

Local coexpression domains in the genome of rice show no microsynteny with Arabidopsis domains

Xin-Ying Ren · Willem J. Stiekema · Jan-Peter Nap

Received: 13 April 2007 / Accepted: 6 July 2007 / Published online: 20 July 2007
© Springer Science+Business Media B.V. 2007

Abstract Chromosomal coexpression domains are found in a number of different genomes under various developmental conditions. The size of these domains and the number of genes they contain vary. Here, we define local coexpression domains as adjacent genes where all possible pair-wise correlations of expression data are higher than 0.7. In rice, such local coexpression domains range from predominantly two genes, up to 4, and make up ~5% of the genomic neighboring genes, when examining different expression platforms from the public domain. The genes in local coexpression domains do not fall in the same ontology category significantly more than neighboring genes that are not coexpressed. Duplication, orientation or the distance between the genes does not solely explain coexpression. The regulation of coexpression is therefore thought to be regulated at the level of chromatin structure. The characteristics of the local coexpression domains in rice are strikingly similar to such domains in the Arabidopsis genome. Yet, no microsynteny between local coex-

pression domains in Arabidopsis and rice could be identified. Although the rice genome is not yet as extensively annotated as the Arabidopsis genome, the lack of conservation of local coexpression domains may indicate that such domains have not played a major role in the evolution of genome structure or in genome conservation.

Keywords Arabidopsis · Coexpression domain · Higher-order chromatin structures · Microsynteny · Rice

Abbreviations

BLAST	Basic local alignment search tool
GO	Gene ontology
MA	Microarray expression data
MPSS	Massively parallel signature sequencing expression data
mya	Million years ago
PSCD	Partially syntenic coexpression domain
PSND	Partially syntenic non-coexpressed domain
TAIR	The Arabidopsis information resource
TIGR	The Institute of Genomic Research

Electronic supplementary material The online version of this article (doi:10.1007/s11103-007-9209-0) contains supplementary material, which is available to authorized users.

X.-Y. Ren · J.-P. Nap (✉)
Applied Bioinformatics, Plant Research International, Plant Sciences Group, Wageningen University and Research Centre, P.O. Box 16, 6700 AA Wageningen, The Netherlands
e-mail: janpeter.nap@wur.nl

X.-Y. Ren · W. J. Stiekema
Laboratory of Bioinformatics, Plant Sciences Group, Wageningen University and Research Centre, 6703 HA Wageningen, The Netherlands

W. J. Stiekema · J.-P. Nap
Centre for BioSystems Genomics, 6700 AB Wageningen, The Netherlands

Introduction

The fast-growing data sets on genome annotation and genome-wide gene expression facilitate the study and comparison of gene activity between and among genomes. The genomic context of genes is supposed to play an important role in the regulation of gene expression (van Drunen et al. 1997). Non-random clusters of similarly expressed (co-regulated, coexpressed, highly expressed and/or broadly expressed) genes have been described in

almost all organisms, ranging from prokaryotes to eukaryotes. In eukaryotes, from yeast to Arabidopsis to human, both short-range co-regulated/coexpressed clusters of two to five genes (Cohen et al. 2000; Ren et al. 2005; Zhan et al. 2006) and longer-range coexpression domains of up to 30 genes spanning up to 100 kb and more (Lercher et al. 2003; Ma et al. 2005; Spellman and Rubin 2002; Zhan et al. 2006) have been described. Duplicated genes (Lercher et al. 2003), shared promoter regions (Kruglyak and Tang 2000), shorter gene distance (Cohen et al. 2000; Roy et al. 2002; Semon and Duret 2006; Williams and Bowles 2004) and/or functional relatedness (Cohen et al. 2000; Lee et al. 2004; Spellman and Rubin 2002; Williams and Bowles 2004) were found to account for only part of the coexpression between genes. Therefore, most studies postulate that the occurrence of coexpression domains, small or large, is regulated on the level of higher-order chromatin structure (Cohen et al. 2000; Hershberg et al. 2005; Ren et al. 2005; Spellman and Rubin 2002; Williams and Bowles 2004), although alternative views exist (Semon and Duret 2006).

Previously we have defined and demonstrated the existence of local coexpression domains in the genome of Arabidopsis (Ren et al. 2005). A local coexpression domain was defined as any set of physically adjacent genes that are highly coexpressed with a pair-wise Pearson's correlation coefficient larger than 0.7. It was shown that a small (5–10%) yet significant fraction of genes in the Arabidopsis genome is organized in such local coexpression domains. Genes in such local domains were for the major part not categorized in the same functional category (GOslim). Neither tandemly duplicated genes nor shared promoter sequence, nor gene distance, explained the occurrence of coexpression of genes in such chromosomal domains. This indicates that other parameters in genes or gene positions are important to establish coexpression in local domains of Arabidopsis chromosomes. Here it is analyzed whether a similar situation exists in the genome of the monocotyledonous model plant rice (*Oryza sativa*), for which an earlier study had shown the existence of longer-range domains (Ma et al. 2005). We combined the whole genome rice annotation data (TIGR version 3; www.tigr.com) with Massively Parallel Signature Sequencing (MPSS; mpss.udel.edu/rice) expression data as well as Affymetrix array expression data (GEO, GSE4438; www.ncbi.nlm.nih.gov/geo) for the rice cultivar japonica in a way similar to the analysis performed for the Arabidopsis genome (Ren et al. 2005). The results show that the characteristics of the two genomes with respect to the occurrence and configuration of local expression domains are remarkably similar.

Also in the rice genome, a small yet significant fraction of genes is organized in local coexpression domains that predominantly consist of two, up to 4, genes that are not

categorized in the same functional category, irrespective of the expression platform used for analyses. The presence of tandemly duplicated genes, shared promoter sequence or gene distance is not fully explaining the occurrence of coexpression of genes in such chromosomal domains. Therefore, the regulation of local coexpression domains is postulated to be at the level of higher-order of chromatin structure. Given the similarities in the characteristics and occurrence of coexpression domains between Arabidopsis and rice, we investigated whether the genes involved showed microsynteny between the two genomes. These analyses did not identify the presence of syntenic local coexpression domains between Arabidopsis and rice.

Material and methods

Genome data

The rice (*Oryza sativa*) genome was obtained from the website of The Institute of Genomic Research (TIGR; www.tigr.org). The rice TIGR version 3 [Jan. 2005] annotation has 57,915 gene loci. In case of alternative splicing, the longest variant of the gene was used. The genes along each chromosome were sorted based on ascending start coordinates and were numbered consecutively. These rank numbers (rank ID) helped to eliminate any discontinuity in the unique Os gene identifiers of the annotated genes and facilitated analyzing physically adjacent genes. In case of overlapping gene loci, the smaller one of the overlapping genes was removed from the data set. This way the order of both gene and rank ID numbers was maintained.

