Current Topics in Plant Biology
Vol. 12, 2011

Original Communication

Genome-wide analysis of the intergenic regions in *Arabidopsis thaliana* suggests the existence of bidirectional promoters and genetic insulators

Xiaohan Yang, Cara M. Winter, Xiuying Xia, and Susheng Gan*

Department of Horticulture, Cornell University, 134A Plant Science, Ithaca, New York 14853-5904, USA

ABSTRACT

The short regions flanked by divergent genes provide a good opportunity for exploring mechanisms of transcriptional regulation because of the confined nature of the upstream regulatory elements of both genes. We performed a genome-wide analysis of coexpression levels of divergent gene pairs in *Arabidopsis thaliana*, along with convergent and parallel gene pairs as controls for comparison, and found that for adjacent genes ≤0.4 kb apart, there was a significantly higher portion of gene pairs showing coexpression in divergent configuration than in parallel or convergent configuration. We divided the different expression patterns in adjacent divergent *Arabidopsis* gene pairs into three categories: coexpression, independent expression, and antiexpression. Our studies on the relationship between coexpression and gene function indicate that similar functionality is not the main cause for the existence of coexpressed divergent gene pairs in *Arabidopsis*. Comparative analysis revealed some conserved intergenic regions flanked by divergent genes ≤0.4 kb apart between *Arabidopsis* and rice. Additionally, we identified overrepresented motifs in the intergenic regions flanked by independently-expressed divergent genes, as compared with the intergenic sequences flanked by coexpressed divergent genes. Our analysis suggests that specific intergenic regions contain potential bidirectional promoters or genetic insulators, offering guidance for future experimental efforts to isolate those regulatory elements.

KEYWORDS: *Arabidopsis*, bidirectional promoter, gene regulation, genetic insulator, intergenic region, rice

INTRODUCTION

The availability of complete genome sequence and gene annotation information and increasingly robust expression data sets for some model higher eukaryotes has made it possible to perform large-scale analyses of genome structure, gene function and gene regulation. Two neighboring genes can be in divergent (← →), convergent (→ ←), or parallel (→ → or ← ←) configuration, and the configuration appears to correlate with the expression patterns of the paired genes, for example, the level of co-expression of adjacent genes is higher for gene pairs in the divergent, or parallel configuration than in the convergent configuration in *Arabidopsis thaliana* [1]. The higher level of co-expression of divergent gene pairs may be attributed to shared regulatory elements (enhancers or silencers) in the intergenic regions that separate the genes, i.e., the intergenic region of a pair of co-expressed divergent genes may serve as a bidirectional promoter. It has been shown that engineered bidirectional promoters can direct two genes in divergent configuration and

*Corresponding author
sg288@cornell.edu

that the two genes are co-expressed [2, 3]. In the process of studying a single gene, more than 10 bidirectional promoters have been identified between divergent genes in animals including mouse, rat, human, and chicken [4, 5, 6, 7, 8, 9]. A genome-wide analysis of divergent gene pairs in the human genome has revealed many more bidirectional promoters between divergent genes that initiate transcription in both directions [10].

An independent expression pattern has also been found in animal divergent gene pairs less than 400 bp apart [11, 12]. Divergent genes have also been shown to be antiregulated in microorganisms and animal species [10, 13, 14, 15, 16]. It is well known that enhancers of one gene promoter can influence the expression of a neighboring gene [3]. What is the underlying mechanism(s) that ensures independent or antiregulated expression of closely linked divergent genes? One possible mechanism is genetic insulator. Genetic insulators have been found between divergent genes in animals to define independent domains of transcriptional activity by blocking the activity of enhancers and silencers when inserted between these regulatory elements and a promoter [17, 18, 19]. But, to our knowledge, no genetic insulators have been documented in plants.

The short regions flanked by divergent genes provide a very good opportunity for exploring transcriptional regulatory mechanisms due to the confined nature of the upstream regulatory elements of both genes, and the available *Arabidopsis* genome sequence [20] can facilitate genome-wide analyses of these intergenic regions. Recently, microarray analyses were carried out to profile expression of most of the predicted genes in the *Arabidopsis thaliana* genome (http://www.arabidopsis.org/). Integration of the genome sequence data with microarray expression data can be very useful for identifying intergenic regions flanked by coexpressed, independently-expressed, and antiexpressed divergent genes, and thus provide guidance for experimentally identifying bidirectional promoters and genetic insulators in *Arabidopsis*.

Cross-species DNA sequence comparison is a fundamental method for identifying biologically important elements, because functional sequences are evolutionarily conserved, whereas nonfunctional sequences drift Conserved noncoding sequences among cultivated cereal genomes have been surveyed to identify candidate regulatory sequence elements [21]. Rice and *Arabidopsis*, which are models for monocotyledonous and eudicotyledonous plants, respectively, diverged from a common ancestor about 200 million years ago [22, 23]. It is highly possible that conserved functional elements can be distinguished from nonconserved sequences in the comparison of intergenic DNA sequences between the two diverged species.

In this study, we performed a genome-wide analysis of the coexpression levels of divergent gene pairs in *A. thaliana*, along with convergent and parallel gene pairs as controls for comparison. Our results show that for the adjacent genes ≤0.4 kb apart, there is a significantly higher portion of gene pairs showing coexpression in divergent configuration than in parallel or convergent configuration. We categorized the short intergenic regions (≤0.4 kb) flanked by divergent genes into three groups: those exhibiting coexpression, independent expression, and antiexpression. By comparing DNA sequences flanked by divergent genes ≤0.4 kb apart between *Arabidopsis* and rice, we identified conserved regions between the two diverged species. We explored the relationship between coexpression and gene function, and our results suggest that common functionality is not the main cause for the existence of bidirectional promoters in *Arabidopsis*. It is likely that most of the bidirectional promoters are ancestral sequences that have persisted through evolution. To identify potential genetic insulators, we analyzed overrepresented motifs in the intergenic regions flanked by independently-expressed divergent genes, using the intergenic sequences between coexpressed divergent genes as a background.

METHODS

Expression and genome data sources

The *Arabidopsis* microarray data files (47 tissue samples with three biological replicates, Supplemental Table S1) for the Developmental Affymetrix Gene Expression Atlas as part of the AtGenExpress Project, supplied by the MPI Tübingen (Schmid, Lohmann and Weigel labs),

were obtained from the TAIR Web site (ftp://ftp.arabidopsis.org/home/tair/Microarrays/). Expression data were present for 22,080 genes across all tissue samples. *Arabidopsis* gene orientation and intergenic sequence data were also obtained from the TAIR Web site (ftp://ftp.arabidopsis.org/home/tair/Maps/seqviewer_data/ and ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/). *Arabidopsis* parallel gene duplication data were obtained from the TIGR Web site (http://compbio.dfci.harvard.edu/tgi/). All of the *Arabidopsis* genome data were based on the TIGR annotation release 5.0 (January 2004). The rice intergenic sequence, gene function annotation, and gene model information were downloaded from the TIGR Web site (ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_2.0/).

Data processing

A total of 15,998 adjacent gene pairs with information on intergenic distance, gene orientation and gene expression were assembled from the microarray expression data and the intergenic sequence and gene orientation data. We removed genes that were not expressed in any of the 47 samples as well as gene pairs identified as duplicates according to the *Arabidopsis* gene duplication data, and obtained 12,625 pairs of adjacent, non-duplicated genes for which expression was detected in at least one sample.

The detection call value in the microarray data files was used to determine whether or not a gene was expressed in any given tissue sample. The value "1" was assigned to a detection call of "P (present)", "-1" to "A (absent), and "0" to "M (marginal) for the detection call in the microarray data files. The sum of these values for the three biological replicates was taken as the final detection call. A gene was considered to be expressed if the final detection call was ≥2, and not expressed if the final detection call was ≤-2. Data were considered to be ambiguous and excluded if the final detection call was ≥-1 and ≤1.

Among the 59,710 adjacent rice gene pairs, 444 divergent rice gene pairs with an intergenic distance of 61 - 400 bp were assembled from the rice gene model data. A customized database RiceRF61_400 was established, which contained DNA sequences between the 444 divergent rice gene pairs extracted from the rice intergenic sequence data file.

Analysis of coexpression levels of adjacent genes

We defined the coexpression rate between gene A and B as r = (Number of samples with both A and B expressed or both A and B not expressed)/(Number of samples with unambiguous data for both genes). Adjacent genes were defined as those genes located immediately next to each other, with no intervening genes, according to the latest TIGR annotation release 5.0.

