Towards the molecular basis of heterosis

Frank Hochholdinger and Nadine Hoecker

University of Tuebingen, Center for Plant Molecular Biology, Auf der Morgenstelle 28, 72076 Tuebingen, Germany

Heterosis describes the superior performance of heterozygous hybrid plants over their homozygous parental inbred lines. Despite therediscovery of this phenomenon a century ago and its paramount agronomic importance, the genetic and molecular basis of heterosis remains enigmatic. Recently, various pioneer studies described differences in genome organization and gene expression of hybrids and their parental inbred lines. At the genomic level, a significant loss of colinearity at many loci between different inbred lines of maize was observed. At the level of gene expression, complex transcriptional networks specific for different developmental stages and tissues were monitored in maize (Zea mays), rice (Oryza sativa) and Arabidopsis (Arabidopsis thaliana). Integration of this complex expression data might contribute to improve our understanding of the molecular basis of heterosis.

The phenomenon of heterosis

Heterosis (see Glossary), or hybrid vigour, describes the superior performance of heterozygous F1-hybrid plants in terms of increased biomass, size, yield, speed of development, fertility, resistance to disease and to insect pest, or to climatic rigours of any kind compared to the average of their homozygous parental inbred lines [1,2]. Self-pollination of hybrids over several generations leads to a gradual reduction of heterozygosity and vigour in these plants, a phenomenon that is known as inbreeding depression (see Glossary). Hence, heterosis and inbreeding depression are different aspects of the same phenomenon [3]. Heterosis was first described by Charles Darwin in 1876 after he observed that progeny of cross-pollinated maize (Zea mays) were 25% taller than progeny of inbred maize [4]. The phenomenon was rediscovered independently by George H. Shull and Edward M. East in 1908 [5,6]. Since then, heterosis has been extensively exploited in plant breeding, particularly in maize. Today, ca. 95% of the maize acreage in the USA and 65% of the maize acreage worldwide is planted to hybrids [7]. Heterosis is most evident for adult traits (Figure 1a) but is already manifested during embryo [8] and early seedling development [9] (Figure 1b). The terms midparent heterosis (MPH) and best parent heterosis (BPH) describe the degree of phenotypic difference of a trait in a hybrid (F1) compared to its parental inbred lines (P1, P2). MPH indicates that a trait displays a hybrid performance that is significantly better than the average (midparent) value of the two parental inbred lines (MPH = F1 – [(P1 + P2)/2]). BPH on the other hand indicates that a hybrid trait performs significantly better than the better (Pb) of the two homozygous parental inbred lines (BPH = F1 – Pb).

Genetic hypothesis to explain heterosis

Despite the rediscovery of heterosis about a century ago and the suggestion of various genetic models to explain this phenomenon, little consensus has yet been reached about the genetic basis of heterosis [10–13]. The most prominent genetic hypotheses to explain heterosis are the ‘dominance’ (see Glossary) and ‘overdominance’ (see Glossary) hypotheses. Both describe nonadditive phenotypic behaviour as a consequence of genetic differences between distinct homozygous parental inbred lines and their heterozygous hybrids. The dominance hypothesis explains heterosis by the complementing action of superior dominant alleles from both parental inbred lines at multiple loci over the corresponding unfavourable alleles leading to improved vigour of hybrid plants [14–17]. The overdominance hypothesis [5] attributes heterosis to allelic interactions at one or multiple loci in hybrids that result in superior traits compared to the homozygous parental inbred lines. Finally, the epistasis (see Glossary) hypothesis [18], considers epistatic interactions between non-allelic genes at two or more loci as main factor for the superior phenotypic expression of a trait in hybrids. However, it has recently been demonstrated in tomato introgression lines that heterosis is manifested even in the absence of epistasis [19].

In summary, all these genetic hypotheses make the combination of a considerable number of genes responsible for the more vigourous phenotypes of hybrids over inbred

---

Glossary

**Colinearity**: Genetic concept that genes correspond to each other within different individuals of a species and are arranged in the same linear sequence.

**Dominance**: Term from quantitative genetics explaining improved vigour of hybrid plants by the complementing action of superior dominant alleles from both parental inbred lines at multiple loci over the corresponding unfavourable alleles.

**Epistasis**: Term from quantitative genetics explaining the superior phenotypic expression of a trait in hybrids by interactions between non-allelic genes at two or more loci.

**Heterosis**: Improved agronomic performance of heterozygous F1-hybrids in comparison to their genetically different homozygous parents.

