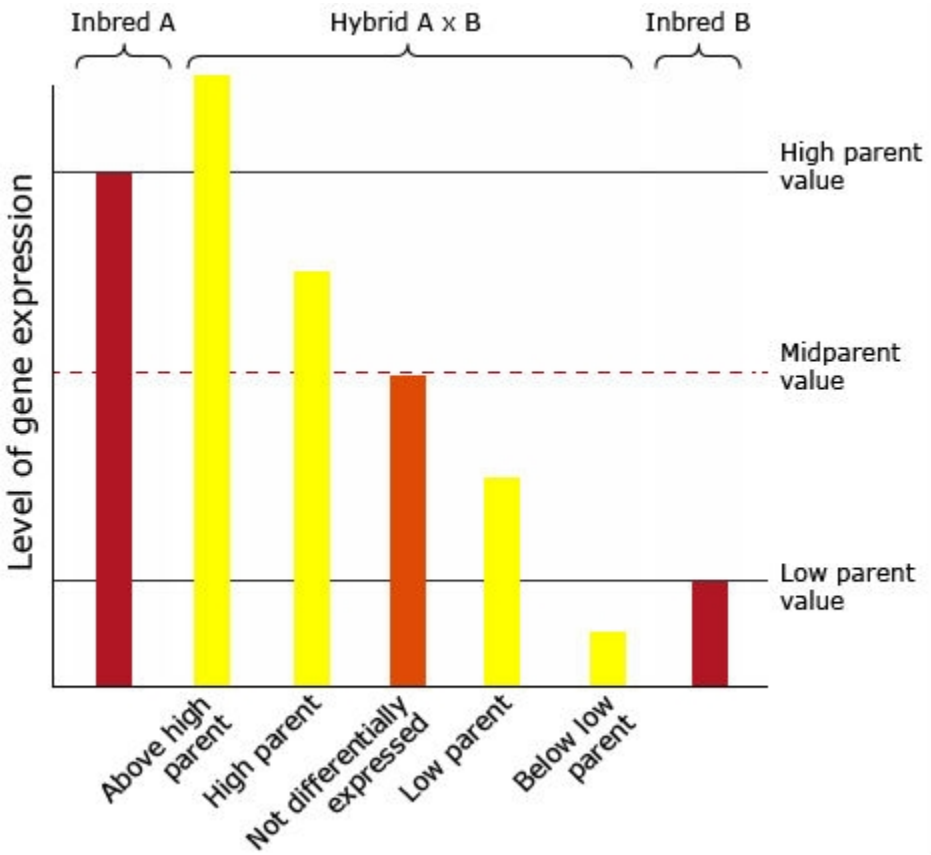


Fig. 19 depicts relative levels of gene expression with parental lines (Inbred A and Inbred B) and their F_1 hybrid (Hybrid A x B). Adapted from Hochholdinger and Hoecker, 2007.

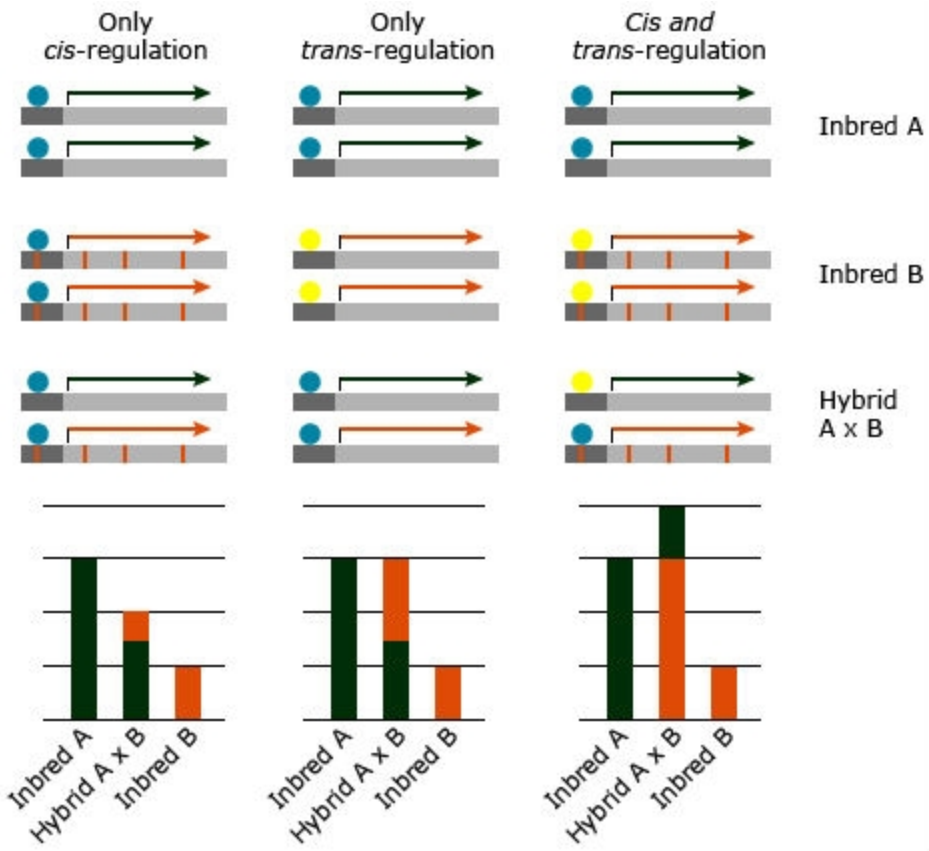


Fig. 20 depicts regulation of allele-specific gene expression in hybrids. Adapted from Hochholdinger and Hoecker 2007.

Gene Expression Studies

These studies (Table 2) analyzed heterosis- associated gene expression in various species by comparing expression patterns of selected genes in inbred lines and hybrids.

Table 2 Expression analyses show either additivity or nonadditivity or both, depending on the approach, developmental stage, and tissue. Source: Hochholdinger and Hoecher, 2007.

