Genome Analysis

Comparative Analysis of Divergent and Convergent Gene Pairs and Their Expression Patterns in Rice, Arabidopsis, and Populus

Nicholas Krom and Wusirika Ramakrishna*
Department of Biological Sciences, Michigan Technological University, Houghton, Michigan 49931

Comparative analysis of the organization and expression patterns of divergent and convergent gene pairs in multiple plant genomes can identify patterns that are shared by more than one species or are unique to a particular species. Here, we study the coexpression and interspecies conservation of divergent and convergent gene pairs in three plant species: rice (Oryza sativa), Arabidopsis (Arabidopsis thaliana), and black cottonwood (Populus trichocarpa). Strongly correlated expression levels between divergent and convergent genes were found to be quite common in all three species, and the frequency of strong correlation appears to be independent of intergenic distance. Conservation of divergent or convergent arrangement among these species appears to be quite rare. However, conserved arrangement is significantly more frequent when the genes display strongly correlated expression levels or have one or more Gene Ontology (GO) classes in common. A correlation between intergenic distance in divergent and convergent gene pairs and shared GO classes was observed, in varying degrees, in rice and Populus but not in Arabidopsis. Furthermore, multiple GO classes were either overrepresented or underrepresented in Arabidopsis and Populus gene pairs, while only two GO classes were underrepresented in rice divergent gene pairs. Three cis-regulatory elements common to both Arabidopsis and rice were overrepresented in the intergenic regions of strongly correlated divergent gene pairs compared to those of noncorrelated pairs. Our results suggest that shared as well as unique mechanisms operate in shaping the organization and function of divergent and convergent gene pairs in different plant species.

Gene rearrangements occur frequently during the evolution of prokaryotic and eukaryotic genomes. The number of rearrangements appears to be a function of the phylogenetic distance between the organisms being studied. Rice (Oryza sativa) and Arabidopsis (Arabidopsis thaliana) are the model monocot and dicot genomes that have been fully sequenced (Arabidopsis Genome Initiative, 2000; International Rice Genome Sequencing Project, 2005). Recently, a second dicot plant genome, Populus trichocarpa, has been sequenced (Tuskan et al., 2006). Divergence time between Populus and Arabidopsis is estimated to be 100 to 120 million years ago and that of Arabidopsis and rice is 130 to 200 million years ago (Wolfe et al., 1989; Chaw et al., 2004; Tuskan et al., 2006). Very little collinearity in gene order has been observed between Arabidopsis and rice due to the large evolutionary distance that separates them (Devos et al., 1999; Liu et al., 2001; Vandepoele et al., 2002). Despite this lack of collinearity, at the level of single genes, 71% of protein coding rice genes had homologs in Arabidopsis genome compared to 90% of Arabidopsis genes with homologs in the rice genome (International Rice Genome Sequencing Project, 2005). Eukaryotic genes appear to be distributed in a non-random fashion with clustered genes exhibiting coordinated expression patterns (Hurst et al., 2004). Different trends of coexpression were observed depending on the types of genes and organisms. Strong positive correlation was observed in the expression patterns of divergent gene pairs compared to weak or no correlation in those of convergent gene pairs in Caenorhabditis elegans (Chen and Stein, 2006). This was attributed to RNA transcripts from convergent genes obstructing each other by base pairing at their 3’ ends (Katayama et al., 2005). Although coexpression patterns were observed in both divergent as well as convergent genes in yeast (Saccharomyces cerevisiae), divergent gene pairs displayed higher correlation than convergent gene pairs (Cohen et al., 2000; Kruglyak and Tang, 2000). Significant numbers of pairs of adjacent genes have been found to have strongly correlated expression levels in Arabidopsis (Williams and Bowles, 2004). Local domains of two to four highly coexpressed genes have also been identified in Arabidopsis (Ren et al., 2005), as have higher-order domains corresponding to regions of euchromatin (Zhan et al., 2006). Additionally, correlated expression of neighboring genes appears to be more common when both genes in a pair are classified in the same functional category (Williams and Bowles, 2004). Correlated expression patterns of divergent or convergent genes

---

1 This work was supported by the National Research Initiative of the U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service (grant no. 2007–35301–18036).

* Corresponding author; e-mail wusirika@mtu.edu.

The author responsible for the distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Wusirika Ramakrishna (wusirika@mtu.edu).

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.108.122416

Plant Physiology, August 2008, Vol. 147, pp. 1763–1773, www.plantphysiol.org © 2008 American Society of Plant Biologists

Downloaded from on July 23, 2018 - Published by www.plantphysiol.org

Copyright © 2008 American Society of Plant Biologists. All rights reserved.
might result due to cis-acting enhancers and/or their involvement in the same or related biological process as determined by Gene Ontology (GO) classification. Furthermore, chromatin organization can regulate coexpression as seen in the case of coordinated expression of two transgenes in tobacco (Nicotiana tabacum) due to an artificial chromatin domain (Mlynarova et al., 2002). Although the tendency for neighboring genes to be coexpressed is well documented in Arabidopsis, little is known about this phenomenon in other plant species.

In this study, bioinformatic analysis was performed to identify divergent and convergent gene pairs, using the three completely sequenced plant genomes, rice, Arabidopsis, and Populus. Coexpression of gene pairs was determined based upon Pearson correlation coefficients calculated using massively parallel signature sequencing (MPSS) and microarray expression data. Gene pair conservation of each species’ divergent and convergent genes with the whole genome sequences of the other two species was determined using BLASTP and TBLASTN. Furthermore, the effect of intergenic distance on the likelihood of both genes in a pair to be expressed (as evidenced by MPSS and/or microarray data) was investigated. Subsequently, GO classification of these gene pairs was used to identify over- and underrepresented classes. Finally, we identified regulatory elements overrepresented in the intergenic regions of gene pairs whose expression levels are strongly correlated to determine the basis of the observed coexpression.

RESULTS

Differential Variation in Divergent and Convergent Gene Numbers with Intergenic Distances in Rice, Arabidopsis, and Populus

Rice, Arabidopsis, and Populus gene annotation data were analyzed for pairs of adjacent genes arranged divergently (→→) and convergently (←←). Release 4 of The Institute for Genomic Research (TIGR) rice (subsp. japonica) pseudomolecules contains a total of 56,563 annotated genes. Discarding hypothetical or transposon-related genes leaves 28,287 genes for further analysis. Among these, a total of 8,742 divergent and 8,772 convergent gene pairs were identified. Only in a minority of these pairs are the two genes separated by a short distance, with approximately one-seventh of divergent pairs and one-third of convergent pairs having 1 kb or less between them (Table I).

