Regulatory DNA binding peptides as novel tools for plant functional genomics

Lindhout, B.I.

Chapter 5

VP16-HRU, a zinc finger artificial transcription factor affecting somatic homologous recombination

Beatrice I. Lindhout, Johan E. Pinas, Paul J.J. Hooykaas and Bert J. van der Zaal

SUMMARY

In previous work we selected a particular transcription factor, designated VP16-HRU, from a pool of zinc finger artificial transcription factors. When expressed in *Arabidopsis thaliana* under control of the ribosomal protein S5A promoter, the *RPS5A::VP16-HRU* construct led to a 200- to 300-fold increase in the frequency of somatic homologous recombination. Here we report on further phenotypic characterization and transcriptome analyses of Arabidopsis mutants containing *RPS5A::VP16-HRU*. Apart from the previously characterized increase in recombination events, the macroscopically visible phenotype was remarkably mild. Mutant seedlings exhibited slightly epinastic leaves and increased anthocyanin production under high light intensity and flowered late. Transcriptome analysis demonstrated that numerous genes were up- or downregulated in the mutant lines. Some of these differentially expressed genes encoded proteins known to be involved in DNA repair, such as RAD51 and PARP2. By careful analysis of a series of micro-array experiments we selected a number of genes that were tested for their individual effects on somatic homologous recombination. *RPS5A* promoter-mediated expression of cDNAs of At1g66300, an F-box protein, and At5g60250, a C3HC4 zinc finger protein, had a stimulatory effect of the somatic HR frequency in 10-day-old primary transformants. When 20-day-old primary transformants were assayed for HR events, an increase in HR frequency could be found for several other cDNA expression constructs as well.

INTRODUCTION

The Cys_2His_2-type of DNA-binding zinc finger (ZF) domain is commonly found within eukaryotic transcription factors, each ZF of this type typically binding 3 contiguous bases of DNA (Pavletich and Pabo, 1991). By large efforts of several research groups, most notably the one led by Barbas at the Scripps Institute, the recognition code for ZF binding sites is now almost complete. By linking individual ZF domains to form a polydactyl zinc finger (PZF) domain, increased binding specificity for a particular DNA sequence can be achieved. Fusion of PZF domains with effector domains, that either stimulate or repress gene transcription, subsequently allows the construction of zinc finger artificial transcription factors (ZF-ATFs) (http://www.zincfingers.org; Wright *et al.*, 2006).

In theory, it is nowadays possible to design a ZF-ATF for binding and regulation of almost any genomic target sequence.

Another intriguing possibility to employ ZF-ATFs in genomic research is the introduction of libraries of different ZF-ATFs into an organism of interest. When each individual ATF is equipped with a relatively low complexity PZF domain, consisting of three to four ZFs, each member of the ZF-ATF library can potentially regulate numerous endogenous genomic loci. Introducing such a library into host cells has been referred to as 'genome interrogation', a kind of brute force approach to reveal hidden phenotypic possibilities as novel dominant traits (reviewed in Beltran *et al*., 2006). Once a desired phenotype is selected, the causal ZF-ATF can easily be identified, thereby bringing the desired novel trait under experimental control. We recently demonstrated that this approach is also applicable at the level of a multi-cellular eukaryotic organism, in this case the model plant *Arabidopsis thaliana*. In that study (Lindhout *et al*., 2006) a ZF-ATF was identified capable of increasing intra-chromosomal somatic homologous recombination (HR) by on average 200- to 300-fold when assayed in an HR indicator line containing an integrated interrupted *GUS* reporter gene. Moreover, mutant seedlings were more resistant to the DNA double strand break (DSB) inducing compound bleomycin. This specific ZF-ATF, consisting of a VP16 transcriptional activation domain fused with a 3-fingered ZF domain (3ZF) theoretically recognizing the sequence 5'GTGGAGGCT3', was designated VP16-HRU, with HRU referring to the HR-up regulated phenotype.

The original *ZF-ATF* gene constructs used for Arabidopsis mutagenesis (Lindhout *et al*., 2006) were driven by the promoter of the ribosomal protein gene *RPS5A*, which is particularly active in dividing cells, such as present in meristems and during early stages of embryo formation, while much less active in fully differentiated cells (Weijers *et al*., 2001). In this study we report on the further characterization of seedlings containing a *RPS5A::VP16-HRU* transgene, both on a phenotypic level as well as on the level of gene expression since it was reasoned that the spectrum of differentially expressed genes might give some clues which types of processes were stimulated or repressed in mutants with a HRU phenotype. Also such seedlings may be used to analyze altered gene expression which might lead to the identification of the still enigmatic endogenous master regulators of HR as these are now directly or indirectly under *RPS5A::VP16-HRU* control. Apart from experiments with *RPS5A::VP16-HRU* containing plant lines, we also expressed VP16-HRU under control of a tamoxifen-inducible promoter (Friml *et al*., 2004), as this might allow easier identification of the immediate target genes of VP16-HRU. We demonstrate that a large number of genes is differentially expressed due to the presence of

RPS5A::VP16-HRU, including genes known to be involved in homologous recombination and/or DNA repair like *Rad51* and *PARP2* (De Block *et al.*, 2005; Li *et al.*, 2004). Furthermore, we selected a series of genes and tested them for their individual HR-inducing potential via expression of their cognate cDNA sequences via the *RPS5A* promoter in an HR-indicator line. We demonstrated that expression of several of these gene constructs caused an increase in the number of somatic HR events. Although still modest in comparison to the *RPS5A::VP16-HRU* mutant lines, the increase in HR frequency in primary transformants expressing cDNA constructs reached up to five-fold compared to control plants transformed with empty vector. This indicates that PZF-mediated mutagenesis and subsequent analysis of mutants forms a valuable approach to identify novel genes involved in fundamental processes like HR.

RESULTS

Phenotypic characterization of RPS5A::VP16-HRU containing plants

Despite the high frequency of somatic HR, plants expressing *RPS5A::VP16-HRU* never developed severe developmental defects (Lindhout *et al.*, 2006). More careful comparison with the parental line revealed that *RPS5A::VP16-HRU* seedlings were initially developing somewhat slower and were slightly reduced in size (Figure 1, compare A, B with D, E). At later stages, leaves and cotyledons of young *RPS5A::VP16-HRU* seedlings were more epinastic and their petioles were shorter than observed for the paternal line (Figure 1, C and F) Bolting was delayed by about one week (Figure 1G). The delay in flowering time was accompanied by the formation of a larger rosette (data not shown). Furthermore it was noticed that *RPS5A::VP16*-HRU plants contained a two- to three-fold higher level of anthocyanin when grown in increased light intensity (data not shown).