Plant organ	Developmental stage	Approach	Genetic background	Global expression trend
Maize				
Embryo	6 DAP	12K cDNA microarrays SSH	UH005 UH301	Additivity
Endosperm	10, 14, 21 DAP	GeneCalling	7 Pioneer® inbred lines	Nonadditivity
Endosperm	18 DAP	RT-PCR	B73 BSSS53	Nonadditivity
Embryo	19 DAP	13.5 microarrays	Mo17	Additivity
Seedling	11 DAG		B73	
Immature ear				
Seedling	14 DAG	14K cDNA microarrays qRT-PCR	Mo17 B73	Additivity
Shoot apical meristem	21-23 DAP	12K cDNA microarrays qRT-PCR	UH002 UH005 UH250 UH31	Nonadditivity
Adult leaves of di- and triploids		Quantitative Northern blotting	Mo17 B73	Nonadditivity
Arabidopsis				
First Leaves	21, 24 DAG	6KcDNA	Col Ler	Nonadditivity

Plant organ	Developmental stage	Approach	Genetic background	Global expression trend
			Cvi	
Rice				
Panicle	Stage III, IV, V	9K cDNA microarrays	Zhenshan97 Minghui63	Additivity

Molecular Insight

The molecular basis of heterosis is not well understood. However, continuing efforts to understand heterosis at the molecular level are providing new insights. In comparative genomics, colinearity describes the conservation of the gene order within a chromosomal segment between different species, resulting in linear arrangement of DNA, mRNA, and the resulting protein sequence. However, when two different cultivars of the same species are mated, chromosome pairing during meiosis allows crossover between colinear genes resulting in meiotic products that could differ in gene content and colinearity (noncolinearity). Some studies have identified several hundreds of genes that display presence/absence variation among investigated lines indicating a very high level of noncolinearity.

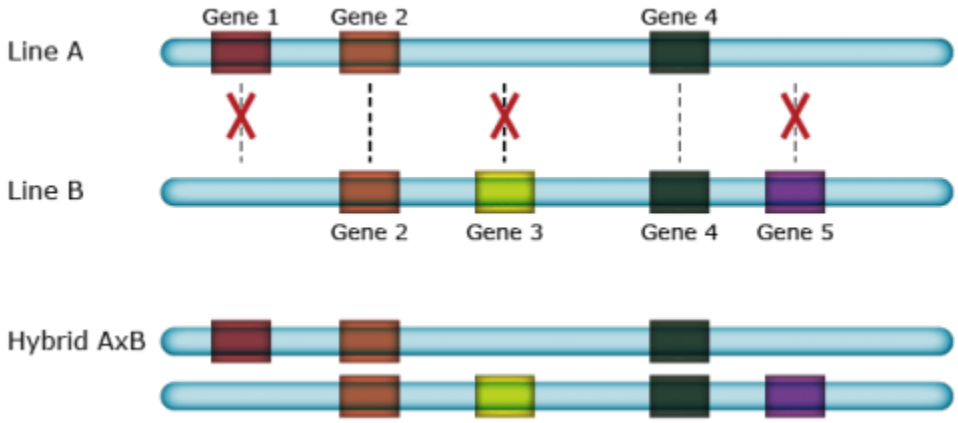


Fig. 21 Genes will stay in the same order on their chromosomes when hybrids are bred.

Hemizygous Complementation

Hemizygous complementation of many genes with minor quantitative effects in hybrids might lead to superior performance of F_1 hybrid plants over their parental inbred lines (Fig. 22). Moreover, given that genes are present in one but absent in other inbreds, any hybrid will have a larger number of different genes (albeit only in one copy), than each of the two inbred parents. The presence of hemizygous genes with minor effect could also explain the inbreeding depression after many generations of selfing due to the loss of hemizygous genes (Fu and Dooner 2002), and /or the lower number of different genes, compared to heterozygous genotypes.

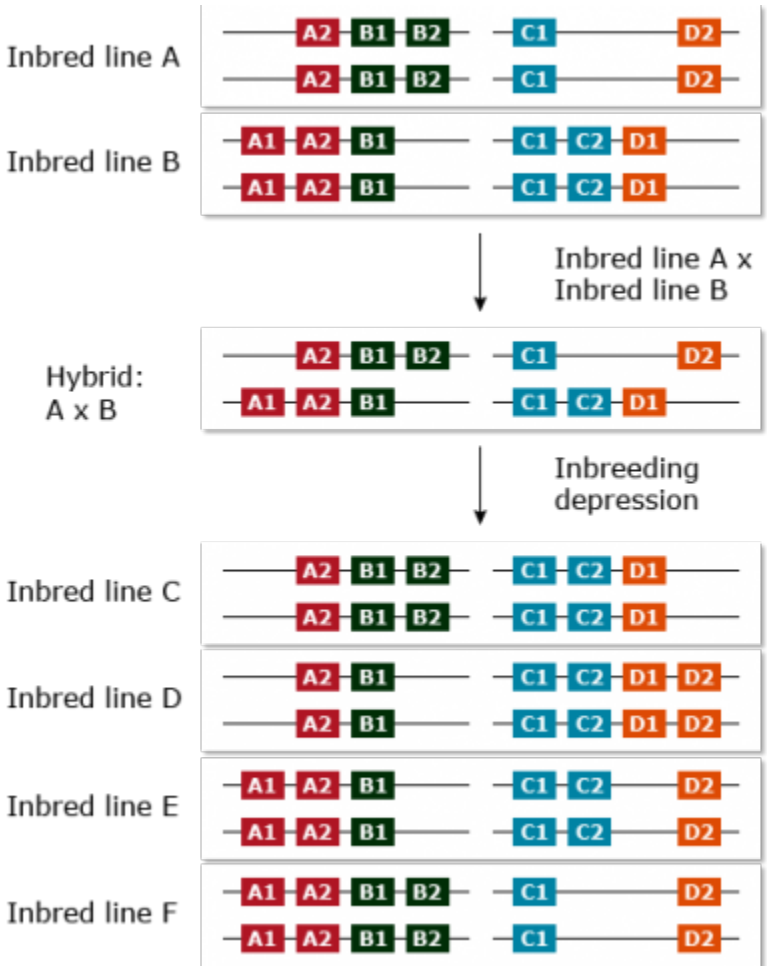


Fig. 22 Hemizygous complementation in maize hybrids. Adapted from Hochholdinger and Hoecker, 2007.

Genetic Similarity Analysis

Marker Applications For Heterotic Pool Formation And Assignment

DNA markers have been found to be useful for description or establishment of heterotic groups in various crops, and to assign inbred lines to those groups, including maize (Fig. 24), rice, sunflower, sorghum, wheat, triticale, and oat. Subsequently, crosses can be restricted to combinations among divergent groups to maximize hybrid performance.