In Arabidopsis, the analysis was performed on 24,019 genes after filtering out hypothetical and transposon-related genes from 30,001 annotated genes. A total of 5,763 divergent gene pairs were identified, of which about 36% are separated by 1 kb or less. Among the 4,949 convergent pairs discovered, 71% were separated by less than 1 kb. Version 1.1 of the Joint Genome Institute (JGI) annotation of the Populus genome lists 45,554 genes. This dataset was not filtered for hypothetical or transposon-related genes, as no predicted functions were given. In all, 8,823 divergent gene pairs were identified, accounting for 39% of the genome. Of these, 613 pairs (7%) were separated by less than 1 kb. A total of 8,967 convergent gene pairs were identified, of which 2,212 (25%) were separated by less than 1 kb. These results show a similar trend in the decrease in the fraction of divergent genes with decreasing intergenic distance from <1 kb to <250 bp in all three species. However, Populus showed significant decrease in the fraction of convergent genes when compared to rice and Arabidopsis when the intergenic distance was decreased from <1 kb to <250 bp. Similarly, rice showed a significantly smaller decrease in the fraction of convergent genes compared to Arabidopsis and Populus. Furthermore, convergent genes were found to be 2- to 4-fold higher compared to divergent genes separated by <500 bp in all three plant genomes.

An interesting observation was made when the results for the three species were compared. The fraction of gene pairs separated by a small distance (<1 kb) appears to be proportional to genome size. In Arabidopsis, with a 115-Mb genome, more than one-third of divergent pairs are separated by <1 kb, compared to rice, with a 450-Mb genome, where only one-seventh of divergent gene pairs show the same pattern, despite there being only 14% more genes under consideration in rice. This trend is even more pronounced in Populus, where about 7% of divergent gene pairs are separated by <1 kb, almost one-half as many as in rice, despite having far more genes (45,554 versus 28,287) and a substantially larger genome (550 Mb). Similar observations were made when comparisons involved convergent gene pairs. This relation-

|         | Divergent |          |          | Convergent |          |          |
|---------|-----------|----------|----------|------------|----------|----------|
|         | Total <1 kb | <500 bp | 250 bp   | Total <1 kb | <500 bp | <250 bp   |
|         | No. | %   | No. | %   | No. | %   | No. | %   | No. | %   | No. | %   |
| Rice    | 8,742 | 14.2 | 532 | 6.1 | 212 | 2.4 | 8,772 | 34.9 | 1,626 | 18.5 | 734 | 8.4 |
| Arabidopsis | 5,763 | 36.5 | 1,145 | 19.9 | 462 | 8.0 | 4,949 | 71.3 | 2,650 | 53.5 | 1,758 | 35.5 |
| Populus | 8,823 | 6.9 | 276 | 3.1 | 141 | 1.6 | 8,967 | 24.7 | 824 | 9.2 | 320 | 3.6 |

Table I. Divergent and convergent gene pairs identified in rice, Arabidopsis, and Populus

Total represents number of pairs among all genes analyzed. <1 kb represents number of gene pairs with <1,000 bp between transcription start sites (divergent pairs) or transcription stop sites (convergent pairs). <500 bp and <250 bp columns are similar.
Expression and Coexpression Patterns of Divergent and Convergent Genes in Rice, Arabidopsis, and Populus

Several types of expression data were compiled for divergent and convergent gene pairs. Our analysis had both qualitative and quantitative aspects. The qualitative analysis confirmed that both genes were in fact expressed and not annotation artifacts. The goal of the quantitative analysis was to determine which gene pairs showed correlated expression levels across multiple tissues and treatments.

Divergent and convergent gene sequences were aligned with EST and full-length cDNA (fl-cDNA) sequences using BLASTN. In rice, the fraction of gene pairs for which matching EST/fl-cDNA sequences were found for both genes increases with decreasing intergenic distance (Fig. 1, A and B). This trend is more pronounced in the case of divergent gene pairs. Both the strong negative correlation between intergenic distance and EST/fl-cDNA matches seen in rice and the weak correlation in Arabidopsis are nonexistent in Populus. This phenomenon may be due to differences in regulatory mechanisms or the availability of fewer EST/fl-cDNA sequences for Populus compared to rice and Arabidopsis.

Analysis of MPSS and microarray data revealed that the fraction of divergent and convergent pairs with expression data for both genes increases significantly with reduced intergenic distance in rice (Tables II and III). Interestingly, a pronounced increase in the fraction (32%–65% for MPSS and 53%–81% for microarray data) of divergent pairs with expression data for both genes was observed compared to a modest increase in the fraction (43%–58% for MPSS and 69%–79% for microarray data) of convergent genes in rice when the intergenic distance was reduced from 1 kb to 250 bp. A similar trend is seen in Arabidopsis, although the increases observed are not as pronounced as in rice and are only statistically significant for MPSS data. For Populus, microarray expression data coverage actually decreases somewhat for divergent pairs and increases only slightly for convergent pairs. Altogether, there was a significant increase in the fraction of rice divergent and convergent gene pairs with fl-cDNA/EST, MPSS, or microarray expression data with decreasing intergenic distance. This trend was not seen in the other two genomes except in the case of Arabidopsis gene pairs with MPSS data.

Correlated expression of genes in divergent and convergent pairs was examined based on the Pearson correlation of their MPSS expression levels. Gene pairs with correlation coefficients $>0.5$ were considered to be significantly coexpressed, while those with coefficients $<-0.5$ were considered antiregulated (i.e. expression of one gene precludes expression of the other). Strong positive correlation was observed in approximately 2% of rice divergent and convergent pairs. In Arabidopsis, 12% of divergent pairs and 10% of convergent pairs showed strong positive correlation, while $<1\%$ of either pair type was antiregulated. No statistically significant connection between intergenic distance and frequency of correlated expression was noted in either species. The mean Pearson correlation of all rice divergent pairs for which MPSS data were available was 0.112, and the same figure for convergent pairs was 0.108. The mean correlation for 3,000 randomly paired rice genes was found to be 0.013, far lower than that of divergent or convergent pairs. In Arabidopsis, the average correlation for divergent pairs was 0.247 and 0.235 for convergent pairs. The mean correlation of 3,000 random gene pairs was 0.098, again significantly lower than the average correlation of divergent and convergent gene pairs. These data support the hypothesis that genes in divergent or convergent arrangement are more likely to be coexpressed than random pairs of genes.