**Inbreeding depression**: Gradual reduction of heterozygosity and vigour in hybrid plants after self-pollination of these plants over several generations.

**Overdominance**: Term from quantitative genetics attributing superior hybrid traits compared to their parental inbred lines to allelic interactions at one or multiple loci.
lines. This also implies that positive and negative effects of various loci might compensate each other, which makes it difficult to support one hypothesis over the other [20]. It is important to keep in mind that the quantitative genetics terms ‘dominant’ and ‘overdominant’ cannot be directly associated with the quantitative behaviour of phenotypic traits or with molecular principles.

**Molecular approaches to study heterosis**

With novel molecular tools at hand, several laboratories have recently started to analyze different molecular aspects of heterosis. In the past, quantitative trait loci (QTL) analyses were a first step towards the molecular understanding of heterosis [21]. Although these studies were able to demonstrate that heterosis is defined by a limited number of individual genes inherited in a complex way, none of the QTLs associated with heterotic traits have yet been cloned. However, with novel molecular tools that allow for a comprehensive QTL-based phenotyping followed by map-based cloning, it might be possible to identify loci controlling heterotic traits in the near future [21].

More recently, several studies initiated the dissection of heterosis at the gene expression and genome organization level. These analyses can be classified in three categories. First, genome organization has been analyzed in maize at various loci for different inbred lines, and expression of a limited number of genes in these regions has been tested. Second, transcriptome-wide gene expression profiles have been determined for different developmental stages and tissues in different species including maize, rice (*Oryza sativa*) and *Arabidopsis* (*Arabidopsis thaliana*). Finally, allele-specific contribution to gene expression has been analyzed for a selected number of genes. These studies will be summarized in the following sections.

**Maize genomes of different inbred lines display significant aberrations from genetic colinearity**

One of the basic notions of genetics is the concept of genetic colinearity (see Glossary), which describes the observation that genomes of individuals in a given species contain the same gene content. Recently, several studies observed significant aberrations from colinearity on the microcolinearity level between different inbred lines of maize. First, it was demonstrated that among ten genes in the *bz* region of the McC inbred line, only the proximal six have counterparts in the B73 inbred line [22]. Similarly, whereas 22 gene copies of the α-zein storage protein subfamily *z1C* were detected in the BSSSS3 inbred line, only 15 *z1C* genes were present in the B73 genomic region [23]. Only seven of the BSSSS3 and six of the B73 *z1C* genes had an intact coding region, and only two intact genes were present in both inbred lines. Finally, among 72 genes identified in various chromosomal regions of the inbred lines B73 and Mo17, 27 genes were absent in one of the inbred lines [24].

In maize, many genes are members of small gene families. Therefore, gene deletions in inbred lines might have only minor quantitative effects on plant performance because these genes might often be functionally compensated by duplicate copies elsewhere in the genome [22]. However, hemizygous complementation of many genes with minor quantitative effects in hybrids might lead to a significantly increased performance of hybrid plants and would be consistent with the dominance hypothesis [22] (Figure 2). This concept would also explain inbreeding depression by the loss of functional genes in subsequent generations when hemizygous genes get lost after several rounds of self-pollination of the hybrids (Figure 2). The high degree of non-colinearity in the genomes of different inbred lines of maize might thus, in part, explain the exceptionally high degree of heterosis in this species. However, it is probable that other molecular mechanisms might also be involved in heterosis because it is unlikely that all species that display heterosis contain a degree of non-colinearity in their genome as high as that of maize.

**Heterosis-associated gene expression in maize**

Recently, several studies have analyzed heterosis-associated gene expression in maize, rice and *Arabidopsis* by comparing expression patterns of selected genes in inbred lines and hybrids [8,23,25], or by performing high-throughput gene expression analyses via microarray profiling or GeneCalling [7,26–31]. These studies analyzed various aspects of plant development including very young embryos (six and 19 days after pollination) [8,27],

