Theor Appl Genet (2006) 113:831–845
DOI 10.1007/s00122-006-0335-x

ORIGINAL PAPER

Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis

Mei Guo · Mary A. Rupe · Xiaofeng Yang · Oswald Crasta · Christopher Zinselmeier · Oscar S. Smith · Ben Bowen

Received: 16 February 2006 / Accepted: 3 June 2006 / Published online: 26 July 2006
© Springer-Verlag 2006

Heterosis, or hybrid vigor, has been widely exploited in plant breeding for many decades, but the molecular mechanisms underlying the phenomenon remain unknown. In this study, we applied genome-wide transcript profiling to gain a global picture of the ways in which a large proportion of genes are expressed in the immature ear tissues of a series of 16 maize hybrids that vary in their degree of heterosis. Key observations include: (1) the proportion of allelic additively expressed genes is positively associated with hybrid yield and heterosis; (2) the proportion of genes that exhibit a bias towards the expression level of the paternal parent is negatively correlated with hybrid yield and heterosis; and (3) there is no correlation between the over- or under-expression of specific genes in maize hybrids with either yield or heterosis. The relationship of the expression patterns with hybrid performance is substantiated by analysis of a genetically improved modern hybrid (Pioneer® hybrid 3394) versus a less improved older hybrid (Pioneer® hybrid 3306) grown at different levels of plant density stress. The proportion of allelic additively expressed genes is positively associated with the modern high yielding hybrid, heterosis and high yielding environments, whereas the converse is true for the paternally biased gene expression. The dynamic changes of gene expression in hybrids responding to genotype and environment may result from differential regulation of the two parental alleles. Our findings suggest that differential allele regulation may play an important role in hybrid yield or heterosis, and provide a new insight to the molecular understanding of the underlying mechanisms of heterosis.

Communicated by F. van Eeuwijk

Electronic Supplementary Material Supplementary material is available to authorised users in the online version of this article at http://dx.doi.org/10.1007/s00122-006-0335-x.

M. Guo (✉) · M. A. Rupe · X. Yang · C. Zinselmeier · O. S. Smith · B. Bowen
Pioneer Hi-Bred International, Inc., 7250 NW 62nd Avenue, Johnston, IA 50131, USA
e-mail: mei.guo@pioneer.com

O. Crasta
CuraGen Corp., New Haven, CT, USA

Present Address:
O. Crasta
Virginia Bioinformatics Institute, Virginia, VA 24061, USA

Present Address:
B. Bowen
Gene-fx, Inc., Berkeley, CA 94705, USA

Introduction

Heterosis is a term first introduced by Shull (1908) to describe the superior performance of hybrid progeny compared to their inbred parents. Heterosis in plants is associated with increases in grain yield, vegetative growth rate, tolerance to pests and environmental stress, accelerated maturity, and many other changes in desirable agronomic characteristics. In maize breeding, heterosis has been widely exploited for many decades, but there is still a very limited understanding of the underlying genetic or molecular mechanisms.

Two fundamental questions regarding the molecular mechanism of heterosis are the relationship between

yield heterosis and gene expression in the hybrids, and how the two different alleles brought together in the hybrids are expressed. The superior performance of the hybrid over the mean of the inbred parents may result from the altered regulation of gene expression in the hybrids, either at the global level or for specific classes of genes. One possible scenario is that the two different alleles brought together in the hybrid create a combined allelic expression pattern in hybrids. Complementation of the allelic expression differences may result in different overall developmental expression patterns in the hybrid compared to the inbred parents. For example, the timing of gene expression may differ in the parents such that the expression pattern in the hybrid results in an extended time of gene expression. Alternatively, at some loci, allelic interaction or a change in the spectrum of trans-acting factors causes gene expression in the hybrid to deviate from simple additive allelic expression patterns of the parents (Birchler et al. 2003; Gibson and Weir 2005).

Early studies reported examples of increased mRNA quantity or protein amount expressed in hybrids as compared to their inbred parents (Leonardi et al. 1991; Romagnoli et al. 1990; Tsaftaris et al. 1995, 1999) and suggested that increased gene expression level in the hybrids may contribute to heterosis. More recent studies of gene expression in hybrid maize have shown both allelic additive types of expression regulation and deviation from it in the triploid endosperm tissue (Guo et al. 2003; Son and Messing 2003) and in leaf tissue (Auger et al. 2005). However, there is still a paucity of data demonstrating any relationship between any expression patterns in maize or those documented in other plants (Bao et al. 2005; Vuylsteke et al. 2005) and yield heterosis.

