JIPB Journal of Integrative Plant Biology

Research Article

Clustering of Pathogen-Response Genes in the Genome of *Arabidopsis thaliana*

Olga A. Postnikova[1,2], Natalia Y. Minakova[2], Alexander M. Boutanaev[2] and Lev G. Nemchinov[1*]

[1]United States Department of Agriculture, Agricultural Research Service, Plant Sciences Institute, Molecular Plant Pathology Laboratory, Beltsville 20705, USA

[2]Institute of Basic Biological Problems, Russian Academy of Sciences, Pushchino 142292, Russia

[*]Corresponding author

Tel: +301 504 5099; Fax: +301 504 5449; E-mail: lev.nemchinov@ars.usda.gov

Available online on 2 August 2011 at www.jipb.net and www.wileyonlinelibrary.com/journal/jipb

doi: 10.1111/j.1744-7909.2011.01071.x

Abstract

Previously, we used heterologous expressed sequence tag (EST) mapping to generate a profile of 4 935 pathogen-response genes of *Arabidopsis thaliana*. In this work, we performed a computer analysis of this profile, revealing 1 594 non-homologous clustered genes distributed among all *A. thaliana* chromosomes, whose co-regulation may be related to host responses to pathogens. To supplement computer data, we arbitrarily selected two clusters and analyzed their expression levels in *A. thaliana* ecotypes Col-0 and C24 during infection with the yellow strain of *Cucumber mosaic virus* CMV(Y). Ecotype Col-0 is susceptible to CMV(Y), whereas C24 contains the dominant resistance gene *RCY1*. Upon infection with CMV(Y), all clustered genes were significantly activated in the resistant ecotype C24. In addition, we demonstrated that posttranslational histone modifications associated with trimethylation of histone H3 lysine 27 are most likely involved in regulation of several cluster genes described in this study. Overall, our experiments indicated that pathogen-response genes in the genome of *A. thaliana* may be clustered and co-regulated.

Keywords: CMV; H3K27me3; heterologous EST mapping; clustering of defense-related genes.

Postnikova OA, Minakova NY, Boutanaev AM, Nemchinov LG (2011) Clustering of pathogen-response genes in the genome of *Arabidopsis thaliana*. *J. Integr. Plant Biol*. doi :10.1111/j.1744-7909.2011.01071.x

Introduction

The functional and physical clustering of genes, known as operons, are characteristic of prokaryotic genomes (Lee and Sonnhammer 2003). The concepts and consequences of gene clusters in eukaryotic genomes are largely unexplored (Brugge-man et al. 2008; Makino and McLysaght 2009). Nevertheless, accumulating evidence indicates that gene order in eukaryotic genomes is not completely random and that genes with similar expression levels tend to be clustered within the same genomic neighborhoods (Boutanaev et al. 2002; Michalak 2008; Roy et al. 2002). Although clusters may be assembled from paralo-gous genes as a result of gene duplication events, the growing number of examples points to the existence of eukaryotic gene clusters with operon-like features containing functionally related non-homologous genes (Osbourn and Field 2009; Osbourn 2010a). Those include gene clusters responsible for choosing different carbon and nitrogen sources in yeast, for production of antibiotics, toxins, and virulence determinants in filamentous fungi, for generation of defense compounds in plants and for innate and adaptive immunity in animals (Osbourn and Field 2009).

To date, there are several possible molecular mechanisms explaining co-expression of neighboring genes and formation of chromosomal clusters (Williams and Bowles 2004; Sproul et al. 2005). First, the homologous neighboring genes that have arisen through tandem duplications may have similar expres-sion patterns (Williams and Bowles 2004; Sproul et al. 2005). Second, the presence of opened chromatin structures can enable activation and co-expression of non-homologous genes

that are located next to each other on chromosomes. That is, there is a direct relationship between chromatin modifications and co-expression of clustered genes (Williams and Bowles 2004; Sproul et al. 2005). Third, sharing common regulatory elements between non-homologous clustered genes may allow their simultaneous transcription (Ishihara et al. 2004; Sproul et al. 2005). Finally, there might be a predisposition for genes from the same metabolic pathway or genes associated with a certain function, to cluster (Hurst et al. 2004).

Clustering of functionally-related non-homologous genes in the same chromatin domains would indicate that natural selection toward clustering is important (Lee and Sonnhammer 2003). Indeed, one of the possible advantages associated with clustering is co-regulation through a common transcription factor (Osbourn 2010b). Positioning of co-regulated genes in close proximity to each other would require specialized architecture of the respective chromosomal domains to allow contained transcriptional activity. Consequently, unique gene clusters may be activated through alterations of these functional chromatin domains that are responsive to specific environmental factors.

Following recognition of environmental stimuli, chromatin opening activity occurs, leading to the formation of chromatin loops and domains, which are mediated by nucleosome remodeling, histone modifications, enhancers, and locus control regions (Dean 2006). In turn, histone modifications can carry important and complex consequences originating from the presence of histone, or epigenetic code (Nightingale et al. 2006).

It was reported that neighboring genes in *A. thaliana* are co-expressed and could be involved in the same biological processes (Williams and Bowles 2004). Concurrently, authors suggested that common functionality may not be the main cause of co-expression of neighboring genes. Zhan et al. (2006) concluded, based upon the analysis of 128 *A. thaliana* gene-expression profiles from the online AFFYMETRIX microarray data repository, that neighboring genes are located within large, co-expressed chromosomal regions and may comprise over 10% of the *A. thaliana* genome. Among them, highly co-expressed non-homologous genes may form small domains that share functional similarity and are probably regulated by nearby regulatory sequences. These discrete, small domains of neighboring genes, as well as those of higher order chromatin domains, are thought to be responsive to diverse environmental treatments in a genome-wide, programmed manner (Zhan et al. 2006).

In an earlier study, we used heterologous expressed sequence tag (EST) mapping as an alternative approach to microarray technology to generate a profile of 4 935 genes associated with changes caused by infection or defense responses in *A. thaliana* (Boutanaev et al. 2009). To develop an expression profile of pathogen-response genes, publicly available EST databases derived from different plant species infected with a variety of pathogens were used. Resulting profiles obtained by mapping of heterologous ESTs represented putative *Arabidopsis* homologs of the corresponding species.

As a continuation of this work, we performed a computer analysis of the chromosomal distribution of genes associated with response to pathogens in *A. thaliana.* This revealed numerous functionally-related, non-homologous and co-expressed genes that were co-localized in close proximity to each other. These clusters of genes, whose co-regulation may depend on infection with a variety of plant pathogens, were distributed among all chromosomes of *A. thaliana.* Experimental assessment of computer prediction demonstrated that co-regulation of the neighboring gene clusters may be affected by the presence of resistance (*R*)-genes.