The Pearson correlation of divergent and convergent gene pairs was also calculated using the mean

![Figure 1](https://www.plantphysiol.org)
intensity levels of each gene's corresponding probes across multiple microarray hybridizations. In rice, 26% of divergent pairs with microarray data for both genes had correlation coefficients >0.5 compared to 10% and 49%, respectively, of Arabidopsis and Populus pairs (Table III). Similar results were found for convergent gene pairs, with 26%, 9%, and 48% of rice, Arabidopsis, and Populus pairs, respectively, showing high levels of correlation. While slight increases in the fraction of pairs showing strongly correlated expression can be seen as intergenic distance decreases, these changes are not statistically significant.

A great deal of variation between species can be noted with regard to the frequency with which gene pairs are strongly correlated. About one-half of Populus divergent and convergent gene pairs show strong correlation, compared to one-fourth of rice and one-tenth of Arabidopsis gene pairs. A partial list of strongly correlated gene pairs is given in Supplemental Tables S1 and S2. Strong negative correlation appears to be quite rare, with only 1.6% of all Arabidopsis divergent gene pairs and 1.5% of Populus divergent gene pairs having Pearson correlation coefficients <−0.5. Similar percentages of convergent gene pairs display strong negative correlation in Arabidopsis and Populus.

To determine the degree to which divergent and convergent arrangement affects coexpression, the mean correlation levels of divergent and convergent gene pairs were calculated and compared with that of a set of 8,000 randomly selected pairs of genes. The mean correlation of rice divergent gene pairs calculated with microarray expression data was 0.390 and for convergent pairs the same figure was 0.392. The mean correlation of the random set was 0.103, approximately one-quarter that of either type of gene pair. In Arabidopsis, the mean correlation for divergent pairs was 0.163 and 0.144 for convergent pairs. Both values are significantly higher than the mean correlation for the Arabidopsis random pairs, which was calculated to be 0.044. This pattern was repeated in Populus, where the mean correlation for divergent and convergent gene pairs was 0.486 and 0.481, respectively, compared to 0.155 for the set of random gene pairs. These results indicate that in all three species, divergent and convergent gene pairs display significantly higher levels of correlated expression than randomly paired genes. For all three organisms, the mean correlation of both pair types is about 3- to 4-fold higher than that of the random sets, suggesting that the degree to which divergent and convergent arrangement affects coexpression of neighboring genes compared to random sets does not vary greatly among these species. Interestingly, the mean Pearson correlation for the expression of divergent or convergent genes are about 3 and 2.5 times higher in Populus and rice, respectively, compared to Arabidopsis. While it is possible that this may reflect biological differences between the species, it is more likely an artifact of the variation in the number of microarray hybridizations analyzed for each species (2,829 in Arabidopsis, 446 in rice, and 150 in Populus). Because the Pearson correlation was calculated using paired data points from each gene across all hybridizations, a larger number of hybridizations would lower the probability of obtaining a high correlation coefficient.

### Differential Interspecies Conservation of Divergent and Convergent Gene Arrangement

Conserved divergent or convergent arrangement of genes across species separated by vast evolutionary distances suggests conserved functional interaction between the proteins encoded by the genes in a pair. Six sets of BLASTP and TBLASTN searches were performed, aligning divergent and convergent genes from rice, Arabidopsis, and Populus with the genomes of the other two species. Conserved rice divergent gene pairs were found to be rare in both Arabidopsis and Populus, with only 26 pairs conserved in Arabidopsis.

| Divergent | Convergent |
|-----------|------------|
| Total | Total | | Total | Total |
| No. | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % |
| Rice | | | | | | | | | | | | | |
| Pairs with MPSS data | 2,483 | 28.4 | 401 | 12.3 | 3.0 | 242 | 45.5 | 6.5 | 138 | 65.1 | 5.7 | 2,568 | 29.3 | 1311 | 42.8 | 16.5 | 813 | 50.0 | 5.8 | 429 | 58.4 | 4.6 |
| R > 0.5 | 46 | 1.9 | 12 | 3.0 | 1.7 | 8 | 3.3 | 0.3 | 3 | 2.2 | −0.7 | 50 | 1.9 | 20 | 1.5 | −1.1 | 9 | 1.1 | −1.0 | 5 | 1.2 | 0.1 |
| R < −0.5 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| Arabidopsis | | | | | | | | | | | | | |
| Pairs with MPSS data | 3,406 | 59.1 | 1,292 | 61.3 | 2.1 | 773 | 67.5 | 4.3 | 350 | 75.8 | 3.8 | 2,880 | 58.2 | 2,517 | 61.1 | 3.5 | 1,736 | 65.5 | 4.6 | 1,226 | 69.7 | 3.7 |
| R > 0.5 | 410 | 12.0 | 133 | 10.3 | −1.9 | 78 | 10.1 | −0.2 | 0 | 0.0 | −0.9 | 298 | 10.1 | 211 | 9.8 | −0.9 | 164 | 9.4 | −0.5 | 117 | 9.5 | 0.1 |
| R < −0.5 | 9 | 0.3 | 5 | 0.4 | 0.9 | 4 | 0.5 | 0.6 | 3 | 0.9 | 0.9 | 11 | 0.4 | 10 | 0.5 | 0.6 | 9 | 0.5 | 0.3 | 7 | 0.6 | 0.3 |
and 77 in Populus (Table IV). For convergent pairs, conservation levels were found to be slightly higher at 42 pairs in Arabidopsis and 111 pairs in Populus. Examining only those pairs with short intergenic distances showed slight increase in conserved divergent pairs, while the fraction of conserved convergent pairs nearly doubled (0.8% in Arabidopsis and 2.6% in Populus) when intergenic distance is <250 bp. The frequency of Arabidopsis gene pair conservation varied greatly in rice and Populus. In rice, only 29 divergent gene pairs were found to be conserved compared to 52 convergent pairs. Arabidopsis gene pairs conserved in Populus were found to be far more common, with 355 divergent pairs and 401 convergent pairs having conserved gene order and orientation. Comparison of Populus gene pairs with rice and Arabidopsis identified 58 and 267 divergent pairs conserved in rice and Arabidopsis, respectively. Among Populus convergent gene pairs, 114 were conserved in rice, while 421 were conserved in Arabidopsis. In each of these comparisons, the number and fraction of conserved convergent gene pairs were higher than those of conserved divergent pairs. These results suggest that the exact spatial arrangement of the gene pair is a necessary regulatory factor in only a small fraction of all such pairs.