We have previously shown with a subset of randomly selected genes that the two parental alleles in maize hybrids can be regulated differentially, at the cumulated transcript level, in different tissues and in different environments (Guo et al. 2004). The study shows that transcript regulation can be allele-specific and that specific expression patterns of the two parental alleles in the hybrid manifest during development and in different environments is possibly associated with hybrid performance.

In the present study, we used GeneCalling technology (Shimkets et al. 1999) to profile mRNA expressed in the immature ear tissue (before pollination) of a series of 16 maize hybrids that varied in yield heterosis. GeneCalling technology is an open-ended, gel-based method that permits comprehensive profiling of mRNA abundance for both known and novel genes in an unbiased way (Crasta and Folkerts 2003), although identification of corresponding genes of profiled cDNA fragments requires additional process (such as isolation, cloning and sequencing). This technology detects 80–90% of expressed genes in a given tissue (Shimkets et al. 1999). GeneCalling mRNA profiling has been successfully used in various maize gene expression studies, identifying genes involved in flavonoid biosynthesis (Bruce et al. 2000), root-lodging resistance (Bruce et al. 2001), stress response during maize seed maturation and germination (Kollipara et al. 2002), and demonstrating allelic expression in the maize endosperm of reciprocal hybrids and inbred parents (Guo et al. 2003).

In the present study, we used GeneCalling mRNA profiling technology to analyze a series of 16 maize hybrids that shared a common female parent and vary in the degree of yield heterosis. The objectives of this study are (1) to gain a global view of the level of expression of genes during one stage of development (immature ear tissue at stage V19) in a series of maize hybrids and their inbred parents; (2) to determine whether specific gene expression patterns in the hybrid are associated with differences in heterosis and/or hybrid yield both among genotypes and in different environments with varying levels of stress; (3) whether up- or down-regulated gene expression in the hybrid correlates with heterosis; and (4) what are the effects of differential regulation of the two parental alleles on gene expression in the hybrid background.

Materials and methods

Experimental design and tissue sampling

A series of 16 maize (*Zea mays* L.) hybrids and their respective inbred parents were selected from the collection at Pioneer Hi-Bred International, Inc. (Table 1). The parental inbreds are either Stiff Stalk Iowa Synthetic (S) or Non-Stiff Stalk (NS) lines, which comprise two heterotic pools used widely in maize breeding (Labate et al. 1997). Hybrids were produced by crossing a common female inbred (S1) with a series of male inbreds that share different percentages of pedigree relationship with the female parent. The crosses were either between different heterotic pools (S/NS), or within a heterotic pool (S/S). The resulting hybrids therefore range from highly heterotic to hybrids exhibiting little heterosis. Yield trials were conducted for all 16 hybrids and 17 inbreds in 1997 and 1999. There were four locations and two replicates per location in each year. At one of these four locations where yield trials

Table 1 Description of inbred parents, their pedigree relationships with inbred S1, and the grain yield of their hybrids when crossed as a male to inbred S1

	Characteristics	Pedigree relationship to S1 inbred (%)	Hybrid yield (bu/acr)	Heterosis (bu/acr)
Male parent				
S3	Mid maturity SS (YD) type Central/eastern US CB adaptation	85.5	67.7	24.9
S4	Mid maturity SS (YD) type Central/western US CB adaptation	45.9	75.3	15.6
S5	Mid/early maturity SS (YD) type	47.3	82.4	36.1
S6	Public line from Iowa SS (100%), YD type	63.9	93.8	52.4
S7	Public line from Iowa SS (100%), YD type	60.5	98.6	60.6
S8	Mid maturity SS (YD) type Eastern/central US CB adaptation	48.8	105.9	56.0
S9	Mid maturity SS (YD) type B37 derivative	62.4	112.4	44.3
S10	Mid maturity SS (YD) type B73 derivative	48.1	113.5	55.2
S11	Mid maturity SS (YD) type Western US CB adaptation	71.0	118.0	66.0
B73	Public US Iowa SS (100%), YD type	20.7	131.9	68.1
NS3	Mid maturity NS (YD) type Central US adaptation	3.1	136.7	53.6
NS4	Mid maturity NS (YD) type Central US adaptation	1.1	142.2	87.1
NS1	Mid maturity NS (YD) type Central US adaptation	0.8	142.6	87.4
NS5	Mid maturity NS (YD) type Central US adaptation	4.5	148.0	88.5
NS6	Mid maturity NS (YD) type Central US adaptation	1.5	156.0	90.8
NS7	Mid maturity NS (YD) type Central US adaptation	2.4	157.8	98.2
Female parent				
S1	Mid maturity Stiff Stalk (YD) type Central/eastern US CB adaptation			