Adjacent pairs were considered to be coexpressed, independently expressed, and antiexpressed if they had a coexpression rate of ≥0.95, 0.20 - 0.63, and ≤0.05, respectively.

To determine if the observed number of coexpressed adjacent pairs was significant, we used the cumulative binomial distribution, given by the formula,

$$P(n \geq y) = \sum_{n=y}^{N} \frac{N!}{y!(N-y)!} p^{y} (1-p)^{N-y}$$

where N is the total number of adjacent pairs sampled, y is the observed number of coexpressed adjacent pairs, and p is the observed probability of two randomly picked non-adjacent genes having a coexpression rate of ≥0.95.

To compare the number of coexpressed adjacent pairs between different configuration groups (divergent, parallel and convergent), the same formula was used, except where p is the observed probability of two adjacent genes in the control group having a coexpression rate of ≥0.95.

Gene function annotation

The Gene Ontology (GO) classification system at TAIR (http://www.arabidopsis.org) was used to define functional similarity.

Sequence analysis

Pairwise comparison of intergenic sequences was performed using the Blastall program [24].

The default parameter settings were used for all parameters except for the scoring matrix (set to "PAM40") and word size (set to "7"). The GC content calculation was carried out using Bioedit [25]. Multiple alignment analyses of DNA sequences were performed using ClustalW [26]. TATA box motifs were surveyed using PLACE Signal Scan Search [27].

Analysis of overrepresented motifs

Overrepresented oligos in the intergenic regions were identified using the oligo-analysis program in Regulatory Sequence Analysis Tools (RSAT) [28]. The motif localization map was drawn using the DNA-pattern program in RSAT [28].

RESULTS AND DISCUSSION

Adjacent divergent gene pairs in *Arabidopsis*

Approximately 25% of the intergenic regions in the *Arabidopsis* genome are flanked by divergent genes, and the lengths of these regions range from less than 0.1 kb to more than 10 kb (Data not shown). The number of divergent gene pairs decreases with intergenic distance on a 1-kb scale. The intergenic regions of less than 1 kb flanked by divergent genes accounts for about 9% of all the intergenic regions in *Arabidopsis* (Fig. 1A). Recently, it was discovered that 11% of the intergenic regions in the human genome are flanked by divergent genes less than 1 kb

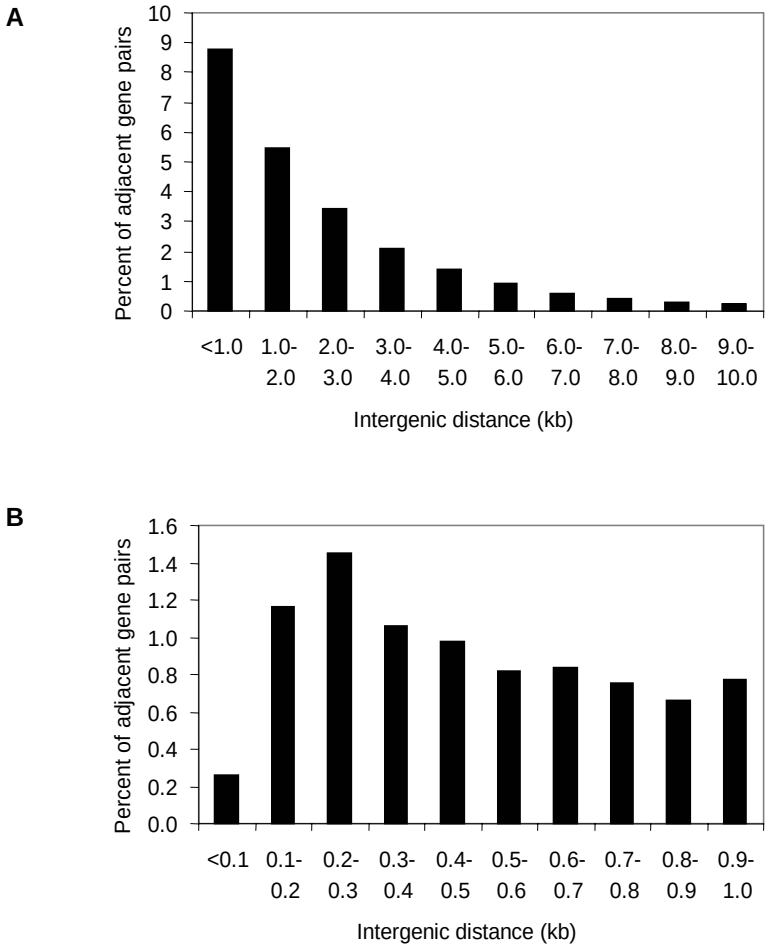

Figure 1. The distribution of intergenic distances between divergent genes as a percentage of adjacent gene pairs in the *Arabidopsis* genome.
(A) The distribution of intergenic distances between divergent genes on a 1-kb scale;
(B) The distribution of intergenic distances between divergent genes on a 0.1-kb scale.

apart [10]. The human genome size (~3 Gb) is much bigger than that of *Arabidopsis* (125 Mb) whereas both species have a comparable number of genes (25,000 - 27,000) [20, 29]. It is interesting that the percentage of divergent gene pairs separated by ≤1 kb in *Arabidopsis* is similar to that in humans whereas there is a large difference in gene density between the two species. Similarly, there was no large difference in the percentage of divergent gene pairs between human chromosomes 21 and 22 which are gene-poor and gene-rich, respectively, indicating that the divergent feature in the human genome has no correlation with gene density *per se* [30].

For the intergenic regions of ≤1 kb in *Arabidopsis*, the number of divergent gene pairs varies with intergenic distance on a 0.1-kb scale, with the lowest number for the region of ≤0.1 kb and the highest number for the region of 0.2-0.3 kb. The intergenic regions of ≤0.4 kb flanked by divergent genes accounts for about 4% of all the intergenic regions (Fig. 1B).

Coexpressed divergent gene pairs

There are 1,940 adjacent gene pairs in *Arabidopsis* containing duplicated genes, with the majority of them in parallel configuration (data not shown). To avoid the interference of gene duplication with coexpression analysis, all of the duplicated adjacent gene pairs were excluded in this study. We calculated the coexpression rate for each adjacent gene pair and divided the adjacent gene pairs into 3 groups: coexpressed, independently expressed, and antiexpressed, with coexpression rates of ≥0.95, 0.20-0.63, and ≤0.05, respectively. We set the coexpression rate of 0.63 as the upper threshold for independent expression because the average coexpression rate of 1,000 pairs of randomly picked non-adjacent genes was 0.63. Our analysis revealed that for each of the three adjacent gene pair configurations, the fraction of coexpressed pairs decreased with an increase in intergenic distance (Fig. 2A). This is consistent with the coexpression pattern of adjacent genes in *Caenorhabditis elegans* in which there was a general correlation between pair distance and coexpression level [31].

To determine if the coexpression level of adjacent gene pairs was significantly higher than the