To examine if detected gene clusters share similar epigenetic properties that could have a direct influence on their co-regulation and transcription, we also studied posttranslational histone modifications (PTM) in two arbitrarily selected clusters. PTMs represent one of the mechanisms involved in epigenetic control of gene expression and play a critical role in processes affecting chromatin structure and chromatin-mediated epigenetic regulation of transcription in plants (Osbourn and Field 2008; Chinnusamy and Zhu 2009; Alvarez et al. 2010). Using chromatin immunoprecipitation (ChIP), we showed that methylation of the lysine residue at position 27 of histone H3 (H3K27me3) plays a major role in the regulation of some of the clustered genes in response to virus infection in both susceptible and resistant ecotypes.

Results

Clustering of non-homologous, pathogenesis-related genes along the chromosomes of *A. thaliana*

Computer-generated analysis of the chromosomal distribution of genes using EST profiling revealed 1 594 non-homologous clustered genes distributed among all chromosomes of *A. thaliana*, whose co-regulation may be related to defense responses or general host responses to different pathogens. To investigate whether the observed distribution of genes differed from random, we generated a model of stochastic distribution using a random number generator, as described in Materials and Methods.

A comparison between the size distribution of pathogenesis-related gene clusters revealed by EST profiling and estimated for the stochastic distribution is shown on **Figure 1A** and in Supplementary Table S2.

Differences between the stochastic distribution (STD) and the distribution revealed by EST profiling (EPD) were observed both in the number of genes included in clusters and in cluster size. The EPD size distribution of the gene clusters also differs

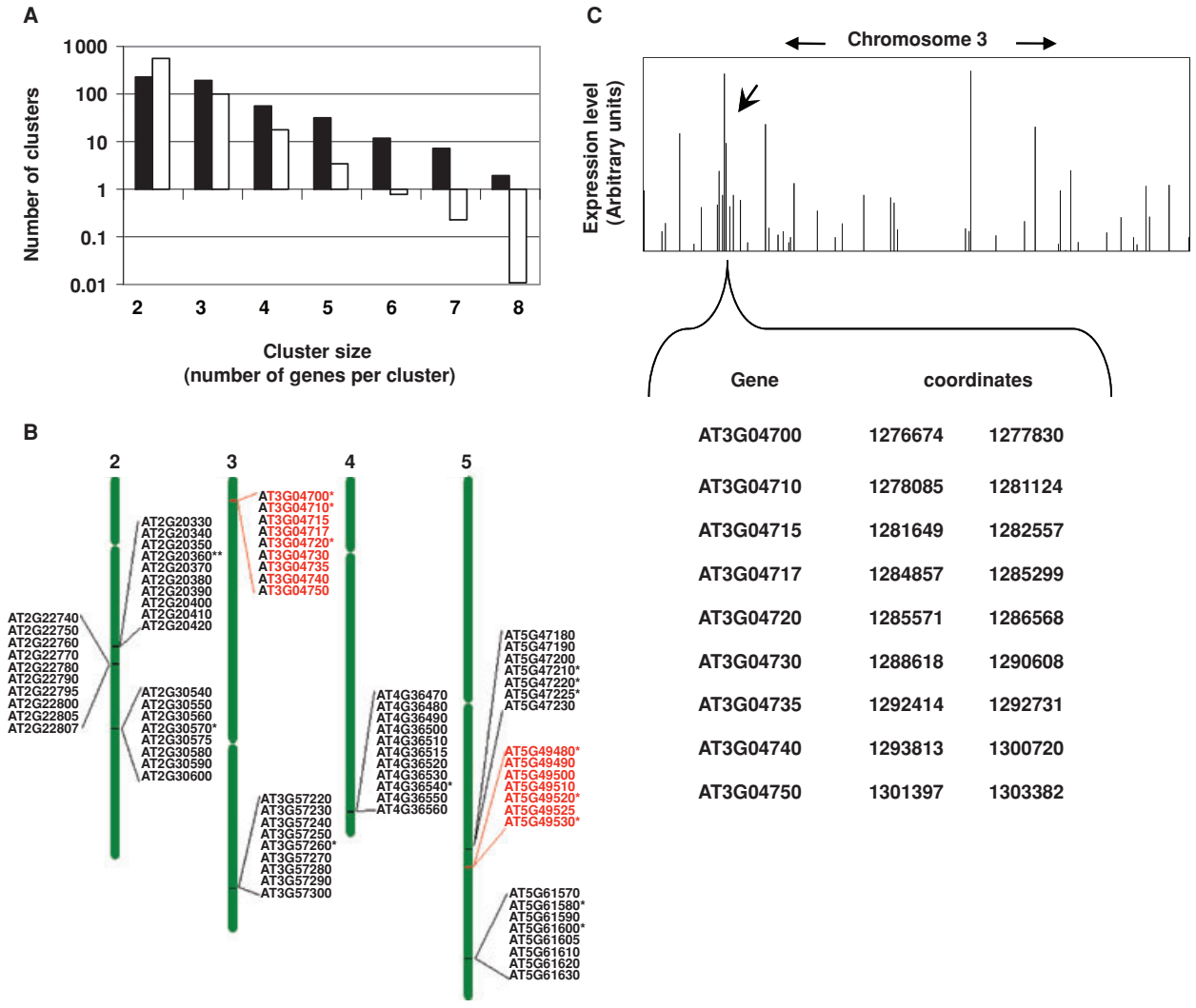

Figure 1. (A) Clusters of non-homologous, pathogen-response genes in the genome *Arabidopsis thaliana*. The height of the bars represents the number of gene clusters of corresponding size revealed by expressed sequence tag (EST) profiling (black) or estimated by stochastic distribution (white). (B) Examples of the large co-expressed, pathogenesis-related clusters located on chromosomes 2, 3, 4 and 5 of *A. thaliana,* as revealed by EST profiling (visualized by Chromosome Map Tool available at The Arabidopsis Information Resource, TAIR). Gene clusters selected for experimental confirmation of computer profiling by quantitative real-time polymerase chain reaction (qRT-PCR) are in red font. Asterisks indicate that this gene was found to be pathogenesis-related according to microarray data presented elsewhere. (C) Expression profile of the pathogenesis-related genes located on chromosome 3 of *A. thaliana,* as revealed by EST profiling. The size of the area is 1.0 Mb. Upregulated genes have positive profile values. The arrow indicates a location of a large cluster of genes, putatively related to defense and/or pathogen response. This cluster was selected to investigate expression levels of its gene-members.

from the STD prediction, especially for clusters of three or more genes. The number of three-gene clusters in the EPD prediction was two times higher than expected by chance. There were three times more four-gene clusters and nine times more five-gene clusters in the EPD as compared with the STD prediction. Six-gene, seven-gene and eight-gene clusters were obtained by EST profiling only; clusters of this size were not predicted to form by chance. A total of 544 clustered pathogenesis-related genes were found by EPD in excess of that expected for STD, of which 57% are included in the clusters of three genes or larger. Evidently, larger clusters correlate to a bigger difference between profiles obtained by EPD and those estimated for the STD of genes participating in response to pathogens.