S stiff stalk synthetic, *NS* non-stiff stalk, *YD* reid yellow dent, *CB* corn belt, *S1* the common female parent for the 16 hybrids. Pedigree was used to estimate percentage relationship to the common female parent, S1. Yield heterosis (F1 yield−mid-parent yield) is expressed as bushels/acre (bu/acr)

were conducted, tissue samples were collected for RNA analysis (see below).

For tissue sampling and RNA profiling in 1997, hybrids and inbreds were grown in the field at Johnston, IA and planted at three different times, in the months of April, May and June, approximately one month apart. Such an experimental design was intended to create growing environment variations; thus, we refer to these different plantings as three environmental replicates. Three primary, immature ears of pre-pollination (approximately 6–8 cm in length) at the stage of V19 were collected from three individual plants and pooled as one biological sample from each environmental replicate. One biological sample from each environmental replicate was used for RNA profiling in 1997 except for the common female S1 inbred, in which three biological replicates were submitted for GeneCalling. In 1999, five of the 16 hybrids and their respective inbred parents: S1/S3, S1/S4, S1/S11, S1/NS1, and S1/NS4 were grown in the field at Johnston, IA. Reciprocal hybrids (made by crossing the S1 inbred as a common male parent) of these five were also grown in the same location. Three primary immature ears of pre-pollination at the stage of V19 were collected from three individual plants and pooled as one biological replicate, and three biological replicates were sampled for GeneCalling RNA profiling.

In the density stress treatment experiment, we selected two commercial maize hybrids, S1/NS1 (Pioneer® hybrid 3394) and S2/NS2 (Pioneer® hybrid 3306) that were developed in the 1990s and 1960s, respectively. The parental inbreds S1 and NS1 are described in Table 1. S2 is in the parentage of S1, and both are derived from the same public line plus other Stiff Stalk public and proprietary lines. The NS2 inbred is a Non-Stiff Stalk line (Labate et al. 1997) and is derived from

a cross of two first-cycle lines out of Midwestern dent open-pollinated populations. The NS1 inbred is a line of complex parentage involving NS2. The inbred parents of both hybrids are adapted to the central US Corn Belt and have very similar maturity. The hybrid 3306 is one of the earliest single cross hybrids released commercially by Pioneer Hi-Bred International and is typical of the germplasm that farmers used in the mid-late 1960s, whereas hybrid 3394 was grown most widely in the early 1990s. In contrast to 3306, the improved yield characteristics of 3394 were selected using very different agronomic management practices, including higher plant density and increased levels of nitrogen fertilizer application. The two hybrids, 3394 and 3306 were grown at Johnston, IA in 1999 at three plant densities: 4,000, 18,000, and 35,000 plants per acre, respectively. The same tissue sampling protocol as above was used for RNA expression profiling. Three primary, immature ears of pre-pollination at the stage of V19 were collected from three individual plants and pooled as one biological sample and three biological replicates were used for GeneCalling analysis.

RNA isolation and transcript profiling

Protocols for RNA isolation and GeneCalling profiling have previously been described (Guo et al. 2003, 2004) The tissue was ground to a fine powder in liquid N_2. Total RNA was extracted using TriPure reagent (Roche Molecular Biochemicals, Indianapolis, IN, USA) according to the manufacturer's protocol. Poly-A^+ RNA was purified from total RNA using oligo (dT) magnetic beads (PerSeptive, Cambridge, MA, USA) and quantified by fluorometry. Poly-A^+ RNA was then subjected to GeneCalling analysis as described in Shimkets et al. (1999). Briefly, the following steps were involved in the GeneCalling process. Double-stranded cDNA was synthesized from the mRNA and digested with 48 different pairs of restriction enzymes (6-bp recognition sites). Adapters were ligated to the cDNA fragments, which was then PCR amplified for 20-cycles using adapter-specific primers. After size fractionation on an electrophoresis gel, the fluorescamine (FAM)-labeled PCR products were quantified by a laser scanner into digital tracefiles. The fluorescent intensity from FAM-labeled cDNA fragments is proportional to the abundance of the corresponding mRNA expressed in the given tissue.