Expression data

The MPSS expression data for rice (cv. Japonica) were obtained from the rice MPSS database (<http://mpss.udel.edu/rice>). Only the unique MPSS tags (mapping to the genome only once) and those mapping to unique gene identifiers in TIGR v3 were used in the analyses. The MA expression data were obtained from the Gene Expression Omnibus repository (GEO; www.ncbi.nlm.nih.gov/geo). The data with accession number GSE4438 were chosen for analysis, because they represent the only Affymetrix chip-based data set in the public domain with a reasonable coverage of the *Oryza sativa* cv. japonica transcriptome under four different experimental conditions (Walia et al. 2007). The earlier analysis in Arabidopsis also concerned Affymetrix chip data (Ren et al. 2005). The MA expression data were mapped to the TIGR v3 annotation based on gene accession numbers.

The expression values in libraries representing biological replicates, i.e., the same tissues under the same experimental

conditions, were averaged. This way, for the MPSS data 18 different libraries were generated, that cover expression in 9 tissues (callus, panicle, leaves, root, germinating seed and seedling meristem, ovary and stigma, pollen, stem) under different experimental treatments or in different developmental stages. From the microarray (MA) data, only the expression data from cv. japonica were used to allow comparison with the MPSS data. Here, 4 different libraries were generated: a sensitive japonica genotype under control conditions and under salt stress; and a tolerant japonica genotype under control conditions and under salt stress (Walia et al. 2007). The expression data is from crown and growing point tissue.

Identification of local coexpression domains

Pearson's correlation coefficient (R) was calculated between all adjacent pairs (duplets) of genes using the expression data from all 18 libraries. If R was higher than 0.7, the gene pair concerned was considered to be coexpressed. The value of $R > 0.7$ is generally considered a rule-of-thumb threshold (see for example bbc.botany.utoronto.ca/affydb/BAR_instructions) and is used in various analyses (Cohen et al. 2000; Lee et al. 2004; Ren et al. 2005). The number of coexpressed adjacent pairs was counted. To evaluate the statistical significance of these numbers, they were compared with the number of coexpressed pairs from 100 randomizations of the population of expressed genes using the cumulative binomial distribution (Cohen et al. 2000). Previous analyses indicated that more than 100 randomizations did not result in significant changes in the numbers obtained (Ren et al. 2005). In each round of randomization, non-adjacent pairs of genes were randomly selected with replacement from the list of expressed genes that have expressed neighbors till the same total number of pairs was obtained. Similarly, coexpressed adjacent triplets, quadruplets and pentaplets were identified as series of genes with consecutive IDs in which all possible (that is, $(n!/(n-2)!)*2$; where n is the number of genes involved) pair-wise R 's should be above the cut-off of 0.7. The significance of results was evaluated with randomizations equivalent to the procedure used in case of duplets.

Duplicated genes

Duplicated genes were identified by local pair-wise protein BLAST (BLASTP 2.2.6 [Apr-09-2003]; (Altschul et al. 1997)), on all gene pairs in the rice genome. A gene pair was considered to be duplicated (dup) if BLASTP yielded an E -value < 0.2 (Fukuoka et al. 2004; Lercher et al. 2003; Williams and Bowles 2004). To determine duplicated triplets, quadruplets and pentaplets, it was required that any pair of the genes concerned had a BLASTP E -value < 0.2 .

Analyses of gene orientation and gene distance

Adjacent gene pairs were separated into tandemly, divergently and convergently transcribed pairs according to their relative direction of transcription. The number of coexpressed pairs in each orientation group was expressed as percentage relative to the total number of adjacent pairs in that group. Random pairs were made by randomly picking two non-adjacent genes from the list of expressed genes represented in pairs, analyzed for their orientation and compared with the real genome using a variant of the two-sample t test for proportions for determining the significance of a difference between two population proportions (Ott and Longnecker 2001). The test statistic is based on the z statistic from the normal distribution and is given by $(p_1 - p_2) / \sqrt{(p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2)}$, with p_1 and p_2 the two sample proportions, n_1 and n_2 the two sample sizes, under the condition that $n_1 * p_1$, $n_1 * (1-p_1)$, $n_2 * p_2$ and $n_2 * (1-p_2)$ are all larger than 5. The z value is converted to a p value using standard normal tables.

To determine the gene distance, the intergenic distance is used. This distance is defined as the length in nucleotides from the annotated end of one gene to the annotated start of the next gene, including the UTRs when known, otherwise the translation start and stop sites were taken. The data sets excluding the duplicated gene pairs were analyzed. For each data set, gene pairs were sorted based on gene distance from short to long and bins of 1,000 pairs were taken and analyzed, excluding the last bin with less than 1,000 pairs. The advantage of using equal pair bin is that it avoids unequal number of gene pairs in different distance categories. Per 1,000-pair bin, gene distance was calculated as the average over all 1,000 pairs. For each 1,000-pair bin, the fraction of coexpressed pairs relative to the total number of pairs in each orientation group in each bin was calculated and plotted.

Functional categorization of genes

TAIR's GOSlim, the Gene Ontology (GO) developed for plants (Berardini et al. 2004) was used to classify the genes present in local coexpression domains. The three aspects of GOSlim, molecular function, biological process and cellular component, were analyzed in parallel. With Python scripts, the number of pairs of which both members could be classified in GOSlim was determined, and the number of pairs of which both members fall into the same well-defined GOSlim category was also determined. The GOSlim categories of 'unknown' and 'other' were not included into well-defined categories, because they give less (or no) information about functional categorization. The percentage of coexpressed pairs falling into the same well-defined category was compared with that of non-coexpressed pairs

to determine whether coexpressed genes are more enriched in the same functional category than non-coexpressed genes.

Assessing synteny between Arabidopsis and rice

The Inparanoid Eukaryotic Orthologous groups database (inparanoid.cgb.ki.se; O'Brien et al. 2005) was used to download all known orthologous and inparalogous clusters between Arabidopsis and rice. Inparanoid defines inparalogs as paralogs that arose through gene duplication after speciation. Inparalogs can form a group of genes that together are orthologous to a gene in another species. There are 9,044 orthologous clusters between Arabidopsis (from Ensemble) and rice (from the Model Organism database) and all of them were taken into account. These clusters were downloaded on Dec. 12, 2005. In the orthologous clusters, 15,544 sequences (proteins) from Arabidopsis are inparalogs and 14,807 sequences (proteins) from rice are inparalogs. More than half of Arabidopsis and rice inparalogs are many-to-many or many to-one orthology cases. Less than half of the cases are one-to-one orthology cases.

The Ensembl protein IDs (for Arabidopsis) and the Model organism database protein IDs (for rice) in Inparanoid were first translated to their unique gene identifiers in the respective TIGR annotation by BLASTP using an E -value $< e^{-20}$. This yielded 14,753 unique Arabidopsis genes and 12,428 unique rice genes as inparalogs. The pairs of genes in local coexpression domains were analyzed to determine which genes in a rice local coexpressed pair have orthologs in an Arabidopsis local coexpressed pair, and vice versa. Because of the larger coverage of genes and tissues, only the MPSS expression data were used for this analysis. As coexpressed triplets and quadruplets are always combinations of coexpressed pairs, they were not further analyzed. For comparison, the pairs of genes that are not coexpressed were analyzed to determine how many non-coexpressed pairs, or one of their member genes, have orthologs in the other plant species. The numbers were then compared between coexpressed pairs and non-coexpressed pairs to determine the significance of occurrence of syntenic local coexpression domains.