Clusters with a large number of genes, located on chromosomes 2, 3, 4 and 5 of *A. thaliana,* are shown in **Figure 1B**

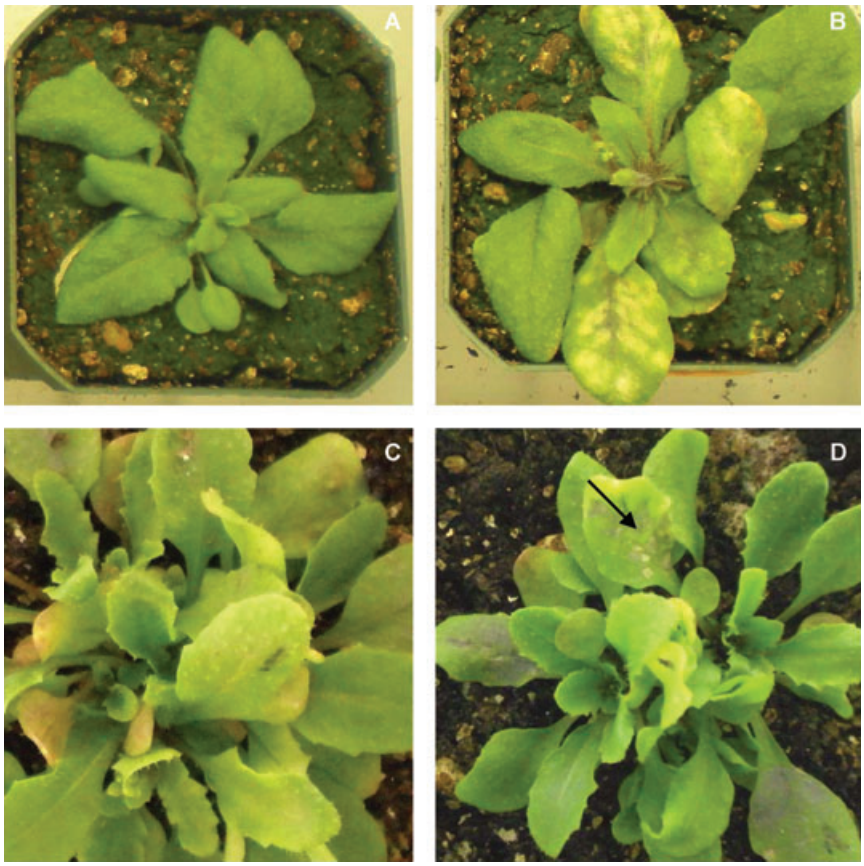

Figure 2. Symptoms on CMV(Y)-infected *Arabidopsis thaliana* plants of ecotypes Col-0 and C24.

(A) Mock-inoculated plant of Col-0.
(B) CMV(Y)-infected plant of Col-0, 7 d post-inoculation (dpi).
(C) Mock-inoculated plant of C24.
(D) CMV(Y)-infected plant of C24, 3 dpi. An arrow indicates the necrotic lesions characteristic of a hypersensitive response (HR) on resistant ecotype C24.

and Supplemental Table S3. Upregulated clustered genes had positive profile values as exemplified by one of the clusters on chromosome 3, which was selected for experimental confirmation of the computer profiling by quantitative real-time polymerase chain reaction (qRT-PCR) (**Figure 1C**). Interestingly, this cluster contains well-known pathogenesis-related genes *PR4* (*AT3G04720*) and *IAA 16* (*AT3G04730*) (Ascencio-Ibáñez et al. 2008; Mukherjee et al. 2010).

Experimental assessment of clustered genes by quantitative real-time polymerase chain reaction

We arbitrarily selected two groups of neighboring genes unveiled by EST profiling and analyzed their expression levels in *A. thaliana* ecotypes Col-0 and C24 during infection with the yellow strain of *Cucumber mosaic virus* CMV(Y) using quantitative real-time polymerase chain reaction (qRT-PCR). *Arabidopsis*/CMV(Y) model (Takahashi et al. 1994) has been established in our laboratory and we have successfully used it in the related study by Boutanaev et al. (2009). Besides, availability of the resistant and susceptible genotypes offered a good possibility to look into the potential role of the *R*-gene in co-regulation of the gene clusters.

The first selected cluster contains nine gene-members and is located on chromosome 3. The second is assembled from seven neighboring genes on chromosome 5. The reason behind selection of these two multi-gene clusters for experimental confirmation of computer data was based on the results of stochastic distribution: clusters of this size were not predicted to form by chance and thus were of particular interest.

After inoculation with a purified preparation of CMV(Y), *A. thaliana* ecotype Columbia 0 (Col-0) plants developed chlorotic symptoms on inoculated leaves 3 d post inoculation (dpi). Symptoms became systemic 5–7 dpi (**Figure 2B**). Virus

presence in the inoculated *A. thaliana* plants was confirmed by reverse transcription polymerase chain reaction (RT-PCR) with CMV-specific primers (not shown). Plants of the resistant ecotype C24 developed characteristic necrotic spots at 3 dpi on inoculated leaves only (**Figure 2D**). After about 7 dpi, inoculated leaves began to dry up and most of them eventually died off (not shown).

We compared expression levels of EST predicted cluster genes in both ecotypes at 1, 3, 5 and 7 dpi. Our preliminary data showed that genes were activated the most at the 5 dpi time point (not shown).

As shown in **Figure 3**, all of the neighboring genes in each of the selected computer-predicted clusters were significantly upregulated in ecotype C24 in response to CMV(Y), as evidenced by the CMV/mock ratio. In susceptible ecotype Col-0, expression levels were on the same level in the selected cluster on chromosome 5, except for two upregulated genes, *ATCP* and WRKY*48*. In contrast, a majority of the genes located in the chromosome 3 cluster were upregulated but half of them on a significantly lower level as compared to C24. Also, two genes of this cluster were downregulated (*IAA16* and *AT3G04740*).

We found that two cluster genes, *AT3G04715* (similar to *MAP3K* alpha 1 protein kinase) and *AT3G04717* (wound-induced protein, WIP-family protein), defined as pseudogenes in The Arabidopsis Information Resource (TAIR), are expressed and upregulated in response to viral infection in both ecotypes, according to our qRT-PCR data.