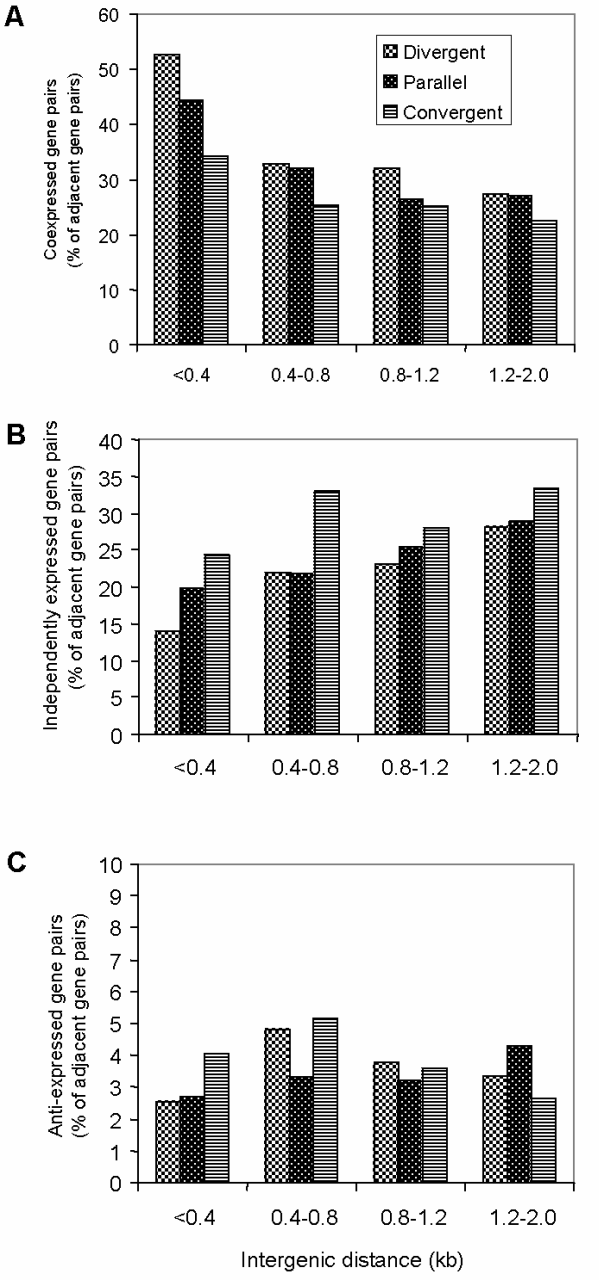

Figure 2. Comparison of coexpression of adjacent gene pairs among three different gene pair configurations in *Arabidopsis*.
(A) The fractions of adjacent gene pairs exhibiting coexpression (with a coexpression rate of 0.95 or higher);
(B) The fractions of adjacent gene pairs exhibiting independent expression (with a coexpression rate of between 0.20 and 0.63);
(C) The fractions of adjacent gene pairs exhibiting antiexpression (with a coexpression rate of 0.05 or lower).

Table 1. Analysis of coexpressed groups of adjacent genes ≤ 400 bp apart.

Configuration	Observed	Expected	*P*-value
Divergent vs Random	312	190	4.65×10^{-14}
Parallel vs Random	695	502	0.00
Convergent vs Random	478	449	0.044
Divergent vs Parallel	312	261	8.98×10^{-6}
Divergent vs Convergent	312	202	1.37×10^{-14}
Parallel vs Convergent	695	533	0.00

Note: For each gene pair configuration, the expected number of coexpressed gene pairs is shown along with the observed number of coexpressed gene pairs and the *P* value for obtaining such a result calculated using the cumulative binomial distribution.

coexpression level observable by chance in random gene pairs distributed throughout the genome, we analyzed the coexpression rate of 1,000 pairs of randomly picked non-adjacent genes, out of which ~320 pairs were shown to be coexpressed. There was a significantly higher portion of coexpressed gene pairs among the adjacent gene pairs ≤0.4 kb apart than among the random non-adjacent gene pairs (Table 1).

Recently, Williams and Bowles [1] determined the relationship between intergenic distance and degree of coexpression on a 1-kb scale (0-1 kb, 1-2 kb, 2-3 kb, etc.), and observed a significant correlation between coexpression and intergenic distance of gene pairs up to 12 kb apart. However, our results show that the portion of coexpressed gene pairs among the adjacent gene pairs >0.4 kb apart was not significantly higher than among the random non-adjacent gene pairs. Further, for the adjacent gene pairs ≤0.4 kb apart, there was a significantly higher portion of coexpressed gene pairs in divergent configuration than in either parallel or convergent configuration (Fig. 2A; Table 1). These results suggest that bidirectional promoters might be enriched in the regions between divergent gene pairs less than 0.4 kb apart. This is consistent with the individual examples of bidirectional promoters identified in mammals, most of which are found in the intergenic regions of less than 0.4 kb [9, 32].

We identified 312 intergenic regions flanked by coexpressed divergent gene pairs ≤0.4 kb apart in the *Arabidopsis* genome (Supplemental Table S2). These intergenic regions containing potential bidirectional promoters, are distributed throughout the *Arabidopsis* genome except in the centromere regions, which are gene-poor (Fig. 3).

Most of the bidirectional promoters found in mammalian genomes are TATA-less and GC-rich [9]. However, most of the potential *Arabidopsis* bidirectional promoters identified in this study contain TATA box motifs, and are not GC-rich (Table 2).

Independently expressed divergent gene pairs

For each of the three gene pair configurations, the fraction of independently expressed pairs increases with intergenic distance. For the gene pairs ≤0.4 kb apart, there was a smaller portion of gene pairs showing independent expression in divergent configuration than in either parallel or convergent configuration (Fig. 2B). Some divergent gene pairs less than 0.4 kb apart have also been found to be independently expressed in animals [11, 12]. Given that the close proximity of divergent genes is a mechanism of co-regulation driven by bidirectional promoters in intergenic regions, as demonstrated in nature [33] as well as in artificial gene constructs [2, 3], and it is known that enhancers of one gene may exert their effects on neighboring genes, how is the independent expression of some closely linked divergent genes achieved? One possible mechanism is the existence of genetic insulators. Genetic insulators are naturally-occurring DNA sequences that define gene boundaries and buffer

against the influence of elements of neighboring gene promoters, and thus play an important role in the regulation of gene expression. More than 20 different genetic insulators have been found in animals to define independent domains of transcriptional activity [17, 19]. To date no genetic insulators of plant origin have been reported. Recently, we identified a 16-bp sequence as a genetic insulator in *Arabidopsis* (unpublished data). It is possible that genetic insulators exist in the intergenic regions flanked by divergent genes showing an independent expression pattern.

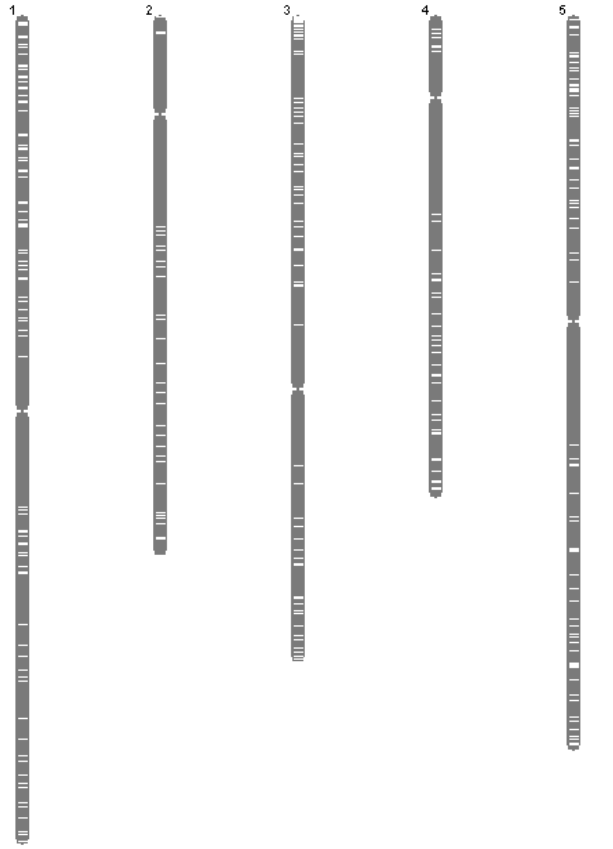

Figure 3. Distribution of potential bidirectional promoters on the *Arabidopsis* chromosomes.
The white tickmarks indicate the positions of bidirectional promoters.

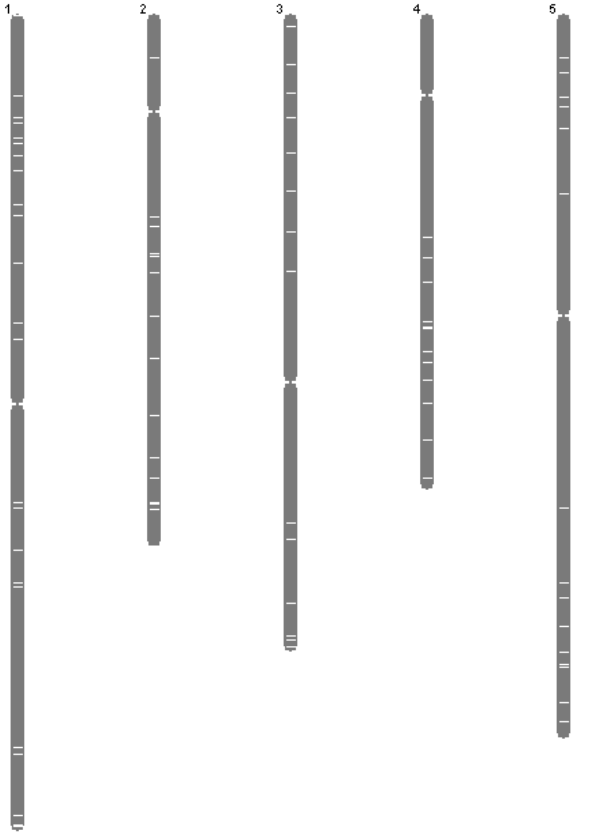

Figure 4. Distribution of potential genetic insulators on the *Arabidopsis* chromosomes.
The white tickmarks indicate the positions of genetic insulators.