Divergent and convergent gene pairs were found to be conserved in some species more frequently when both genes shared one or more GO terms. Rice divergent and convergent gene pairs with shared GO terms were found to be more likely to be conserved in Arabidopsis or Populus compared to all divergent and convergent pairs (Table V). The fraction of Populus gene pairs with shared GO terms organized in a divergent manner in rice and convergent manner in Arabidopsis increased significantly compared to all gene pairs. This trend was not observed in Arabidopsis gene pairs conserved in the other two plant genomes. These results suggest that divergent and convergent genes with shared GO are more likely to be conserved compared to all conserved gene pairs.

Strongly correlated expression also raises the probability of a gene pair being conserved. While the increases in conservation frequency are seldom as great as those caused by shared GO terms, they are nonetheless quite significant, especially in the case of rice divergent gene pairs conserved in Arabidopsis or Populus, where up to 3-fold increases were observed (Table V). Similarly, a 2-fold change was observed on the conservation of rice convergent genes with correlated expression. This trend of few fold changes was not observed in other comparisons. These data indicate that the effect of strongly correlated expression on the conservation of divergent and convergent genes varies based on the organisms being examined. A partial list of conserved gene pairs, and conserved gene pairs displaying correlated expression, is given in Supplemental Tables S3 to S6.

### GO Classification of Divergent and Convergent Genes

GO classification data were downloaded for all genes included in our analysis. While at least one GO classification was found for about 99% of Arabidopsis divergent and convergent genes, in rice, only 41% of divergent genes and 45% of convergent genes had similar data available due to the ongoing process of GO classification of the rice genome. A similar situation exists in Populus, where approximately 45%
of both divergent and convergent genes have at least one GO classification. The full GO vocabulary was used in classifying the *Populus* genes, while the Plant GOslim vocabulary was used for rice and Arabidopsis.

Two analyses were performed on those pairs for which GO classification data were found for both genes. The first of these was a search for pairs in which both genes were grouped into the same GO class, as shared or related function could be a contributing factor in the coordinated expression of neighboring genes. Approximately 4.9% of rice divergent genes separated by \(<1 \text{ kb}\) have at least one GO class in common, and this percentage increased to 7.5% for genes separated by \(<250 \text{ bp}\). Among rice convergent genes, this percentage rose from 11.3% for genes separated by \(<1 \text{ kb}\) to 15% for genes separated by \(<250 \text{ bp}\). A similar pattern is seen in *Populus*, but the effect of decreased intergenic distance is much weaker, with the fraction of pairs with shared GO classes increasing only from 2.9% to 4.3% among divergent pairs and from 2.1% to 2.2% among convergent pairs.

In Arabidopsis, the fraction of pairs with common GO classifications remained nearly constant at about 45% across different intergenic distances for both divergent and convergent genes. These results suggest that the likelihood of the genes in a pair sharing the same GO class increases greatly in rice and to a lesser degree in *Populus*, but not in Arabidopsis, if the genes are physically closer to each other.

The second analysis sought out GO classes that were disproportionately represented among divergent or convergent genes relative to the whole genome. Over- or underrepresentation was determined using the binomial test (normal approximation, \(P < 0.0001\)). In rice, two GO classes were found to be significantly underrepresented among divergent genes. Genes whose protein products are involved in secondary metabolic and biosynthetic processes are underrepresented in rice divergent pairs. Among others, several cytochrome P450 and glycosyl hydrolase family proteins are part of GO class of secondary metabolic process. No overrepresented GO classes were identified among rice divergent or convergent genes. Several over- and underrepresented GO classes were found in Arabidopsis and *Populus* divergent and convergent gene pairs. GO class nucleic acid binding, which includes

| Gene pair conservation | Total | \(<1 \text{ kb}\) | \(<500 \text{ bp}\) | \(<250 \text{ bp}\) | Total | \(<1 \text{ kb}\) | \(<500 \text{ bp}\) | \(<250 \text{ bp}\) |
|------------------------|-------|----------------|----------------|----------------|-------|----------------|----------------|----------------|
| Rice                   |       |                |                |                | Arabidopsis |                |                |                |
| Conserved in Arabidopsis | 26    | 0.3            | 4              | 0.3            | 1       | 0.4            | 1              | 0.5            |
| Conserved in *Populus*  | 77    | 0.9            | 13             | 1.0            | 5       | 0.9            | 2              | 0.9            |
| Arabidopsis            |       |                |                |                | *Populus* |                |                |                |
| Conserved in rice       | 29    | 0.5            | 5              | 0.2            | 1       | 0.1            | 1              | 0.2            |
| Conserved in *Populus*  | 355   | 6.2            | 125            | 5.9            | 78      | 6.8            | 24             | 5.2            |
| *Populus*              |       |                |                |                |         |                |                |                |
| Conserved in rice       | 58    | 0.7            | 7              | 1.1            | 4       | 1.4            | 1              | 0.7            |
| Conserved in Arabidopsis| 267   | 3.0            | 17             | 2.8            | 8       | 2.9            | 5              | 3.5            |

| Table IV. Divergent and convergent gene pairs conserved in other species |
|------------------------|-------|----------------|----------------|----------------|-------|----------------|----------------|----------------|
| Gene pair conservation | Total | \(<1 \text{ kb}\) | \(<500 \text{ bp}\) | \(<250 \text{ bp}\) | Total | \(<1 \text{ kb}\) | \(<500 \text{ bp}\) | \(<250 \text{ bp}\) |
| Rice                   |       |                |                |                | Arabidopsis |                |                |                |
| Conserved in Arabidopsis | 26    | 0.3            | 4              | 0.3            | 1       | 0.4            | 1              | 0.5            |
| Conserved in *Populus*  | 77    | 0.9            | 13             | 1.0            | 5       | 0.9            | 2              | 0.9            |
| Arabidopsis            |       |                |                |                | *Populus* |                |                |                |
| Conserved in rice       | 29    | 0.5            | 5              | 0.2            | 1       | 0.1            | 1              | 0.2            |
| Conserved in *Populus*  | 355   | 6.2            | 125            | 5.9            | 78      | 6.8            | 24             | 5.2            |
| *Populus*              |       |                |                |                |         |                |                |                |
| Conserved in rice       | 58    | 0.7            | 7              | 1.1            | 4       | 1.4            | 1              | 0.7            |
| Conserved in Arabidopsis| 267   | 3.0            | 17             | 2.8            | 8       | 2.9            | 5              | 3.5            |