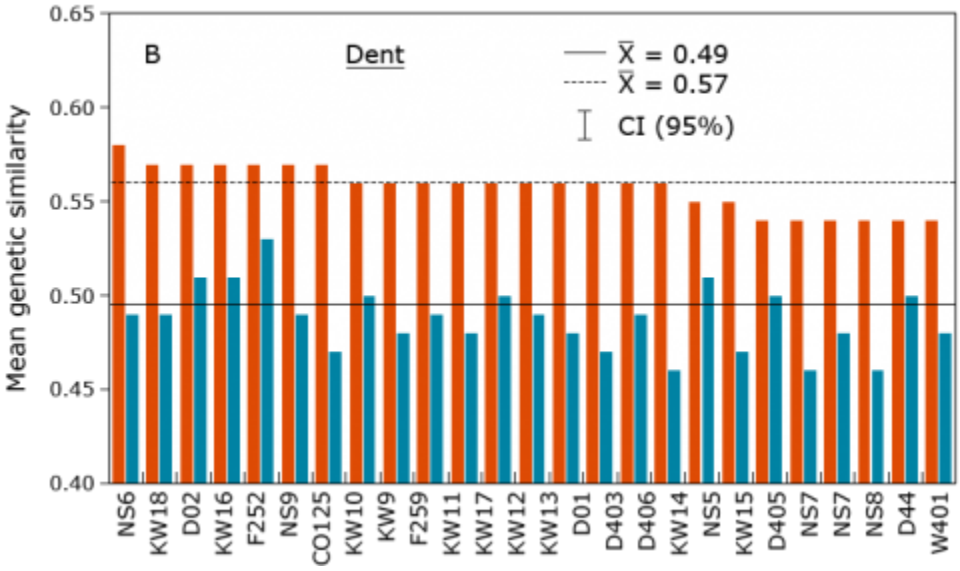


Fig. 23 Mean genetic similarity (GS) calculated from AFLP data for European Dent inbred lines to unrelated lines within the group. White and solid bars refer to mean GS in combination with lines from same heterotic group. Adapted from Lübberstedt et al., 2000.

Genomic Tools to Understand Heterosis

Heterosis, commonly referred to as hybrid vigor, can be expressed in two ways.

- **Mid-parent heterosis** is when the performance of the hybrid exceeds the mean performance of its parents.
- **High-parent heterosis** is when the hybrid performs better than either parent.

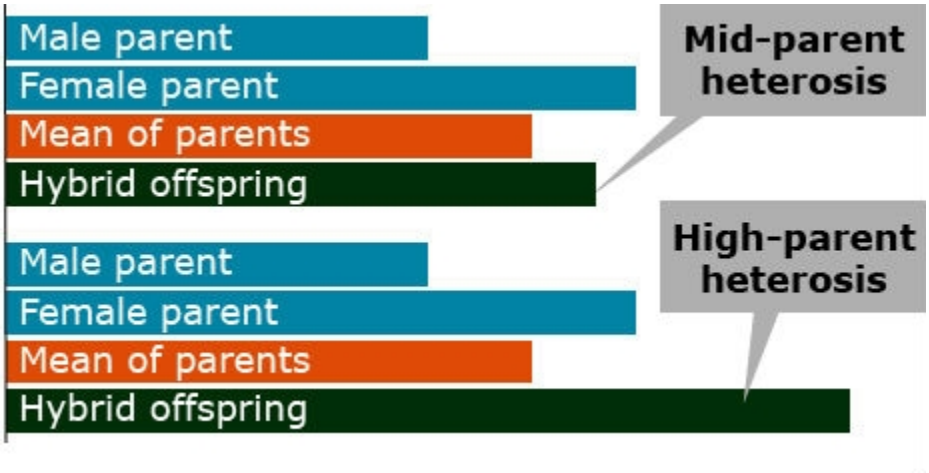


Fig. 24 Expressions of heterosis.

Both types of heterosis are not of commercial interest because they measure the relative performance of hybrids to their inbred parents. If parents are poor performing, heterosis may be high, but the hybrids with the highest heterosis might not be the most superior genotypes. From an agronomic perspective, hybrid performance is most critical, which is the hybrid grain yield (or any other target trait) irrespective of the parental performance.

Maize is an example of a species in which heterotic groups are important for maximizing the performance of hybrid cultivars. One heterotic group in the Midwestern U.S. is referred to as the Iowa Stiff Stalk Synthetic, which was developed by corn breeders of the USDA-ARS and Iowa State University. The other heterotic groups are referred to as non-Stiff Stalk. They include the maize populations Lancaster and Reid Yellow Dent. The best hybrid performance has generally been obtained by crossing inbreds from the Stiff Stalk Synthetic with those from one of the other heterotic groups.

Genomic Approaches For Predicting Hybrid Performance

Field trials to assess hybrid performance are laborious, time consuming and expensive. Testing all possible combinations for a large number of inbred lines to select the best inbred combinations is not feasible in a breeding program. Thus, prediction of hybrid performance and heterosis based on inbred line information is of great interest for plant breeders to evaluate only a small fraction of available inbred lines in the field.

Fig. 25 Hybrid corn seed is obtained by detasseling, as these teenage workers are seen doing in a field near New Ulm, Minnesota in 1974. Photo by Flip Schulke, U.S. National Archives and Records Administration.

DNA-Based Markers (1)

Genomic Approaches For Predicting Hybrid Performance

Example 1: DNA-Based Markers

Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses (Schrag et al., 2009)

In contrast to the work by Frisch et al. (2010) below that used non-DNA markers (mRNA), Schrag et al. (2009) utilized DNA-based markers (AFLP) to estimate hybrid performance in maize.

The following marker-based methods were used:

A. MLR-H: The prediction of hybrid performance is regarded as a multiple linear regression (MLR) problem and the hybrid performance effects ("-H") of the genotypic classes are computed at each AFLP marker locus
B. MLR-LM: Is a hybrid performance prediction approach that uses DNA-based markers and combines line per se performance with mid-parent heterosis ("-LM").
C. TEAM-H: Total effect of associated markers (TEAM) is the sum of marker class effects across AFLP markers that show significant association with a trait of interest. Hybrid performance values ("-H") are regressed on the TEAM values across all hybrids in the experiment.
D. TEAM-LM: Analogous to MLR-LM and used to predict hybrid performance by adding mid-parent heterosis predicted by TEAM and the mid-parent performance estimated from mean of linear regression models of the corresponding parental lines per se performance.

DNA-Based Markers (2)

Genomic Approaches For Predicting Hybrid Performance

Example 1: DNA-Based Markers

From their analyses, Schrag et al. (2009) concluded that DNA-based markers can be used to efficiently predict hybrid performance (Fig. 26).