Results

Local coexpression domains consist of two to four neighboring genes

The TIGR version 3 of the rice genome has 57,915 predicted genes. This is about twice the number of genes predicted for Arabidopsis (28,952 genes; TIGR5 annotation). The coverage of the MPSS expression data for the

rice genome is 40%. The expression coverage in the Affymetrix array data set we used is 26% (Table 1). This is about half or less of the expression coverage for the Arabidopsis genome (72% in the TIGR5 update; Ren et al. 2005). This difference can reflect the more advanced annotation of the Arabidopsis genome at this time, and/or the more complex (duplicated) organization of the rice genome. The rice genome annotation has more genes that are physically overlapping than the Arabidopsis genome. Excluding the smaller overlapping genes from the analyses, we were able to identify 12,920 gene pairs with MPSS expression data and 6,032 pairs with MA expression data in rice (Table 1; see also Materials and methods). Of these, 584 (4.5%) in MPSS and 320 (5.3%) in MA were identified to represent a local coexpression domain as defined as being coexpressed with a pair-wise Pearson's correlation coefficient larger than 0.7 (Table 1). This percentage is similar to what we have found previously for Arabidopsis (Ren et al. 2005) and agrees well with other findings that ~3–5% of a genome is tightly coexpressed (Semon and Duret 2006).

Table 1 Description of rice expression data used for whole-genome local coexpression analysis

	MPSS	MA
<i>Genes with expression</i>		
Excluding overlapping genes	23,146	14,789
Without expressed neighbor(s)	5,081	5,438
represented in pairs	18,065	9,351
<i>Adjacent pairs</i>		
Total	12,920	6,032
Tandemly duplicated pairs (td)	1,663 (12.9%) ^a	573 (9.5%) ^a
Coexpressed	584 (4.5%) ^b	320 (5.3%) ^b
Total excluding td	11,257	5,459
Coexpressed excluding td	438 (3.9%) ^c	288 (5.3%) ^c
<i>Coexpressed adjacent pairs</i>		
Total	584	320
Tandemly duplicated pairs	146 (25%) ^d	32 (10%) ^d
<i>Tandemly duplicated pairs</i>		
Total	1,663	573
Coexpressed	146 (8.8%) ^e	32 (5.6%) ^e

^a Percentage of tandemly duplicated pairs relative to the total number of adjacent pairs

^b Percentage of coexpressed adjacent pairs relative to the total number of adjacent pairs

^c Percentage of coexpressed adjacent pairs excluding td relative to the total number of adjacent pairs excluding tandemly duplicated pairs

^d Percentage of coexpressed tandemly duplicated pairs relative to the total number of coexpressed adjacent pairs

^e Percentage of coexpressed tandemly duplicated pairs relative to the total number of tandem duplicated pairs

Notably duplicated genes are supposed to influence coexpression statistics due to their common origin (Lercher et al. 2003), although a surprising finding for the Arabidopsis coexpression domains was that only a minor fraction of duplicated genes were actually coexpressed (Ren et al. 2005). The occurrence of duplicated pairs in the rice set was determined with pair-wise protein BLAST using a cut-off of $E < 0.2$ (Fukuoka et al. 2004; Lercher et al. 2003; Williams and Bowles 2004). This identified 1,663 (12.9%) duplicated gene pairs in the MPSS gene pair data set and 573 (9.5%) duplicated gene pairs in the MA gene pair data set.

Of these, only 146 (8.8%) in the MPSS data and 32 (5.6%) in the MA data were coexpressed (Table 1). Although this percentage is somewhat higher than the percentage of coexpression in non-duplicated pairs (3.9%), the majority of all duplicated pairs (91.2% in MPSS data and 94.4% in MA data) are not coexpressed. This shows that also in rice gene duplication does not correlate well with coexpression and suggests that expression divergence is a common phenomenon after duplication (Williams and

Bowles 2004). Excluding the duplicated pairs from the coexpressed sets, there are 438 gene pairs in the MPSS data and 288 gene pairs in the MA data coexpressed in rice. This accounts for 75% (=438/584) in MPSS and 90% (=288/320) in MA of all coexpressed pairs. Therefore, also in rice the occurrence of duplicated genes cannot explain the occurrence of local coexpression domains. Extending the size of the local coexpression domain to triplets, quadruplets, pentaplets and on, requiring that all pair-wise combinations of genes have a tightly correlated expression, shows that few larger local coexpression domains exist (Table 2). No quadruplet domains could be identified when tandemly duplicated genes were excluded (Table 2). To assess the significance of the occurrence of the various local coexpression domains, we compared the number of coexpressed pairs, triplets and quadruplets with the average of such domains in 100 randomly generated genomes using the cumulative binomial distribution (Cohen et al. 2000). Such comparisons revealed that local coexpression pairs occur in the rice genome significantly more often than expected by chance alone (Table 2). However, when

Table 2 Local coexpression domains in the rice genome

	Rice genome		Random genome (100×)	
	Total ^a	Coexpressed ^b	Average ^c	<i>P</i> -value ^d
<i>Pairs</i>				
MPSS + td ^e	12,920	584 (4.52%)	408 ± 17	1.46 × 10 ⁻¹⁷
MPSS-td ^f	11,257	438 (3.89%)	356 ± 21	2.17 × 10 ⁻⁶
MA + td ^g	6,032	320 (5.30%)	301 ± 17	0.012
MA-td ^h	5,459	288 (5.28%)	271 ± 16	0.014
<i>Triplets</i>				
MPSS + td	7,775	23 (0.30%)	8.78 ± 2.9	2.95 × 10 ⁻⁵
MPSS-td	6,831	13 (0.19%)	7.74 ± 3.0	0.025
MA + td	2,461	5 (0.20%)	6.54 ± 2.7	n.s.
MA-td	2,149	3 (0.14%)	5.10 ± 2.4	n.s.
<i>Quadruplets</i>				
MPSS + td	4,887	3 (0.06%)	0.24 ± 0.47	1.81 × 10 ⁻³
MPSS-td	4,318	0 (0%)	0.18 ± 0.39	n.s.
MA + td	1,079	0 (0%)	0.14 ± 0.37	n.s.
MA-td	nd ⁱ	nd	nd	nd

^a Total number of pairs, triplets, quadruplets in each data set

^b Coexpressed pairs, triplets, quadruplets in each data set. Percentages in brackets are coexpressed relative to the total

^c Average plus/minus standard deviation from 100 randomizations

^d *P*-value according to the cumulative binomial distribution (Cohen et al. 2000) for obtaining such a result by chance. $P < 0.05$ is considered significant; n.s.: not significant

^e MPSS data set including tandemly duplicated genes

^f MPSS data set excluding tandemly duplicated genes

^g MA data set including tandemly duplicated genes

^h MA data set excluding tandemly duplicated genes

ⁱ Not determined

excluding the duplicated genes, triplets and quadruplets do not occur significantly more often than by chance (at $P < 0.05$) in both expression datasets. Local coexpression domains therefore consist of at most 2 genes when duplicated genes are not taken into consideration. This number appears smaller than in the Arabidopsis genome (Ren et al. 2005), but this may reflect the lesser coverage of the annotation of the rice genome.