The same 48 pairs of PCR primers were used for all the samples in this study, and these cover 80–90% of the expressed genes represented in the mRNA pool from the tissue analyzed (Shimkets et al. 1999). For each primer pair, three independent PCRs were made from an individual mRNA sample. A composite trace is calculated based on the average peak height and variance of the three PCR reactions from each sample (Shimkets et al. 1999). One mRNA sample from each genotype per environmental replicate was analyzed (three PCR reactions), except for one genotype (the S1 inbred) in which three experimental replicates consisting of nine PCR reactions were profiled. We analyzed the three biological replicates of the S1 genotype and did not find a significant difference between the replicates (see Fig. 3d for representative traces).

Data analysis

Using the original GeneCalling data, we selected inter-parental differentially expressed (IPDE) cDNA fragments based on the criteria that the average intensities of the cDNA fragments differed by at least two fold and a *P*-value of 0.01. The IPDE cDNA fragments could be due to expression level differences as well as allelic sequence polymorphism. We chose two fold as an empirical cut-off based on CuraGen's prior analysis on the reproducibility, sensitivity and false-positive rate of the GeneCalling technology (Shimkets et al. 1999). The two fold cut-off is somewhat arbitrary, but is used here, as it has also been used in other studies from maize (Bruce et al. 2000, 2001; Guo et al. 2003). Depending on the expression level of each gene, however, a lower fold change cut-off would favor false positives and a higher fold change cut-off would increase the chance of missing differentially expressed genes.

In order to obtain a quantitative measurement of the F_1 hybrid expression level relative to the mid-parental level for each IPDE cDNA fragment, we adapted the d/a ratio from quantitative genetics as a metric. In this measurement, d (dominant gene action) = F_1 (hybrid) − μ (average of the parents); a (additive gene action) = P_1 (parent 1) − μ. In the case of a complete dominant gene action of the P_1 allele, $F_1 = P_1$ then, $d/a = 1$; similarly, $d/a = -1$ if the other parental allele (P_2) is dominant. In the case of additive gene action, $F_1 = \mu$, then, $d/a = 0$. Using this concept, we considered the RNA expression level as a phenotype of each gene and measured the F1 hybrid expression level relative to the allelic additive expression (Guo et al. 2003). Transcript expression may be affected by multiple loci, the estimates for "a" and "d" for each transcript of interest would represent composite additive and dominance genetic effects, respectively. In the hybrid, the two parents contribute one dose each to the genetic constitution. Additive allelic expression in the hybrid would give a mid-parent level (μ) of $(P_{female} + P_{male})/2$. For each IPDE cDNA fragment, we

first calculated the deviation of the actual hybrid expression level from the average of the parents, as $d = F_1 - \mu$, and then calculated the deviation of the male parent from the average as $a = P_{male} - \mu$. The d/a ratio was then used to measure the hybrid expression level relative to the average of the parental level. If the hybrid expression level is equal to the average expression level, then $d = F_1 - \mu = 0$, which results in $d/a = 0$. Therefore, a zero value of the d/a ratio indicates that the level of expression in the hybrid is the same as the mid-parent value and fits the prediction of allelic additive expression. If the hybrid expression deviates from the average expression level and is biased towards the male parent's level, then the values $d = F_1 - \mu$ and $a = P_{male} - \mu$, would be both negative or both positive, resulting in $d/a > 0$. Likewise, $d/a < 0$ will be obtained if the hybrid expression is biased towards the female parent's level, where the values $d = F_1 - \mu$ and $a = P_{male} - \mu$, would be opposite in sign, one is negative and the other is positive. While the absolute value of the d/a ratio indicates the degree of the deviation from allelic additive expression, the sign of the d/a ratio indicates the direction of the deviation, maternal or paternal.

Results

Hybrid yields and heterosis

In order to examine gene expression patterns in hybrids in relation to heterosis, we crossed a Stiff Stalk (S1) line as a common female parent with 16 different inbred lines as males, that include Stiff Stalk and non-stiff stalk (NS) lines sharing different percentages of pedigree relationship with the female parent (Table 1). This experimental design created a series of hybrids that share a common parent with a wide range in yield and heterosis. Grain yield of hybrids and their inbred parents was tested in 2 years, 1997 and 1999 in four locations with two replicates per location in each year. The grain yield of all hybrids was lower in 1997 than in 1999. In general, 1997 was a low yielding season due to various environmental stresses, including a combination of low temperature and frost early in the season, prolonged local flooding, and water-deficient conditions later in the season. The growing season of 1999 in contrast, was a mild season and resulted in a higher grain yield (Fig. 1). The average yield and heterosis data from the 2 years are shown in Table 1. Yield heterosis was calculated by the yield advantage of the hybrid over the mid-parent values.

