

Transcriptomic analysis of growth heterosis in larval Pacific oysters (*Crassostrea gigas*)

Dennis Hedgecock^{*†}, Jing-Zhong Lin^{*‡}, Shannon DeCola[‡], Christian D. Haudenschild[‡], Eli Meyer^{*}, Donal T. Manahan^{*}, and Ben Bowen^{*†¶}

^{*}Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089-0371; and [‡]Solexa, Inc., 25861 Industrial Boulevard, Hayward, CA 94545

Communicated by George N. Somero, Stanford University, Pacific Grove, CA, December 8, 2006 (received for review May 6, 2006)

Compared with understanding of biological shape and form, knowledge is sparse regarding what regulates growth and body size of a species. For example, the genetic and physiological causes of heterosis (hybrid vigor) have remained elusive for nearly a century. Here, we investigate gene-expression patterns underlying growth heterosis in the Pacific oyster (*Crassostrea gigas*) in two partially inbred ($f = 0.375$) and two hybrid larval populations produced by a reciprocal cross between the two inbred families. We cloned cDNA and generated 4.5 M sequence tags with massively parallel signature sequencing. The sequences contain 23,274 distinct signatures that are expressed at statistically nonzero levels and show a highly positively skewed distribution with median and modal counts of 9.25 million and 3 transcripts per million, respectively. For nearly half of these signatures, expression level depends on genotype and is predominantly nonadditive (hybrids deviate from the inbred average). Statistical contrasts suggest ≈ 350 candidate genes for growth heterosis that exhibit concordant nonadditive expression in reciprocal hybrids; this represents only $\approx 1.5\%$ of the $>20,000$ transcripts. Patterns of gene expression, which include dominance for low expression and even underdominance of expression, are more complex than predicted from classical dominant or overdominant explanations of heterosis. Preliminary identification of ribosomal proteins among candidate genes supports the suggestion from previous studies that efficiency of protein metabolism plays a role in growth heterosis.

MPSS | nonadditive expression | factorial cross of partially inbred lines | potency | protein metabolism

Whole-genome sequencing has yielded rich insights into the number of genes required to make complex eukaryotic animals (1–4). Surprisingly few genes, however, can effect major changes in body design and shape. For example, elaboration and mutation of a fairly small number of developmentally important regulatory genes, such as the *Hox* gene cluster, appear to have driven the evolution of the major metazoan body plans (5). Profound changes in shape or morphology, underlying adaptive differences among closely related species, also can be caused by mutations in a very few genes (6–8).

Compared with our growing understanding of the evolution of biological shape and form, our understanding of what regulates body size and the rates and efficiencies of physiological functions is much less developed. For example, the genetic and physiological causes of heterosis (hybrid vigor) and its converse, inbreeding depression, have remained elusive for nearly a century, despite the economic and scientific significance of these phenomena (9–12). Major genetic explanations for heterosis are overdominance, the superiority of heterozygotes at genes affecting fitness or economic traits, dominance, the masking, in hybrids, of deleterious recessive mutations by dominant alleles inherited from one or the other inbred parent, and epistasis, the interaction of alleles at different loci. Under the first two hypotheses, inbreeding depression results from the increased frequency of less fit homozygous genotypes.

Plants, particularly crops and conifers, and bivalve molluscs exhibit high levels of inbreeding depression and heterosis. Initially, the evidence for heterosis in bivalves was indirect, a widely observed correlation of marker heterozygosity and fitness-related traits, such as growth (13, 14). Direct evidence for growth heterosis was demonstrated in the Pacific oyster by means of controlled crosses between partially inbred families (15). Further study of F_2 hybrid populations revealed a large load of highly deleterious recessive alleles, consistent with the dominance hypothesis (16). Physiological causes of heterosis have received less attention than genetic causes (10), but classical methods for quantifying energy and nutrient budgets are advancing understanding of the physiological underpinnings of growth heterosis in bivalves. Both larval and adult hybrid oysters have higher feeding rates and efficiencies than their inbred counterparts (17, 18). A large proportion of growth efficiency is achieved by lower rates of protein turnover in hybrids or more highly heterozygous individuals (19).