Local coexpression domains seem randomly distributed over the genome (Fig. 1; MPSS data only). Only 10 coexpressed pairs are common between the MPSS and MA coexpressed sets, out of a total of 4,643 common pairs (excluding tandemly duplicated pairs). Detailed information about locations, orientations, expressions and gene distances (only for the pairs) for all pairs, triplets and quadruplets in both the MPSS and MA data sets are given in the supplementary data. All subsequent analyses were focused on domains consisting of non-duplicated gene pairs, unless stated differently.

Orientation and distance do not solely explain the occurrence of local coexpression

In yeast, there are several examples that divergently transcribed promoter regions are the cause of co-regulated neighboring genes (Korbel et al. 2004; Kruglyak and Tang 2000). If promoter sharing is an important mechanism for coexpression in the rice genome, divergently transcribed gene pairs should be over-represented in the sub-population of coexpressed pairs, compared to coexpressed pairs that are tandemly or convergently transcribed. For all three-orientation groups, the number of pairs and the number of coexpressed pairs in the rice genome were determined

(Table 3). For each orientation group, the fraction of coexpressed pairs relative to the total number of pairs in that group was calculated (Table 3). None of the fractions are significantly different from each other using a statistical test for comparing population proportions (Ott and Longnecker 2001). The fraction of coexpressed divergent pairs in the MPSS data is the lowest of the three groups (Table 3). Therefore, shared promoter regions cannot solely explain the coexpression of adjacent genes.

The physically closer two genes are, the higher the likelihood is that they are coexpressed due to either *cis* or *trans*-activation (Hershberg et al. 2005). Therefore, we determined the intergenic distance, defined as the sequence length in nucleotides from the annotated end of one gene to the annotated start of the neighboring gene, including UTRs when known, otherwise taking the start and stop site for translation. This distance was used to investigate whether it would explain the characteristics of local coexpression domains. In Fig. 2, the fraction of coexpressed pairs is plotted for each orientation and for each 1000-pair bin after sorting based on intergenic distance. The results show that the fraction of coexpressed pairs, irrespective of gene orientation, does not decrease with larger gene distance. When gene distance is defined as the sequence length from the start of one gene till the start of the next gene (Ren et al. 2005), the result is similar (data not shown). As a consequence, increasing intergenic distances do not seem to be a barrier for the occurrence of local coexpression and short intergenic distances do not favor coexpression. Therefore, intergenic distance does not solely explain local coexpression in the rice genome, as it did not in the Arabidopsis genome (Ren et al. 2005).

Fig. 1 Distribution of local coexpression domains over all 12 rice chromosomes.

Rectangles are schematic representation of chromosomes 1–12 from top to bottom. The numbers on the top show the scale in million bases along the chromosomes. Each gene in a local coexpression domain is depicted with a black bar. Only MPSS datasets excluding tandemly duplicated genes are shown. The orders of the drawings in each rectangle are: first lane, coexpressed pairs; second lane, coexpressed triplets; third lane, coexpressed quadruplets, fourth lane, partially syntenic coexpression domains (PSCDs) between Arabidopsis and rice

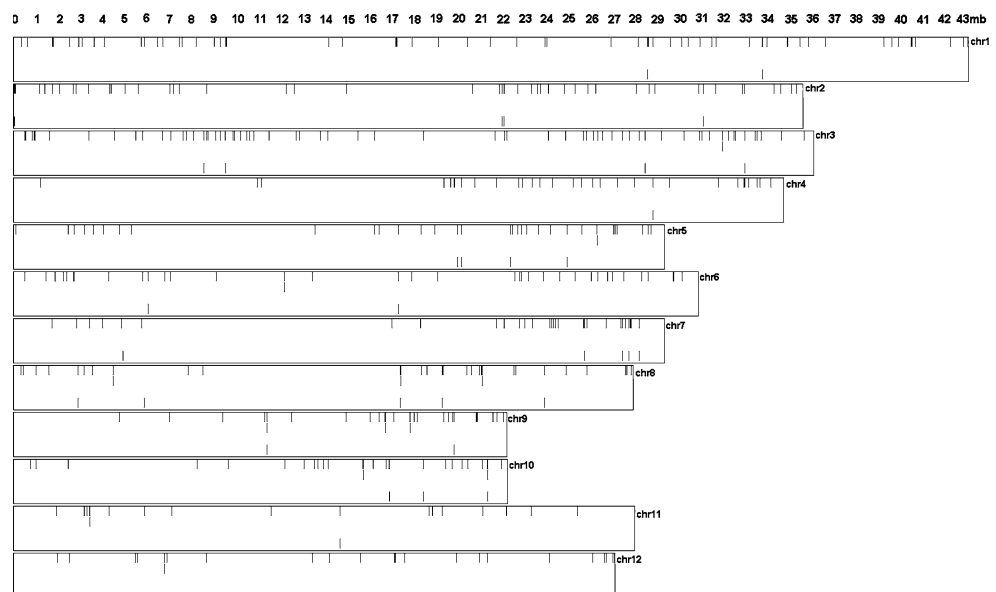


Table 3 Orientation of coexpressed gene pairs

Orientation groups ^a	Total ^b	Coexpressed ^c
<i>MPSS</i>		
tan-td	5,621	239 (4.25%)
div-td	2,418	82 (3.39%)
con-td	3,218	117 (3.64%)
<i>MA</i>		
tan-td	2,707	143 (5.28%)
div-td	1,224	72 (5.88%)
con-td	1,528	73 (4.78%)

^a tan-td, div-td, con-td, respectively are the sub-groups of tandemly, divergently, convergently transcribed pairs excluding tandem duplicates

^b Total number of pairs in each direction group

^c Number of coexpressed pairs in each direction group. Percentages in the brackets are number of coexpressed pairs relative to the total number of pairs. None of the proportions are significantly different from each other according to the z test for comparing population proportions

Functional categorization of coexpressed genes

To characterize the kind of genes that are present in the rice coexpression domains, the gene ontology (GO) developed

for plants (GOSlim; Berardini et al. 2004) was used. The GOSlim ontology provides a controlled vocabulary to describe gene and gene product attributes in plants, focusing on three aspects of annotation: molecular function, biological process and cellular component. Each aspect has 15–16 categories with 4–5 categories having terms like “unknown” or “other”. To all pairs of genes, each aspect of the GOSlim annotation was assigned. For each aspect, the number of pairs was determined for which both member genes were covered by a GOSlim assignment. In addition, the number of pairs for which both member genes fall into the same well-defined categories (excluding the “unknown” and “other” subcategories) was determined. The fraction of the latter was compared between coexpressed pairs and non-coexpressed pairs to determine whether coexpressed pairs were enriched in the same categories (Table 4). In the MPSS data, the GOSlim annotation coverage for both member genes in a pair is 22% for molecular function, 12% for biological process and only 3.4% for cellular component. In the MA data, these 3 percentages are: 25%, 13% and 3.7%, respectively. Comparing these figures with the GOSlim coverage of the Arabidopsis genes (data from Ren et al. 2005, but updated to the TIGR5 annotation), which is ~94% for all 3 aspects, shows that currently the rice genome is considerably less