As described previously, introduction of the *RPS5A::VP16-HRU* gene construct into the HR indicator line 1406, containing an integrated interrupted *GUS* reporter gene (Gherbi *et al.*, 2001), led to a dramatic increase in the amount of recombination events (Lindhout *et al.*, 2006). When analysed in more detail, it was found that very young seedlings rarely exhibit signs of recombination events. However, after 7 days of growth mitotic HR events started to become visible in *RPS5A::VP16-HRU* seedlings whereas no GUS positive spots were yet detected in more than 180 seedlings of parental line 1406. At 10 days of growth the average HR frequency found within the selected *RPS5A::VP16-HRU* line had risen dramatically, up to a level over 200-fold higher than observed in

parental seedlings (Figure 2). Throughout seedling development most GUS positive spots were observed within the cotyledons, appearing in leaves predominantly when the leaf primordia started to expand. HR events remained relatively rare in hypocotyls and were hardly found in roots. At later stages of seedling development, the increase in the number HR events seemed to be less pronounced. However, detailed numerical analysis was complicated by the gradual loss of GUS staining in ageing cells, an effect already noticed in cotyledons after 20 days of development (data not shown). Nevertheless, the appearance of numerous HR events at well stained leave margins indicated that the HRU phenotype was present at later stages as well (data not shown).

Most GUS positive spots in the cotyledons and leaves appeared to be constrained to single cells or just a few neighbouring cells. In Arabidopsis and other higher plants, cotyledons develop almost exclusively by means of cell division during embryogenesis while their further growth during and after germination mostly depends on cell expansion (Tsukaya, 2002). The fact that we very rarely found large blue sectors of connected cells proved that most recombination events triggered by the *RPS5A::VP16-HRU* gene occurred well after the wave of cell divisions during organogenesis, probably when cells had already differentiated. Considering the fact that the *RPS5A* promoter is predominantly active during early stages of embryogenesis and subsequently in meristematic regions (Weijers *et al.*, 2001), somatic recombination events thus predominantly occur when *RPS5A::VP16-HRU* gene expression levels have already declined. This implies that the HR phenotype is likely to be an indirect effect of *VP16-HRU* action.

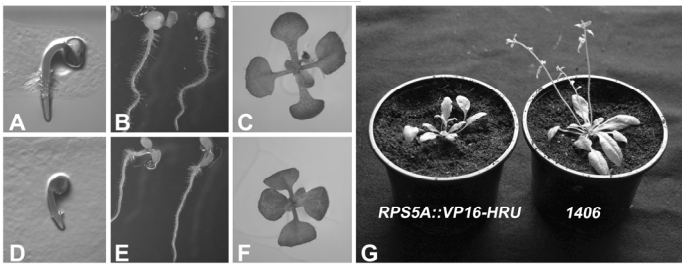


Figure 1: Phenotypic characterization of *RPS5A::VP16-HRU* expressing plantlets compared with the parental line 1406. (A, B and C) 1406 seedlings grown on ½ MS. (D,E and F) *RPS5A::VP16-HRU* seedlings grown on ½ MS. *RPS5A::VP16-HRU* seedlings are slightly delayed and show mild epinastic growth of the cotyledons (F). (G) *RPS5A::VP16-HRU* plants are delayed in flowering time (left plant) compared to the parental line 1406 (right plant).

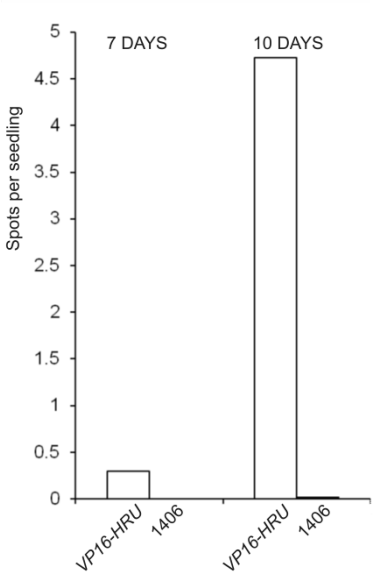


Figure 2: Intra-chromosomal HR in plant lines *RPS5A::VP16-HRU* and parental line 1406. Seedlings were grown on solidified 1/2MS medium at 21 ºC and stained for GUS expression after 7 and 10 days of growth. Recombination frequency is shown as the mean number of GUS positive spots per seedling. At least 50 seedlings were analyzed for each time point.

VP16-HRU effects on the transcriptome, a first analysis

In order to investigate changes in the transcriptome brought about by the expression of the artificial transcription factor VP16-HRU, we performed five micro-array experiments (Table 1) using Agilent® ArabidopsisIII arrays. These arrays, first released in August 2004, contained specific oligonucleotide tags (60 mers) for all the Arabidopsis genes that were known by AGI codes (Atg locus) together with over 10,000 oligonucleotide tags for un-annotated and small non-coding transcripts corresponding to CHR numbers. For reasons of simplicity we will use the term locus identifier (ID) to refer to either of these two. The Agilent® ArabidopsisIII arrays are suitable for dual colour hybridization which allows direct comparison of gene expression in two samples without inter-array variation. The rationales for each of the micro-array experiments are given in further detail below.

Microarrays #1 and #2 can be considered as pilot experiments aimed to get an indication by which strategy the most meaningful information could be obtained from generated data sets. For the first array (array #1), RNA samples were extracted from 10-day old seedlings grown in tissue culture. In this case the *VP16-HRU* gene was driven by the *RPS5A* promoter and the global gene

expression pattern was evaluated against that of seedlings of parental line 1406. This array should in principle indicate to which extent the transcriptome was changed due to the presence of the *RPS5A::VP16-HRU* gene. For micro-array #2, a two-component expression system, in fact forming a tamoxifen-inducible *CaMV35S* promoter (Friml *et al*., 2004) was used to express the VP16-HRU protein (*TAM35S::VP16-HRU*) and the 'free' VP16 protein (*TAM35S::VP16)*, which lacks zinc fingers but still contains the other features like the N-terminal FLAG-tag and the SV40 nuclear localization signal. This experiment was designed to investigate whether or not primary target genes can be found that are up- or downregulated due to a short term presence of the transcription factor VP16-HRU. By inducing 'free' VP16 as a control, differential gene expression patterns should be mainly due to the presence of the PZF domain of VP16-HRU which in principle transforms the VP16 domain into a targeted transcriptional activator. Any 'VP16' effects should therefore be filtered out. In order to dampen any fortuitous signals due to effects related to the positional integration of transgenes we isolated RNA from mixtures of T2 seedlings originating from 50 primary transformants, both for the *TAM35S::VP16-HRU* construct as well as for the *TAM35S::VP16* construct. To determine how long a tamoxifen treatment should last in order to obtain a substantial level of tamoxifen-induced protein within transgenic seedlings, a plant line containing the gene construct *TAM35S::GFP* was used in a time course experiment in liquid culture. It was evident that after 4 hours of induction, protein accumulated to a detectable level (Figure 3) and further increased over time. For the microarray experiment an induction time of 7 hours was used, which seemed to be a reasonable compromise between the shortest possible induction time and the maximal levels of induced VP16-HRU or VP16 protein. Prior to the transcriptome analysis, *TAM35S::VP16-HRU* seedlings were analysed for HR frequency and it was found that neither a 7 hrs induction time nor a prolonged treatment by growing seedlings in the presence of tamoxifen for two weeks could induce a significant increase in the number of HR events comparable with the frequency found in *RPS5A::VP16-HRU* seedlings (data not shown). Although this indicated that a particular spatio-temporal expression of VP16-HRU might be important for the eventual HRU phenotype as will be discussed below, it was reasoned that a transcriptome profile of *TAM35S::VP16-HRU* might still give clues about direct genomic targets of VP16-HRU.