| Table V. Effect of shared GO terms and high correlation on gene pair conservation |
|------------------------|-------|----------------|----------------|----------------|-------|----------------|----------------|----------------|
| Gene pair conservation | Total | \(<1 \text{ kb}\) | \(<500 \text{ bp}\) | \(<250 \text{ bp}\) | Total | \(<1 \text{ kb}\) | \(<500 \text{ bp}\) | \(<250 \text{ bp}\) |
| Rice                   |       |                |                |                | Arabidopsis |                |                |                |
| Conserved in Arabidopsis | 26    | 0.3            | 4              | 0.3            | 1       | 0.4            | 1              | 0.5            |
| Conserved in *Populus*  | 77    | 0.9            | 13             | 1.0            | 5       | 0.9            | 2              | 0.9            |
| Arabidopsis            |       |                |                |                | *Populus* |                |                |                |
| Conserved in rice       | 29    | 0.5            | 5              | 0.2            | 1       | 0.1            | 1              | 0.2            |
| Conserved in *Populus*  | 355   | 6.2            | 125            | 5.9            | 78      | 6.8            | 24             | 5.2            |
| *Populus*              |       |                |                |                |         |                |                |                |
| Conserved in rice       | 58    | 0.7            | 7              | 1.1            | 4       | 1.4            | 1              | 0.7            |
| Conserved in Arabidopsis| 267   | 3.0            | 17             | 2.8            | 8       | 2.9            | 5              | 3.5            |
zinc finger family proteins and translation initiation factors, was found to be underrepresented in both divergent and convergent gene pairs in Arabidopsis. However, genes belonging to the same GO class are overrepresented in *Populus* divergent gene pairs. GO class signal transduction is overrepresented in Arabidopsis divergent genes, which includes several Leu-rich repeat family proteins and ethylene-responsive factors. Interestingly, GO classes apoptosis, defense response, and transmembrane receptor activity were underrepresented in both divergent and convergent genes of *Populus*. Gene pairs overrepresented in specific GO classes suggest that they are more likely to be organized in a divergent or convergent manner. Similarly, underrepresented GO classes suggest that genes belonging to these classes do not tend to be organized in a specific orientation (divergent or convergent). Although the reason for this bias is not known, it is possible that functional relationships exist among these genes. A full listing of these classes can be found in Supplemental Table S7, and the number of each species’ correlated or conserved pairs in each GO class is given in Supplemental Tables S8, S9, and S10.

Regulatory Elements Overrepresented in Intergenic Regions of Divergent Genes with Correlated Expression

The intergenic regions of all divergent and convergent pairs separated by 1 kb or less were analyzed for known regulatory elements using the Plant Cis-acting Regulatory DNA Element (PLACE) database (Higo et al., 1999; http://www.dna.affrc.go.jp/PLACE/index.html). In addition, 1-kb regions upstream of convergent genes were examined for the presence of regulatory elements. For each species and pair type, the gene pairs were divided into two subsets: those displaying strongly correlated expression and those with weak or no correlation. The fractions of sequences in these two sets containing each element were then compared, and their differences were tested for statistical significance using the binomial test (*P* < 0.0001).

In Arabidopsis and rice divergent gene pairs, several elements were found to be overrepresented among strongly correlated pairs (Table VI). This differs significantly from the results obtained for convergent pairs, where none of the elements were found to be overrepresented. These results suggest that correlated expression in divergent gene pairs is at least in part caused by the presence of specific regulatory elements in the intergenic region, where they can influence the expression of both genes in the pair. While similar numbers of elements were found in the intergenic regions of divergent and convergent pairs, we found no significant difference in the elements found between correlated and noncorrelated convergent pairs. Therefore, it seems likely that correlated expression due to shared regulatory elements is a feature only of divergent gene pairs. A complete list of all overrepresented regulatory elements identified can be found in Supplemental Table S11.

The results for *Populus* were quite different from those for rice and Arabidopsis. Although many elements were identified in the *Populus* sequences, very few showed any significant difference in frequency between correlated and noncorrelated pairs. This is most likely a reflection of the composition of the PLACE database, which contains regulatory elements gleaned from recent publications. As rice and Arabidopsis have been more thoroughly studied than *Populus*, there may be many regulatory elements in the *Populus* genome involved in correlated expression, as we hypothesized there to be in rice and Arabidopsis, that are not represented in the PLACE database.

Three of the regulatory elements identified were overrepresented in the intergenic regions of coexpressed divergent gene pairs in both rice and Arabidopsis. These three elements are CGACG element required for the expression of rice α-amylase Amy3D gene (Hwang et al., 1998), E2F consensus sequence recognized by E2F transcription factors and present in promoters of target genes that regulate cell cycle, DNA replication, DNA repair, and chromatin structure (Vandepeoele et al., 2005), and a sulfur-responsive element core sequence present in the promoter region of a sulfate transporter gene of Arabidopsis (Maruyama-Nakashita et al., 2005). This last element overlaps largely with an auxin response element. Furthermore, one overrepresented element, PRECONSCRHSP70A, is shared by *Populus* and Arabidopsis promoters flanked by coexpressed divergent genes. This is the consensus sequence of a plastid response element found in the promoter of nuclear gene *HSP70A* in *Chlamydomonas* and induced by a chlorophyll precursor, Mg-protoporphyrin, and light (Von Gromoff et al., 2006). Furthermore, the most overrepresented element

| Table VI. Regulatory elements overrepresented in intergenic regions of correlated pairs |
|---------------------------------|----------------|----------------|
| Divergent | No. of Elements | Overrepresented | Convergent | No. of Elements | Overrepresented |
|----------|----------------|----------------|------------|----------------|----------------|
| Arabidopsis | 286 | 39 | Arabidopsis | 248 | 0 |
| Rice | 283 | 16 | Rice | 282 | 0 |
| Populus | 262 | 1 | Populus | 288 | 0 |
in promoters of Arabidopsis-correlated divergent gene pairs is UP2 motif, which is found upstream of genes up-regulated on main stem decapitation (Tatematsu et al., 2005). GCC-box core found in many pathogen-responsive genes (Brown et al., 2003) was the most overrepresented element in rice promoters with correlated divergent gene pairs. The occurrence of these elements between strongly correlated genes suggests that they play a role in regulating both genes in the pair, with either the elements being shared as part of a single bidirectional promoter or having a similar set of regulatory elements present in each gene’s separate promoter.