Table 2. Features of intergenic sequences flanked by coexpressed, independently expressed and antiexpressed divergent genes with an intergenic distance of ≤0.4 kb.

Features	Coexpressed	Independent	Anti-expressed
Number of sequences	312	83	15
Sequences with TATA box	199 (64%)	58 (70%)	8 (53%)
Sequences with CAAT box	271 (87%)	80 (96%)	15 (100%)
GC%*	36.70 ± 0.05	37.33 ± 0.04	35.94 ± 0.03

*Average ± Standard deviation.

We identified 83 intergenic regions flanked by independently expressed divergent gene pairs ≤0.4 kb apart in *Arabidopsis* (Supplemental Table S3). These intergenic regions may harbor potential genetic insulators. Their distribution in the *Arabidopsis* genome is illustrated in Figure 4.

Antiexpressed divergent gene pairs

For each of the three gene pair configurations, the fraction of antiexpressed pairs is relatively small (less than 6%) as compared with coexpressed or independently expressed ones, and there was no large difference in the fraction of antiexpressed pairs among the three configurations (Fig. 2C). We identified 15 pairs of antiexpressed genes in divergent configuration with distances of ≤0.4 kb in the *Arabidopsis* genome (Supplemental Table S4). Similarly, divergent gene pairs in the human genome are not always coexpressed, and a minority of them exhibit a mutually exclusive expression pattern [10]. It is not clear how the closely-linked divergent genes are antiregulated. As discussed above, it is possible that genetic insulators may be implicated in the transcriptional regulatory mechanism of closely linked antiexpressed divergent genes. This possibility needs to be experimentally explored.

Functional classes of divergent genes

We used the Gene Ontology (GO) classification system at TAIR (http://www.arabidopsis.org) to define functional similarity between members of gene pairs. Among the 312 pairs of co-expressed divergent genes, only 72 pairs have both flanking genes annotated, and only ~21% of the 72 pairs were involved in the same biological processes (Table 3). Similarly, Williams and Bowles [1] found that in *Arabidopsis* common functionality is not a major cause for the coexpression of neighboring genes, although some genes in the same pathway are coexpressed. The correlations between expression patterns from hundreds of experiments for both *Saccharomyces cerevisiae* and *Caenorhabditis elegans* generally provide only a very weak signal for the prediction of functional interactions; at a correlation threshold of 0.6, the fraction of annotated proteins that are part of the same pathway is only 54% in *S. cerevisiae* and 34% in *C. elegans* [34]. Although bidirectional promoters are often shared by genes encoding either related products or proteins required in common metabolic or regulatory pathways [33], our analysis and other published data indicate that similar functionality is not a major cause for the existence of coexpressed divergent gene pairs.

Among the 83 pairs of independent divergent genes, only 19 pairs have both flanking genes annotated. About 16% of the 19 pairs were involved in the same biological process (Table 3). As for the antiexpressed divergent genes, because none of the 15 pairs of antiexpressed divergent genes ≤0.4 kb apart (Supplemental Table S4) have both flanking genes annotated at this time, we did not perform an analysis of functionality.

Overrepresented motifs in the regions between independent divergent genes

To search for potential insulators with enhancer-blocking activity in the intergenic regions between independently expressed divergent genes ≤0.4 kb apart (Supplemental Table S3), we analyzed these regions for overrepresented motifs using the intergenic regions flanked by coexpressed genes (Supplemental Table S2) as a background. We identified 25 overrepresented motifs which are potentially part of genetic insulators with enhancer-blocking activity (Table 4). In *Drosophila*, GAGA sites are essential for

Table 3. GO annotation analysis of adjacent genes in divergent configuration with an intergenic distance of 400 bp or less.

	Expression pattern of adjacent genes		
	Coexpressed	Independent	Antiexpressed
Total gene pairs	312	83	15
Number of gene pairs with known process	72	19	N/A
Pairs of genes involved in the same process	15	3	N/A

the enhancer-blocking activity of the SF1/b3 minimal insulator, and the highly conserved SF1 core sequence contains multiple GAGA site [18, 35]. It is very interesting that in *Arabidopsis* the GAGA motif is overrepresented in the intergenic regions flanked by independently expressed divergent genes as compared with the intergenic regions flanked by coexpressed divergent genes (Table 4). We found that eleven out of the 83 intergenic regions between independent divergent *Arabidopsis* genes contain multiple GAGA sites in the middle region (Fig. 5). These eleven

Table 4. Overrepresented oligos in the intergenic regions flanked by independently expressed divergent genes with an intergenic distance of 400 bp or less.

4 bp	5 bp	6 bp	7 bp	8 bp
agag	agaga	tggaca	tactaga	aggtttat
aaga	ctccc	cttaag	ggcggac	ggacaaac
gaga	aaaag	aaccgc	atggttg	aagcttca
acag	ctctc	agagag	acttaag	cgaattcg
aaag	aagag	ctcttc	gaagaga	gccggaac

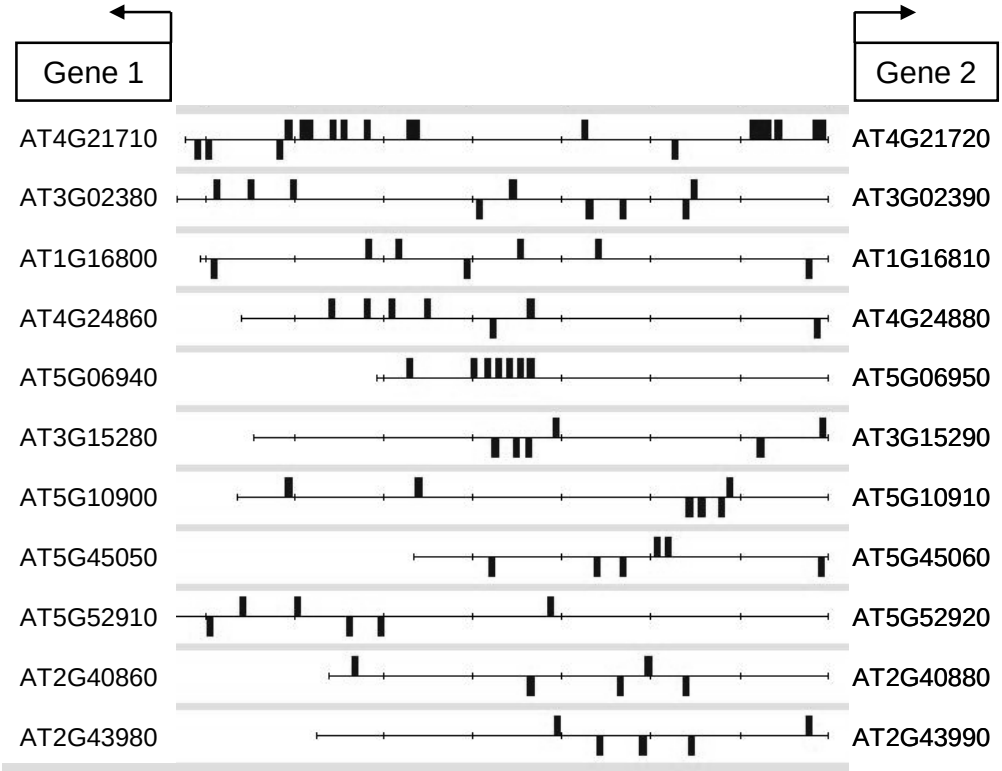

Figure 5. The locations of overrepresented GAGA sites in the intergenic regions flanked by independently expressed genes in Table 2.
Only the intergenic sequences having GAGA sites in the middle region are shown.