Fig. 2 Gene distance does not solely explain the occurrence of coexpression. Gene distance, defined as the length in nucleotides from the annotated end of one gene to the annotated start of the next gene relative to the strand the genome that is given, with annotated start always smaller than the annotated end. X-axis is the averaged gene distance (in base pair) in each 1,000-pair bin. The Y-axis depicts the number of pairs (A, D), number of coexpressed pairs (B, E) and the fraction of coexpressed pairs (C, F), relative to the total number of pairs in each orientation (tan: tandem pairs; div: divergent pairs; con: convergent pairs) in each 1,000-pair bin

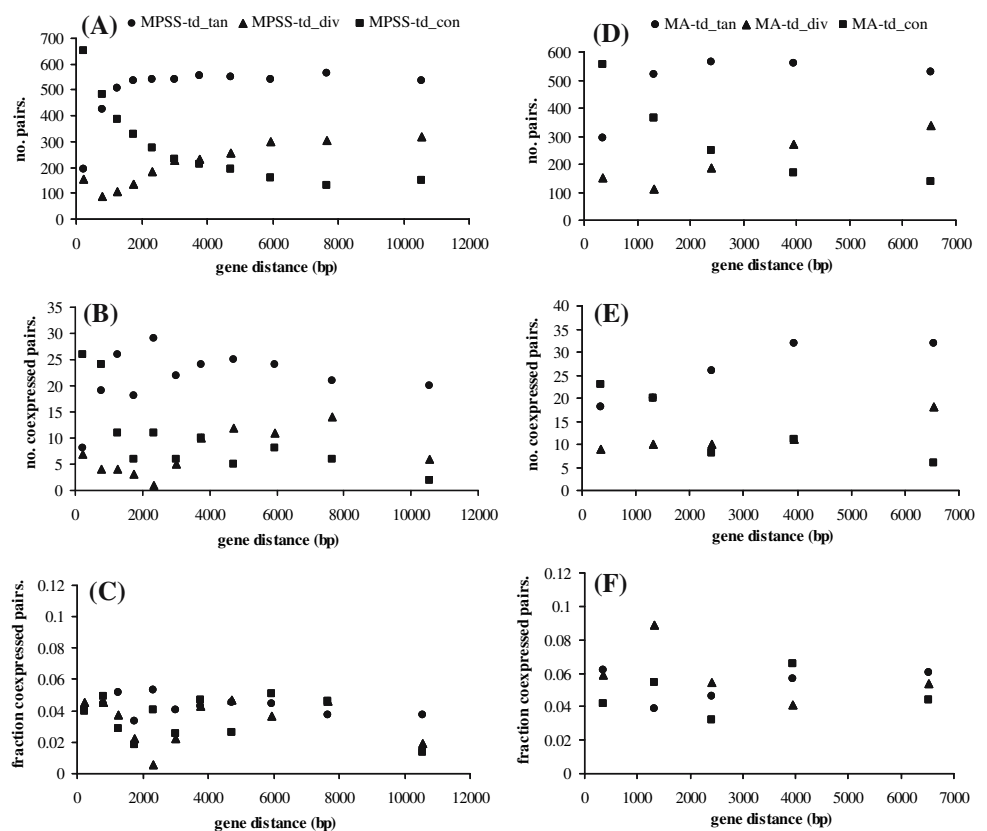


Table 4 Distribution of gene pairs over GOslim categories (Non-duplicated pairs)

	All ^a	Coexpressed ^b	Non-coexpressed ^c	<i>P</i> -value ^d
MPSS				
GO_func				
covered ^e	2502	100	2402	0.42
sameKnCat ^f	365 (14.6%)	12 (12.0%)	353 (14.7%)	
GO_proc				
Covered	1366	50	1316	0.47
sameKnCat	144 (10.5%)	7 (14%)	137 (10.4%)	
GO_comp				
Covered	383	17	366	0.60
sameKnCat	113 (29.5%)	6 (35.3%)	107 (29.2%)	
MA				
GO_func				
covered ^e	1365	83	1282	0.13
sameKnCat ^f	177 (13.0%)	7 (8.43%)	170 (13.3%)	
GO_proc				
Covered	707	43	664	0.13
sameKnCat	67 (9.48%)	2 (4.65%)	65 (9.79%)	
GO_comp				
Covered	202	10	192	0.24
sameKnCat	45 (22.3%)	4 (40%)	41 (21.4%)	

^a Number of neighboring pairs excluding td. All other pairs are all duplicate-free, unless stated otherwise

^b Number of coexpressed pairs

^c Number of non-coexpressed pairs

^d *P* value from the standard normal tables of the *z* statistic for the difference of the population proportion between coexpressed pairs and non-coexpressed pairs in the rice genome; *, significant (two-tailed; *P* < 0.05). The *P* value is the probability under the null hypothesis that the two population proportions are the same

^e Number of pairs of which both members are assigned (covered) with GOslim categories

^f Number of pairs of which both members fall into the same “known” GOslim category (excluding the categories with the indications ‘unknown’ and ‘other’). Percentage is the number of pairs relative to the number of pairs covered

well annotated than the Arabidopsis genome. When comparing the coexpressed and non-coexpressed pairs in rice for the fraction of gene pairs falling into the same well-defined GOslim category, there is no significant difference (Table 4). Therefore, also in the rice genome coexpressed gene pairs are not enriched for the same functional category.

Microsynteny of local coexpression domains between rice and Arabidopsis

The structural characteristics of local coexpression domains in rice and in Arabidopsis (Ren et al. 2005) are remarkably similar. This prompts the question whether such domains also share functional characteristics and possibly consist of the same or related genes. Microsynteny in local expression domains of these two genomes would reflect conservation of such domains. The Inparanoid

Eukaryotic Orthologous database (O’Brien et al. 2005) was used to retrieve the current list of genes that are supposed to be orthologous between Arabidopsis (14,753 genes) and rice (12,428 genes), including all many-to-many relationships. The genes establishing coexpressed pairs based on MPSS expression data in either Arabidopsis (944 pairs including 116 duplicated pairs; data from Ren et al. 2005, but updated to the TIGR 5 annotation) or rice (584 pairs, including 146 duplicated pairs) were searched against these lists. This way, we aimed to identify the pairs of which both genes in the pair have an ortholog in the other plant and these orthologs are also coexpressed. The analyses showed that there was not a single coexpressed pair in either Arabidopsis or rice of which both genes are orthologous to a gene of a coexpressed pair in the other species. Therefore, given the current annotation of the two genomes, there are no syntenic local coexpression domains between Arabidopsis and rice.