Next, we made an attempt to determine variation in gene expression outside of the experimentally confirmed clusters in order to define their boundaries. Our data demonstrated that the size of the clusters on chromosomes 3 and 5 can be extended up to 100 kb (**Figure 4**), due to similar expression levels as compared to the other gene neighbors. Nevertheless, these data do not contradict results published elsewhere. For instance, Williams and Bowles (2004) found that there are local clusters of up to 20 genes in the genome of *A. thaliana* with an overall median cluster size of 100 kb. In addition, formation of the gene clusters may also be related to the changes in specific chromatin structure, since shared chromatin environment can facilitate coordinated expression (Wit and Steensel 2009).

Thus, we experimentally confirmed the computer-predicted clustering of pathogen-responsive genes in a random model of host-pathogen interactions. Two selected groups of neighboring genes comprising respective clusters on chromosomes 3 and 5 were constitutively expressed in the resistant ecotype C24 in response to the viral infection.

Changes in histone H3 lysine 27 trimethylation patterns induced by infection with CMV(Y)

Recent studies indicated that regulation of stress-responsive genes often depends on chromatin remodeling (Chinnusamy and Zhu 2009). To find out what kind of chromatin modifications participate in regulation of pathogenesis-related genes and to assess their potential role in cluster formation, we studied the dynamics of H3K27me3 in the coding regions of several genes that were experimentally confirmed to be part of the two described clusters on chromosomes 3 and 5. Our data showed that a decrease in H3K27me3 of the majority of the analyzed genes was a characteristic response to viral infection in both susceptible and resistant ecotypes (**Figure 5**). The decrease in H3K27me3 clearly correlates with the results on gene expression in two defined clusters. This observation agrees with a known role of H3K27me3 as a key marker of epigenetic repression (Zhang et al. 2007).

According to Zhang et al. (2007) who performed a genome-wide identification of regions containing H3K27me3 in *A. thaliana* using ChIP and high-density Affymetrix whole-genome tiling microarrays (ChIP-on-chip), only WRKY*48* and AGAMOUS-like 83 (*AGL83*) are the targets for H3K27me3. However, we revealed in this study that H3K27me3 is normally present at higher levels at *AT3G04715* as well, and that its level significantly decreased (fourfold) in response to CMV(Y). The disagreement between our ChIP/Q-PCR results and ChIP-microarray data of Zhang et al. (2007) may possibly be explained by different methodology, the developmental stage at which tissues were harvested, and growth conditions. Interestingly, according to our qRT-PCR data, *AT3G04715* is overexpressed in both ecotypes in response to CMV(Y).

We noticed that for some of the genes we studied, for example *IAA16* in Col-0, a decreased level of H3K27me3 in response to CMV(Y) does not cause elevation of their expression. Presumably, this fact may be explained by the low basal level H3K27me3 in these genes. Besides, according to Zhou et al. (2010), *IAA16* and *At3G04740* are targets for histone H3 lysine 9 acetylation, which can possibly play a more dominant role in their regulation than H3K27me3.

We also demonstrated that the degree of modifications in H3K27me3 is not related to the changes in number of nucleosomes since according to ChIP there are no significant differences between abundance of H3 in virus-infected and control plants (Supplemental Figure S1).

Therefore, our results suggest that PTMs associated with H3K27me3 occur in the majority of the clustered genes in both ecotypes and most likely take part in regulation of the genes responsive to the viral infection.

Discussion

Gene expression in eukaryotes is thought to be coordinated at two basic levels: locally, at the level of individual genes through cis-regulatory sequences and corresponding regulatory proteins, and epigenetically or globally through the non-random

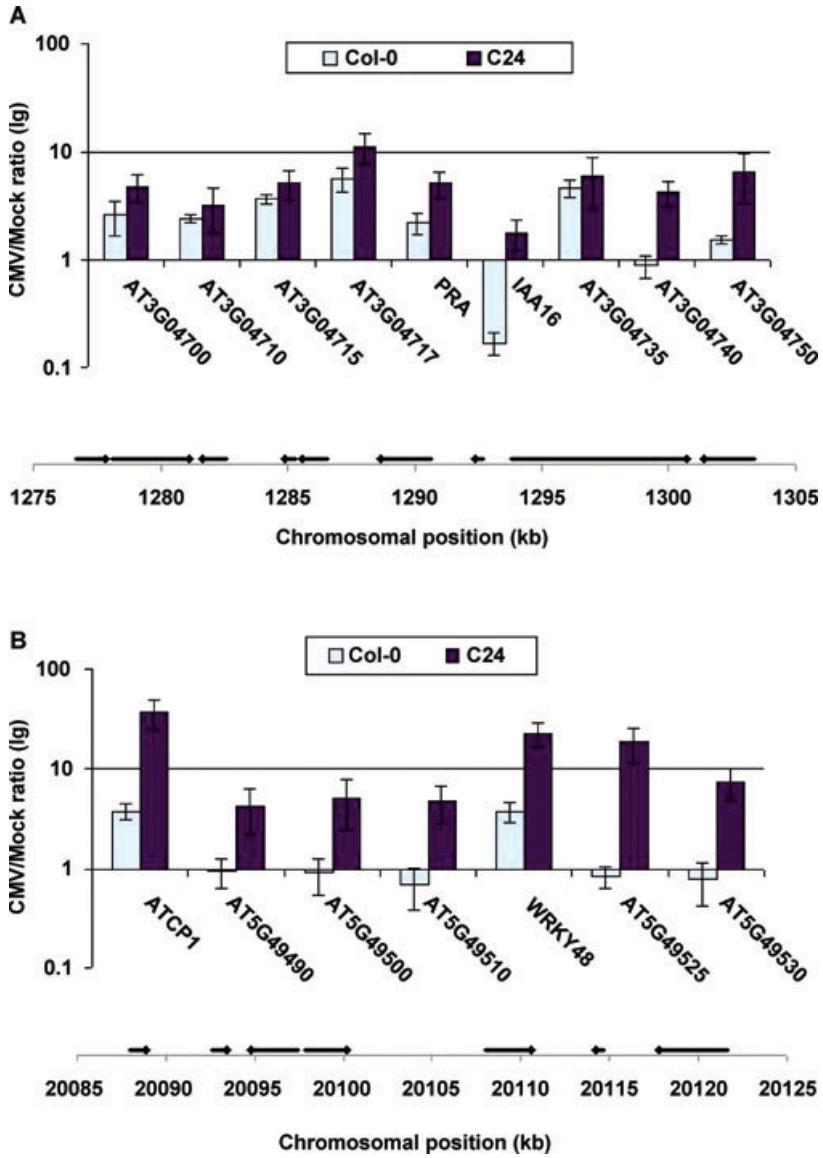

Figure 3. Experimental assessment of computer profiling by quantitative real-time polymerase chain reaction (qRT-PCR) using the yellow strain of *Cucumber mosaic virus* and *Arabidopsis thaliana* ecotypes C24 (contains dominant resistance gene *RCY1*) and Col-0 (susceptible to the virus), as a model of host-pathogen interaction.