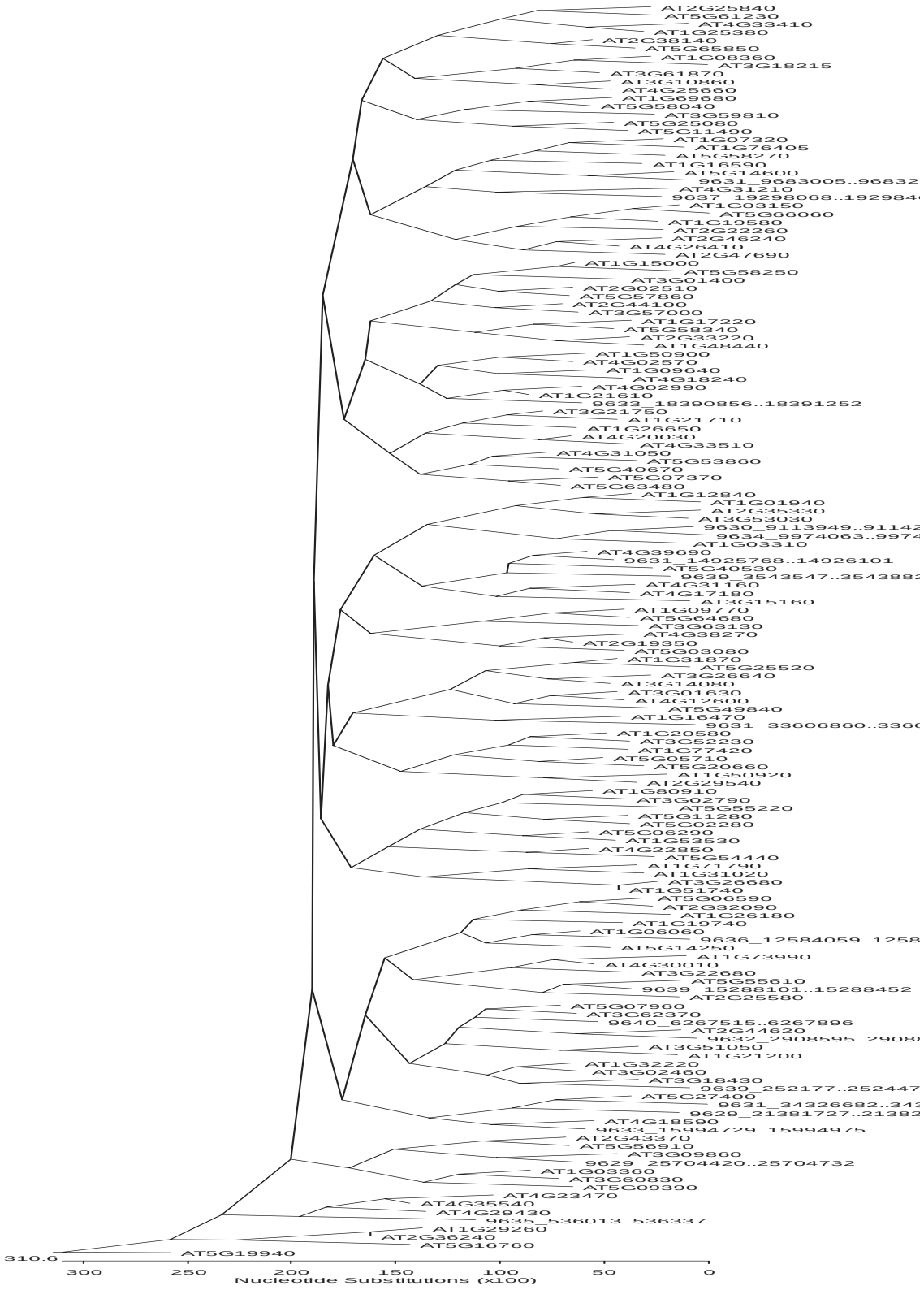
AT2G25840
AT5G61230
AT4G33410
AT1G25380
AT2G38140
AT5G65850
AT1G08360
AT3G18215
AT3G61870
AT3G10860
AT4G25660
AT1G69680
AT5G58040
AT3G59810
AT5G25080
AT5G11490
AT1G07320
AT1G76405
AT5G58270
AT1G16590
AT5G14600
9631_9683005..96832
AT4G31210
9637_19298068..192984
AT1G03150
AT5G66060
AT1G19580
AT2G22260
AT2G46240
AT4G26410
AT2G47690
AT1G15000
AT5G58250
AT3G01400
AT2G02510
AT5G57860
AT3G44100
AT3G57000
AT1G17220
AT5G58340
AT2G33220
AT1G48440
AT1G50900
AT4G02570
AT1G09640
AT4G18240
AT4G02990
AT1G21610
9633_18390856..18391252
AT3G21750
AT1G21710
AT1G26650
AT4G20030
AT4G33510
AT4G31050
AT5G53860
AT5G40670
AT5G07370
AT5G63480
AT1G12840
AT1G01940
AT2G35330
AT3G53030
9630_9113949..91143
9634_9974063..9974
AT1G03310
AT4G39690
9631_14925768..14926101
AT5G40530
9639_3543547..3543882
AT4G31160
AT4G17180
AT3G15160
AT1G09770
AT5G64680
AT3G63130
AT4G38270
AT2G19350
AT5G03080
AT1G31870
AT5G25520
AT3G26640
AT3G14080
AT3G01630
AT4G12600
AT5G49840
AT1G16470
9631_33606860..3360
AT1G20580
AT3G52230
AT1G77420
AT5G05710
AT5G20660
AT1G50920
AT2G29540
AT1G80910
AT3G02790
AT5G55220
AT5G11280
AT5G02280
AT5G06290
AT1G53530
AT4G22850
AT5G54440
AT1G71790
AT1G31020
AT3G26980
AT1G51740
AT5G06590
AT2G32090
AT1G26180
AT1G19740
AT1G06060
9636_12584059..1258
AT5G14250
AT1G73990
AT4G30010
AT3G22680
AT5G55610
9639_15288101..15288452
AT2G25580
AT5G07960
AT3G62370
9640_6267515..6267896
AT2G44620
9632_2908595..2908
AT3G51050
AT1G21200
AT1G32220
AT3G02460
AT3G18430
9639_252177..252447
AT5G27400
9631_34326682..343
9629_21381727..213382
AT4G18590
9633_15994729..15994975
AT2G43370
AT5G66910
AT3G09860
9629_25704420..25704732
AT1G03360
AT3G60830
AT5G09390
AT4G23470
AT4G35540
AT4G29430
9635_536013..536337
AT1G29260
AT2G36240
AT5G16760
AT5G19940
310.6
300
250
200
150
100
50
0
Nucleotide Substitutions (x100)

Figure 6

intergenic regions may be worth our attention in future efforts to locate genetic insulators in plants.

Conserved intergenic regions flanked by divergent genes between *Arabidopsis* and rice

We performed a blast search against the customized database RiceRF61_400, which contains 444 DNA sequences between divergent rice gene pairs 61 - 400 bp apart, using the 312, 83, and 15 DNA sequences between divergent *Arabidopsis* gene pairs ≤0.4kb apart exhibiting coexpression, independent expression, and antiexpression, respectively (Supplemental Tables S2-S4). About 6% of the DNA sequences between divergent *Arabidopsis* gene pairs ≤0.4kb apart exhibiting coexpression or independent expression had significant hits (identity ≥60%; E-value ≤0.0001) in the database RiceRF61_400 whereas no hits were found for the DNA sequences between divergent *Arabidopsis* gene pairs ≤0.4kb apart exhibiting antiexpression.

We also carried out pairwise comparisons using blast analysis between sequences flanked by divergent *Arabidopsis* genes ≤0.4 kb apart. Among the 48,516 pairwise comparisons of the 312 sequences flanked by coexpressed divergent genes, 0.3% showed significant homology (identity of ≥60% with E-value ≤0.0001). Similarly, among the 3,403 pairwise comparisons of the 83 sequences flanked by independent divergent genes, 0.3% showed significant homology (identity of ≥60% with E-value ≤0.0001). No homology was found among the 15 DNA sequences flanked by antiexpressed divergent *Arabidopsis* gene pairs ≤0.4kb apart.

152 conserved DNA sequences (*Arabidopsis*, 134; rice, 18) were obtained from pairwise comparisons between sequences flanked by coexpressed divergent *Arabidopsis* genes ≤0.4 kb apart as well as from blast searches against the database RiceRF61_400 using the DNA sequences flanked by coexpressed divergent *Arabidopsis* gene pairs ≤0.4kb apart. Multiple alignment analysis show that the intergenic regions containing the conserved DNA sequences can be grouped into several clusters (Fig. 6). This suggests that the divergent transcription of functionally unrelated *Arabidopsis* genes in Supplemental Table S2 are driven by bidirectional promoters that are ancestral sequences that have persisted through evolution. Likewise, it is suggested that the known bidirectional promoters in the mammalian genome are preserved ancestral sequences [9].