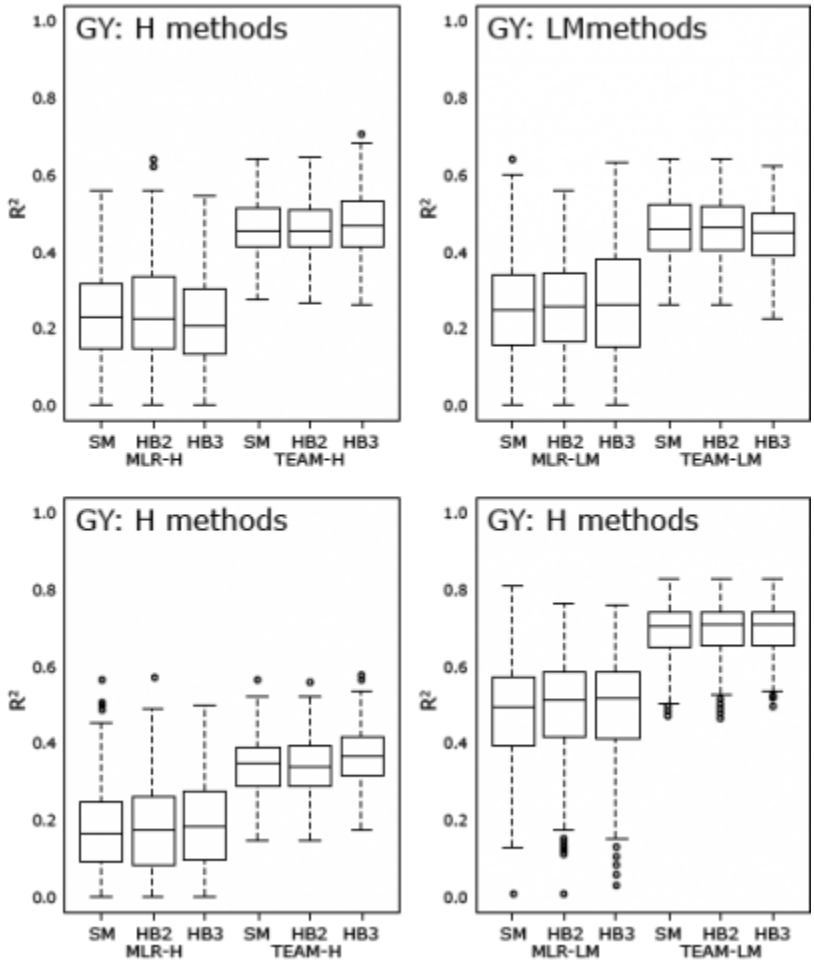


Fig. 26 Efficiency of DNA marker-based methods (MLR-H, MLR-LM, TEAM-H, TEAM-LM) applied to single AFLP marker data (SM) and haplotype blocks (HB2, HB3) for prediction of grain yield (GY) and grain dry matter content (GDMC) of hybrids of which no (Type 0) or only one (Type 1) parental line was evaluated for testcross performance. Adapted from Schrag et al., 2009.

Genomic Approaches For Predicting Hybrid Performance

Example 2: Non-DNA Markers

Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize (Frisch et al., 2010). Frisch et al. (2010) conducted a gene expression study to determine hybrid performance in maize (Fig. 28). In this study, transcription profiles from seedlings of 21 day old parental maize lines of a 7 × 14 factorial with a 46-k oligonucleotide array were analyzed to predict the performance 98 hybrid combinations based on the transcriptome-based distances. Five seedlings per entry were pooled for RNA extraction. The maize 46-k array from the maize oligonucleotide array project (http://www.maizearray.org, University of Arizona, USA) that contain 43381 oligonucleotides (in total 46,128 features) printed on a glass-slide was used for hybridization analyses.

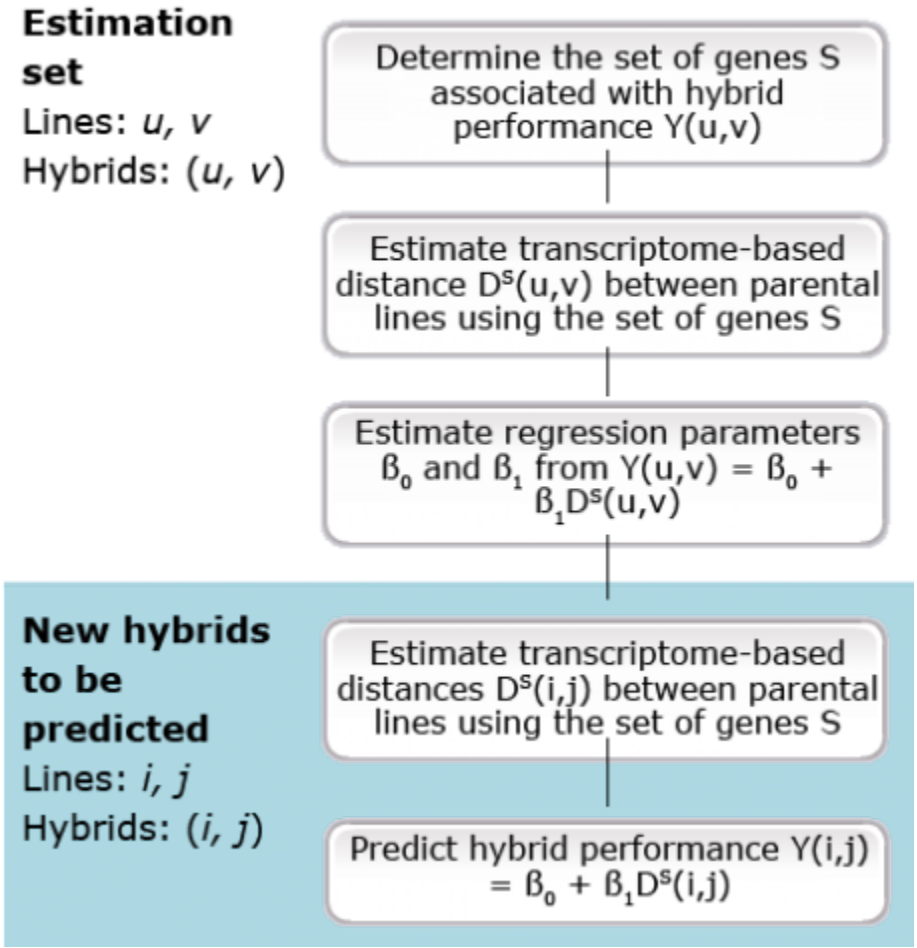


Fig. 27 A transcriptome-based approach to predict hybrid performance. Adapted from Frisch et al., 2010.