The two clusters located on chromosomes 3 **(A)** and 5 **(B)** were randomly selected for the experiment. Gene order corresponds to their actual position on the chromosomes. In **(A)** and **(B)**, black columns (purple columns for online version) represent the ratio of gene expression levels between CMV infected and mock inoculated plants in ecotype C24 whereas grey columns (blue columns for online version) represent expression ratio of the same genes in Col-0. Bars designate confidence intervals ($P = 0.05$). The number of bioreplicates was no less than 3. Gene expression levels were examined on the 1st, 3rd, 5th and 7th dpi with CMV(Y), although only data obtained from 5dpi plants are presented here. On the bottom is shown the position of genes corresponding to their actual position on the chromosomes while arrows indicate the direction of transcription.

distribution of genes, their specific chromosomal locations, structural variations of chromatin domains, and changes in the physical properties of entire chromosomes (Grewal and Moazed 2003; Zhan et al. 2006; Bruggeman et al. 2008).

Our work demonstrates, both by computer-generated analysis of the chromosomal distribution of genes and by experimental assessment of specific host-pathogen interaction, that non-homologous genes associated with response to pathogens in

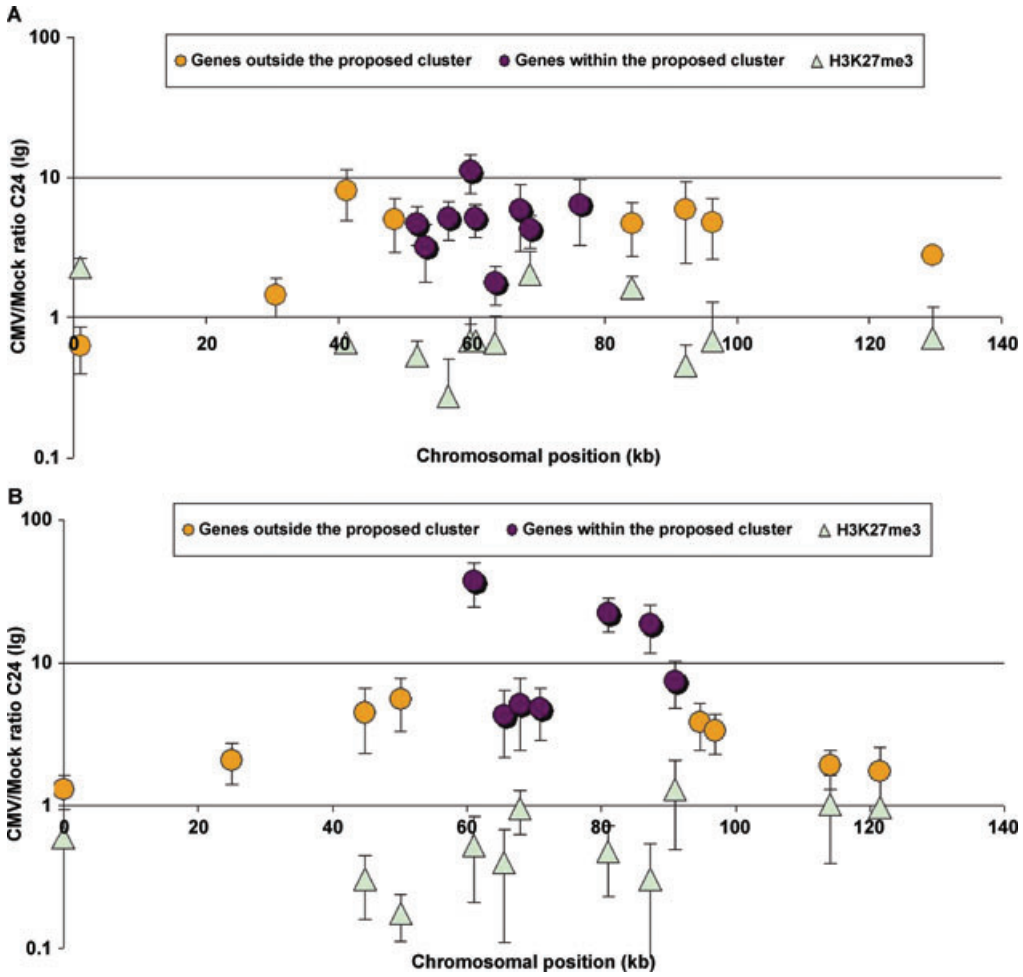

Figure 4. Identification of cluster boundaries by quantitative real-time polymerase chain reaction (qRT-PCR) using CMV(Y) and *Arabidopsis thaliana* ecotype C24.

The ratio of gene expression level between CMV infected and mock inoculated plants within clusters found by expressed sequence tag (EST) profiling and located on chromosomes 3 (**A**) and 5 (**B**) are shown as black circles (purple circles for online version). Grey circles (orange circles for online version) represent their neighboring genes. The circles' locations correspond to their actual position on the chromosomes. Triangles represent the ratio of trimethylation of histone H3 lysine 27 (H3K27me3) between CMV infected and mock inoculated plants. The abscissa zero point is the leftmost coordinate of a minimally responded gene adjacent to the chromosomal cluster. Bars designate confidence intervals ($P = 0.05$).

A. thaliana are co-expressed and located in close proximity to each other. Numerous clustered genes, whose co-regulation is related to defense and/or pathogenesis, were distributed among all chromosomes of *A. thaliana*. The clustering of genes to a greater extent than is predicted by chance may reflect underlying mechanisms of transcriptional co-regulation (Boutanaev et al. 2002).

According to our estimates, the boundaries of the prospective pathogenesis-related gene clusters are more likely to be diffuse rather than distinct and may be directly related to the changes in the epigenetic chromatin environment or in gene expression level. Future investigation may shed additional light on the exact size of the described gene clusters and on the mechanisms of their regulation. This work is currently in progress in our laboratory.