CONCLUSION

Genome-wide analysis of sequence and microarray expression data revealed three different expression patterns in adjacent divergent gene pairs in *Arabidopsis*: coexpression, independent expression, and antiexpression. We identified specific intergenic regions containing potential bidirectional promoters or genetic insulators, offering guidance for future experimental efforts to isolate those regulatory elements.

ACKNOWLEDGMENTS

We thank Drs. Steven Tanksley and Lukas Mueller of Cornell University for advice and critical reading of this manuscript, Drs. June Nasrallah, Susan McCouch (Cornell University) and Thomas Brutnell (Boyce Thompson Institute) for useful discussions. This work was supported by the US Department of Energy's Energy Biosciences grant DE-FG02–02ER15341 and by a new faculty startup fund at Cornell to S. G. C. M. W. was a recipient of the Cornell Presidential Genomics Graduate Fellowship.

Legend to Figure 6. Phylogenetic tree revealed by multiple alignment analysis of 152 conserved DNA sequences obtained from pairwise comparisons between sequences flanked by coexpressed divergent *Arabidopsis* genes ≤0.4 kb apart as well as from a blast search against the database RiceRF61_400 using the DNA sequences flanked by coexpressed divergent *Arabidopsis* gene pairs ≤0.4kb apart.
The names that begins with "AT" are intergenic sequences from *Arabidopsis*, otherwise from rice. Each intergenic sequence is represented by one gene flanking the intergenic region.

SUPPLEMENTARY TABLES

Supplementary Table S1. Microarray expression data used in this study.

Slide name	Sample description	Slide name	Sample description
ATGE_1	cotyledons	ATGE_33	flowers stage 12
ATGE_2	hypocotyl	ATGE_34	flowers stage 12, sepals
ATGE_3	roots	ATGE_35	flowers stage 12, petals
ATGE_4	shoot apex, vegetative + young leaves	ATGE_36	flowers stage 12, stamens
ATGE_6	shoot apex, vegetative	ATGE_37	flowers stage 12, carpels
ATGE_8	shoot apex, transition (before bolting)	ATGE_39	flowers stage 15
ATGE_9	roots	ATGE_40	flowers stage 15, pedicels
ATGE_10	rosette leaf #4, 1cm long	ATGE_41	flowers stage 15, sepals
ATGE_12	rosette leaf # 2	ATGE_42	flowers stage 15, petals
ATGE_13	rosette leaf # 4	ATGE_43	flowers stage 15, stamen
ATGE_14	rosette leaf # 4	ATGE_45	flowers stage 15, carpels
ATGE_15	rosette leaf # 8	ATGE_73	mature pollen
ATGE_16	rosette leaf # 10	ATGE_75	siliques, w/ seeds stage 2
ATGE_17	rosette leaf # 12	ATGE_77	siliques, w/ seeds stage 4
ATGE_19	leaf 7, petiol	ATGE_79	seeds, stage 6, w/o siliques
ATGE_20	leaf 7, proximal half	ATGE_82	seeds, stage 8, w/o siliques
ATGE_21	leaf 7, distal half	ATGE_84	seeds, stage 10, w/o siliques
ATGE_25	senescing leaves	ATGE_93	Root
ATGE_26	cauline leaves	ATGE_94	Root
ATGE_27	stem, 2nd internode	ATGE_95	Root
ATGE_28	1st node	ATGE_98	Root
ATGE_29	shoot apex, inflorescence (after bolting)	ATGE_99	Root
ATGE_31	flowers stage 9	ATGE_100	seedling, green parts
ATGE_32	flowers stage 10/11		

Supplemental Table S2. Intergenic regions flanked by divergent coexpressed genes with an intergenic distance of 400 bp or less.

Gene1	Intergenic distance (bp)	Gene2	Gene1	Intergenic distance (bp)	Gene2
AT1G01725	191	AT1G01730	AT1G27510	366	AT1G27520
AT1G01910	284	AT1G01920	AT1G29250	87	AT1G29260
AT1G01930	166	AT1G01940	AT1G29700	231	AT1G29710
AT1G02090	174	AT1G02100	AT1G30290	279	AT1G30300
AT1G03140	110	AT1G03150	AT1G30880	201	AT1G30890
AT1G03300	392	AT1G03310	AT1G31010	244	AT1G31020
AT1G03350	167	AT1G03360	AT1G31860	139	AT1G31870

Supplemental Table S2 continued..

Gene1	Intergenic distance (bp)	Gene2	Gene1	Intergenic distance (bp)	Gene2
AT1G04070	189	AT1G04080	AT1G32210	170	AT1G32220
AT1G04340	313	AT1G04350	AT1G33970	356	AT1G33980
AT1G04960	301	AT1G04970	AT1G48200	127	AT1G48210
AT1G06050	389	AT1G06060	AT1G48430	207	AT1G48440
AT1G06200	273	AT1G06210	AT1G48830	232	AT1G48840
AT1G06460	192	AT1G06470	AT1G50430	137	AT1G50440
AT1G07310	171	AT1G07320	AT1G50500	150	AT1G50510
AT1G07360	294	AT1G07370	AT1G50890	166	AT1G50900
AT1G07820	354	AT1G07830	AT1G50910	299	AT1G50920
AT1G08350	187	AT1G08360	AT1G51590	115	AT1G51600
AT1G08470	133	AT1G08480	AT1G51730	81	AT1G51740
AT1G09140	322	AT1G09150	AT1G52220	138	AT1G52230
AT1G09630	206	AT1G09640	AT1G52590	251	AT1G52600
AT1G09760	176	AT1G09770	AT1G53520	197	AT1G53530
AT1G10500	219	AT1G10510	AT1G54050	399	AT1G54060
AT1G12640	270	AT1G12650	AT1G54100	325	AT1G54110
AT1G12830	348	AT1G12840	AT1G54140	213	AT1G54150
AT1G13870	286	AT1G13880	AT1G59960	386	AT1G59970
AT1G14140	219	AT1G14150	AT1G61780	248	AT1G61790
AT1G14300	258	AT1G14310	AT1G62730	224	AT1G62740
AT1G14990	352	AT1G15000	AT1G63970	198	AT1G63980
AT1G15270	212	AT1G15280	AT1G64510	146	AT1G64520
AT1G16460	166	AT1G16470	AT1G64850	344	AT1G64860
AT1G16570	194	AT1G16590	AT1G67930	249	AT1G67940
AT1G17210	326	AT1G17220	AT1G69670	362	AT1G69680
AT1G19570	307	AT1G19580	AT1G71210	173	AT1G71220
AT1G19730	250	AT1G19740	AT1G71780	218	AT1G71790
AT1G20575	142	AT1G20580	AT1G73177	61	AT1G73180
AT1G21190	313	AT1G21200	AT1G73980	396	AT1G73990
AT1G21600	104	AT1G21610	AT1G74030	239	AT1G74040
AT1G21700	259	AT1G21710	AT1G74370	220	AT1G74380
AT1G21770	155	AT1G21780	AT1G76020	295	AT1G76030
AT1G24040	235	AT1G24050	AT1G76400	212	AT1G76405
AT1G24350	248	AT1G24360	AT1G77410	234	AT1G77420
AT1G25375	277	AT1G25380	AT1G78620	161	AT1G78630
AT1G26170	312	AT1G26180	AT1G78895	209	AT1G78900
AT1G26640	188	AT1G26650	AT1G79550	255	AT1G79560

Supplemental Table S2 continued..