Genetic Distance Formula

Genomic Approaches For Predicting Hybrid Performance

Example 2: Non-DNA Markers

D_A = Genetic distance between inbred lines i and j as depicted in the Equation; D_A is used with molecular marker data; in Frisch et al. (2010), AFLP analyses resulted in 1,835 markers.

Invalid Equation

Equation 1

where:
b_m (i) = indicator variable for inbred line i; value = 0 or 1
b_m (j) = indicator variable for inbred line j; = 0 or 1
n_m = number of AFLP bands
SM (i,j) = single matching coefficient

In the equation for genetic distance or DA, b_m(i) and b_m(j) are indicator variables taking the value one (1), if AFLP band m is observed in inbred line i or inbred line j, respectively, and zero (0) otherwise. SM(i,j) is the single matching coefficient.

Example 2: Non-DNA Markers

D_E = Euclidean distance between inbred lines i and j as depicted in the equation; DB is used with gene expression data.

$$D_E(i,j) = \sqrt{\sum_{g-1}^{n_g}[l_g(i) - l_g(i)]^2}$$

Equation 2

where:
l_g (i) = base - two logarithm of transcript abundance
l_g (j) = base - two logarithm of transcript abundance of gene g inbred line j
n_g = number of genes

Genomic Approaches For Predicting Hybrid Performance

Example 2: Non-DNA Markers

D_B = Binary distance between inbred lines i and j as depicted in the equation; DB is used with gene expression data

$$D_B(i,j) = \sqrt{\frac{1}{n_g}\sum_{g=1}^{n_g}[x_g(i) - x_g(j)]^2}$$

Equation 3

where:
x_g (i) = indicator variable or inbred line i; value = 0 or 1
x_g (j) = indicator variable for inbred line j; value = 0 or 1
n_g = number of genes

In the equation for binary distance or DB (Equation 2), xm(i) and xm(j) are indicator variables taking the value 1 or 0, depending on differential gene expression of gene g in inbred lines i and j. If gene g is differentially expressed in lines i and j, then

$x_g(i)$ = 1 and $x_g(i)$ = 0 for $l_g(i) > l_g(j)$, and $x_g(i)$ = 0 and $x_g(i)$ = 1 for $l_g(i) \leq l_g(j)$

If gene g is not differentially expressed, then

$x_g(i) = x_g(i) = 0$

In the latter case, the equation 3 simplifies to, $$D_u(i,j) = \sqrt{\frac{n_s(i,j)}{n_g}}$$ where n_s(I,j) is the number of genes differentially expressed in line i and j.

Genomic Approaches For Predicting Hybrid Performance

Example 2: Non-DNA Markers

The distances $\mathbf{D_B}$ and $\mathbf{D_E}$ were determined from the subset of genes $\mathbf{S_P}$, comprising 10,810 differentially expressed genes. $\mathbf{S_P}$ is the subset of genes that were differentially expressed in at least one pair of parental lines. For the r value in Fig. 28, ns = $P>0.05$ and *** = $P\leq0.001$. The performance of the 98 hybrids was assessed in the field. Multivariate analyses for germplasm grouping was used and showed that the transcriptome-based distances were powerful as other DNA based markers to separate flint from dent inbred lines (Fig. 28). Note that the differentially expressed genes associated with hybrid performance and/or heterosis were identified in an estimation set, and then used to predict new hybrids. The correlations presented in Fig. 28 are for hybrids, which have not been used to pick the yield associated genes.

Frisch et al. (2010) suggested that the close positive significant correlations between the transcriptome-based distances with hybrid performance and heterosis (Fig. 28) may be explained by: (i) the high density of transcriptome loci, which was as a consequence of a high number of differentially expressed genes, indicating good coverage of the genes underlying grain yield, (ii) RNA expression profiling investigates directly the genes, and does not rely on LD between marker alleles and trait of interest, therefore, it is not affected by different linkage phases in different heterotic pools and directly quantifies functional genes between two lines, and (iii) the contribution of additive–additive interactions, which may increase the proportion of phenotypic variance explained by the transcriptome-based distances (Frisch et al., 2010).

According to Frisch et al. (2010), transcriptome-based selection is a promising procedure to predict hybrid peformance in the future. Two main advantages could be attained from RNA expression profiling: (i) enhancing the efficiency of the hybrid breeding program by selecting seedlings directly after inbred line production rather than testing inbred line combinations for many seasons and/or analyzing specific tissues, and (ii) with the reduction in the transcriptome analysis cost in the future, pre-selection at the seedling stage can improve the cost efficiency of hybrid plant breeding programs. It view of high correlations between transcriptome-based distances and hybrid performance ($r\approx 0.80$), it could be concluded that indirect selection based on transcriptome-based distances has the same efficiency as that of direct selection under field conditions (Frisch et al., 2010).

For the prediction of hybrid performance and heterosis, transcriptome data have two advantages over DNA marker data: (i) they do not rely on linkage disequilibrium between marker alleles and QTL alleles, and (ii) they quantify directly the expression of genes, since this analysis not only determines if specific genes are present, but also the degree to which the genes are up or down-regulated. Consequently, transcriptome-based approaches may be superior to DNA marker-based approaches in some situations.