To examine if clusters of co-expressed genes share similar epigenetic properties that could have a direct influence on their co-regulation and transcription, we studied trimethylation of Lys 27 on histone 3, an epigenetic mark of repressive chromatin state and gene expression. Our experiments showed that a decrease in H3K27me3 was a characteristic response to viral infection in both susceptible and resistant ecotypes in the majority of the genes we analyzed. This indicated that a decrease in the level of trimethylation of cluster genes

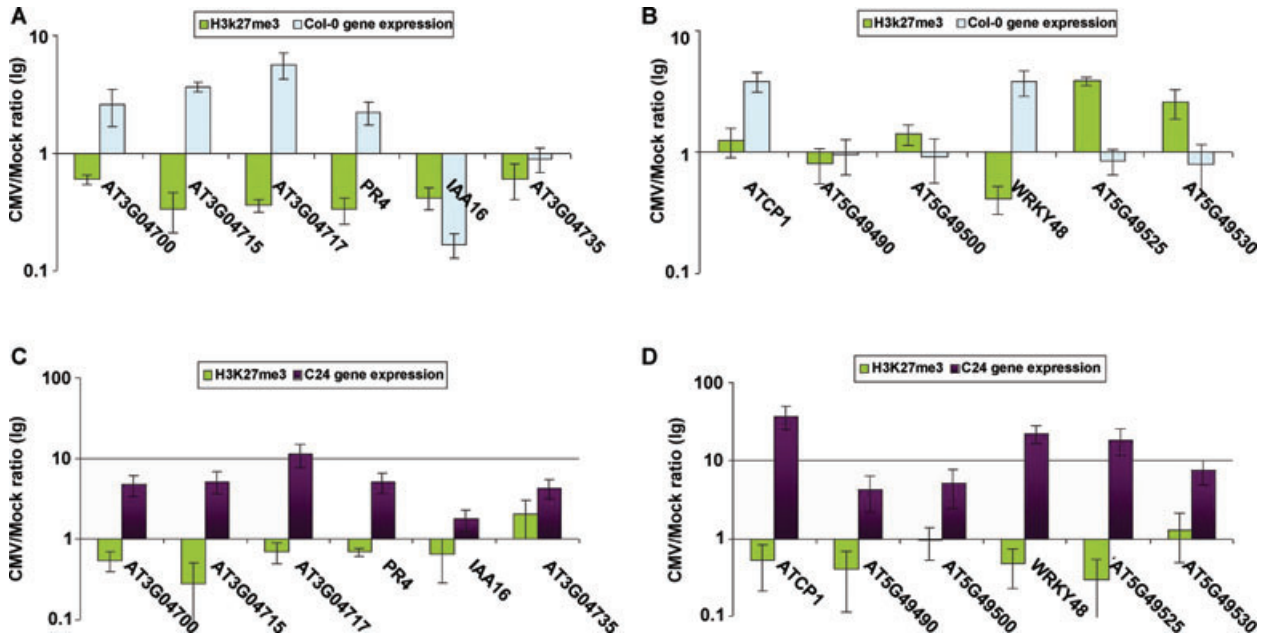

Figure 5. A comparison between changes in gene expression and changes in trimethylation of histone H3 lysine 27 (H3K27me3) in response to CMV(Y) infection.

Light grey columns (Blue columns for online version) in **(A)** and **(B)** represent expression of cluster genes in Col-0; black columns (purple columns for online version) in **(C)** and **(D)** represent expression of the same genes in C24. The level of H3K27me3 is shown in **(A, B, C** and **D)** as dark grey columns (green columns for online version). Bars designate confidence intervals ($P = 0.05$). Gene order corresponds to their actual position on the chromosomes.

correlates with their overexpression and may be reflective of involvement of this particular modification in regulation of responses to pathogenesis.

Formation of gene clusters and their regulation may be directly related to structural arrangements of chromatin within the nucleus. According to one model of nuclear organization, eukaryotic genomes are functionally compartmentalized into chromatin domains by attachment of specific chromosomal DNA regions called scaffold/matrix attachment regions (S/MARs) to a supporting structure that consists of protein and RNA (Bode et al. 2003). The nuclear matrix plays an important role in regulation of nuclear activities, including DNA replication and transcription, nuclear and chromatin organization, cell cycle regulation, cell development, and differentiation (Mattout-Drubezki and Gruenbaum 2003; Tetko et al. 2006). To look into the potential role of S/MARS in regulation of the two selected clusters, we examined the distribution of S/MARS in the chromosome region of interest by using the coordinates of start and end position of predicted S/MARs element in *A. thaliana* (Rudd et al. 2004). We revealed the presence of two large S/MAR elements, *At5SMAR3683* 2.5 kb and *At5SMAR3656* 2.2 kb, at both ends of the experimentally-tested cluster in chromosome 5. S/MARS elements (At5SMAR3517 3,6kb and At5SMAR3526 1,1 kb) were also located at the borders of another computer-predicted cluster on the same chromosome. Distribution of S/MARs was random in the chromosome 3 cluster. Although subject to further tests and verification, these data may suggest a direct involvement of the nuclear matrix in structural organization and regulation of clustered genes.

Previously, Marathe et al. (2004) suggested that genes participating in *R* gene-mediated response to bacterial and viral pathogens may form small, pathogen-responsive clusters in the *A. thaliana* genome. They identified a total of 444 differentially regulated genes as belonging to the resistome associated with the *RCY1* gene in ecotype C24. Non-random chromosomal distribution and transcriptional co-regulation of these genes were implied but not clearly proven (Marathe et al. 2004). Global gene expression in ecotypes resistant and susceptible to CMV(Y) was analyzed using a microarray analysis of approximately 9 000 genes (Ishihara et al. 2004). The authors attempted to group genes into clusters based on their altered expression pattern. However, the roles and implications of chromosomal gene clustering in molecular mechanisms of host-pathogen interactions were neither discussed nor mentioned.

Here, we provide direct evidence that non-homologous genes associated with response to pathogens in *A. thaliana* appeared to be co-expressed and physically clustered. Our

results also suggest a potential role of the *R*-gene in regulation of clustered genes since most of them, at least in two experimentally tested clusters, were upregulated in the resistant ecotype C24. The *R* genes are presumed to activate a signaling cascade that coordinates plant defense responses to block pathogen spread, resulting in an incompatible interaction (Kang et al. 2005). We are currently using transgenic Col-0 line transformed with the *RCY1* gene to elucidate the exact role of the *R*-gene in these processes (Postnikova and Nemchinov 2011) as well as to reveal a possible influence of natural variations in the genetic background of *A. thaliana* ecotypes on the regulation of clustered genes (Chen et al. 2005).

It is important to emphasize that gene-profiling used in this study is based on the EST database derived from different plant species infected with a variety of pathogens – fungi, bacteria, and viruses (Boutanaev et al. 2009). Consequently, a possibility exists that each type of host-pathogen interaction involving different groups of microorganisms or even individual pathogens may have a unique pattern of chromosomal gene clustering, which may or may not overlap with the one reported in this work.