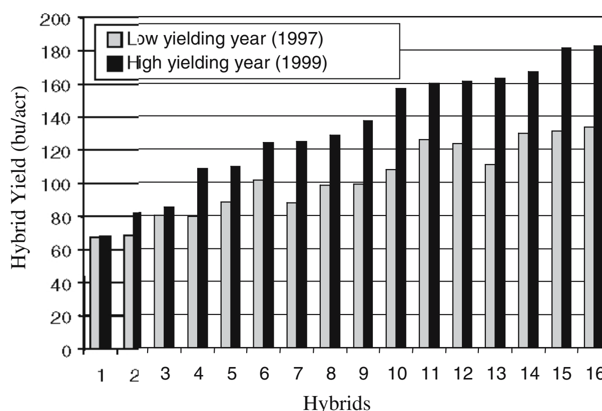

Fig. 1 Hybrid yields from a high yielding year (1997) and a low yielding year (1999). The hybrids are listed in Table 1. Yield was measured in bushels/acre (*bu/acr*) at four locations per year and in two replicates per location. Hybrids are ranked by yield data from 1999. All hybrids yielded higher in 1999 than in 1997

Gene expression differences between the inbred parents

Poly $(A)^+$ RNA from non-pollinated immature ear tissues (V19 stage) of 16 hybrids and 17 inbred parents harvested in 1997 was subjected to GeneCalling analysis as described in Shimkets et al. (1999, Materials and methods). In this study, we used the original GeneCalling data, the profile of each sample consisted of approximately 70,000 cDNA fragments. As described by Shimkets et al. (1999), each transcript can frequently be represented by more than one cDNA fragment using the GeneCalling approach since several restriction fragments from each cDNA can be amplified by this method. In comparing cDNA fragments from different genotypes using GeneCalling, cDNA sequence polymorphism can further complicate the expression analysis. Such polymorphisms can exist at the restriction site used prior to amplification or between sites to generate a restriction fragment length polymorphism. When comparing two inbreds of different genotypes with their F_1 hybrid, a cDNA sequence with no polymorphism will appear as one fragment size in all three genotypes. In contrast, a cDNA fragment that exhibits a sequence polymorphism detected by GeneCalling will yield two fragments of different lengths in the F1 hybrid, but only one or the other of these two fragments in each parental inbred. In either case, it is still valid to determine if the expression level of the hybrid bands deviates from the mid-parent. Indeed, sequence polymorphisms provide an opportunity to determine allele-specific levels of expression in the F_1 hybrid rather than the total expression level produced by the two combined alleles at a given locus.