DISCUSSION

With the recent completion of several plant genomes and the availability of genome-wide quantitative expression data, it is possible to unravel many of the unexplained aspects of the inner workings of complex organisms. Here, we investigated the organization of convergent and divergent genes in three plant genomes, their expression patterns, and the degree of coexpression exhibited by them. Our study not only identified similar patterns with respect to the organization of divergent and convergent genes with decreasing intergenic distance in the three plant genomes but also in cases where a pattern is unique to only one of the three plant genomes. It is very likely that some of these divergent trends are linked to the biology of the specific organism. This is further illustrated by over- and/or underrepresented GO classes either shared by divergent and convergent genes or unique to them in one or more species, which is related to their function.

In rice, it was observed that the fraction of divergent and convergent gene pairs for which expression data exists for both genes increases as the distance between the two genes decreases. However, this phenomenon was not observed in either Arabidopsis or Populus, which may be caused by biological differences between monocots and dicots. This needs to be confirmed by the study of several other monocot and dicot genomes. Some of these differences can also be attributed to the source of the expression data based on different results obtained for Arabidopsis with the three data sets of fl-cDNA/EST, MPSS, and microarray.

Our comparative analysis identified a number of divergent and convergent gene pairs in rice, Arabidopsis, and Populus that possess homologs in the same orientation in at least one other species. The fraction of conserved gene pairs ranges from 0.3% to 8.1% across species and pair types, which is in accord with the results of earlier studies. Seoighe and colleagues (2000), performing a similar analysis on two yeast species, found that only 9% of yeast gene pairs remained adjacent, and of those 65% maintained the same orientations, leaving only 5.85% of all gene pairs conserved with regard to both gene order and relative orientation. Comparisons between rice and Arabidopsis (Liu et al., 2001) identified a rate of 5.5% for conservation of gene pair order and a probability of only 0.005 for the pair to be conserved without additional genes being inserted between them. Ren and colleagues (2007) found no local coexpression domains in rice that were conserved in Arabidopsis. However, their criteria for coexpression ($R > 0.7$) were different than those used here, and a coexpression domain was considered conserved only if the homologous domain was also coexpressed. Therefore, there may be coexpressed divergent or convergent gene pairs that are conserved in our study but not in their study. Together with our findings, these results indicate that exact conservation of gene pair order and orientation between species is quite rare. This rarity, however, would seem to imply that when divergent or convergent arrangement is conserved, there is likely to be some regulatory aspect to that arrangement necessary for proper gene function, such as bidirectional promoters or enhancer sequences in the pair’s intergenic region. Bidirectional promoters have been identified and characterized in relatively large numbers in the human genome (Adachi and Lieber, 2002; Trinklein et al., 2004; Lin et al., 2007) yet have received little attention in plants. The overrepresentation of some regulatory elements in the intergenic regions of strongly correlated divergent gene pairs supports the hypothesis that shared elements are responsible, at least in part, for the coordinated expression observed in many divergent pairs. These elements may have novel mechanisms for regulating these genes. This explanation, however, does not apply to convergent gene pairs, despite the similar frequency of correlated expression observed among them. Other factors such as local chromatin organization may be responsible for the coexpression observed, especially in the case of convergent gene pairs. This study provides a foundation for more detailed studies of the regulatory elements involved in coordinating the expression of divergent and convergent gene pairs.

Two factors have been identified that affect the probability of a divergent or convergent gene pair being conserved in other species. Gene pairs that have one or more GO classifications in common are more likely to be conserved in another species. The second factor that increases the likelihood of a gene pair being conserved is strong coexpression. This association is most likely due to a shared or similar function of the genes in a pair.

The functional basis for the high level of coexpression observed in many divergent and convergent gene pairs can take on myriad forms. One of the most straightforward is involvement in the same biological process, a situation observed in numerous gene pairs based on the frequency of shared GO classifications. One such divergent gene pair found in rice consists of a phospho-2-dehydro-3-deoxyheptonate aldolase 1 and a cytokinin-O-glucosyltransferase 2 gene. Both genes are in the GO class “amino acid and derivative metabolic process,” and the pair is strongly correlated ($R = 0.62$).
and conserved in Arabidopsis. Another cause of gene pair coexpression is shared regulatory elements, which would induce the expression of both genes in response to a single stimulus. An example of such a gene pair is found on chromosome 1 in Arabidopsis. The genes in this divergent gene pair code for two auxin-responsive/indoleacetic acid-induced proteins, IAA3 and IAA17, which display correlated expression levels ($R = 0.65$) and are conserved in rice. One commonly observed trend is shared or similar molecular functions among genes in a divergent or convergent pair. The rice convergent gene pair consisting of a Ser/Thr protein phosphatase PP2A catalytic subunit and a phosphatidic acid phosphatase family protein is an example of this. Both genes, in addition to being annotated as phosphatases, have the GO classification "hydrolase activity" and are very strongly correlated ($R = 0.78$). Convergent arrangement of this pair is conserved in both Arabidopsis and Populus, which, along with all other such indicators, suggests very compellingly that these genes have some type of functional relationship and that their convergent arrangement is an essential part of their regulation. A similar set of circumstances surrounds the rice divergent gene pair consisting of a sugar transporter family protein and a protein kinase domain containing protein. According to their GO classifications, the products of both genes are located in the nuclear membrane. The pair is conserved in Arabidopsis and has a Pearson correlation of 0.73, which suggests that a functional relationship exists between the two genes. No such relationship is indicated in the available data, so the data relating to this gene pair generated in this study could serve as inspiration for further study of this and other similar pairs.