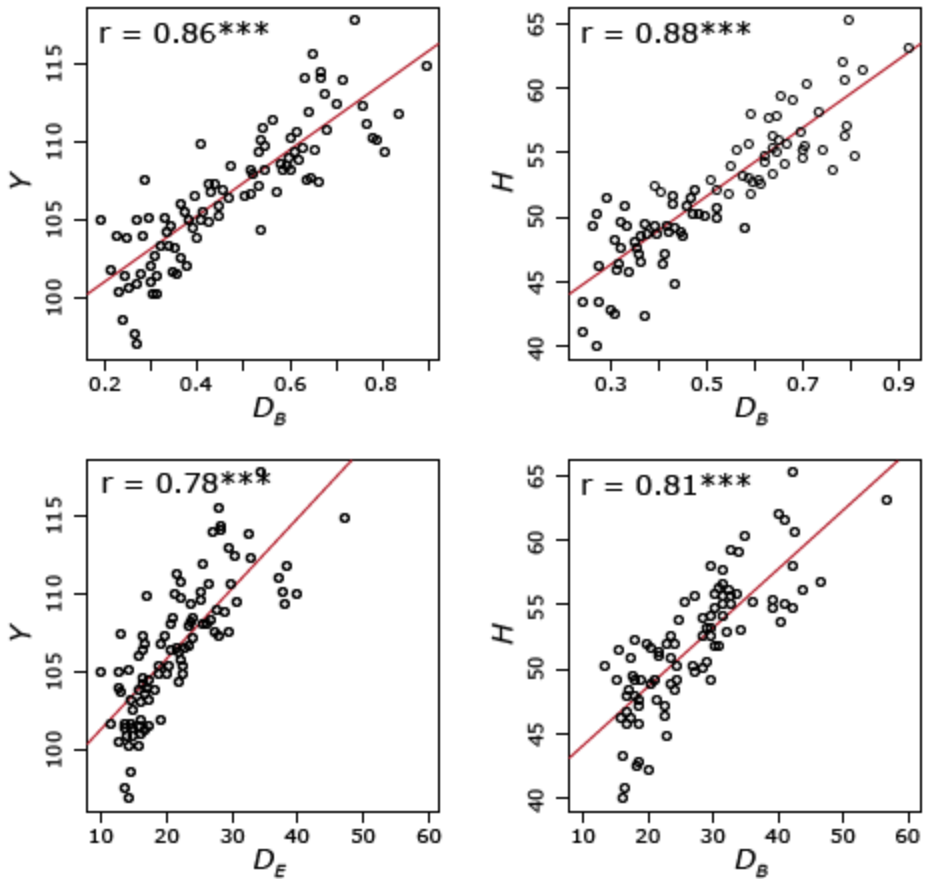


Fig. 28 Correlation of hybrid performance (Y) and mid-parent heterosis (H) for grain yield with the binary distance DB and Euclidean distance DE. The distances were determined from a subset of genes (Sy) containing 1,424 genes whose expression pattern is associated with hybrid performance and another (Sh) containing 1,763 genes associated with heterosis. Adapted from Frisch, et al., 2010.

Genomic Approaches For Predicting Hybrid Performance

Example 3

Correlation between parental transcriptome and field data for the characterization of heterosis in maize (Thiemann et al., 2010)

The study by Thiemann et al. (2010) compared parental inbreds in a mixed pool crosses using microarray analysis. The study also examined correlation of gene transcript abundance to mid-parent heterosis and hybrid performance for grain yield and grain dry matter concentration. The third objective of the study was to perform gene ontology (GO) analyses for functional comparison of gene groups correlated in their parental expression level for hybrid performance for grain yield and grain dry matter concentration. Lastly, Thiemann et al. (2010) characterized the function of gene groups correlated with mid-parent heterosis for grain yield.

Fig. 29 The study by Thiemann et al. examined maize crops from the University of Hohenheim in Germany. Photo by

Christian Fischer; licensed under CC BY-SA 3.0 via Wikimedia Commons.

Interwoven Loop Design

Genomic Approaches For Predicting Hybrid Performance

Example 3

Correlation between parental transcriptome and field data for the characterization of heterosis in maize (Thiemann et al., 2010)

The objective of the study by Thiemann et al. (2010) was to compare parental inbreds in a mixed pool crosses using microarray analysis. The study also examined correlation of gene transcript abundance to mid-parent heterosis and hybrid performance for grain yield and grain dry matter concentration. The third objective of the study was to perform gene ontology (GO) analyses for functional comparison of gene groups correlated in their parental expression level for hybrid performance for grain yield and grain dry matter concentration. Lastly, Thiemann et al. (2010) characterized the function of gene groups correlated with mid-parent heterosis for grain yield.

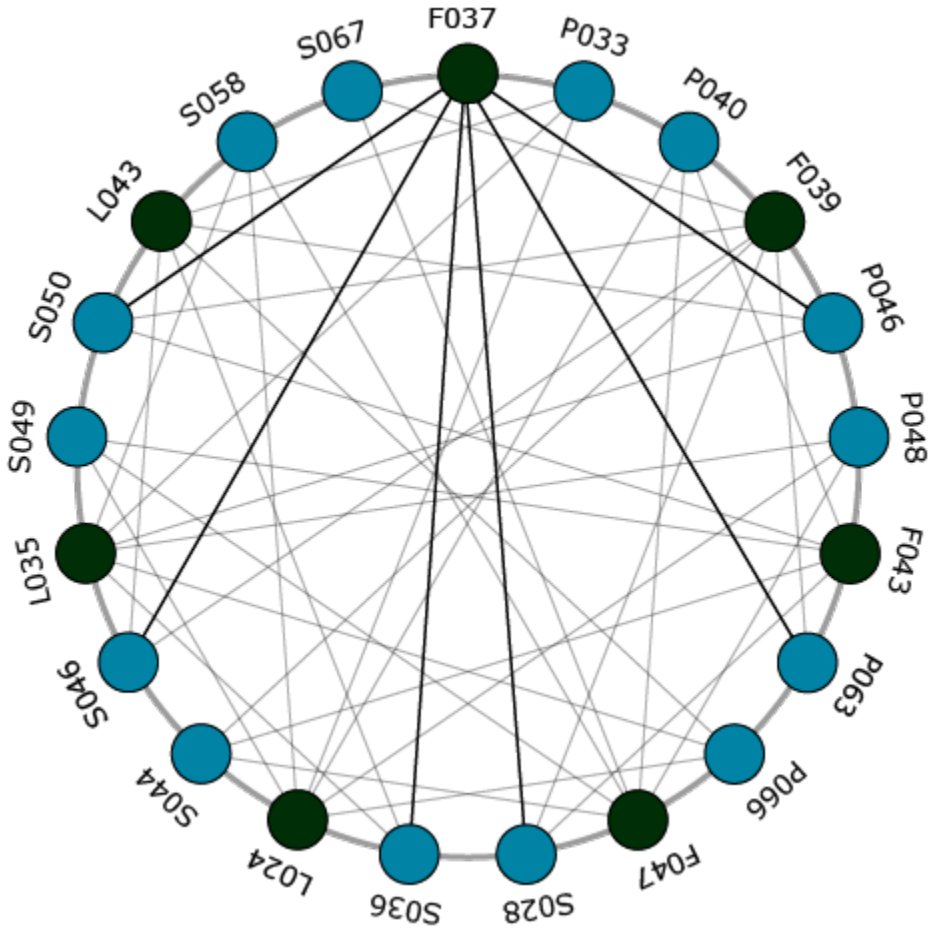


Fig. 30 Interwoven loop design of a microarray experiment. The blue and green circles show 7 flint and 14 dent inbred

lines, respectively. The lines represent the crossing schemes and the bold lines show the general scheme of the mixed-pool hybridizations. Adapted from Thiemann et al., 2010.