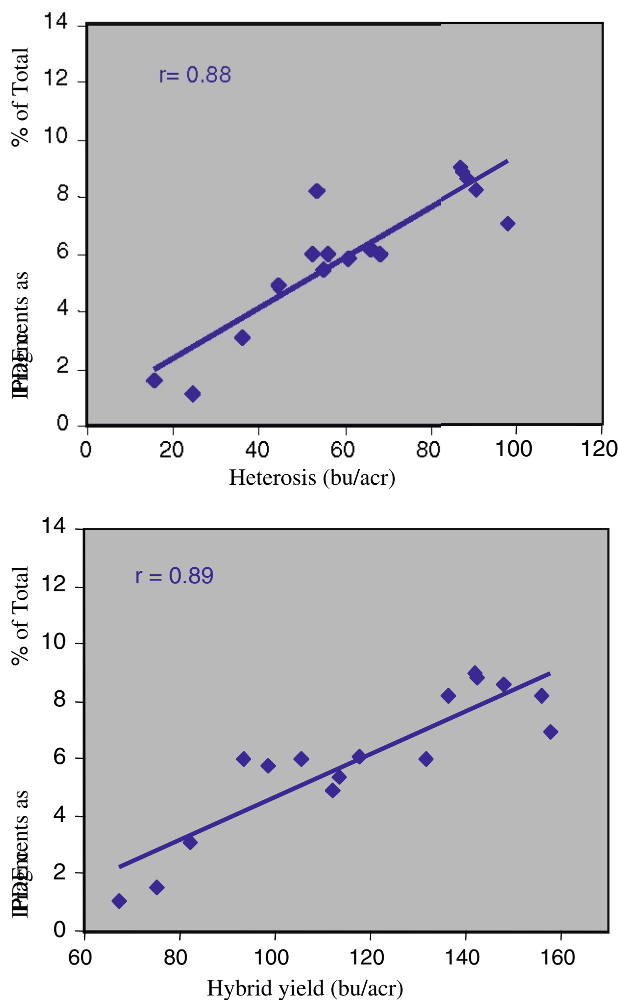

Fig. 2 Correlation between the proportion of Inter-Parental Differentially Expressed (IPDE) cDNA fragments with yield heterosis (*upper panel*) and hybrid yield (*lower panel*). Heterosis was calculated as yield of single cross minus mid-parent yield as shown in Table 1. Yield was measured in bushels/acre (*bu/acr*) in two years, 1997 and 1999, at four locations per year, and two replicates per location. Yield and heterosis data shown are the mean values of the two years

mRNA profiles of the hybrids and inbreds grown in 1997 were obtained from three environmental replicates, and each environmental replicate was sampled as three RT-PCR replicates in GeneCalling analysis. The cDNA fragments that are differentially expressed between the inbred parents by at least two fold were selected for expression analysis in the hybrids, and are hereafter referred to as IPDE (Inter-Parental Differentially Expressed) cDNA (Materials and methods). We found that the proportion of IPDE cDNA fragments varied between 1 and 10% among the different pairs of inbreds chosen as hybrid parents. The percentage of IPDE cDNA fragments correlated positively with hybrid yield and yield heterosis (Fig. 2).

Gene expression in F_1 hybrids relative to the mid-parent prediction

In order to obtain a quantitative measurement of the F_1 hybrid expression level relative to the average of the parental levels for each IPDE cDNA fragment, the *d*/*a* ratio, a metric often employed in quantitative genetics, was used, where "*d*" and "*a*" stand for a dominant and an additive gene action, respectively (Comstock and Robinson 1952; Gardner et al. 1953). This metric has been used for quantitative analysis of transcript level relative to allelic dosage regulation in the maize endosperm (Guo et al. 2003). The *d*/*a* ratio is a measure of the level of gene expression in the hybrid in relation to the allelic dosage, i.e. the mid-parent value. The farther the deviation of the F_1 expression from the mid-parent, the greater the absolute *d*/*a* value. For example, when the F_1 hybrid expresses at the same level as the maternal parent, the *d*/*a* ratio is -1. When the F_1 hybrid expression level is the same as the paternal parent, the *d*/*a* ratio is $+1$. If the hybrid expression level is outside the range of the parents, the *d*/*a* ratio is < -1, if the deviation is towards the maternal parent, and $> +1$ if towards the paternal parent. The *d*/*a* ratio is 0 when the F_1 expression level is the same as the mean of the expression level of the two parents.

If most of the IPDE cDNA fragments exhibit allelic additive expression in the F_1 hybrid, the *d*/*a* ratios should exhibit a normal distribution with the peak at zero and a narrow variance (mostly due to inherent variation of the GeneCalling method). We calculated the *d*/*a* ratio for every IPDE cDNA fragment in each individual hybrid and the distributions of the *d*/*a* ratios from all 16 hybrids (approximately 125,000 data points total) are shown in Fig. 3a. The majority ($\sim 80\%$) of the IPDE cDNA fragments have a *d*/*a* ratio between -1 and $+1$, indicating that, for most genes, the expression level in the hybrid was within the range of the parents. To estimate the proportion of genes that expressed at the approximate mid-parent level, we set an arbitrary cutoff of a *d*/*a* ratio between -0.5 and $+0.5$ and found that among the different hybrids 35–55% of IPDE cDNAs fell within this range of approximate mid-parent expression. The remaining IPDE cDNAs ($\sim 20\%$) were expressed at a level in the F_1 hybrid beyond the range of the two parents, either higher than the higher parent, or lower than the lower parent (*d*/*a* ratio $> +1$ or < -1), but none of these fragments was uniformly under- or over-expressed in all or even a majority of the hybrids analyzed.

We also looked for IPDE cDNA fragments that exhibited similar *d*/*a* ratios between -1 and $+1$ among all hybrids analyzed, but none were conclusively identi-

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/268077132005006121