Gene1	Intergenic distance (bp)	Gene2	Gene1	Intergenic distance (bp)	Gene2
AT1G27390	133	AT1G27400	AT1G80670	259	AT1G80680
AT1G80900	395	AT1G80910	AT3G03330	160	AT3G03340
AT2G01060	134	AT2G01070	AT3G03600	245	AT3G03610
AT2G02400	64	AT2G02410	AT3G04780	227	AT3G04790
AT2G02500	149	AT2G02510	AT3G04820	284	AT3G04830
AT2G17560	385	AT2G17570	AT3G05060	206	AT3G05070
AT2G17975	203	AT2G17980	AT3G09800	230	AT3G09810
AT2G18400	172	AT2G18410	AT3G09850	268	AT3G09860
AT2G19340	262	AT2G19350	AT3G10210	215	AT3G10220
AT2G19720	197	AT2G19730	AT3G10850	161	AT3G10860
AT2G20725	119	AT2G20740	AT3G11240	244	AT3G11250
AT2G21260	169	AT2G21270	AT3G11620	143	AT3G11630
AT2G22250	216	AT2G22260	AT3G12260	257	AT3G12270
AT2G25570	324	AT2G25580	AT3G14075	396	AT3G14080
AT2G25830	214	AT2G25840	AT3G14900	152	AT3G14910
AT2G27330	121	AT2G27340	AT3G15150	226	AT3G15160
AT2G29530	280	AT2G29540	AT3G16000	196	AT3G16010
AT2G31190	290	AT2G31200	AT3G16640	263	AT3G16650
AT2G32080	228	AT2G32090	AT3G18210	278	AT3G18215
AT2G33210	150	AT2G33220	AT3G18380	165	AT3G18390
AT2G35320	128	AT2G35330	AT3G18420	150	AT3G18430
AT2G36230	128	AT2G36240	AT3G18850	342	AT3G18860
AT2G37240	173	AT2G37250	AT3G19630	204	AT3G19640
AT2G38130	167	AT2G38140	AT3G21280	60	AT3G21290
AT2G38650	265	AT2G38660	AT3G21740	334	AT3G21750
AT2G40650	180	AT2G40660	AT3G22670	269	AT3G22680
AT2G43360	138	AT2G43370	AT3G23570	360	AT3G23580
AT2G43750	396	AT2G43760	AT3G23700	264	AT3G23710
AT2G44040	234	AT2G44050	AT3G24820	184	AT3G24830
AT2G44090	398	AT2G44100	AT3G26360	372	AT3G26370
AT2G44610	209	AT2G44620	AT3G26630	236	AT3G26640
AT2G44630	259	AT2G44640	AT3G26670	121	AT3G26680
AT2G45980	259	AT2G45990	AT3G29280	163	AT3G29290
AT2G46230	201	AT2G46240	AT3G44740	295	AT3G44750
AT2G47680	357	AT2G47690	AT3G46210	181	AT3G46220
AT3G01160	205	AT3G01170	AT3G49080	125	AT3G49100
AT3G01360	279	AT3G01370	AT3G49990	212	AT3G50000

Supplemental Table S2 continued..

Gene1	Intergenic distance (bp)	Gene2	Gene1	Intergenic distance (bp)	Gene2
AT3G01390	232	AT3G01400	AT3G51040	178	AT3G51050
AT3G01620	290	AT3G01630	AT3G52220	231	AT3G52230
AT3G01770	274	AT3G01780	AT3G53020	258	AT3G53030
AT3G02080	300	AT3G02090	AT3G53570	244	AT3G53580
AT3G02190	200	AT3G02200	AT3G53690	214	AT3G53700
AT3G02450	203	AT3G02460	AT3G56830	30	AT3G56840
AT3G02555	211	AT3G02560	AT3G56990	210	AT3G57000
AT3G02780	169	AT3G02790	AT3G57560	185	AT3G57570
AT3G03150	307	AT3G03160	AT3G58170	155	AT3G58180
AT3G58600	288	AT3G58610	AT4G31200	365	AT4G31210
AT3G59800	340	AT3G59810	AT4G33400	354	AT4G33410
AT3G60740	354	AT3G60750	AT4G33500	98	AT4G33510
AT3G60820	144	AT3G60830	AT4G34660	195	AT4G34670
AT3G61070	363	AT3G61080	AT4G35530	182	AT4G35540
AT3G61860	317	AT3G61870	AT4G35760	280	AT4G35770
AT3G62360	215	AT3G62370	AT4G36140	296	AT4G36150
AT3G62800	195	AT3G62810	AT4G36390	116	AT4G36400
AT3G63120	373	AT3G63130	AT4G38150	114	AT4G38160
AT3G63390	54	AT3G63400	AT4G38260	99	AT4G38270
AT3G63410	198	AT3G63420	AT4G39680	139	AT4G39690
AT4G01310	260	AT4G01320	AT5G01450	248	AT5G01460
AT4G01560	242	AT4G01570	AT5G02150	275	AT5G02160
AT4G01897	200	AT4G01900	AT5G02270	245	AT5G02280
AT4G02560	224	AT4G02570	AT5G03070	327	AT5G03080
AT4G02620	266	AT4G02630	AT5G04790	165	AT5G04800
AT4G02980	381	AT4G02990	AT5G05000	230	AT5G05010
AT4G12010	262	AT4G12020	AT5G05200	304	AT5G05210
AT4G12590	113	AT4G12600	AT5G05700	175	AT5G05710
AT4G14790	135	AT4G14800	AT5G06260	190	AT5G06265
AT4G14880	362	AT4G14890	AT5G06280	351	AT5G06290
AT4G16700	239	AT4G16710	AT5G06580	105	AT5G06590
AT4G17040	219	AT4G17050	AT5G07360	262	AT5G07370
AT4G17170	154	AT4G17180	AT5G07950	109	AT5G07960
AT4G18230	212	AT4G18240	AT5G08040	249	AT5G08050
AT4G18580	199	AT4G18590	AT5G08280	183	AT5G08290
AT4G20020	218	AT4G20030	AT5G08415	280	AT5G08420
AT4G21100	341	AT4G21105	AT5G08530	169	AT5G08535

Supplemental Table S2 continued..

Gene1	Intergenic distance (bp)	Gene2	Gene1	Intergenic distance (bp)	Gene2
AT4G21140	206	AT4G21150	AT5G08580	342	AT5G08590
AT4G21860	344	AT4G21865	AT5G09380	397	AT5G09390
AT4G22300	289	AT4G22310	AT5G10700	181	AT5G10710
AT4G22840	272	AT4G22850	AT5G11030	192	AT5G11040
AT4G23460	98	AT4G23470	AT5G11270	173	AT5G11280
AT4G24490	394	AT4G24500	AT5G11480	262	AT5G11490
AT4G25540	163	AT4G25550	AT5G14040	219	AT5G14050
AT4G25650	209	AT4G25660	AT5G14240	231	AT5G14250
AT4G26400	173	AT4G26410	AT5G14590	209	AT5G14600
AT4G28200	169	AT4G28210	AT5G16050	310	AT5G16060
AT4G29420	169	AT4G29430	AT5G16750	331	AT5G16760
AT4G29480	159	AT4G29490	AT5G16940	185	AT5G16950
AT4G30000	201	AT4G30010	AT5G18110	274	AT5G18120
AT4G30890	256	AT4G30900	AT5G18790	157	AT5G18800
AT4G31040	207	AT4G31050	AT5G19930	275	AT5G19940
AT4G31150	306	AT4G31160	AT5G20170	165	AT5G20180
AT4G31170	230	AT4G31180	AT5G20650	197	AT5G20660
AT5G22330	278	AT5G22340	AT5G55600	171	AT5G55610
AT5G23080	228	AT5G23090	AT5G56130	376	AT5G56140
AT5G25070	180	AT5G25080	AT5G56900	251	AT5G56910
AT5G25510	253	AT5G25520	AT5G57850	318	AT5G57860
AT5G27390	229	AT5G27400	AT5G58030	140	AT5G58040
AT5G38880	297	AT5G38890	AT5G58240	123	AT5G58250
AT5G40190	394	AT5G40200	AT5G58260	200	AT5G58270
AT5G40520	235	AT5G40530	AT5G58330	225	AT5G58340
AT5G40660	384	AT5G40670	AT5G58440	38	AT5G58450
AT5G43130	400	AT5G43140	AT5G59740	271	AT5G59750
AT5G45010	307	AT5G45020	AT5G61220	263	AT5G61230
AT5G45250	188	AT5G45260	AT5G61760	183	AT5G61770
AT5G47680	376	AT5G47690	AT5G63470	169	AT5G63480
AT5G47760	178	AT5G47770	AT5G63830	264	AT5G63840
AT5G47880	174	AT5G47890	AT5G64670	150	AT5G64680
AT5G49830	253	AT5G49840	AT5G65260	358	AT5G65270
AT5G51110	245	AT5G51120	AT5G65560	189	AT5G65570
AT5G52200	240	AT5G52210	AT5G65840	253	AT5G65850
AT5G53850	397	AT5G53860	AT5G66055	216	AT5G66060
AT5G54430	177	AT5G54440	AT5G67490	210	AT5G67500
AT5G55210	227	AT5G55220	AT5G67580	325	AT5G67590

Supplemental Table S3. Intergenic regions flanked by divergent independently expressed genes with an intergenic distance of 400 bp or less.