Genomic Approaches For Predicting Hybrid Performance

Example 3

Correlation between parental transcriptome and field data for the characterization of heterosis in maize (Thiemann et al., 2010)

The objective of the study by Thiemann et al. (2010) was to compare parental inbreds in a mixed pool crosses using microarray analysis. The study also examined correlation of gene transcript abundance to mid-parent heterosis and hybrid performance for grain yield and grain dry matter concentration. The third objective of the study was to perform gene ontology (GO) analyses for functional comparison of gene groups correlated in their parental expression level for hybrid performance for grain yield and grain dry matter concentration. Lastly, Thiemann et al. (2010) characterized the function of gene groups correlated with mid-parent heterosis for grain yield.

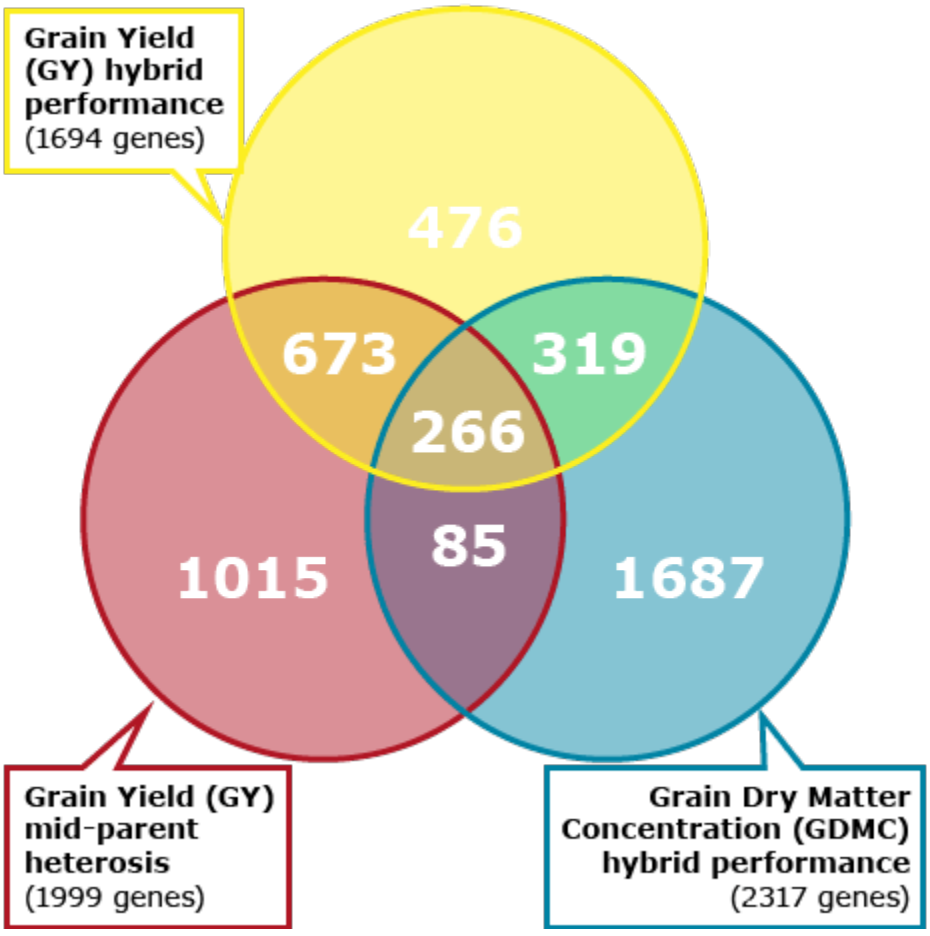


Fig. 31 Venn diagram of trait-correlated genes showing the number of genes whose mid-parent expression level is

correlated to hybrid performance for grain yield and grain dry matter concentration, as well as the genes correlated to mid-parent heterosis for grain yield. Adapted from Thiemann et al. 2010.

Genomic Approaches For Predicting Hybrid Performance

Example 3

Correlation between parental transcriptome and field data for the characterization of heterosis in maize (Thiemann et al., 2010)

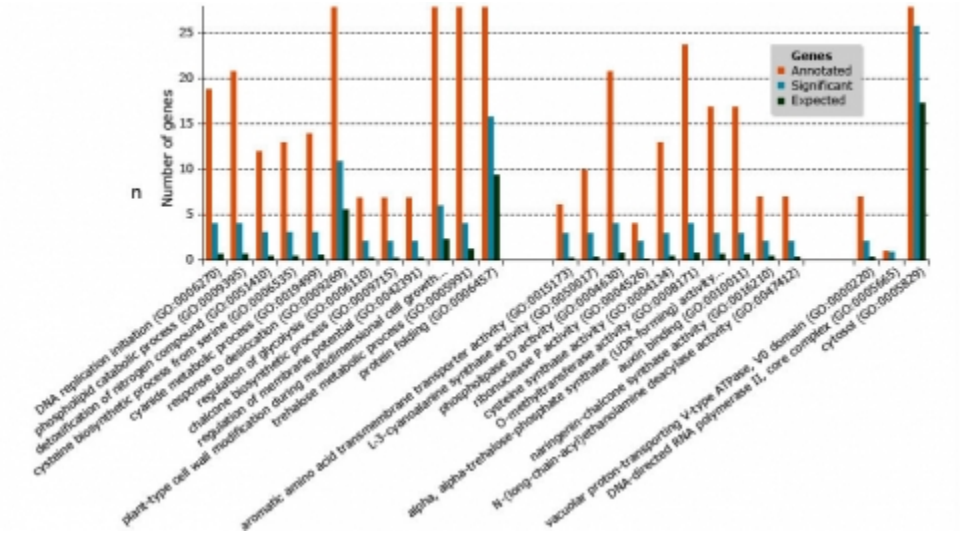


Fig. 32 Overrepresented GO terms among genes correlated to hybrid performance for grain yield. Adapted from Thiemann et al. 2010.