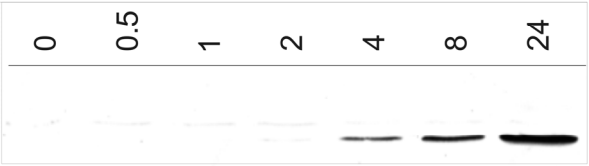


Figure 3: Induction of *TAM35S::GFP* with tamoxifen in Arabidopsis. GFP protein was detected by means Western Blot analysis with an anti-FLAG antibody at 0; 0.5; 1; 2; 4; 8 and 24 hrs post induction with tamoxifen (1µM) in liquid medium

Analyzing the results of microarrays #1 and #2, we found that a large set of locus IDs was up- or downregulated in the seedlings containing *RPS5A::VP16-HRU* (array #1). The number of up- and downregulated locus IDs was 225 and 195, respectively, based on a minimum of 2-fold change with a confidence of 95% ($p \leq 0.05$) (Table 1). However, when *VP16-HRU* was expressed via tamoxifen induction (array #2) only 32 locus IDs were upregulated and 150 were downregulated. We carefully ruled out that any dye bias could have accounted for this lower number of significantly induced locus IDs. As for the idea of identifying primary VP16-HRU target genes, irrespective of a possible role in somatic HR, none of the 32 genes that seemed to be upregulated by induction of VP16-HRU possessed the cognate 9-bp target site for the 3ZF domain, also not within their untranslated regions. Functional categorization of those 26 out of 32 upregulated genes for which data were available in the public domain did not link any of them with DNA or RNA metabolism.

Remarkably, several of the most conspicuously repressed mRNAs after tamoxifen-mediated VP16-HRU expression encoded seed storage proteins: At5g44120, At4g28520, At1g03880, corresponding to 12S seed storage proteins, and At4g27140, At4g27150, At4g27160 and At4g27170, corresponding to a group of 2S seed storage proteins apparently tandemly arranged within the genome. The immediate flanking regions of these genes lack the supposedly optimal 9-bp binding site for the 3ZF moiety of VP16-HRU. Closer inspection of the raw intensity values of the hybridization signals corresponding to these locus IDs in array #1 and #2 revealed that the apparent repression of the seed storage protein genes by tamoxifen induction was largely caused by a high level of mRNA expression induced in *TAM35S::VP16* containing seedlings. The fact that the expression pattern of functionally related genes shares a similar behaviour can be taken as evidence for their common regulation, but in this particular case the reason for downregulation of the seed storage genes remains obscure. Further examination of the 150 genes that were found to be repressed after VP16-HRU induction revealed only one gene that could be linked to DNA

metabolism (At1g30010, encoding an intron maturase). However, evaluation of intensity values indicated that the large repression ratio was based on low intensity, which made the differential expression unreliable.

Analysis of the combined data derived from array #1 and #2 revealed that three annotated genes could be identified in the overlapping fraction of upregulated locus IDs. Two of these genes are involved in the light reaction of photosynthesis (At1g15820 and At2g303570) and, as such, a role in homologous recombination seemed to be rather far-fetched. Moreover, their non-induced level of gene expression was already very high, suggesting that the protein products were unlikely to be rate limiting factors for recombination. A function for the small hypothetical protein putatively encoded by At1g44414 (101 amino acid residues) still has to be found, just as for the 112 amino acid ortholog known from rice. *In silico* analysis using genevestigator (https://www.genevestigator.ethz.ch, Zimmermann *et al.,* 2004) and the bio-array resource for Arabidopsis functional genomics (http://bar.utoronto.ca/, Toufighi *et al*., 2005) indicated that the gene was normally expressed at a low level, but its expression was relatively high in developing embryos, during seed development, as well as within the root endodermis. Furthermore this gene was annotated as being induced by heat and hypoxia. Homologous EST sequences from other species than Arabidopsis and rice had not been reported per November 2007. Concerning locus IDs referring to unannotated transcripts and non-coding RNAs (identified by CHR numbers), hybridization signals were mostly extremely low and therefore within noise level (below 150 fluorescence intensity values, regardless of the sample's origin). The hybridization signals obtained for CHR3v012120 suggested that this particular oligomer recognized a VP16-HRU-induced transcript, but further leads towards its identity were unavailable.

As for the relatively large group of locus IDs for which the transcript levels were possibly downregulated after tamoxifen-induced VP16-HRU expression, only two annotated genes also appeared to be downregulated in array #1; At2g44840, encoding a putative ethylene responsive element binding protein, and At1g09090, encoding the respiratory burst oxidase protein B. While At2g44840 was downregulated about 2-fold in both arrays, At1g09090 showed a 2.2-fold downregulation when VP16-HRU was expressed via tamoxifen induction and apparently a 129-fold repression when VP16-HRU was expressed via the *RPS5A* promoter. The latter value was primarily caused by a seemingly high expression level of At1g09090 in 1406 control seedlings, which was found to be an artefact when other microarray experiments were performed.

Table 1: Experimental design and results of microarray experiments. Numbers of differentially expressed locus IDs are based on at least 2-fold regulation and 95% confidence ($p \leq 0.05$).

	samples			differential expression		
array ID	Cy3	Cy5	plant material and growth conditions	induced	repressed	total
1	*RPS5A::VP16-HRU*	1406	10-day old, selected line, 1/2MS tissue culture	225	195	420
2	*pGPINTAM::VP16-HRU*	*pGPINTAM::VP16*	10-day old T2 seedlings derived from 50 T1 seedlings, liquid B5 +7 hrs induction with1 µM tamoxifen	32	150	182
3	*RPS5A::VP16-HRU*	1406	20-day old, green parts, selected line, greenhouse	553	417	970
4	*RPS5A::VP16-HRU*	1415	20-day old, green parts, selected line, greenhouse	395	296	691
5	*RPS5A::VP16-HRU*	*RPS5A::VP16(1406)*	20-day old, green parts, selected line, greenhouse	1547	380	1927

Design of additional microarray experiments

Comparing the gene expression systems used in micro array experiments #1 and #2, it seems as if a peak of *RPS5A*-mediated VP16-HRU expression in early embryos and dividing cells will be followed by a dramatic increase in somatic HR events, albeit only after a relatively long lag phase, thus when the cells will finally be differentiated. However, using the *TAM35S::VP16-HRU* system it might be concluded that cells which have already differentiated cannot simply be triggered to initiate somatic HR upon receiving a VP16-HRU stimulus. In order to be able to find a causal link between gene expression patterns and HR events, it was reasoned that a minimum of three additional micro array experiments were required using transgenic *RPS5A::VP16-HRU* seedlings rather than seedlings harbouring the *TAM35S::VP16-HRU* system.