CONCLUSION

We identified patterns of expression and coexpression patterns of divergent and convergent gene pairs in rice, Arabidopsis, and Populus. Strongly correlated expression was observed in significant numbers of gene pairs in all three species and at significantly higher levels than randomly paired genes. Cross-species conservation of divergent and convergent arrangement was found to be rare, although the frequency of conservation was significantly higher among pairs with strongly correlated expression or shared GO classifications. We identified several coexpressed gene pairs with shared GO terms suggesting functional correlation. Furthermore, we identified a few regulatory elements that may be involved in coordinating the expression of divergently arranged genes. In all, patterns of divergent and convergent gene pair coexpression and conservation were characterized, and several factors that influence these phenomena were identified, providing a foundation for more detailed study of the various mechanisms of regulating these genes.

MATERIALS AND METHODS

Identification of Divergent and Convergent Gene Pairs

Sequence and annotation data for the rice (Oryza sativa subsp. japonica ‘Nipponbare’ genome were downloaded from the Rice Genome Annotation Database at TIGR (http://www.tigr.org/tgi/rice/[OS1]). Similar data for the Arabidopsis (Arabidopsis thaliana) genome were obtained from The Arabidopsis Information Resource (TAIR) (ftp://ftp.arabidopsis.org/home/tair/Genes/TIGR5_genome_release) and JGI (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html) Web sites, respectively. A Perl script was used to parse this data and identify pairs of adjacent genes on opposite strands, designating genes arranged head-to-head as divergent pairs and those arranged end-to-end as convergent pairs. Pairs containing genes annotated as hypothetical or transposon related were excluded from all later analyses.

Analysis of Gene Pair Expression

EST data were downloaded for Arabidopsis (TAIR EST FTP site: ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_data_sets/), rice (Rice Full-length cDNA Consortium: http://cdna01.dna.affrc.go.jp/cDNA), and Populus (PopulusDB: http://poppel.fysbot.umu.se/proj_downl.php) and converted into BLAST databases. Genes in convergent and divergent pairs were aligned with the EST/B-cDNA data using BLASTN. Hits with at least 95% identity were deemed significant and used, along with other types of expression data, to determine if annotated genes were actually expressed or false positives from gene prediction. In addition, Arabidopsis EST and B-cDNA alignment data were downloaded from the Salk Institute Genomic Analysis Laboratory (http://signal.salk.edu/data) and were used to assign additional matches to Arabidopsis divergent and convergent genes.

MPSS (Meyers et al., 2004) data were collected for rice (http://mpss.udel.edu/rice/) and Arabidopsis (http://mpss.udel.edu/at/) genes. Only 17-bp signatures of classes 1, 2, 5, and 7 that mapped to a single gene were used, and abundance values $< 5$ were ignored as background interference. When multiple signatures had significant abundance values in the same library, the average abundance was used. Correlated expression between genes in divergent and convergent gene pairs was examined by calculating the Pearson correlation coefficient using each gene’s average abundance values across multiple libraries (17 in Arabidopsis, 72 in rice).

Microarray data for all three species were compiled from several sources (rice, Yale rice project [http://bioinformatics.med.yale.edu/rc/overview.jsp]; Arabidopsis, Nottingham Arabidopsis stock center microarray database [http://affymetrix.arabidopsis.info/narrays/help/usefulfiles.html] and Stanford Microarray Database [http://genome-www5.stanford.edu/]; and Populus, National Center for Biotechnology Information Gene Expression Omnibus [http://www.ncbi.nlm.nih.gov/geo/]). Expression data were collected for a total of 2,820 hybridizations in Arabidopsis, 446 hybridizations in rice, and 150 hybridizations in Populus. Both the rice and Arabidopsis datasets included mappings of microarray spots to gene locus identifiers, while probe sequences on the Populus oligo arrays were aligned with the coding region sequences of Populus divergent and convergent genes using BLASTN. Oligos which aligned uniquely with 100% identity were inferred to be associated with individual genes. Correlated expression was again tested with the Pearson correlation coefficient, this time pairing data points from the same hybridization and channel.

Conservation of Gene Pair Arrangement

The protein sequences of all genes in divergent and convergent pairs from each species (rice, Arabidopsis, and Populus) were aligned with the full set of protein sequences from the two remaining species using BLASTP (Altschul et al., 1997) to identify homologs. If a divergent or convergent gene pair possessed homologs in the same arrangement, then that gene pair was considered conserved.

In an attempt to identify more distantly related homologs, an additional set of alignments was performed, this time aligning the protein sequences of each species with the translated genomes of the other two using TBLASTN (Altschul et al., 1997). When both genes in the original pair had hits with e-values no greater than 1E-20 within 50 kb of each other, in the same orientation (divergent or convergent) as the original and with no other genes between them, then the pair was considered conserved.
GO Classification

GO classification data were downloaded for all rice, Arabidopsis, and Populus divergent and convergent genes (rice, TIGR Rice database; Arabidopsis, TAIR GO FTP site: ftp://ftp.arabidopsis.org/home/tair/Ontologies/ Gene_Ontology; Populus, JGI Poplar Database FTP site: ftp://ftp.jgi-psf.org/pub/JGI_data/Poplar/annotation/v1.1/functional). Rice and Arabidopsis genes were classified using the higher-level Plant COGid vocabulary, while only annotations using the full GO vocabulary were available for Populus. GO class assignments for genes in divergent and convergent pairs were compared to identify pairs in which both genes were in the same class. To identify GO classes in which divergently or conversely arranged genes appeared significantly more or less frequently than genes in that species did overall, we compared the number of genes in each group (e.g. rice divergent genes) using the binomial test. The test statistic Z was computed using the following formula:

\[ Z = \frac{(F_d - F_c)}{\sqrt{\frac{N_d \cdot (1 - \frac{d}{N_d})}{N_c}}} \]

where \( F_d \) is the fraction of divergent or convergent genes in the GO class, \( F_c \) is the fraction of all genes in that class, and \( N_d \) is the total number of divergent or convergent genes in that species. A GO class was considered significantly over- or underrepresented (\( P < 0.0001 \)) when \( |Z| > 3.719 \).