Genomic approaches to understanding the genetic basis of heterosis have included indirect mapping of quantitative trait loci (20–22) and statistical analysis of the gene action implied (23). More recently, attention has turned to gene expression in intraspecific hybrids (24, 25), although large-scale comparisons of patterns of gene expression in maize have just begun (26). Here, we use megacloning and massively parallel signature sequencing (MPSS; 27, 28), coupled technologies that generate short sequence tags from complex RNA samples, to analyze gene expression patterns in inbred and hybrid Pacific oysters produced by controlled reciprocal crosses between two inbred families. MPSS offers several advantages for organisms with poorly characterized genomes, such as the oyster: It requires no previous sequence information (28), it is comprehensive and quantitative (29–33), and it detects rare messages (down to ≈ 3 transcripts per million) and small differences in expression among genotypes (32).

Results

Two partially inbred families of Pacific oyster were crossed to produce inbred (named 33 and 55; inbreeding coefficient $f =$

Author contributions: D.H., D.T.M., and B.B. designed the project, C.D.H. and E.M. performed research; C.D.H. contributed new bioinformatic tools; D.H., J.-Z.L., S.D., and C.D.H. analyzed MPSS data; and D.H., S.D., C.D.H., E.M., and D.T.M. wrote the paper.

Conflict of interest statement: J.-Z.L., S.D., C.D.H., and B.B. were employees of Lynx Therapeutics, Inc., at the time this work was done. C.D.H. remains an employee of Solexa, Inc., which merged with Lynx Therapeutics in 2005.

Freely available online through the PNAS open access option.

Abbreviations: D, dominance; FS, four-stepper; MPSS, massively parallel signature sequencing; OD, overdominance; PD, partial dominance; SNP, single-nucleotide polymorphism; tpm, transcripts per million; TS, two-stepper; UD, underdominance.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (series GSE3596, platform GPL3062, and samples GSM83048–GSM83051). The sequences from the Megaclone libraries have been deposited in the GenBank database (accession nos. DV736295–DV736964).

[†]To whom correspondence should be addressed. E-mail: dhedge@usc.edu.

[‡]Present address: Forestgenix, LLC, 5566 San Juan Way, Pleasanton, CA 94566.

[¶]Present address: Gene-*fx*, Inc., Berkeley, CA 94705.

© 2007 by The National Academy of Sciences of the USA

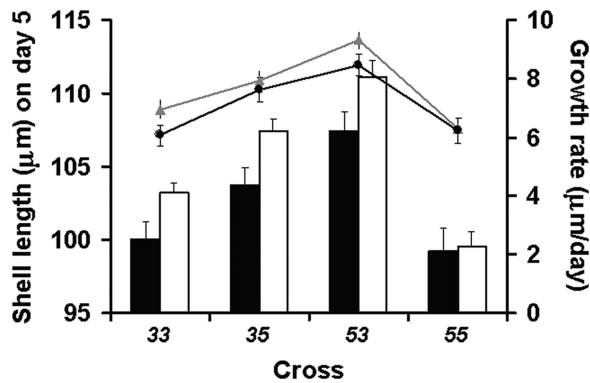


Fig. 1. Shell lengths on day 5 (in micrometers; bars) and growth rates (micrometers per day, between days 2 and 5; lines) for two inbred (33, 55) and two hybrid (35, 53) families produced by a factorial cross between partially inbred lines 35 and 51 of Pacific oysters ($n = 50$ for each family and time point). Black line and filled bars are for replicate A; gray line and open bars are for replicate B; errors are 1 SE. Significant growth heterosis is evident by day 5.

0.375) and reciprocal hybrid (35 and 53, male parent listed first) larvae. Heterosis for growth is evident from differences between shell lengths of 5-d-old inbred and hybrid veliger larvae, as well as from the greater growth rate, from day 2 to day 5, of hybrids compared with inbreds (Fig. 1). Heterosis for larval shell length at day 5 is defined by the potence, $h_p = Q/L > 1.0$, where Q is twice the deviation of a hybrid from the mean of the inbred shell lengths (mid-parent) and L is the absolute difference between the mean trait-values of the two parental inbred lines, i.e., Q is the quadratic and L , the linear contrast estimated from analysis of variance (ANOVA; 10, 15). Potence values for the two reciprocal hybrids are much greater than 1.0 ($h_{35} = 4.56$, $P < 0.0001$; $h_{53} = 7.93$, $P < 0.0001$); hybrid 53 is significantly larger than hybrid 35 ($P = 0.0012$) and inbred family 33 is larger, although not significantly so, than inbred family 55 ($P = 0.054$). Clearly, our experimental material exhibits classical hybrid vigor or growth heterosis, as observed in previous crosses among inbred lines (15, 17, 18).