Since HR events were predominantly found in the green parts of seedlings and rapidly accumulated while the tissues developed, the aerial parts of 20-day old seedlings grown on soil under identical greenhouse conditions were used for RNA extraction. Array #3 can be regarded as a repetition of array #1, but now using the green parts from older seedlings of the parental control line 1406 and from the selected line of seedlings containing the *RPS5A::VP16-HRU* construct. A similar expression pattern in arrays #1 and #3 should provide evidence that the co-regulation of genes is to some extent independent of age and growth conditions. Array #4 is a true biological replicate of array #3 since the cDNA samples were derived from HR reporter line 1415 and from a transgenic 1415 line containing *RPS5A::VP16-HRU*. Also in line 1415 it was found that *RPS5A::VP16-HRU* gene expression resulted in a dramatic upregulation of intra-chromosomal HR (Lindhout *et al*., 2006). For array #5, RNA was extracted from 1406 seedlings expressing either *RPS5A::VP16* or *RPS5A::VP16-HRU*. This array served several purposes. First of all, it should strengthen the data from array #3 since the cDNA sample derived from *RPS5A::VP16-HRU* was used on both arrays, thereby creating an internal technical replicate. In addition to that, data from array #5 should directly point out genes that are induced by the

complete VP16-HRU protein rather than by the VP16 moiety alone. VP16-affected genes can be further identified by comparing Cy5 channels of array #5 and array #3, this inter-array comparison is legitimate since all seedlings were grown under the same conditions and have an identical genetic background.

Combined analysis of microarray data

For each array the number of locus IDs was determined for which at least 2-fold differential signals were obtained with a significance of $p \leq 0.05$ (Table 1). Comparison of these numbers revealed that the fraction of 'significant' output is larger for arrays #3, #4 and #5 than for the first arrays, array #1 and #2. Since arrays #1 and #2 were hybridized in a separate experiment, we cannot exclude that variance in experimental conditions accounts to some part for this difference in number of genes. As already described above, only a small amount of locus IDs shared the same kind of differential regulation in arrays #1 and #2. Pairwise combinations of arrays #3, #4, and #5 with array #2 further corroborated that the *TAM35S::VP16-HRU* system, under the experimental conditions used, only rarely led to shared regulation characteristics for sequence identifiers. Hence, due to array #2, no co-regulated locus IDs could be discovered when the data from all five microarray experiments were combined.

Within the biological replicate, represented by arrays #3 and #4, 36-66% and 29-40% of the up- and downregulated locus IDs, respectively, were identical (Figure 4A). Only for 16 locus IDs a negatively correlated profile was found (data not shown). The combination of arrays #3, #4 and #5 demonstrated that for 111 locus IDs the presence of VP16-HRU led to an enhanced hybridization signal (Fig 4A, left panel) and for 42 locus IDs the RNA levels diminished (Fig 4A, right panel). Remarkably, the number of 1547 significantly upregulated RNAs in array #5 was at least 3-fold more than in any other array. As the Cy3-labelled cDNA sample used for this array was identical to that used in array 3#, the Cy5-labelled cDNA must have created this difference. A significant fraction of the transcriptome must have been affected by expression of 'free' VP16. Inter-array comparison of scaled Cy5 signals from array #5 with those of #3 and #4 indeed supported that notion, since a large fraction (~ 70%) of the locus IDs that were supposedly downregulated by 'free' VP16 were upregulated by VP16-HRU (data not shown). This effect led to enhanced values for the factor by which corresponding RNAs seemed to be induced by VP16-HRU in other arrays; many apparent induction values that were just less than the 2-fold threshold upheld in arrays #3 and #4 were more than 2-fold different in array #5, thus raising the number of significantly upregulated RNAs dramatically.

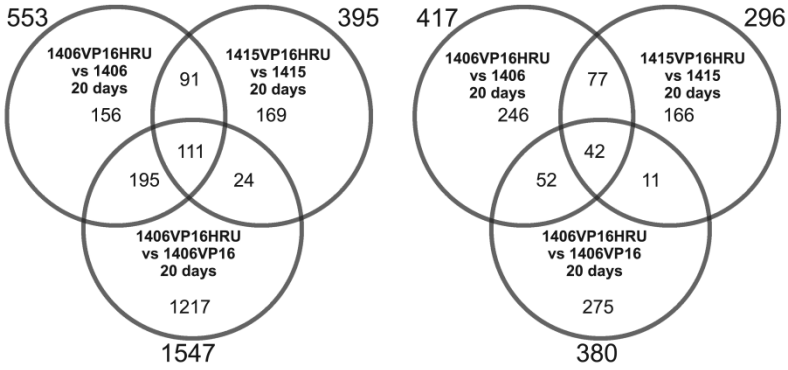
By combining the data set obtained from the 10-day old plants to the data sets of the biological replicate it was found that 77 upregulated locus IDs and 18 downregulated locus IDs still reacted similarly to VP16-HRU expression, thus independent from stage of development and growth conditions (Figure 4B). By combining the data sets from arrays #1, #3, #4 and #5, the number of locus IDs for which the corresponding RNAs were supposedly up- or downregulated due to the presence of the *RPS5A::VP16-HRU* transgene declined to 63 and 8, respectively (Figure 4C). The extent of regulation varied between 2- and 100-fold change for the upregulated locus IDs, while the most strongly repressed signals were up to 7.7-fold down. The microarray data corresponding to these locus IDs are listed in supplementary tables 1 and 2, respectively.

Evidently, it was tried to correlate significant changes in expression levels to the presence of a VP16-HRU binding site nearby or within corresponding genes. Only three of the 63 putatively upregulated genes contained the supposedly optimal 9-bp binding site 5'-GTGGAGGCT-3', namely At4g39510, encoding a cytochrome P450 family protein, At3g08570, encoding a phototropic-responsive protein, and At1g32900, encoding a starch synthase. The mRNA levels corresponding to these genes were not significantly changed by short term VP16-HRU expression (microarray #2). In addition to that, their basal level of expression in the different arrays was already high which made them unlikely rate-limiting factors for somatic HR. For these reasons it was concluded that the three genes mentioned did not deserve any special attention. Moreover, the zinc finger domain of VP16-HRU might bind to sequences deviating from 5'-GTGGAGGCT-3', which makes a focus on exactly this particular sequence an unnecessary risk. From the eight putatively repressed genes in *RPS5A::VP16-HRU* seedlings, none contained the 9-bp predicted binding site.