Gene1	Intergenic distance (bp)	Gene2	Gene1	Intergenic distance (bp)	Gene2
AT1G01150	358	AT1G01160	AT3G09480	288	AT3G09490
AT1G09290	163	AT1G09300	AT3G12070	247	AT3G12080
AT1G11390	322	AT1G11400	AT3G15280	322	AT3G15290
AT1G11880	345	AT1G11890	AT3G18870	234	AT3G18880
AT1G13320	185	AT1G13330	AT3G22770	134	AT3G22780
AT1G13900	211	AT1G13910	AT3G25855	264	AT3G25860
AT1G15140	265	AT1G15150	AT3G50420	47	AT3G50430
AT1G16800	352	AT1G16810	AT3G52090	290	AT3G52100
AT1G20340	349	AT1G20350	AT3G58520	60	AT3G58530
AT1G21160	299	AT1G21170	AT3G61760	389	AT3G61770
AT1G26540	197	AT1G26550	AT3G62220	375	AT3G62230
AT1G31790	383	AT1G31800	AT3G62910	289	AT3G62920
AT1G33030	222	AT1G33040	AT4G14320	309	AT4G14330
AT1G48550	171	AT1G48560	AT4G15810	378	AT4G15820
AT1G49140	282	AT1G49150	AT4G17750	188	AT4G17760
AT1G52920	215	AT1G52930	AT4G21270	266	AT4G21280
AT1G55880	180	AT1G55890	AT4G21710	360	AT4G21720
AT1G56345	316	AT1G56350	AT4G21810	370	AT4G21820
AT1G71730	335	AT1G71740	AT4G23930	395	AT4G23940
AT1G72370	242	AT1G72380	AT4G24860	329	AT4G24880
AT1G78270	137	AT1G78280	AT4G26760	324	AT4G26770
AT1G79210	203	AT1G79220	AT4G29050	376	AT4G29060
AT1G79440	356	AT1G79450	AT4G32420	137	AT4G32430
AT1G79880	276	AT1G79890	AT4G36050	261	AT4G36060
AT2G04650	147	AT2G04660	AT4G38870	252	AT4G38880
AT2G17130	256	AT2G17140	AT5G05510	276	AT5G05520
AT2G17970	308	AT2G17972	AT5G06940	253	AT5G06950
AT2G20440	342	AT2G20450	AT5G09830	175	AT5G09840
AT2G20480	223	AT2G20490	AT5G10900	331	AT5G10910
AT2G20700	273	AT2G20710	AT5G13270	225	AT5G13280
AT2G22400	141	AT2G22410	AT5G19660	188	AT5G19670
AT2G26140	357	AT2G26150	AT5G45050	232	AT5G45060
AT2G29570	219	AT2G29580	AT5G51430	255	AT5G51440
AT2G34960	345	AT2G34970	AT5G52910	383	AT5G52920
AT2G39090	137	AT2G39100	AT5G52950	206	AT5G52960
AT2G40860	280	AT2G40880	AT5G55570	179	AT5G55580

Supplemental Table S3 continued..

Gene1	Intergenic distance (bp)	Gene2	Gene1	Intergenic distance (bp)	Gene2
AT2G43180	186	AT2G43190	AT5G57960	217	AT5G57970
AT2G43390	390	AT2G43400	AT5G59240	159	AT5G59250
AT2G43980	287	AT2G43990	AT5G59600	139	AT5G59610
AT2G47760	200	AT2G47770	AT5G63020	282	AT5G63030
AT3G02380	365	AT3G02390	AT5G65110	382	AT5G65120
AT3G06150	209	AT3G06160			

Supplemental Table S4. Intergenic regions flanked by divergent anti-expressed genes with an intergenic distance of 400 bp or less.

Gene 1	Intergenic distance (bp)	Gene 2
AT1G19920	265	AT1G19930
AT1G20670	377	AT1G20680
AT1G33490	379	AT1G33500
AT1G60710	329	AT1G60720
AT2G43270	206	AT2G43280
AT3G07920	388	AT3G07930
AT3G49010	390	AT3G49020
AT3G56430	191	AT3G56440
AT3G61800	321	AT3G61810
AT3G63220	394	AT3G63230
AT4G04470	390	AT4G04480
AT4G08700	283	AT4G08710
AT5G41850	159	AT5G41860
AT5G64720	321	AT5G64730
AT5G67310	212	AT5G67320

REFERENCES

1. Williams, E. J. and Bowles, D. J. 2004, Genome Res., 14, 1060.
2. He, Y. and Gan, S. 2001, Plant Mol. Biol., 47, 595.
3. Xie, M., He, Y. and Gan, S. 2001, Nat. Biotechnol., 19, 677.
4. Lennard, A., Gaston, K., and Fried, M. 1994, DNA Cell Biol., 13, 1117.
5. Gavalas, A. and Zalkin, H. 1995, J. Biol. Chem., 270, 2403.
6. Ikeda, S., Ayabe, H., Mori, K., Seki, Y., and Seki, S. 2002, Biochem. Biophys. Res. Commun., 296, 785.
7. Ikeda, S., Mochizuki, A., Sarker, A. H., and Seki, S. 2000, Biochem. Biophys. Res. Commun., 273, 1063.
8. Remaley, A. T., Bark, S., Walts, A. D., Freeman, L., Shulenin, S., Annilo, T., Elgin, E., Rhodes, H. E., Joyce, C., Dean, M., Santamarina-Fojo, S., Brewer, H. B., and Jr. 2002, Biochem. Biophys. Res. Commun., 295, 276.
9. Zhang, L. F., Ding, J. H., Yang, B. Z., He, G. C., and Roe, C., 2003, Genomics, 82, 660.

10. Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., and Myers, R. M. 2004, Genome Res., 14, 62.
11. Ame, J. C., Schreiber, V., Fraulob, V., Dolle, P., Murcia, G., and Niedergang, C. P. 2001, J. Biol. Chem., 276, 11092.
12. Otte, D. M., Schwaab, U., and Luers, G. H. 2003, Gene, 313, 119.
13. Hirschman, J. E., Durbin, K. J., and Winston, F. 1988, Mol. Cell Biol., 8, 4608.
14. Christoffersen, C. A., Brickman, T. J., Hook-Barnard, I., and McIntosh, M. A. 2001, J. Bacteriol., 183, 2059.
15. Nieto, C., Puyet, A., and Espinosa, M. 2001, J. Biol. Chem., 276, 14946.
16. Delpy, L., Decourt, C., Le Bert, M., and Cogne, M. 2002, J. Immunol., 169, 6875.
17. West, A.G., Gaszner, M., and Felsenfeld, G. 2002, Genes Dev., 16, 271.
18. Belozerov, V. E., Majumder, P., Shen, P., and Cai, H. N. 2003, Embo J., 22, 3113.
19. Kuhn, E. J. and Geyer, P. K. 2003, Curr. Opin. Cell Biol., 15, 259.
20. The *Arabidopsis* Genome Initiative. 2000, Nature, 408, 796.
21. Guo, H. and Moose, S. P. 2003, Plant Cell, 15, 1143.
22. Wolfe, K. H., Gouy, M., Yang, Y. W., Sharp, P. M., and Li, W. H. 1989, Proc. Natl. Acad. Sci. USA, 86, 6201.
23. Schoof, H. and Karlowski, W. M. 2003, Curr. Opin. Plant Biol., 6, 106.
24. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990, J. Mol. Biol., 215, 403.
25. Hall, T. A. 1999, Nucleic Acids Symposium Series, 41, 95.
26. Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994, Nucleic Acids Res., 22, 4673.
27. Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. 1999, Nucleic Acids Res., 27, 297.
28. Van Helden, J. 2003, Nucleic Acids Res., 31, 3593.
29. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., and Holt, R. A. 2001, Science, 291, 1304.
30. Adachi, N. and Lieber, M. R. 2002, Cell, 109, 807.
31. Lercher, M. J., Blumenthal, T., and Hurst, L. D. 2003, Genome Res., 13, 238.
32. Whitehouse, C., Chambers, J., Catteau, A., and Solomon, E. 2004, Gene, 326, 87.
33. Guarguaglini, G., Battistoni, A., Pittoggi, C., Di Matteo, G., Di Fiore, B., and Lavia, P. 1997, Biochem. J., 325 (Pt 1), 277.
34. Van Noort, V., Snel, B., and Huynen, M. A. 2003, Trends Genet., 19, 238.
35. Ohtsuki, S. and Levine, M. 1998, Gen. Dev., 12, 3325.