![Figure 1](https://www.sciencedirect.com)
patterns were observed in these gene expression studies. Whereas in some studies nonadditive gene expression (Figure 3a) was prevalent between inbred lines and hybrids [23, 25, 26, 28, 30], additive gene expression (Figure 3a) was observed in other studies for most of the genes [7, 8, 27, 31]. Finally, in one analysis, a similar number of genes followed additive and nonadditive expression [29] (Table 1). The discrepancy of these results might be the result of significant differences in genotypes, plant material, experimental designs and statistical procedures applied in the various studies. However, it might also be an indication that in different tissues or developmental stages different global expression patterns might prevail, which might nevertheless be related to heterosis. This notion is supported by the observation that different tissues and organs within a hybrid plant display significant differences in their degree of heterosis [20]. Remarkably, it was also observed that triploid hybrids, in general, displayed higher expression levels compared to their diploid counterparts, although the transcript levels were nonadditive [25], thus underscoring the importance of genomic dosage for gene expression in hybrids. Another important finding in those studies that analyzed more than one inbred versus hybrid combination [28, 29] was that no consensus gene set was found that was differentially expressed between all inbred/hybrid combinations. This might indicate that it was not possible to identify any key genes of heterosis in these studies and that global trends of gene expression might correlate with heterosis. Although global trends of gene expression appeared to be different from tissue to tissue, it was observed in at least one study that the proportion of allelic additivity was positively associated with hybrid yield and heterosis for immature maize ears [29]. Moreover, most heterosis-related gene expression analyses have not revealed a predominant functional category to which differentially expressed genes belong. This might be because only a small fraction of the 25 500 Arabidopsis [32] or the estimated 59 000 maize or rice genes [33] were profiled (Table 1). Therefore, it is remarkable that an open-end differential screening of maize genes revealed that 25% of the genes differentially expressed between maize inbred lines and hybrids during early embryo development were associated with signal transduction [8].

Finally, more significant expression differences are found between parental inbred lines than between reciprocal hybrids. For example, whereas 4–18% of the genes were differentially expressed between two inbred lines in various tissues, no differentially expressed genes were identified in these tissues between the reciprocal hybrids [27]. This is not surprising given that reciprocal hybrids are genetically identical and that gene expression differences in reciprocal hybrids are mainly attributed to epigenetic effects. In the case of reciprocal hybrids selected for cold germination and desiccation tolerance, only 1–2% of the genes were differentially expressed [34]. In summary, these gene expression profiling studies represent a first step towards the definition of the complex gene expression networks that might be relevant in the context of heterosis. However, there is currently no direct link between the classical genetic hypothesis and these gene expression profiles. This might be attributed to the complex molecular

endosperm (10, 14, 18 and 21 days after pollination) [23, 26], 11 and 14 day-old seedlings [7, 27], 21–23 day-old shoot apical meristems [28], immature ear tissue [27, 29], young and adult leaves [25, 30], and rice panicles [31] in a wide variety of different genetic backgrounds (Table 1). Besides the differences in the tissues and genotypes analyzed, these studies also differed significantly in their experimental design and their statistical procedures. In summary, no uniform global expression

Figure 2. Hemizygous complementation in maize hybrids. Different maize inbred lines are characterized by a considerable loss of genetic colinearity, that is, the loss of particular genes. It is hypothesized that gene loss can be partly compensated in hybrids by hemizygous complementation, which also leads to an improved performance of hybrids versus inbred lines. The colored boxes A to D represent different genes. Different members of gene families are indicated by the same colour (e.g. A1 and A2 belong to a gene family). The loss of a member of a gene family can be partly compensated for by the action of other members of this gene family and leads only to minor phenotypic effects in inbred lines. When a hybrid is selfed over several generations inbreeding depression is observed because not all genes present in the hybrid will be maintained. To date, partial loss of colinearity has only been observed in maize. In other species as well as in maize, additional factors might also contribute to heterosis.