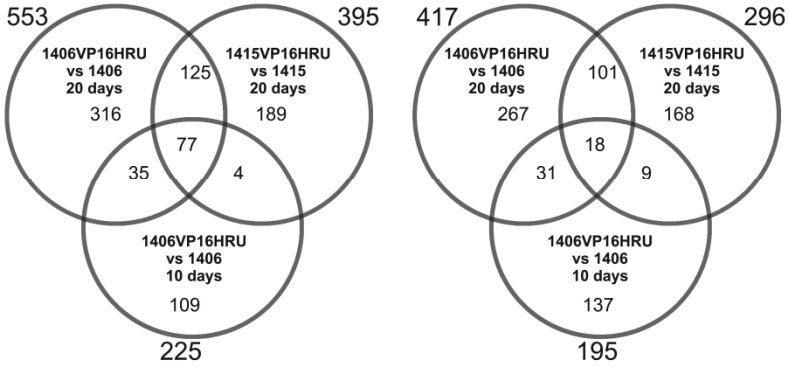
Inspection of the intensity values obtained showed that three of the eight genes found to be repressed in arrays #1,#3,#4 and #5, At2g18050, At3g29370, and At3g46370, appeared to be downregulated due to 'free' VP16 expression (Cy5 channel of array #5 compared to Cy5 channels of array #3 and #4), leaving only a marginal additive effect for the 3ZF domain present in VP16-HRU. Intensity values for At1g55320 (acyl-activating enzyme AAE18), were consistent in all *RPS5A* arrays, but their low levels made the accuracy of these data questionable. For the remaining repressed genes, At1g21460, At1g11170, At1g18710, and At2g43590, additional data regarding their expression characteristics were available in the public domain (http://bar.utoronto.ca/ ; https://www.genevestigator.ethz.ch) but none of them were indicative of any role for these genes in DNA repair and recombination. Expression of the predicted chitinase gene At2g43590 was annotated as being rather root-specific,

which corresponded very well with the fact that the signals for At2g43590 were much higher in arrays #1 and #2, probed with cDNA derived from whole seedlings, than in the arrays #3, #4 and #5, probed with cDNA derived from aerial parts of plantlets.

a



b



c

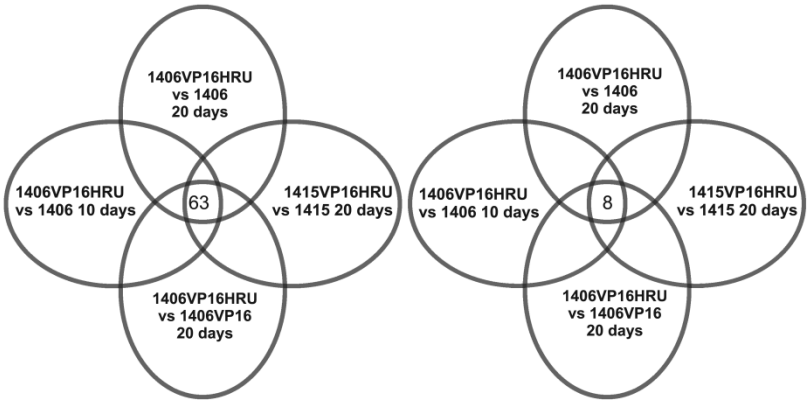


Figure 4: Diagrams showing the number of locus IDs shared between different microarray experiments with at least a 2-fold difference in expression level ($p \leq 0.05$) . Figures on the left represent upregulated locus IDs, figures on the right represent downregulated locus IDs. (A) Pair wise and combined overlap of microarrays #3, #4, and #5, regarding locus IDs exhibiting differential expression levels due to *RPS5A::VP16-HRU* presence referenced against parental lines (biological replicate arrays #3 and #4) and due to the presence of *RPS5A::VP16-HRU* referenced against the presence of *RPS5A::VP16* (array #5). (B) Pair wise and combined overlap of microarrays #1, #3, and #4, regarding locus IDs sharing a similar differential expression pattern in aerial parts due *RPS5A::VP16-HRU* presence under greenhouse conditions at 20 days after germination (biological replicate arrays #3 and #4) and under tissue culture conditions in whole seedlings at 10 days after germination (array #1). (C) Overlap of regulated locus IDs between all '*RPS5A*' arrays (#1,#3,#4 and #5.)

Selection of candidate genes to be tested for HR- inducing capacity.

As a further step toward understanding the modes of action of *RPS5A::VP16-HRU* leading to the strongly enhanced frequency of somatic HR events, we set out to use data from our microarray experiments to select a manageable collection of genes which could be individually tested for their HR-inducing potential. In theory, the differentially expressed locus IDs obtained after short term VP16-HRU expression (array #2) should form a good lead to identify primary VP16-HRU target genes. However, as already mentioned above, we were unable to provide evidence that induction of VP16-HRU with the tamoxifen system could lead to enhanced rates of somatic HR. In contrast to that and although the sequence of events might be complicated, the link between *RPS5A*-mediated VP16-HRU expression and HR was firmly established. We therefore focused on arrays #1, #3, #4 and #5 for selection of candidate HR-inducing genes, taking array #2 in account only for considering its intensity values for particular locus IDs during the final selection.

During the selection procedure, the following criteria were applied. First of all, in order to prevent genes dropping out of the selection procedure prior to evaluation of their expression characteristics no immediate selection pressure for confidence (p-value) was applied. Secondly, a candidate HR-inducing gene should be differentially regulated as a consequence of *RPS5A::VP16-HRU* expression in a manner that cannot be explained by the presence of just the VP16 moiety. Thirdly, genes with very high basal expression levels in one or more arrays were omitted since they are not likely to be rate limiting in somatic HR. Finally, some particular genes of interest that might not have fully met with all aspects of selection, as also described above, were added to the list.