Non-DNA vs. DNA-Based Markers

Recall in the advantage of DNA markers is that they are not affected by environmental factors. However, presence of a particular DNA sequence may not always lead to the expected expression for a trait of interest. This is because the expression of a particular allele depends on environmental conditions, and also interaction with other genes. Thus, even though an allele with a known effect on a particular trait is present, it might not result in the expected phenotype.

Therefore, DNA markers are considered to be a measure of the genetic potential of an individual. The equivalent in human genetics is the **risk concept**. Based on DNA information, it is possible to predict the risk of a patient for showing a particular condition (e.g., 30% to get pancreatic cancer at a certain age). However, whether this condition is expressed, depends on other circumstances. In contrast, if RNA- or metabolite-based biomarkers for this cancer type are available, onset of this condition can be predicted with high accuracy. Thus, non-DNA markers are indicative of the realized potential of an individual.

Since variation in gene expression is the main basis for phenotypic variation, and changes in level of gene expression is observed in hybrids compared to their parents (Hochholdinger and Hoeker, 2007), analysis of gene expression may be a better approach to determine hybrid performance. Recent studies have assessed transcriptome (mRNA expression) data to determine hybrid performance (Frisch et al., 2010; Thiemann et al., 2010). The advantage of transcriptome-based approaches is that transctriptome-based distances directly quantify the expression of genes, which may control the phenotype and do not depend on the linkage between markers and genes, which show weak correlation with heterosis.

Reflection

The Module Reflection appears as the last "task" in each module. The purpose of the Reflection is to enhance your learning and information retention. The questions are designed to help you reflect on the module and obtain instructor feedback on your learning. Submit your answers to the following questions to your instructor.

1. In your own words, write a short summary (< 150 words) for this module.
2. What is the most valuable concept that you learned from the module? Why is this concept valuable to you?
3. What concepts in the module are still unclear/the least clear to you?

References

Abdel-Ghani, A. H., and T. Lübberstedt. 2013. Parent selection – usefulness and prediction of hybrid performance. p. 349-368. In Lübberstedt, T., and R. K. Varshney. (eds.) Diagnostics in plant breeding. Springer, Verlag.

Brunner, S., K. Fengler, M. Morgante, et al. 2005. Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17: 343-360.

Falke, K. C., and S. Mahone. 2013. Non-DNA biomarkers. p. 11-37. In Lübberstedt, T., and R. K. Varshney. (eds.) Diagnostics in plant breeding. Springer, Verlag.

Forster, B. P., E. Herbele-Bors, K. J. Kasha, and A. Touraev. 2007. The resurgence of haploids in higher plants. Trends Plant Sci. 12: 268-375.

Frisch, M., A. Thiemann, J. Fu, et al. 2010. Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. Theor. Appl. Genet. 120: 441-450.

Fu, H., and H. K. Dooner. 2002. Intraspecific violation of genetic colinearity and its implication in maize. Proc. Natl. Acad. Sci. USA 99: 9573-9578.

Hochholdinger, F., and N. Hoecker. 2007. Towards the molecular basis of heterosis. Trends Plant Sci. 12: 427-432.

Jones, R. W., T. Reinot, U. K. Frei, et al. 2012. Selection of haploid maize kernels from hybrid kernels for plant breeding using near-infrared spectroscopy and SIMCA analysis. Appl. Spectrosc. 66: 447-450.

Kibite, Solomon. An Isozyme Marker Linked to the N-1 Gene Governing Nakedness in Oat. Lacombe Research Centre, Lacombe, Alberta. Referenced in USDA Oat Newsletter, October 2002.

Lübberstedt, T. L., A. E. Melchinger, C. Duble, et al. 2000. Relationships among early European maize inbreds: IV. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD, and pedigree data. Crop Sci. 40: 783-791.

Murovec, J., and B. Bohanec. 2012. Haploids and Doubled Haploids in Plant Breeding. p. 87-106. In I. Abdurakhmonov (ed.). Plant Breeding. Available from: **http://www.intechopen.com/books/ plant-breeding/haploids-and-doubled-haploids-in-plant-breeding**.

Prigge, V., and A. E. Melchinger. 2012. Production of haploids and doubled haploids in maize. Methods Mol. Biol. 877: 161-172.

Sanamyan, Marina, Petlyakova, Julia, Rakhmatullina, Emma, and Sharipova, Elnora. 2014. Cytogenetic Collection of Uzbekistan, World Cotton Germplasm Resources, Dr. Ibrokhim Abdurakhmonov (Ed.), ISBN: 978-953-51-1622-6, InTech, DOI: 10.5772/58589. Available from: http://www.intechopen.com/books/ world-cotton-germplasm-resources/cytogenetic-collection-of-uzbekistan

Scholten, S., and A. Thiemann. 2013. Transcriptome-based prediction of heterosis and hybrid performance. p. 265-279. In Lübberstedt, T., and R. K. Varshney. (eds.) Diagnostics in plant breeding. Springer, Verlag.

Schrag, T.A., J. Mohring, H.P Maurer, et al. 2009. Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. Theor. Appl. Genet. 11: 741-751.

Thiemann, A., J. Fu, T.A. Schrag, A.E. Melchinger, M. Frisch, and S. Scholten. 2010. Correlation between parental transcriptome and field data for the characterization of heterosis in Zea mays L. Theor. Appl. Genet. 120(2): 401-413.

Tuvesson, S., C. Dayteg, P. Hagberg, et al. 2007. Molecular markers and doubled haploids in European plant breeding programmes. Euphytica 158: 305-312.

Werner, K., W. Friedt, and F. Ordon. 2005. Strategies for pyramiding resistance genes against the Barley Yellow Mosaic Virus complex (BaMMV, BaYMV, BaYMV-2). Mol. Breeding 16: 45-55.

Acknowledgements

This module was developed as part of the Bill & Melinda Gates Foundation Contract No. 24576 for Plant Breeding E-Learning in Africa.

Molecular Plant Breeding Modern Tools for Line Development and Predicting Hybrid Performance
Author: Thomas Lübberstedt, and Walter Suza (ISU)

Multimedia Developers: Gretchen Anderson, Todd Hartnell, and Andy Rohrback (ISU)

How to cite this module: Lübberstedt, T. and W. Suza. 2016. Modern Tools for Line Development and Predicting Hybrid Performance. *In* Molecular Plant Breeding, interactive e-learning courseware. Plant Breeding E-Learning in Africa. Retrieved from **https://pbea.agron.iastate.edu**.