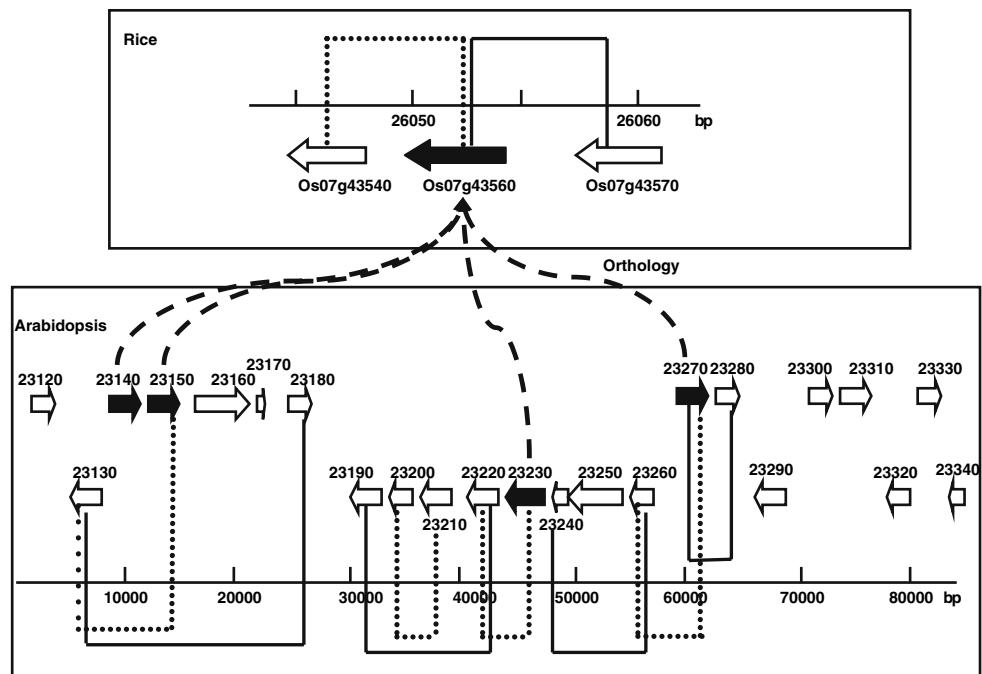


Fig. 3 Schematic representation of the chromosomal regions covering genes involved in a four-to-one orthology between Arabidopsis and rice. Top part of the figure is the chromosomal region from rice (from gene locus *Os07g43540.1* to gene locus *Os07g43570.1*). Bottom part of the figure is the chromosomal region from Arabidopsis, representing 23 genes (from gene locus *At4g23120* to *At4g23340*; the numbers in the picture do not carry “*At4g*”). Black

arrows represent the four Arabidopsis and the one rice gene involved in this orthology, and dashed curved connecting lines show the orthology relationships. Black bracket-like lines depict duplication and genes connected and included within by black bracket line are duplicated to each other. Dotted lines depict coexpression relationship and genes connected and included by dotted line are coexpressed with each other

Partially syntenic local coexpression domains can occur by chance

In 34 cases though, one gene of a coexpressed pair in one plant species was orthologous to at least one gene of a coexpressed pair in the other plant. That is 3.6% of all (944) coexpressed pairs in Arabidopsis and 5.8% of all (584) coexpressed pairs in rice. We will refer to such a case as a partially syntenic coexpression domain (PSCD). To assess the significance of such partially syntenic domains, we evaluated all the genes in non-coexpressed pairs, comparing Arabidopsis (15,629 pairs including 617 duplicated pairs) and rice (12,336 pairs including 1,517 duplicated pairs) to establish whether PSCDs are more enriched in the genome than partially syntenic non-coexpressed domains (PSND). We identified 4,488 PSNDs (72 due to duplicated pairs) between all non-coexpressed pairs of genes in both plant genomes. This is 28.7% of all Arabidopsis non-coexpressed pairs and 36.4% of all rice non-coexpressed pairs. The percentage of PSNDs among non-coexpressed pairs is 6–8 times higher than that of PSCDs from coexpressed pairs. Therefore, PSCDs do not seem to occur more often than expected by chance alone.

A complicating issue in the analysis of synteny is the occurrence of many-to-many orthologs. The Inparanoid database defines so-called inparalogs as paralogs arising through gene duplication after speciation. These can form a group of genes that together are orthologous to a gene in the other species. As a result, there can be many to many, many to one and one to one relationships. Individual member genes in many-to-many or many-to-one relationships may not be the main orthologs. Interestingly, there is one many-to-one case in which four Arabidopsis genes are all orthologs of the same single rice gene (*Os07g43560.1*). These 4 Arabidopsis genes are: *At4g23140.2*, *At4g23150.1*, *At4g23230.1* and *At4g23270.1*. The first two, *At4g23140.2*, *At4g23150.1*, form a local coexpressed pair. The other two genes, *At4g23230.1* and *At4g23270.1*, are not more than ten genes away from the previous two genes on the same chromosomal region. The latter two genes are separated from each other by a few genes. Further analysis shows that gene *At4g23270.1* has a duplicated neighbor, *At4g23280.1*, but is not coexpressed with it. It is, however, coexpressed with its other neighbor *At4g23260.1*, but is not duplicated with it. Orthology is established between *At4g23270.1* and the rice gene *Os07g43560.1*, but

not between any of the neighbors of the *Arabidopsis* genes. The rice gene Os07g43560.1 is also coexpressed with one of its neighboring genes Os07g43540.1, but it is not a duplicate of it, while this rice gene itself is a duplicate of another neighboring gene (Os07g43570), but it is also not coexpressed with it. A schematic representation of the resulting gene configuration is given in Fig. 3. Such detailed analyses may reveal local microsynteny in the twilight zone of statistical significance and evolutionary relevance.

Discussion

Local coexpression domains represent only a small part of the genome

Setting stringent criteria for coexpression using MPSS and microarray expression data, the rice genome was found to contain a small but significant number of local coexpression domains that range from predominantly two, up till 4, genes, irrespective of the expression platform used for analysis. This is similar to the situation in *Arabidopsis* (Ren et al. 2005). It shows that a genome is essentially not a random entity with respect to the occurrence of local coexpression domains. Our results agree with other coexpression studies where strong coexpression was shown to occur only within close proximity of several genes (Cohen et al. 2000; Hershberg et al. 2005; Lercher et al. 2003; Semon and Duret 2006). Although coexpression was shown to extend to regions covering up to 30 genes and more (Spellman and Rubin 2002) and to cover chromosomal regions up to 100 kb (Ma et al. 2005; Spellman and Rubin 2002; Williams and Bowles 2004) and more, there appears to exist a decrease in the strength of coexpression with increasing distances. The local coexpression domains described here represent ~4–5% of the potential coexpression fraction in the whole genome as found in other studies (Semon and Duret 2006). Larger but looser coexpression domains might cover up to ~10% (Cohen et al. 2000; Williams and Bowles 2004) till 20% (Spellman and Rubin 2002) of the genome. The difference in occurrence between local and longer-range weaker but still statistically significant coexpression domains is highly dependent on the method used (Semon and Duret 2006). Moreover, the expression platforms used focus on snapshots of gene expression at the RNA level. Such data ignore various post-transcriptional regulatory mechanisms that can for example result in poor correlations between RNA and protein levels. Although we have shown before that in a transgenic set-up the creation of an artificial local coexpression domain results in markedly improved RNA/protein correlations (Mlynarova et al. 2002), this does not need to be the case

for all endogenous local coexpression domains now identified.