Starting off the selection procedure with the only requirement that for each sequence ID there should at least be a 2-fold upregulation in arrays #1,

#3, #4 and #5, 80 locus IDs were left as being induced by VP16-HRU, a number close to the 63 locus IDs that were found by immediately including a p-value cut-off at 0.05 (see also Figure 4C). In order to apply the second criterion, we set a threshold for the ratio of the signals for *RPS5A::VP16* and 1406 (being the scaled Cy5 channel of array #5 divided by the scaled Cy5 channel #3, data not shown). With this ratio set at a maximum of 1.4, meaning no VP16 induction above an extra 40% added to the basal level, 57 locus IDs remained, four of them referring to un-annotated or non-coding RNAs for which further experimental possibilities were lacking. For the third criterion the intensity values of the remaining 53 locus IDs were carefully examined throughout the different microarrays. It was reasoned that a particular gene was a much more likely candidate for being a HR-regulating gene when there was a rather strict correlation between its mRNA expression levels and the extent of HR events observed in GUS-stained seedlings. Hence, in those cases where the apparent mRNA expression level in a sample derived from a parental line was much higher than the level found in other arrays for the *RPS5A::VP16-HRU* expressing seedlings, it was reasoned that the expression level in *RPS5A::VP16-HRU* seedlings itself was not causally related to the HRU phenotype. Similarly, a very high intensity value in array #2 (short term tamoxifen-induced expression) compared with the other arrays was also interpreted as an indication that the expression level per se was not linked to the HRU phenotype since enhanced HR frequencies were not observed under the conditions used for the inducible expression system. Such analyses narrowed down the list of possible HR-inducing genes from 53 to just 19 (Table 2).

From this list of 19 genes, 9 were removed on rather tentative grounds. At1g78440, a putative gibberellin-2 oxidase, was removed as it displayed an anti-correlated profile within array #2 and exhibited relatively moderate regulation in the other arrays. At1g21050, encoding a protein with unknown function, was removed for rather similar reasons as well as for being abundantly expressed in control samples. Five more genes were removed predominantly for the reason that putative encoded proteins were extremely small without any predicted function or domain, namely At5g13825, At3g27630, At1g59920, At5g59930, and At1g53480. Two more genes, At1g21326 and At1g34315, encoded unknown proteins as well but were primarily removed because their induction profile was not fully consistent and rather moderate in one or more cases. The genes represented by the remaining ten sequence identifiers were all regarded as candidate HR-inducing genes. Further data mining (http://bar.utoronto.ca/; https://www.genevestigator.ethz.ch) revealed that At5g60250, encoding a

C3HC4-type zinc finger protein, was previously found to be responsive to genotoxic stress, which might be indicative of a role in DNA repair.

In addition to the genes that passed all our selective steps, we chose three other candidate genes: At4g02390 encoding PARP2, At5g20850 encoding RAD51, and At5g48720, encoding an unknown protein. Although significantly upregulated in arrays #3, #4, and #5, *PARP2* and *Rad51* were initially eliminated from our selection only for the reason that in array #1 rather marginal folds upregulation were found. In Arabidopsis, both *PARP1* and *PARP2* have been described as genes that are induced by DNA breaks (Doucet-Chabeaud *et al.*, 2001). In addition, *PARP2* upregulation upon a variety of genotoxic treatments has been reported (Nagata *et al.*, 2005; http://bar.utoronto.ca/; https://www.genevestigator.ethz.ch). The RAD51 protein is the eukaryotic homolog of the prokaryotic RecA protein and is conserved from yeast to humans and has an established role in HR. Also for this gene a wealth of biological data has indicated that its mRNA is induced by genotoxic stress (Molinier *et al.*, 2004; Nagata *et.al.*, 2005; http://bar.utoronto.ca/; https://www.genevestigator.ethz.ch). Addition of At5g48720, a protein of unknown function, was based upon a known response to genotoxic stress (http://bar.utoronto.ca/; https://www.genevestigator.ethz.ch). The differential regulation of this gene was consistently induced in all 'RPS5A' arrays, but was left out of the initial selection due to its low induction level in array #1 and due to being just under 2-fold induction level in array #4. Altogether we have thus selected 13 genes to be tested for their HR-inducing potential (Table 2).

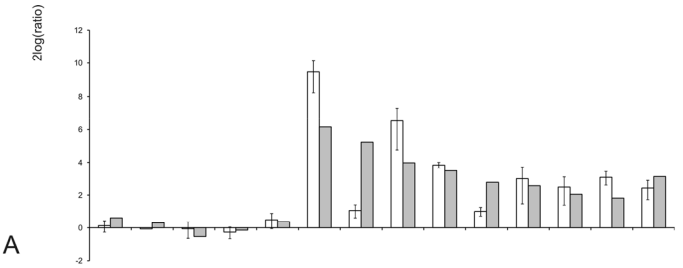
Table 2: Selection of locus IDs (see text for details) used for final selection of candidate HR-inducing genes. Fold change ratios and intensity values for each array are listed. In boldface the genes that were selected to be tested for HR-inducing potential (see text above)

		20 days greenhouse array #3			20 days greenhouse array #4			20 days greenhouse array #5			10 days tissue culture array #1			7 hours induction array #2		
Primary sequence name	**Annotation**	**1406**	**1406-VP16-HRU**	**Fold change**	**1415**	**1415-VP16-HRU**	**Fold change**	**1406-VP16**	**1406-VP16-HRU**	**Fold change**	**1406**	**1406-VP16-HRU**	**Fold change**	**1406-VP16**	**1406-VP16-HRU**	**Fold change**
At1g43780	**serine carboxypeptidase S10 family protein**	**34**	**2447**	**71.0**	**41**	**819**	**20.0**	**33**	**2495**	**76.7**	**36**	**3742**	**104.2**	**230**	**111**	**-2.1**
At5g07610	**F-box family protein**	**30**	**1130**	**37.6**	**43**	**430**	**10.0**	**36**	**1147**	**32.1**	**86**	**196**	**2.3**	**87**	**134**	**1.5**
At1g66300	**F-box family protein**	**32**	**510**	**15.7**	**34**	**224**	**6.5**	**30**	**550**	**18.1**	**27**	**161**	**5.9**	**35**	**25**	**-1.4**
At5g60250	**zinc finger (C3HC4-type RING finger) family protein domain**	**105**	**1180**	**11.2**	**74**	**426**	**5.8**	**112**	**1324**	**11.9**	**202**	**561**	**2.8**	**523**	**670**	**1.3**
At3g55890	**yippee family protein**	**1022**	**7306**	**7.1**	**619**	**2267**	**3.7**	**802**	**6510**	**8.1**	**382**	**1997**	**5.2**	**976**	**646**	**-1.5**
At3g10150	**calcineurin-like phosphoesterase family protein**	**192**	**1313**	**6.9**	**119**	**605**	**5.1**	**244**	**1179**	**4.8**	**155**	**882**	**5.7**	**200**	**92**	**-2.2**
At3g55910	**expressed protein PA26**	**528**	**3369**	**6.4**	**431**	**1108**	**2.6**	**440**	**2854**	**6.5**	**471**	**1463**	**3.1**	**1462**	**964**	**-1.5**
At5g15800	**developmental protein SEPALLATA1 / floral homeotic protein (AGL2) (SEP1)**	**266**	**1578**	**5.9**	**169**	**903**	**5.3**	**293**	**1893**	**6.5**	**52**	**384**	**7.3**	**88**	**106**	**1.2**
At5g46915	**transcriptional factor B3 family protein**	**59**	**185**	**3.2**	**49**	**112**	**2.3**	**75**	**206**	**2.8**	**83**	**180**	**2.2**	**97**	**50**	**-1.9**
At5g35120	**hypothetical protein**	**1107**	**3011**	**2.7**	**898**	**2959**	**3.3**	**863**	**2482**	**2.9**	**545**	**1546**	**2.8**	**605**	**521**	**-1.2**
At4g02390	**poly (ADP-ribose) polymerase**	**381**	**3327**	**8.7**	**365**	**1314**	**3.6**	**401**	**2764**	**6.9**	**804**	**936**	**1.2**	**988**	**1039**	**1.1**
At5g20850	**DNA repair protein RAD51**	**344**	**1424**	**4.1**	**209**	**496**	**2.4**	**312**	**1276**	**4.1**	**934**	**1502**	**1.6**	**1008**	**1031**	**1.0**
At5g48720	**expressed protein**	**1233**	**4339**	**3.5**	**967**	**1869**	**1.9**	**963**	**3650**	**3.8**	**676**	**982**	**1.5**	**797**	**950**	**1.2**
At1g59920	hypothetical protein	545	3204	5.9	470	3022	6.4	632	2720	4.3	198	1264	6.4	295	285	-1.0
At1g59930	hypothetical protein	629	3614	5.7	494	3270	6.6	612	3008	4.9	213	1187	5.6	265	235	-1.1
At1g53480	expressed protein	232	687	3.0	152	526	3.5	230	769	3.3	155	818	5.3	229	352	1.5
At1g21050	expressed protein	13244	28641	2.2	9855	22072	2.2	11178	24207	2.2	11308	25052	2.2	8970	6416	-1.4
At5g13825	hypothetical protein	31	647	21.0	39	180	4.6	37	644	17.5	28	377	13.6	23	62	2.7
At3g27630	hypothetical protein	315	5533	17.6	199	1424	7.2	245	5131	20.9	590	1317	2.2	852	786	-1.1
At1g21326	VQ motif-containing protein	142	672	4.7	126	267	2.1	154	632	4.1	88	270	3.1	104	50	-2.1
At1g34315	expressed protein	838	2071	2.5	531	1075	2.0	647	1791	2.8	295	1103	3.7	630	378	-1.7
At1g78440	gibberellin 2-oxidase / GA2-oxidase (GA2OX1)	133	271	2.0	105	360	3.4	121	391	3.2	184	513	2.8	183	72	-2.5