RNA was extracted from 6-d-old larvae of each larval population and cloned via Megaclone methods (27). Sequences of 666 inserts from these libraries (average length 238 bp) show that 89% begin with the expected DpnII restriction enzyme site, GATC. Of these, 25% match GenBank sequences for the

Crassostrea gigas mitochondrial genome (AF177226), as expected in rapidly growing larvae. Of the 446 nonmitochondrial DNA sequences, 84% are unique but only 40% of these ($n = 151$) have matches in GenBank.

From the Megaclone libraries, we obtain 17-bp sequences by MPSS for each of the four genotypes in replicates of two different sequencing frames [two-stepper (TS) and four-stepper (FS)]. Sequences deriving from the cloning vector, having poor signal-to-noise characteristics, or present in only one of the 15 MPSS runs (a TS replicate for 33 is missing) are removed from the data. The weighted mean relative frequency of a signature within a family is calculated as the sum, over replicate runs, of beads with that signature (the signature count) divided by the total number of sequences obtained in that reading frame for that family. Relative frequency is multiplied by 10^6 to yield transcripts per million (tpm). The reading frame with the highest normalized count across families is selected to represent the relative transcript frequency or expression of each signature. These filtering and selection steps leave 4.5 million sequences, comprising 53,771 distinct 17-bp signature sequences (Table 1). For 283 signatures, we observe matches of 14 or more base pairs to the Pacific oyster mitochondrial genome. One signature, GATCATAGGAGAAGTTA, which corresponds to the large mitochondrial ribosomal RNA subunit, has an observed count of 552,607 or 22.8% of the 2.4 million clones sequenced in the FS reading frame, similar to the 25% of 666 Megaclone library sequences checked. Signatures with matches to oyster mitochondrial DNA are left in the analysis of gene-expression patterns to follow and will be treated in more detail (J. Curole, E.M., D.T.M., and D.H., unpublished data).

Sequence pairs read in the same frame, differing from each other by a single base pair, and absent in one or the other inbred parent line ($n = 943$) are interpreted as expressed single-nucleotide polymorphisms (SNPs). For subsequent analyses, counts for these pairs are pooled arbitrarily under the signature for which expression in inbred family 55 is zero. We detected a more complex SNP family, composed of two SNP signature pairs, which were sequenced in different frames, together with two additional singleton signatures, all differing from each other by a single base pair. The SNP signature pair with the highest expression level is retained and the other signatures are omitted, leaving 52,825 signatures (i.e., $53,771 - 943 - 3$).

Given these expression data, the problem is to identify those signatures that show significant nonadditive expression on the assumption that nonadditive rather than additive gene expres-

Table 1. Clones sequenced and unique signatures detected by MPSS in two inbred (33 and 55) and two hybrid (35 and 53) genotypes

Reading frame	Filtering status	Run	Clones/signatures*	Genotype				Totals
				33	35	53	55	
TS	Before	1	Clones	577,307	531,946	544,843	555,701	2,209,797
			Signatures	22,128	18,009	23,102	17,618	49,889
TS	Before	2	Clones	–	491,956	462,881	499,852	1,454,689
			Signatures	–	15,885	21,213	19,303	37,695
TS	After	Pooled	Clones	563,525	499,169	492,596	510,863	2,066,153
			Signatures	17,708	14,068	17,588	15,297	33,266
FS	Before	1	Clones	598,046	651,271	556,938	681,356	2,487,611
			Signatures	19,292	17,662	21,709	21,191	49,990
FS	Before	2	Clones	559,781	593,744	586,917	705,467	2,445,909
			Signatures	19,636	16,306	24,163	19,897	50,435
FS	After	Pooled	Clones	568,515	612,098	560,488	681,568	2,422,669
			Signatures	9,738	9,122	11,438	10,256	20,505
Chosen	After	Pooled	Clones	1,132,040	1,111,267	1,053,084	1,192,431	4,488,822
			Signatures	27,446	23,190	29,026	25,553	53,771

*Numbers of clones sum to totals, but numbers of signatures do not because a given signature may be present in more than one genotype.