The terms cluster or chromosomal domain and associated terms such as neighboring are generally based on a (much) more loose definition compared to the definition used here to identify local coexpression domains. Local coexpression domains require a pair-wise correlation between the expressions of ALL adjacent genes above 0.7. The larger domains are defined on the basis of the use of a sliding window of either a given sequence length (number of nucleotides) or of a given number of genes (Spellman and Rubin 2002; Williams and Bowles 2004). In such a window, the average correlation is calculated and compared with simulated sets. This allows for the presence of genes within a domain that are not strongly (co)expressed but are “carried along for a ride” in the open chromatin domain (Spellman and Rubin 2002).

The two expression data sets here analyzed show very little overlap in coexpressed pairs (only 10 out of 4,643 all common pairs). The low number of shared pairs is thought to reflect the biological background of the data sets. The MPSS data concern a broad range of tissues and experimental conditions, while the MA data only cover crown and growing point tissue under control and salt stress condition (Walia et al et al. 2007). The number of libraries taken into account, as well as their biological background, obviously influences the possibility of identifying significant numbers of coexpression domains above the number of domains expected by chance. Two genes evaluated in four comparably similar conditions (MA data) are predicted to have a higher likelihood of exhibiting a similar expression pattern, than when examined in 18 different (MPSS) conditions. When considering genome-wide local coexpression of genes, a wide diversity of tissues and/or conditions should be taken for analysis. Highly tissue-specific coexpression may be masked in this approach and should be analyzed by other means.

Parameters shaping local coexpression domains

The existence of local coexpression domains in rice could not be explained solely by gene orientation, such as tandemly, divergently or convergently oriented gene pairs. No relative enrichment of the proportion of coexpressed pairs was seen. The fraction of coexpressed genes in the divergent orientation was even lower than for the other two orientations (Table 3). So shared promoter regions (for divergent pairs) and transcriptional read-through (for tandem pairs) do not explain the local coexpression domains in rice, similar to what we have concluded for *Arabidopsis* (Ren et al. 2005). This is in contrast to some other studies in which shared promoter region (for divergent pairs) and transcriptional read-through established coexpression

domains (Semon and Duret 2006). Whereas we do not detect any preferred orientation to result in coexpression in rice, other studies show a higher degree of coexpression in divergent and tandemly oriented gene pairs (Williams and Bowles 2004; Zhan et al. 2006). The differences in conclusions are most likely due to the different methods used, such as the definition of coexpression of neighboring genes as well as the dataset and/or expression platform used.

Gene distance is not the explanatory factor for the occurrence of local coexpression domains in rice. No significant decrease in the fraction of coexpressed genes was observed with increasing intergenic distance (Fig. 2). The fraction of coexpressed pairs does not decrease even with gene distances up to 12 kb. It shows that at a relatively large distance, neighboring genes can still be coexpressed. Another study reported that when genes >12 kb apart were taken into account, the negative correlation between coexpression and gene distance was gone (Williams and Bowles 2004). In a comparative study of 6 eukaryotic genomes, coexpression was shown to vary at chromosomal distances above 100 kb (Fukuoka et al. 2004). This suggests that considerable coexpression of neighboring genes can occur even at large gene distance, although the coexpression may not be related to the physical distance anymore. While gene distance itself is not predictive for coexpression (Cohen et al. 2000; Kruglyak and Tang 2000), the likelihood of coexpression would favor short gene distances (Hershberg et al. 2005; Hurst et al. 2002; Lercher et al. 2003; Semon and Duret 2006). It should be kept in mind that the rice data set now analyzed is far from complete in terms of its annotation, so what are now far-apart neighboring genes may be no longer directly neighboring the moment the annotation is improved. Because the characteristics of the local coexpression domains in Arabidopsis and rice are remarkably alike, the actual numbers and possibly even genes involved in local coexpression domains may be subject to change, but not the occurrence of such domains in these genomes.

With the gene ontology developed for plants (GOslim), there is no evidence that coexpressed genes are more enriched in the same functional category in comparison to non-coexpressed genes (Table 4). Previous studies suggested that clustering of functionally related genes would occur in all metazoans (Cohen et al. 2000; Lercher et al. 2003). Recent studies demonstrated a significant enrichment for coexpressed genes in the same metabolic pathway (Williams and Bowles 2004) or the same biological processes (Zhan et al. 2006), although this appeared not to be the explanation for the coexpression of neighboring genes (Williams and Bowles 2004). In worm, clusters of similarly expressed genes cover similar biological functions (Roy et al. 2002). In human, coexpression over the whole

genome was shown to correlate with functional relationships between the genes (Lee et al. 2004). Our study found no enrichment of coexpressed gene pairs in the same functional category than non-coexpressed pairs, suggesting that it is not necessarily true that the natural selection maintained regional coexpression by keeping genes with similar functions in adjacent positions (Cohen et al. 2000; Semon and Duret 2006).

The genomic context of genes is supposed to play an important role in the regulation of gene expression (van Drunen et al. 1997). In a number of coexpression studies in various organisms, the occurrence of coexpression domains, whether small (local) or larger (global), sometimes independent of gene orientation and gene distance, were all supposed to be regulated at the level of higher-order chromosomal structure (Cohen et al. 2000; Hershberg et al. 2005; Ren et al. 2005; Spellman and Rubin 2002; Williams and Bowles 2004; Zhan et al. 2006). It will be difficult to formally exclude the possibility of transcriptional regulation as a cause of local coexpression. Co-regulated transcription could occur through shared promoter elements between the neighboring genes in coexpression domains. However, previous experience with transgene expression data indicated that the particular position of neighboring genes in a genome affects the expression of that gene considerably (Mlynarova et al. 1994, 1995). Two physically neighboring transgenes could only be made to into an artificial local expression domain, that is, show correlated expression, when chromatin-organizing elements were placed around the genes (Mlynarova et al. 2002). Although this experimental result indicates the importance of chromatin organization in establishing local coexpression in a plant genome, this result does not need to be the case for all coexpression domains identified. With this caveat, this study of local coexpression domains in rice that are independent of duplication, gene orientation, or gene distance strengthen the notion that the regulation of genes in such domains resides at the level of higher-order chromatin structures. Future studies using for example advanced fluorescence in situ hybridization (FISH) technology on interphase chromosomes (Walter et al. 2006), or the transgenic approach outlined above (Mlynarova et al. 2002), will generate further experimental support for the location and characteristics of local coexpression domains. In addition, the new approaches in and insights from epigenetics and epigenomics (Henderson and Jacobsen 2007), such as the genome-wide mapping and analysis of DNA methylation (Zhang et al. 2006; Zilberman et al. 2007) and future histone modification maps (Esteller 2007) of plants, possibly in combination with genetics (Jansen and Nap 2001), will be useful to get more insight into the occurrence and function of local coexpression domains in plant genomes.