To summarize, clustering of pathogen-response genes may be one of the adaptive mechanisms coordinating a fast reaction to pathogen attack and environmental stress at large. Whether a portion of clustered genes within the same open chromatin domain is directed toward a specific biological function while the remaining genes play a supplemental but essential role or are involved in a different but interdependent pathway remains to be clarified.

Materials and Methods

EST mapping and building of gene expression profiles

Gene profiling was performed using a software package that attributes each EST from the database to a gene in the annotated *A. thaliana* genome. The software also generates a table with the numbers of ESTs found for each gene, as described previously (Boutanaev et al. 2002, 2009). Briefly, EST profiling consisted of the following three steps (Boutanaev et al., 2009): (i) BLAST homology search in the annotated genome using EST data derived from different sources; (ii) assignment of each EST to a corresponding gene based on coordinates of the homology region present in the BLAST output file and building of an expression profile by assigning each gene a fraction of homologous EST (a value of expression level); (iii) building of a specific expression profile by the subtraction of a number of profiles originating from different EST sets from the profile in question.

The following publicly available databases were used: Genbank at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) and The *Arabidopsis* Information Resource (http://www.arabidopsis.org).

Stochastic distribution

The stochastic distribution was generated by producing 1 594 random, non-repetitive numbers in the range of 1–32 943 (the total number of genes in the *A. thaliana* genome). With the assumption that the row of numbers from 1 to 32 943 comprises the order of genes in the genome of *A. thaliana*, each reiteration assigns random genomic positions to the 1 594 pathogenesis-related genes. The proportion of genes found in clusters and the size distribution of clusters were calculated, and the values were averaged for 50 reiterations.

Plant growth, virus purification and inoculation of plants

Arabidopsis thaliana plants (ecotype Columbia-0, catalog # WT-02–38-05, accession Col-0/Redei-L206440 and ecotype C24, catalog # WT-23–02-01, accession L203288) were obtained from Lehle Seeds, Round Rock, TX, USA and grown to 21 d of age in 2-inch square pots at a density of two to three plants per pot in a Percival growth chamber (model AR36LC8, Percival Scientific, Perry, IA, USA), which was set for a 16 h photoperiod and 22 °C, 70% RH. A yellow strain of CMV was kindly provided by Dr H. Takahashi from the Department of Life Sciences in the Graduate School of Agricultural Science of Tohoku University, Sendai, Japan. The virus was propagated in *Nicotiana benthamiana* and purified by differential centrifugation as described by Lot et al. (1972) with minor modifications. The quality of the purified viral preparation was additionally confirmed by transmission electron microscopy and Western blotting with CMV(Y)-specific antibodies (a gift of H. Takahashi, not shown). Virions were resuspended to approximately 0.5 mg/mL in 20 mM potassium phosphate buffer (pH 7.2), and 10 μL of the solution was rub-inoculated onto leaves of 3-week-old plants marked and dusted with carborundum. Control plants were mock inoculated with phosphate buffer alone.

RNA extraction, first-strand synthesis and qRT-PCR

Five inoculated leaves on each of 6–10 plants per time point were harvested at 1, 3, 5, and 7 dpi and immediately processed. Total RNA was extracted using the Fast RNA Pro Green Kit (MP Biomedicals, Solon, OH, USA). Copy DNA was synthesized using the SuperScript First-Strand cDNA Synthesis System according to the manufacturer's directions (Invitrogen). qRT-PCR was performed with iQ SYBR Green Supermix kit (Bio-Rad Laboratories, Inc., Hercules, CA, USA) on the MiniOpticon Real-Time PCR system (Bio-Rad) using the following parameters: 94 °C – 1 min (one cycle); 94 °C – 30 s, 60 °C – 30 s, 72 °C – 30 s (30 cycles). Amplification was performed in three to five biological replicas and two technical replicas. The *A. thaliana* actin gene *UBQ5* (*AT3G62250*) was used as a reference in all qRT-PCR experiments. The specificity of the amplifications

was confirmed by the single-peak melting curves of the PCR products.

Quantitative RT-PCR ratio was calculated as a ratio between each of the CMV-infected samples and a corresponding average of the mock-inoculated samples. To obtain a final ratio for any given gene, we calculated average and *SD* from all biological replicas.

ChIP sample preparation

Chromatin immunoprecipitation was performed as described in Gendrel et al. (2005) and Haring et al. (2007) with the following modifications. Calibration curves were constructed before immunoprecipitation experiments to determine the optimal amounts of chromatin to be used in each experiment. Antibodies used in this study were from Abcam: H3K27me3 (ab6002) and core H3 (ab1791). Chromatin was isolated from 1.5–2 g rosette leaves from experimental and control plants. After reverse cross-linking and proteinase K treatment, the immunoprecipitated DNA and input were purified with Qiagen gel extraction kit. The resulting DNA was subjected to quantitative real-time PCR analysis. Each immunoprecipitation experiment was independently performed two to three times with separately isolated biological samples. ChIP-qRT-PCR results were calculated. Input Control DNA alongside the immunoprecipitated samples was run for each primer pair. The amplification efficiency of each primer pair was determined using 10-fold serial dilution. Normalized input was calculated by the formula: $100 \times \text{AE}^{(\text{CT INPUT} - \text{CT IP})}$, where AE is the amplification efficiency.

Primer design

Primers for cDNA qRT-PCR were designed using sequences of the last two exons of each gene (when possible) to ensure amplification of the mRNA only. On average, the length of the amplified fragments was 200–300 base pairs (Supplementary Table S1). Primers for ChIP real-time PCR were designed within exons and the length of the amplified fragments was 100–150 base pairs.

Acknowledgements

We are grateful to Hideki Takahashi of Tohoku University, Japan for providing the CMV(Y) strain and *RCY1*-transformed lines. We thank Wesley Schonborn for critical reading of the manuscript and language editing. This work was supported by the United States Department of Agriculture, Agricultural Research Service.

Received 11 May 2011 Accepted 14 Jul. 2011

References

Alvarez ME, **Nota F**, **Cambiagno DA** (2010) Epigenetic control of plant immunity. *Mol. Plant Pathol.* **11**, 563–576.

Ascencio-Ibáñez JT, **Sozzani R**, **Lee TJ**, **Chu TM**, **Wolfinger RD**, **Cella R**, **Hanley-Bowdoin L** (2008) Global analysis of *Arabidopsis* gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant Physiol.* **148**, 436–454.

Bode J, **Goetze S**, **Heng H**, **Krawetz SA**, **Benham C** (2003) From DNA structure to gene expression: Mediators of nuclear compartmentalization and dynamics. *Chromosome Res.* **11**, 435–445.