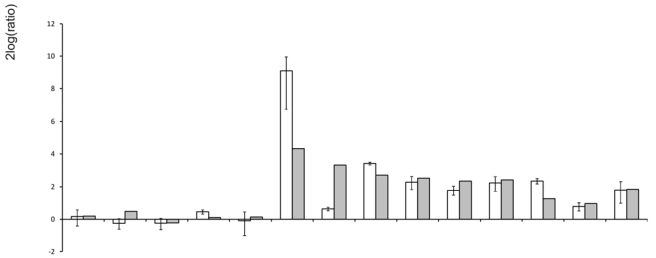
Quantitative reverse transcription PCR analysis

It was decided that the most straightforward approach for assessing the effect of the candidate HR-inducing genes on somatic HR frequency should consist of cloning a cDNA sequence encompassing the open reading frame of the gene of interest behind the *RPS5A* promoter and subsequent introduction of these constructs in HR indicator line 1406. While suitable cDNA sequences for most of the candidate genes were readily obtained, this was not the case for At3g55890, At3g55910, At5g46915, and At5g35120, despite a variety of efforts. Their intended analysis *in planta* thus became impossible.

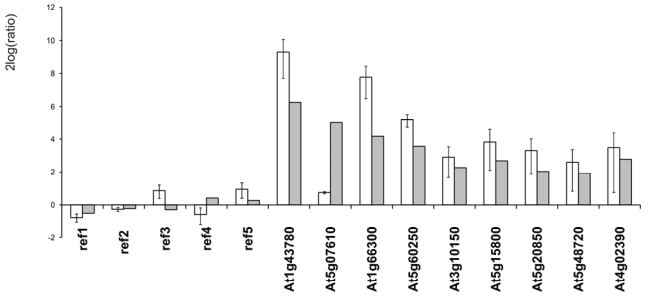
In order to assess the reliability of the microarray data regarding the mRNAs corresponding to the expression constructs, mRNA levels were analysed by Quantitative Reverse Transcription PCR (Q-RT-PCR). Except for reference gene *Roc1* (At4g38740), where this was impossible, amplicons of genes were chosen in such a manner that they did not coincide with the oligonucleotide sequence representing the gene on the microarray (see supplementary Table 3 for primers), thus providing an independent means to verify the identity of the candidate gene. Templates for Q-RT-PCR analyses were derived from independently isolated RNA extracts from 20 day-old greenhouse grown plant material. By normalization of data against at least three out of five reference genes we complied with the stringent requirements of the GeNorm method (Vandesompele *et al.*, 2002). Each Q-RT-PCR reaction was performed in duplicate and the results were compared with the fold change ratios as obtained in microarrays #3, #4 and #5 (Figure 5A, B and C respectively). For eight out of nine candidate genes analyzed, the Q-RT-PCR data agreed very well with the microarray data (Figure 5D). Only for the F-box family protein gene At5g07610 mRNA induction could not be confirmed, although the theoretical amplicon sequence and the oligonucleotide sequence supposed to be present on the array were in principle specific for this gene.


2log(ratio)
12
10
8
6
4
2
0
-2

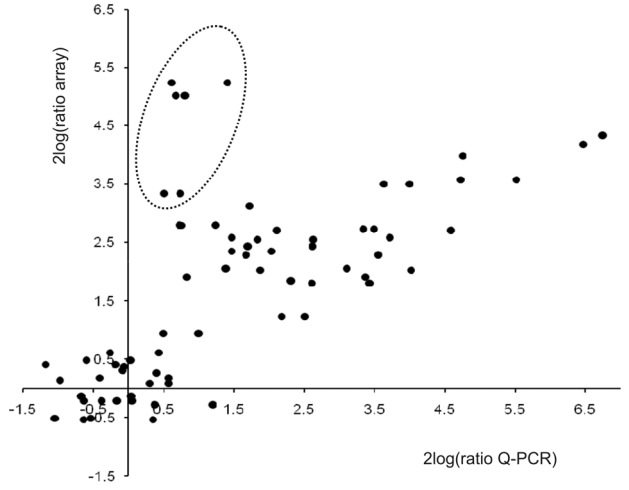
A


2log(ratio)
12
10
8
6
4
2
0
-2

B


2log(ratio)
12
10
8
6
4
2
0
-2
ref1
ref2
ref3
ref4
ref5
At1g43780
At5g07610
At1g66300
At5g60250
At3g10150
At5g15800
At5g20850
At5g48720
At4g02390