Lack of microsyntenic coexpression

From an evolutionary point of view, syntenic regions between species reveal genes for conserved and important traits. Macrosynteny is generally not easily detectable after a long evolutionary time, as colinearity erodes by various mechanisms, such as transposon activity, intra or inter-chromosomal rearrangements, duplications, translocations, inversions and/or individual divergence after speciation (Salse et al. 2002). While macrosynteny may not be detectable any more for genomes that diverged more than 100 million years ago (mya), microsynteny, i.e., conservation of local gene order and orientation, may still exist and be informative (Devos et al. 1999; Salse et al. 2002). Arabidopsis and rice are thought to have diverged about 120–200 mya (Salse et al. 2002). Microsyntenic local coexpression domains between Arabidopsis and rice would indicate the importance of the evolutionary conservation of regulatory systems beyond sequence similarity after the divergence of dicotyledonous and monocotyledonous plants. Analyses show that there is not a single coexpressed pair in either Arabidopsis or rice of which both genes are orthologous to a gene in a coexpressed pair in the other species. Therefore, there are no syntenic local coexpression domains between Arabidopsis and rice. Although the analyses were performed for only one monocot and one dicot and should be extended to many more genomes, the results could be taken to suggest that maintenance of coexpression has not been an important driving force in genome conservation during or after the divergence of dicotyledonous and monocotyledonous plants. Although individual genes in local coexpression domains in either rice or Arabidopsis may have an ortholog in the other species, establishing so-called partially syntenic coexpression domains (PSCDs), this does not seem to occur above chance in the context of whole-genome configurations. Without statistical significance, the occurrence of such PSCDs is unlikely to have any evolutionary relevance on a genome-wide scale. Detailed analyses of individual cases and gene locations may suggest the occurrence of local microsynteny and point to chains of evolutionary events in which the conservation of coexpression could be involved. However, more detailed studies are required to assess the functional relevance, if any, of such genomic constitutions.

Acknowledgements We thank Prof. Blake Meyers (University of Delaware) for early access to the MPSS rice expression data and Dr. Ludmila Mlynarova (Wageningen University) for discussion. This research was supported by a program subsidy from the Netherlands Organization for Scientific Research (NWO) and by the Centre for BioSystems Genomics (CBSG), which is part of the Dutch Genomics Initiative.

References

- Altschul SF, Madden TL, Schaffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Berardini TZ, Mundodi S, Reiser L et al (2004) Functional annotation of the Arabidopsis genome using controlled vocabularies. *Plant Physiol* 135:745–755
- 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
- Devos KM, Beales J, Nagamura Y, Sasaki T (1999) Arabidopsis-rice: will colinearity allow gene prediction across the eudicot-monocot divide? *Genome Res* 9:825–829
- Esteller M (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8:286–298
- Fukuoka Y, Inaoka H, Kohane IS (2004) Inter-species differences of co-expression of neighboring genes in eukaryotic genomes. *BMC Genomics* 5:4
- Henderson IR, Jacobsen SE (2007) Epigenetic inheritance in plants. *Nature* 447(7143):418–424
- Hershberg R, Yeger-Lotem E, Margalit H (2005) Chromosomal organization is shaped by the transcription regulatory network. *Trends Genet* 21:138–142
- Hurst LD, Williams EJ, Pal C (2002) Natural selection promotes the conservation of linkage of co-expressed genes. *Trends Genet* 18:604–606
- Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. *Trends Genet* 17:388–391
- Korbel JO, Jensen LJ, von Mering C, Bork P (2004) Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nat Biotechnol* 22:911–917
- Kruglyak S, Tang H (2000) Regulation of adjacent yeast genes. *Trends Genet* 16:109–111
- Ma L, Chen C, Liu X et al (2005) A microarray analysis of the rice transcriptome and its comparison to Arabidopsis. *Genome Res* 15:1274–1283
- Lee HK, Hsu AK, Sajdak J et al (2004) Coexpression analysis of human genes across many microarray data sets. *Genome Res* 14:1085–1094
- Lercher MJ, Blumenthal T, Hurst LD (2003) Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. *Genome Res* 13:238–243
- Mlynarova L, Jansen RC, Conner AJ et al (1995) The MAR-mediated reduction in position effect can be uncoupled from copy number-dependent expression in transgenic plants. *Plant Cell* 7:599–609
- Mlynarova L, Loonen A, Heldens J et al (1994) Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region. *Plant Cell* 6:417–426
- Mlynarova L, Loonen A, Mietkiewska E et al (2002) Assembly of two transgenes in an artificial chromatin domain gives highly coordinated expression in tobacco. *Genetics* 160:727–740
- O'Brien KP, Remm M, Sonnhammer ELL (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res* 33:D476–D480
- Ott RL, Longnecker M (2001) Categorical data. In: Crockett C, Jackson L, Rockwell P et al (eds) An introduction to statistical methods and data analysis. Duxbury, Pacific Grove, CA 93950 USA, pp 482–485
- Ren XY, Fiers MWEJ, Stiekema WJ, Nap JP (2005) Local coexpression domains of two to four genes in the genome of Arabidopsis. *Plant Physiol* 138:923–934

- Roy PJ, Stuart JM, Lund J, Kim SK (2002) Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* 418:975–979
- Salse J, Piegu B, Cooke R, Delseny M (2002) Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *Nucleic Acids Res* 30:2316–2328
- Semon M, Duret L (2006) Evolutionary origin and maintenance of coexpressed gene clusters in mammals. *Mol Biol Evol* 23:1715–1723
- Spellman PT, Rubin GM (2002) Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J Biol* 1:5
- van Drunen CM, Oosterling RW, Keultjes GM et al (1997) Analysis of the chromatin domain organisation around the plastocyanin gene reveals an MAR-specific sequence element in *Arabidopsis thaliana*. *Nucleic Acids Res* 25:3904–3911
- Walia H, Wilson C, Zeng L et al (2007) Genome-wide transcriptional analysis of salinity stressed japonica and indica rice genotypes during panicle initiation stage. *Plant Mol Biol* 63:609–623
- Walter J, Joffe B, Bolzer A et al (2006) Towards many colors in FISH on 3D-preserved interphase nuclei. *Cytog Genet Res* 114:367–378
- Williams EJ, Bowles DJ (2004) Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res* 14:1060–1067
- 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
- Zhang XY, Yazaki J, Sundaresan A et al (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126:1189–1201
- Zilberman D, Gehring M, Tran RK et al (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* 39:61–69