Boutanaev AM, **Postnikova OA**, **Nemchinov LG** (2009) Mapping of heterologous expressed sequence tags as an alternative to microarrays for study of defense responses in plants. *BMC Genomics* **10**, 273.

Boutanaev AM, **Kalmykova AI**, **Shevelyov YY**, **Nurminsky DI** (2002) Large clusters of co-expressed genes in the *Drosophila* genome. *Nature* **420**, 666–669.

Bruggeman FJ, **Oancea I**, **van Driel R** (2008) Exploring the behavior of small eukaryotic gene networks. *J. Theor. Biol.* **252**, 482–487.

Chen WJ, **Chang SH**, **Hudson ME**, **Kwan WK**, **Li J**, **Estes B**, **Knoll D**, **Shi L**, **Zhu T** (2005) Contribution of transcriptional regulation to natural variations in *Arabidopsis*. *Genome Biol*. **6**, R32.

Chinnusamy V, **Zhu JK** (2009) Epigenetic regulation of stress responses in plants. *Curr. Opin. Plant Biol*. **12**, 133–139.

Dean A (2006) On a chromosome far, far away: LCRs and gene expression. *Trends Genet.* **22**, 38–45.

Gendrel AV, **Lippman Z**, **Martienssen R**, **Colot V** (2005) Profiling histone modification patterns in plants using genomic tiling microarrays. *Nat. Methods* **2**, 213–218.

Grewal SIS, **Moazed D** (2003) Heterochromatin and epigenetic control of gene expression. *Science* **301**, 798–802.

Haring M, **Offermann S**, **Danker T**, **Horst I**, **Peterhansel C**, **Stam M** (2007) Chromatin immunoprecipitation: Optimization, quantitative analysis and data normalization. *Plant Methods* **24**, 3–11.

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

Ishihara T, **Sakurai N**, **Sekine KT**, **Hase S**, **Ikegami M**, **Shibata D**, **Takahashi H** (2004) Comparative analysis of expressed sequence tags in resistant and susceptible ecotypes of *Arabidopsis thaliana* infected with *Cucumber mosaic virus*. *Plant Cell Physiol.* **45**, 470–480.

Kang BC, **Yeam I**, **Jahn MM** (2005) Genetics of plant virus resistance. *Annu. Rev. Phytopathol.* **43**, 581–621.

Lee JM, **Sonnhammer ELL** (2003) Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res.* **13**, 875–882.

Lot H, **Marrou J**, **Quiot JB**, **Esvan C** (1972) A contribution to the study on *Cucumber mosaic virus* (CMV). II. Quick method of purification. *Ann. Phytopathol*. **4**, 25–38.

Makino T, **McLysaght A** (2009) The evolution of functional gene clusters in eukaryote genomes. In: Pontarotti P, ed. *Evolutionary Biology*. Springer-Verlag, Berlin. pp. 185–194.

Marathe R, **Guan Z**, **Anandalakshmi R**, **Zhao H**, **Dinesh-Kumar SP** (2004) Study of *Arabidopsis thaliana* resistome in response to cucumber mosaic virus infection using whole genome microarray. *Plant Mol. Biol*. **55**, 501–520.

Mattout-Drubezki A, **Gruenbaum Y** (2003) Dynamic interactions of nuclear lamina proteins with chromatin and transcriptional machinery. *Cell Mol. Life Sci*. **60**, 2053–2063.

Michalak P (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* **91**, 243–248.

Mukherjee AK, **Carp MJ**, **Zuchman R**, **Ziv T**, **Horwitz BA**, **Gepstein S** (2010) Proteomics of the response of *Arabidopsis thaliana* to infection with *Alternaria brassicicola*. *J. Proteomics* **73**, 709–720.

Nightingale KP, **O'Neill LP**, **Turner BM** (2006) Histone modifications: Signalling receptors and potential elements of a heritable epigenetic code. *Curr. Opin. Genet. Dev*. **16**, 125–136.

Osbourn AE, **Field B** (2008) Metabolic diversification – independent assembly of operon-like gene clusters in different plants. *Science* **320**, 543–547.

Osbourn AE, **Field B** (2009) Operons. *Cell Mol. Life Sci.* **66**, 3755–3775.

Osbourn A (2010a) Secondary metabolic gene clusters: Evolutionary toolkits for chemical innovation. *Trends Genet*. **10**, 449–457

Osbourn A (2010b) Gene clusters for secondary metabolic pathways: An emerging theme in plant biology. *Plant Physiol*. **154**, 531–535.

Roy PJ, **Stuart JM**, **Lund J**, **Kim SK** (2002) Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* **418**, 975–979.

Rudd S, **Frisch M**, **Grote K**, **Meyers BC**, **Mayer K**, **Werner T** (2004) Genome-wide *in silico* mapping of scaffold/matrix attachment regions in *Arabidopsis* suggests correlation of intragenic scaffold/matrix attachment regions with gene expression. *Plant Physiol*. **135**, 715–722.

Sproul D, **Gilbert N**, **Bickmore WA** (2005) The role of chromatin structure in regulating the expression of clustered genes. *Nat. Rev. Genet*. **6**, 775–781.

Takahashi H, **Goto N**, **Ehara Y** (1994) Hypersensitive response in cucumber mosaic virus-inoculated *Arabidopsis thaliana*. *Plant J*. **6**, 369–377.

Tetko IV, **Haberer G**, **Rudd S**, **Meyers B**, **Mewes HW**, **Mayer KF** (2006) Spatiotemporal expression control correlates with intragenic scaffold matrix attachment regions (S/MARs) in *Arabidopsis thaliana*. *PLoS Comput. Biol*. **2**, e21.

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

Wit E, **Steensel B** (2009) Chromatin domains in higher eukaryotes: Insights from genome-wide mapping studies. *Chromosoma* **118**, 25–36.

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

Zhang X, **Clarenz O**, **Cokus S**, **Bernatavichute YV**, **Pellegrini M**, **Goodrich J**, **Jacobsen SE** (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol.* **5**, e129.

Zhou J, **Wang X**, **He K**, **Charron JB**, **Elling AA**, **Deng XW** (2010) Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in *Arabidopsis* reveals correlation between multiple histone marks and gene expression. *Plant Mol. Biol*. **72**, 585–595.

(Co-Editor: Christophe Robaglia)

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Primers used for qRT-PCR.

Table S2. Distribution of pathogenesis-related gene clusters revealed by EST profiling and estimated for the stochastic distribution.

Table S3. GO terms of the cluster genes.

Figure S1. Level of nucleosome density of cluster genes in mock and CMV inoculated plants.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.