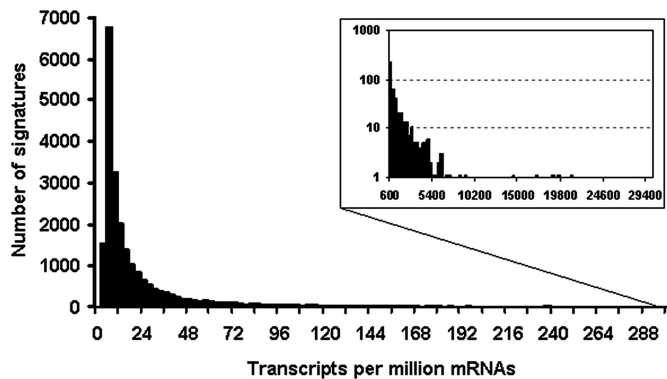


Fig. 2. Transcript abundance (in tpm), averaged over the four genotypes, for 23,273 MPSS signatures with $Np \geq 7$, where N is the number of clones sequenced and p is the mean proportion of clones having a specific signature. (Inset) Shown are 471 signatures with expression levels >600 tpm; note changed scales of axes.

sion must underlie the observed heterosis for growth (Fig. 1). This evokes a statistical comparison of expression levels among families that is identical to the analysis of larval growth. The MPSS expression data for each of the 52,825 signatures are fit to a general linear model, $p = g$, where p is the proportion of clones having a particular signature sequence and g is the genotype (family). We find that expression levels of 10,623 signatures depend on genotype at a nominal significance threshold of $P \leq 0.0001$; this number is lowered to 6,297 if the significance threshold is corrected to $P \leq 0.05$ for multiple tests. Mean signature counts for significant linear models (i.e., $N\bar{p}$, where N is the total number of sequences, 2,442,669 for the FS data and 2,066,153 for the TS data, and \bar{p} is the mean relative frequency over the four genotypes) are all >7 , which is equivalent to ≈ 3 tpm. Of the 42,202 signatures, for which the linear model is not significant, 12,651 signatures have mean counts more than ≈ 3 tpm but no significant variation among genotypes. The remaining 29,551 signatures are very rare, accounting for only 1.8% of the 4.5 million sequences and likely represent artifacts of the MPSS process. Averaged over the four families, the median and modal counts of 23,274 signatures with statistically nonzero expression are 9.25 and ≈ 3.5 tpm, respectively, which suggests that, even with low-frequency artifacts removed, most genes are expressed at low levels (Fig. 2).

Mode of gene expression may be categorized as additive or nonadditive, using linear and quadratic contrasts among genotypic values, just as in the analysis of larval shell length (i.e., L is the difference between the mean expression levels of the two parental inbred lines and Q is twice the deviation of a hybrid from mid-parent expression levels). Expression is additive if $L > 0$ and Q is zero, i.e., hybrid expression is not different from the mid-parent level. Signatures with nonadditive expression are categorized into six classical patterns of gene action: (i) overdominance (OD), i.e., overexpressed in hybrids relative to the higher expressing parent, $h_p > 1.0$; (ii) dominance (D) “+”, i.e., hybrids expressing like the higher expressing parent, $h_p = 1.0$; (iii) partial dominance (PD) +, i.e., hybrids expressing above mid-parent but below the higher expressing parent, $0 < h_p < 1.0$; (iv) D “-”, i.e., hybrids like the lower expressing parent, $h_p = -1$; (v) PD-, i.e., hybrids below mid-parent but above the lower expressing parent, $0 > h_p > -1$; and (vi) underdominance (UD), i.e., underexpressed in hybrids relative to the lower expressing parent, $h_p < -1.0$. The criteria for these patterns of gene expression are mutually exclusive and yield nonoverlapping sets of signatures for each of the hybrids.

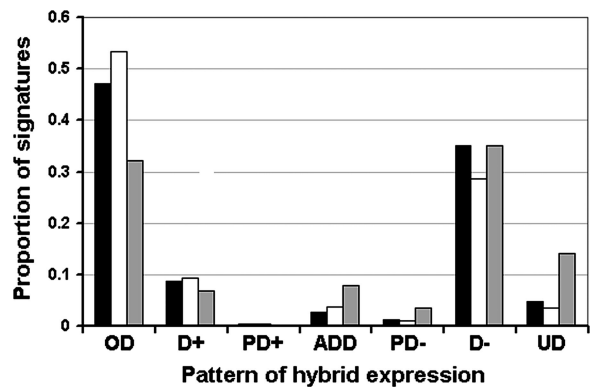


Fig. 3. Proportions of MPSS signatures with modes of expression categorized as overdominant (OD), dominant for high expression (D+), partially dominant for high expression (PD+), additive (ADD), partially dominant for low expression (PD-), dominant for low expression (D-), or underdominant (UD). Numbers of signatures in these seven categories equal 3,762 and 3,705 for the 35 (black bars) and 53 hybrids (open bars), respectively. A total of 394 signatures (gray bars) had common expression patterns in both hybrids; of these, 348 are candidates in the nonadditive categories of OD, D+, D-, and UD.