Regulatory Motif Analysis

Intergenic regions were compiled for all divergent and convergent gene pairs separated by 1 kb or less. These sequences were then scanned for known regulatory elements using the PLACE database (http://www.dna.affrc.go.jp/PLACE). For each element identified, we calculated the number of sequences in which it appeared. Elements represented in less than 30% of the intergenic regions of divergent and convergent genes were not considered for further analysis. We compared the frequency with which each element appeared in strongly correlated gene pairs with that of pairs showing little or no correlation. The normal approximation of the binomial test (cutoff value of 0.0001) was used to test for statistically significant differences in frequency of element occurrence between the two data sets.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Table S1. Coexpressed divergent genes separated by <250 bp with Pearson R > 0.5.

Supplemental Table S2. Coexpressed convergent genes separated by <250 bp with Pearson R > 0.5.

Supplemental Table S3. Divergent genes separated by <250 bp with conserved gene order and orientation.

Supplemental Table S4. Convergent genes separated by <250 bp with conserved gene order and orientation.

Supplemental Table S5. Conserved divergent gene pairs with high Pearson correlation R > 0.5.

Supplemental Table S6. Conserved convergent gene pairs with high Pearson correlation R > 0.5.

Supplemental Table S7. GO categories significantly under- or overrepresented in different gene pair classes.

Supplemental Table S8. Number of highly correlated or conserved rice genes in various GO classes.

Supplemental Table S9. Number of highly correlated or conserved Arabidopsis genes in various GO classes.

Supplemental Table S10. Number of highly correlated or conserved Populus genes in various GO classes.

Supplemental Table S11. Regulatory elements overrepresented in intergenic regions of correlated gene pairs versus noncorrelated pairs.

ACKNOWLEDGMENT

The authors thank Matthew McCormick for his invaluable assistance in the early stages of this project.

Received May 1, 2008; accepted May 27, 2008; published May 30, 2008.

LITERATURE CITED

Adachi N, Lieber MR (2002) Bidirectional gene organization: a common architectural feature of the human genome. Cell 109: 807–809

Altshul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796–815

Brown RL, Kazan K, McGrath KC, Maclean DJ, Manners JM (2003) A role for the GCC-box in jasmonate-mediated activation of the PDF1.2 gene of Arabidopsis. Plant Physiol 132: 1020–1032

Chaw SM, Chang CC, Chen HL, Li WH (2004) Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. J Mol Evol 58: 424–441

Chen N, Stein LD (2006) Conservation and functional significance of gene topology in the genome of Caenorhabditis elegans. Genome Res 16: 606–617

Cohen BA, Mitra RD, Hughes JD, Church GM (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet 26: 183–186

Devis KM, Beales J, Nagamura Y, Sasaki T (1999) Arabidopsis-rice: Will consilience allow gene prediction across the eudicot-monocot divide? Genome Res 9: 825–829

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res 27: 297–300

Hurst LD, Pal C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet 5: 299–310

Hwang YS, Karrer EE, Thomas BR, Chen L, Rodriguez RL (1998) Three cis-elements required for rice alpha-amylase Amy3D expression during sugar starvation. Plant Mol Biol 36: 331–341

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436: 793–800

Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakamichi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawaj j et al. (2005) Antisense transcription in the mammalian transcriptome. Science 309: 1564–1566

Kruglyak S, Tang H (2000) Regulation of adjacent yeast genes. Trends Genet 16: 109–111

Lin JM, Collins PJ, Trinklein ND, Fu Y, Xi H, Myers RM, Weng Z (2007) Transcription factor binding and modified histones in human bidirectional promoters. Genome Res 17: 818–827

Liu H, Sachidanandam R, Stein L (2001) Comparative genomics between rice and Arabidopsis shows scant collinearity in gene order. Genome Res 11: 2020–2026

Maruyama-Nakashita A, Nakamura Y, Watanabe-Takahashi A, Inoue E, Yamaya T, Takahashi H (2005) Identification of a novel cis-acting element conferring sulfur deficiency response in Arabidopsis roots. Plant J 42: 305–314

Meyers BC, Lee DK, Vu TH, Tej SS, Edberg SB, Matvienko M, Tindell LD (2004) Arabidopsis MPSS. An online resource for quantitative expression analysis. Plant Physiol 135: 801–813

Mlynarova L, Loonen A, Mietkiewska E, Jansen RC, Nap JP (2007) Assembly of two transgenes in an artificial chromatin domain gives highly coordinated expression in tobacco. Genetics 165: 727–740

Ren XY, Fiers M, Stiekena WJ, Nap JP (2005) Local coexpression domains of two to four genes in the genome of Arabidopsis. Plant Physiol 138: 923–934

Ren XY, Stiekena WJ, Nap JP (2007) Local coexpression domains in the genome of rice show no microsynteny with Arabidopsis domains. Plant Mol Biol 65: 205–217

Seonghe C, Federspiel N, Jones T, Hansen N, Bivolarovic V, Sureycki R, Tamse R, Komp C, Huizlar L, Davis RW, et al. (2000) Prevalence of small inversions in yeast gene order evolution. Proc Natl Acad Sci USA 97: 14433–14437

Tatematsu K, Ward S, Leyer S, Kamiya Y, Nambara E (2005) Identifica-
tion of cis-elements that regulate gene expression during initiation of axillary bud outgrowth in Arabidopsis. Plant Physiol 138: 757–766

Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP, Myers RM (2004) An abundance of bidirectional promoters in the human genome. Genome Res 14: 62–66

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313: 1596–1604

Vandepoele K, Saeys Y, Simillion C, Raes J, Van de Peer Y (2002) The automatic detection of homologous regions (ADHoRe) and its application to microcolinearity between Arabidopsis and rice. Genome Res 12: 1792–1801

Vandepoele K, Vlieghe K, Florquin K, Hennig L, Beemster GT, Gruissem W, Van de Peer Y, Inze D, De Veylder L (2005) Genome-wide identification of potential plant E2F target genes. Plant Physiol 139: 316–328

Von Gromoff ED, Schroda M, Oster U, Beck CF (2006) Identification of a plastid response element that acts as an enhancer within the Chlamydomonas HSP70A promoter. Nucleic Acids Res 34: 4767–4779

Williams EJG, Bowles DJ (2004) Coexpression of neighboring genes in the genome of Arabidopsis thaliana. Genome Res 14: 1060–1067

Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA 86: 6201–6205

Zhan S, Horrocks J, Lukens LN (2006) Islands of co-expressed neighbouring genes in Arabidopsis thaliana suggest higher-order chromosome domains. Plant J 45: 347–357