www.sciencedirect.com
Figure 3. Relative gene expression levels in hybrids and regulation of allele-specific gene expression in hybrids. (a) Gene expression levels in hybrids are not strictly related to the genetic concepts of dominance and overdominance. Therefore, a systematic terminology that was previously suggested [27] should be used to describe gene expression in hybrids A × B (yellow and orange) compared to their parental inbred lines A and B (red). Gene expression in a hybrid gene can be manifested in an additive manner (orange) as the average expression of the parental inbred lines (no differential expression). Alternatively, gene expression can be significantly different from the midparent value either between the high and low parent value, or above the high or below the low parent value (yellow). (b) Effects of cis- and trans-regulation on the allelic contribution to gene expression in hybrids. Genes that are completely subject to cis-regulation reflect the relative expression levels of the parental inbred lines in the allelic ratio of gene expression in the hybrid. Genes that are exclusively regulated by trans-acting factors show equal expression of the two alleles in the hybrid. Genes that are subject to cis- and trans-regulation fall in between the two classes, that is, the relative allelic contribution to gene expression in hybrids for this class of genes neither displays the relative expression levels of the parental inbred lines nor an equal expression of both alleles. Genes are indicated by gray bars. The dark grey 5' end of the genes represents the regulatory promoter and/or enhancer region, whereas the light-grey part corresponds to the transcribed region of the gene. Distinct trans-acting factors (e.g. transcription factors, repressors) are indicated as blue or yellow circles binding to the promoter and/or enhancer region. Cis-differences of genes (e.g. SNPs, insertions, deletions) are indicated as red bars in the grey gene structure. Size of green and orange arrows indicate expression levels of the allele. Green and orange areas in the histograms indicate allelic contribution to gene expression in the inbred lines and hybrids.
regulation that is required for the manifestation of even a single phenotypical trait on the level of gene expression. These regulatory networks could include genes that are up- and downregulated in the hybrid relative to their parental inbred lines, although they are related to the manifestation of the same phenotypical trait. For example, if a gene that encodes for a repressor of a particular pathway is downregulated, other genes in that pathway could be derepressed and thus be upregulated.

Allele-specific gene expression

Global gene expression analyses via microarray technology do not allow for the quantification of allele-specific contribution to gene expression. Allele-specific analyses can help to distinguish between cis- and trans-acting regulation of gene expression. A gene that is completely subjected to trans-regulation is expected to provide a similar contribution of both alleles to gene expression in the hybrid whereas genes subjected to cis-regulation will show unequal expression of the two alleles in the hybrid [35] (Figure 3b). In this context, allele-specific gene expression of 32 genes whose expression in the hybrid deviated from the midparent value was assayed via quantitative SNP-based Sequenom technology [27]. For 18 genes cis-regulation was prevailing, whereas only one gene was classified as trans-regulated. The remaining 13 genes were classified as a combination of cis- and trans-regulation. In another study of parental transcript accumulation via allele specific RT-PCR, 11 of 15 maize genes displayed an unequal expression of the two alleles [36]. Maternal or parental transmission had little effect on allele-specific expression, implying only minimal parent-of-origin effects. Differential response of the two alleles in a hybrid to environmental stress implied unequal functions of the different alleles that might have an impact on heterosis. Finally, allelic expression of the ZmGrp3 gene (a marker for root initiation) in six reciprocal hybrids was assayed but no significant deviation from a 1:1 expression ratio of the two alleles in the hybrids was detected [37]. In conclusion, allele-specific gene expression was frequently observed in recent studies and might thus represent another regulatory mechanism in the context of the phenotypic manifestation of heterosis. Interestingly, parent-of-origin effects apparently play only a minor role for allele-specific gene expression whereas environmental factors might have a more severe influence on allelic contribution to gene expression [27] and, thus, heterosis.

Concluding remarks

Recently, several heterosis-related molecular studies have been performed. First, a considerable loss of gene colinearity in the maize genome indicates that, at least in maize, these genomic differences might be a key to some aspects of heterosis. The sequencing of complete genomes of various genotypes in different species [38–40] will help to better understand the relevance of genome organization in relation to heterosis. Moreover, various gene expression studies including allele-specific expression analyses revealed significant differences in global gene expression patterns in different tissues and developmental stages. Detailed expression profiling experiments under standardized conditions of experimental design, microarray platforms and data analysis will be required to fully understand the underlying mechanisms of heterosis.
analysis in a multitude of genotypes, developmental stages and organs will further refine the insight in genes that might be differentially expressed between inbred lines and hybrids and, therefore, play a role during heterosis manifestation. These experiments might also be helpful for heterosis predictions based on expression profiles. The functional characterization of such differentially expressed genes will determine their relevance in the context of heterosis. Finally, future progress in the integration of genomic tools and the mapping and cloning of complex heterosis associated QTLs via association mapping and map-based cloning might allow to map and clone simultaneously multiple key, genetically-unrelated genes of heterosis. A major challenge in this context is to monitor and quantify adequately the complex heterotic phenotypes that are related to almost all heterotic traits. In conclusion, the molecular data available thus far, do not indicate a simple correlation between one of the genetic hypotheses and the molecular events leading to heterosis. Thus, the availability of novel genetic and genomic tools that allow for the integrated study of the complex interactions between genome organization and expression might contribute to a better understanding of heterosis.

Acknowledgement
This work was supported in part by the DFG framework program SPP1149 ‘Heterosis in plants’ (award HO2249/6).

References