Of the 3,762 signatures that have significant linear models for hybrid 35, 97.4% have nonadditive expression and only 2.6% have additive expression. Of 3,705 significant genotype-dependent signatures for hybrid 53, 96.2% are nonadditive and only 3.7% are additive. Equally noteworthy is the uneven distribution of signatures across the categories of nonadditive expression (Fig. 3). Approximately half of the signatures show overexpression in hybrids (OD), with the next highest category being dominance for low expression (D-), followed by dominance for high expression (D+), with underexpression (UD) being slightly more common than additive expression. Less than 1% is in the partially dominant categories.

We next consider signatures, for which expression levels in hybrids resemble those in the paternal or maternal lines. To infer parental effects, we require that the linear model be significant (at $P \leq 0.0001$), that the parental inbred lines differ in expression (L significant at $P \leq 0.001$), and that hybrids resemble paternal or maternal parents within a $P > 0.01$ threshold of nonsignificance. The criteria for parental patterns of expression do not yield sets of signatures that are mutually exclusive to those in Fig. 3. For example, we find 531 signatures with paternal patterns of expression; of these, 169 are already counted in the four categories of dominance (D-35, D+53, D+35, and D-53) and 362 are newly categorized. Likewise, of the 552 signatures showing maternal effects on expression at the $P > 0.01$ level of similarity, 170 overlap with the dominance categories and 382 are newly categorized. Removing overlap between the parental-effect and dominance categories, we find that $\approx 70\%$ of the 10,623 signatures with expression levels that depend significantly on genotype can be assigned to additive, nonadditive, or parental-effect categories of expression in one or both hybrids. Overdominance or overexpression in one or more hybrids is by far the most common pattern of expression.

In the classical dominance and overdominance models of heterosis, there should be no genotypic difference for nuclear genes between reciprocal hybrids. This is strictly true only if parent lines are completely inbred, i.e., AA and BB, such that $AB = BA$. Our lines are only three-eighths inbred, so residual heterozygosity within parent lines will make reciprocal hybrids nonequivalent at many loci. Nevertheless, growth heterosis is observed in both hybrids (Fig. 1). To produce a tractable number of candidate heterosis genes for functional analysis, we focus on those signatures showing the same mode of expression in both

reciprocal hybrids. The intersections of the sets of signatures categorized into modes of expression also show predominance of nonadditive expression, 875 of 906 signatures (96.6%), but the distribution of signatures across categories is now dominated by signatures in the D⁻ category (71.7%), followed by the OD (14.0%), UD (6.2%), D⁺ (3.0%), additive (ADD) (3.4%), PD⁺ (0.1%), and PD⁻ (1.5%) categories. Nearly half of the 650 signatures with D⁻ expression patterns have zero counts for both hybrids and one inbred family. Limiting D⁻ patterns to those signatures with nonzero expression in all four larval populations reduces the number of D⁻ candidates to 138 and the total number of candidates in nonadditive categories (OD, 127; D⁺, 27; D⁻, 138; and UD, 56) to 348 (Fig. 3).

Discussion

Despite nearly a century of work on heterosis, little consensus has emerged on the genetic or physiological causes of this important phenomenon. Functional genomics offers a novel approach to unraveling the molecular and physiological causes of heterosis. In this study, we categorize 70% of genotype-dependent patterns of gene expression into classical modes of gene action. We find few additive patterns of gene expression (2%), in which hybrid expression is the average of inbred expression levels, a result differing markedly from that of Swanson-Wagner *et al.* (26), who report 78% additive patterns in maize, but consistent with levels of whole genome nonadditive patterns of expression observed in crosses of isogenic lines of *Drosophila melanogaster* (34). The pervasive nonadditivity of whole-genome expression patterns in *Drosophila* and now in our study is surprising in light of the classical view that additive gene action underlies variation in most quantitative traits. For growth heterosis in oyster larvae, a surprisingly small number of genes, ≈ 350 , only $\approx 1.5\%$ of the $>20,000$ transcripts, appears to control the striking variation in body size and physiological function. Although larger than the number of genes that can effect major changes in morphology, the number of genes controlling growth heterosis is on the same scale as the number of genes known to regulate the physiology of life span and lipid accumulation in *Caenorhabditis elegans* (35).

MPSS has been successfully applied and tested in model species with sequenced genomes (28–33). These studies show that MPSS provides excellent coverage of the transcriptome, to the point of revealing previously uncharacterized transcripts (30, 32), and that the expression of most transcripts can be accurately and precisely quantified (33). Oyster MPSS signatures matching the mitochondrial genome sequence (GenBank accession NC_001276) correspond to 26 of 27 DpnII restriction sites in the genome, the majority being on the coding strand with a substantial fraction of unknown function on the noncoding strand (J. Curole, E.M., D.T.M., and D.H., unpublished data). In model species, the major sources of error and bias with MPSS have been identified and quantified (30, 33). Imperfectly matched signatures from the oyster mitochondrial genome yield a sequencing error estimate of 2.4%, compared with the 3.2% estimated for *Arabidopsis* (30). Furthermore, we observe a significant negative correlation between expression and Megaclone insert length, previously identified as a bias with “classical” Megaclone libraries (32). These subsequent observations strongly imply that MPSS data for the oyster has the same coverage, accuracy, precision, sequencing error, and bias as those characterized in model systems.

With a genome sequence, one can immediately identify $>90\%$ of the 17-bp signature sequences from MPSS (27, 29, 31). Our study, on the other hand, illustrates how MPSS can be usefully applied to an organism without a sequenced genome, to address a significant biological problem that has been under study for nearly a century. Key to the success of MPSS for a nonmodel species, however, is the availability of clear phenotypic contrasts,

which are afforded by the classical expression of growth heterosis in hybrids produced by crossing of inbred lines of oysters. In this experimental context, phenotypic contrasts among genotypes reveal a set of candidate genes for further functional and genetic analyses of growth heterosis.

To obtain candidate heterosis genes, we fit the expression profile for each of nearly 53,000 signature sequences for inbred and hybrid oysters to a general linear model. Just over half (57%) of the signatures are expressed at insignificant levels and are not associated with genotype. Twenty percent (10,623) of all signatures are expressed at levels that depend on genotype ($P < 0.0001$), and another 24% are expressed at adequate levels (mean expression in excess of ≈ 3 tpm) but independently of genotype. At the 5% level of significance adjusted for multiple testing, we expect 2,650 of the nearly 53,000 linear models to be significant by chance, so the 6,297 signatures with significant genotype dependent expression is 2.4 times greater than the number expected by chance.

Among the signatures with identifiable patterns of genotype-dependent expression, we, like Swanson-Wagner *et al.* (26), observe patterns expected under the classical hypotheses of heterosis (i.e., D⁺, dominance for high expression and OD, overexpression of genes in hybrids), as well as patterns of expression not anticipated by the classical hypotheses (D⁻, dominance for low expression, and UD, underexpression of genes in hybrids). Considering each of the hybrids separately, overexpression (OD) of genes is the most prevalent pattern (47% and 53% of categorized signatures in hybrids 35 and 53, respectively), followed by the D⁻, D⁺, and UD patterns. We note that the hybrid family with the greater degree of growth heterosis, 53, shows a greater proportion of OD patterns. The prevalence of over- and underdominant expression for oyster larvae differs from what has been reported for either maize (26) or *Drosophila* (34) hybrid-inbred comparisons, in which only 3.2% and 5% of expression patterns, respectively, show over- or underexpression in hybrids.

The genetic basis of OD or UD phenotypes may be interaction between cis-acting regulatory elements or differences in levels of trans-acting factors. It should be possible to distinguish between cis and trans regulation of expression levels by contingency tests of the association between expression and genotype in the next generation. With cis regulation, the association between heterozygosity at the candidate locus and expression level should remain unchanged; with trans regulation, recombination between the candidate locus and trans-acting elements should eliminate or reduce over- or underexpression in heterozygotes. The D⁺ and D⁻ patterns of expression, on the other hand, may result from allelic dosage effects (12, 25), monoallelic expression, or possibly trans-acting factors. The likely importance of trans-acting elements in regulating many nonadditive patterns of gene expression suggests that growth heterosis, in general, may result largely from epistatic gene interaction. Untangling the networks of gene regulation and interaction that underlie growth heterosis will require simultaneous application of quantitative genetic and physiological genomic approaches (34, 36, 37).

The functional genomic approach to understanding growth heterosis afforded by MPSS reveals patterns of gene expression, UD and D⁻, which are not anticipated by classical quantitative genetic theory. These patterns of expression may be related to physiological differences that have been observed between fast- and slow-growing oyster larvae with different genotypes. Pace *et al.* (18) showed that the physiological bases of growth heterosis in larvae were enhanced feeding rate and metabolic efficiency likely realized through mechanisms of protein synthesis and degradation (turnover). For adult bivalves, Hawkins *et al.* (19) showed that much of the difference in resting metabolic rates between fast- and slow-growing mussels could be ascribed to differences in whole-body protein turnover. Thus, low expression

because various artifacts, such as palindromic sequences, are likely to bias counts downward.

The algorithm used to find expressed SNPs starts with signatures, for which expression in inbred family 55 is zero, listed in descending order of abundance in inbred family 33. The algorithm then searches for a SNP partner from a list of signatures sequenced in the same reading frame, for which expression in inbred family 33 is zero, listed in descending order of abundance in inbred family 55. Counts for both signatures were combined under the signature for which inbred family 55 had zero expression, and the pooled data were subjected to statistical analysis.

Statistical Tests. MPSS data are categorical and binomial. However, the normal distribution provides a good approximation of the binomial distribution, if the interval $[\mu \pm 3\sigma]$, where $\mu = np$ and $\sigma = np(1-p)$, lies between 0 and n . This interval will not include zero, if $n > \approx 9/p$, which means that, for sample sizes in this study (2,066,153 for TS, 2,422,669 for FS), p across genotypes has to be greater than ≈ 4 tpm.

We compare simultaneously the expression profiles of two inbred and two hybrid genotypes, signature by signature, under the simple model, $p = g$, by using the general linear model procedure (Proc GLM) of SAS (SAS Institute, Cary, NC). The raw data for each signature are the numbers of ones (presence of the signature) and zeros (absence of the signature), which sum to the grand total of sequences in that reading frame. Type III sums of squares are used to determine model significance,

because sample size varies by genotype. We estimate the linear and quadratic contrasts for genotypes 33, 35, 53 and 55 with design vectors $L = [-1\ 0\ 0\ 1]$, $Q_{35} = [-1\ 2\ 0\ -1]$, and $Q_{53} = [-1\ 0\ 2\ -1]$. We also test the significance of the following post hoc contrasts: $35 = 33 [-1\ 1\ 0\ 0]$, $53 = 33 [-1\ 0\ 1\ 0]$, $35 = 55 [0\ 1\ 0\ -1]$, $53 = 55 [0\ 0\ 1\ -1]$, $35 = MP [-0.5\ 1\ 0\ -0.5]$, $53 = MP [-0.5\ 0\ 1\ -0.5]$, and $35 = 53 [0\ 1\ -1\ 0]$, where MP is the average of expression levels in the two inbred parental genotypes. Calculated potency ($h_p = Q/|L|$), the linear contrast, and the significance of pairwise contrasts are used to categorize expression patterns, when the linear model is significant at the $\alpha = 0.0001$ level. This level of significance is higher than the $\alpha = 0.05$ level adjusted for simultaneous testing of 52,825 signatures, which requires a threshold probability of $1 - (1 - 0.05)^{1/52825} = 9.71 \times 10^{-7}$, but we focus on patterns of expression rather than significance of any particular signature. Classifying a pattern as OD, for example, requires $h_p > 1.0$ and the contrast between hybrid and highest expressing parent to be significant at $\alpha = 0.001$. Categories, for which parent-hybrid similarity is required, such as D+ or D-, are assigned at a significance level of $\alpha > 0.01$ for the appropriate contrast.

This research was made possible by funding from the U.S. Department of Agriculture National Research Initiative Competitive Grants Program (to D.H.), the USDA Western Regional Aquaculture Center (to D.H. and D.T.M.), the W. M. Keck Foundation (to D.T.M.), and the National Science Foundation (Gen-En; to D.H. and D.T.M.).

- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, et al. (1996) *Science* 274:546–567.
- The *C. elegans* Sequencing Consortium (1998) *Science* 282:2012–2018.
- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. (2000) *Science* 287:2185–2195.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. (2001) *Science* 291:1304–1351.
- Lewis EB (1978) *Nature* 276:565–570.
- Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D & Kingsley DM (2004) *Nature* 428:717–723.
- Wu P, Jiang T-X, Suksawang S, Widelitz RB, Chuong C-M (2004) *Science* 305:1465–1466.
- Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ (2004) *Science* 305:1462–1465.
- Gowen JW, ed (1952) *Heterosis* (Iowa State College Press, Ames, IA).
- Griffing B (1990) *Genetics* 126:753–767.
- Crow JF (1998) *Genetics* 148:923–928.
- Birchler JA, Auger DL, Riddle NC (2003) *Plant Cell* 15:2236–2239.
- Zouros E, Romerodorey M, Mallet AL (1988) *Evolution (Lawrence, Kans)* 42:1332–1341.
- Britten HB (1996) *Evolution (Lawrence, Kans)* 50:2158–2164.
- Hedgecock D, McGoldrick DJ, Bayne BL (1995) *Aquaculture* 137:285–298.
- Launey S, Hedgecock D (2001) *Genetics* 159:255–265.
- Bayne BL, Hedgecock D, McGoldrick D, Rees R (1999) *J Exp Mar Biol Ecol* 233:115–130.
- Pace DA, Marsh AG, Leong P, Green A, Hedgecock D, Manahan DT (2006) *J Exp Mar Biol Ecol* 353:188–209.
- Hawkins AJS, Bayne BL, Day AJ (1986) *Proc R Soc London B* 229:161–176.
- Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) *Genetics* 132:823–839.
- Xiao JH, Li JM, Yuan LP, Tanksley SD (1995) *Genetics* 140:745–754.
- Li ZK, Luo LJ, Mei HW, Wang DL, Shu QY, Tabien R, Zhong DB, Ying CS, Stansel JW, Khush GS, et al. (2001) *Genetics* 158:1737–1753.
- Cockerham CC, Zeng ZB (1996) *Genetics* 143:1437–1456.
- Song R, Messing J (2003) *Proc Natl Acad Sci USA* 100:9055–9060.
- Auger DL, Gray AD, Ream TS, Kato A, Coe EH, Jr, Birchler JA (2005) *Genetics* 169:389–397.
- Swanson-Wagner RA, Jia Y, DeCook R, Borsuk LA, Nettleton D, Schnable PS (2006) *Proc Natl Acad Sci USA* 103:6805–6810.
- Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C, Mao JI, Luo S, Kirchner JJ, Eletr S, et al. (2000) *Proc Natl Acad Sci USA* 97:1665–1670.
- Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, et al. (2000) *Nat Biotech* 18:630–634.
- Meyers BC, Vu TH, Tej SS, Ghazal H, Matvienko M, Agrawal V, Ning J, Haudenschild CD (2004) *Nat Biotech* 22:1006–1011.
- Meyers BC, Tej SS, Vu TH, Haudenschild CD, Agrawal V, Edberg SB, Ghazal H, Decola S (2004) *Genome Res* 14:1641–1653.
- Jongeneel CV, Iseli C, Stevenson BJ, Riggins GJ, Lal A, Mackay A, Harris RA, O'Hare MJ, Neville AM, Simpson AJG, et al. (2003) *Proc Natl Acad Sci USA* 100:4702–4705.
- Jongeneel CV, Delorenzi M, Iseli C, Zhou DX, Haudenschild CD, Khrebtsukova I, Kuznetsov D, Stevenson BJ, Strausberg RL, Simpson AJG, et al. (2005) *Genome Res* 15:1007–1014.
- Stolovitzky GA, Kundaje A, Held GA, Duggar KH, Haudenschild CD, Zhou D, Vasicek TJ, Smith KD, Aderem A, Roach JC (2005) *Proc Natl Acad Sci USA* 102:1402–1407.
- Gibson G, Riley-Berger R, Harshman L, Kopp A, Vacha S, Nuzhdin S, Wayne M (2004) *Genetics* 167:1791–1799.
- Ashrafi K, Chang FY, Watts JL, Fraser AG, Kamath RS, Ahringer J, Ruvkun G (2003) *Nature* 421:268–272.
- Wayne ML, McIntyre LM (2002) *Proc Natl Acad Sci USA* 99:14903–14906.
- Anholt RRH, Dilda CL, Chang S, Fanara JJ, Kulkarni NH, Ganguly I, Rollmann SM, Kamdar KP, Mackay TFC (2003) *Nat Genet* 35:180–184.
- Meyer E (2006) PhD dissertation (University of Southern California, Los Angeles).
- Langdon C, Evans F, Jacobson D, Blouin M (2003) *Aquaculture* 220:227–244.