C


2log(ratio array)
6.5
5.5
4.5
3.5
2.5
1.5
0.5
-0.5
-1.5
-1.5
-0.5
0.5
1.5
2.5
3.5
4.5
5.5
6.5
2log(ratio Q-PCR)

D

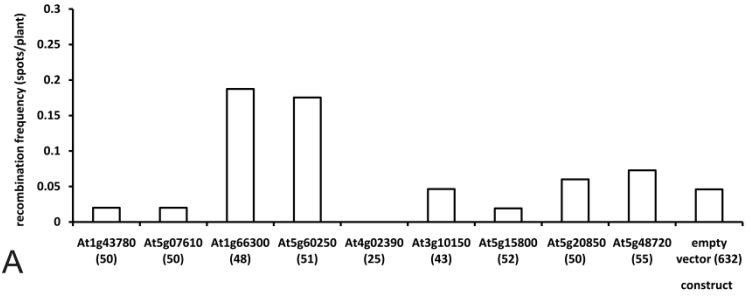
Figure 5: Relative gene expression levels as determined by Q-RT-PCR analysis (open bars) in duplicate in comparison with data derived from individual microarrays (gray bars). (A) Q-RT-PCRs in comparison with microarray #3 (*RPS5A::VP16-HRU*, referenced against parental line 1406), (B) Q-RT-PCRs in comparison with microarray #4 (*RPS5A::VP16*-HRU referenced against parental line 1415), and (C) Q-RT-PCRs in comparison with microarray #5 (*RPS5A::VP16-HRU* referenced against *RRPS5A::VP16*). Error bars indicate the variation between duplicate Q-RT-PCRs. (D) Correlation of relative gene expression values as determined by Q-RT-PCR analysis (x-axis) and microarray analysis (y-axis). The selected outliers within the ellipse represent data points for At5g07610 for which Q-RT-PCR could not confirm the data from the microarray experiments. Note that relative expression levels are on a ^{2}log scale.

Homologous recombination in plants overexpressing candidate genes

For those genes for which plant expression constructs could be made, primary transformants with the various *RPS5A* promoter-driven cDNA constructs were generated in which the frequency of HR events after 10 days of growth was compared to the frequency observed within similarly grown empty vector controls (Figure 6A). While the average HR frequency in most primary transformants was low at this stage of development, comparable to the control frequency, *RPS5A* promoter-mediated expression of At1g66300 (encoding an F-box protein) and At5g60250 (encoding a C3HC4 type zinc finger containing protein) led to a conspicuous 4-fold rise in HR frequency. When HR frequency was determined after 20 days of growth (Figure 6B), the overall pattern of relative HR frequencies changed remarkably compared to that found after 10 days of growth (Figure 6A). While seedlings expressing the At1g66300 construct continued to accumulate somatic HR events in much higher frequency than controls, such a difference had disappeared for primary transformants containing the At5g60250 construct (Figure 6B). Although limited numbers of primary transformants were obtained with *RPS5A::PARP2* (At4g02390), the strongly elevated HR frequency found within these transformants stained after 20 days of growth indicated that *RPS5A* promoter-mediated expression of PARP2 had a positive effect on HR frequency. Also primary transformants harbouring *RPS5A::SEP1* (At5g15800, *Sepallata1*) and the construct for *RPS5A* promoter-mediated expression of At4g48720 (encoding an unknown protein) exhibited an increased HR frequency compared with the empty vector control after 20 days of growth. It was noted that many transformants containing *RPS5A::SEP1* displayed an extreme early flowering phenotype (data not shown), very similar to that observed due to overexpression of the related gene *Sepallata3* (Honma and Goto, 2001). Since HR frequencies displayed by the reporter line 1406 were noticed to rise during development (empty vector controls in Figure 6), an accelerated transgene-induced progress through development might be expected

to lead to an extra increase in HR frequencies. In this respect, the data regarding HR frequency within *RPS5A::SEP1* containing seedlings should be interpreted with caution as they might have been linked with the early flowering phenotype rather than with a direct effect upon the frequency of HR events. Concerning At3g10150, a putative calcineurin-like phosphoesterase, it might be concluded that its *RPS5A* promoter-mediated expression can result in an elevated HR frequency after 20 days of growth although less conspicuous than for At1g66300 (F-box), At4g02390 (PARP2), At4g48720 (unknown protein), and At5g15800 (SEP1). Although it is tempting to formulate a similar conclusion for At1g43780, the serine carboxypeptidase S10 family gene which also exhibited the most dramatic upregulation due to VP16-HRU presence, the rise in HR frequency is just over two-fold compared to empty vector controls and is based upon a relatively limited number of transgenic seedlings.

It was of interest to note that transformation with the construct for At5g07610 did not result in higher HR frequencies in primary transformants. As mentioned above, differential expression of At5g07610 mRNA could not be verified with Q-RT-PCR. The samples of seedlings containing this construct might therefore be regarded as extra internal negative controls. Concerning At5g20850 (RAD51), no evidence was obtained that *RPS5A*-driven expression resulted in enhanced HR frequencies in somatic cells.



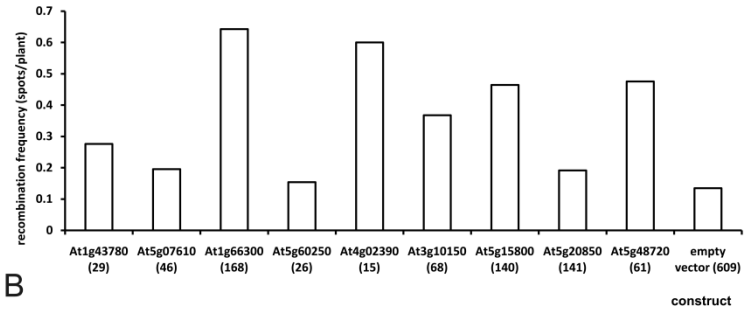


Figure 6: Homologous recombination frequency, quantified as the mean number of GUS positive spots per seedling, in 10-day old primary transformants (A) and 20-day old primary transformants (B). Gene identity numbers on the x-axis refer to the origin of the open reading frame (cDNA) that is present within the *RPS5A* promoter-driven expression cassette. In parentheses on the x-axis are the numbers of primary transformants that were analysed. This number was low for At4g02390 (*PARP2*) due to low transformation efficiency. Note the different scales on the y-axis.

DISCUSSION

The HRU phenotype

In this work we have primarily focussed on further phenotypic characterization and molecular analysis of Arabidopsis plants expressing the artificial transcription factor VP16-HRU under control of the *RPS5A* promoter, as a consequence of which an unprecedented 200- to 300-fold rise in the frequency of somatic HR was observed previously (Lindhout *et al.*, 2006). While the much less hyper recombinogenic Arabidopsis mutants *recq14a, caf-1, rad50*, *rad17*, *rad9*, *centrin2*, and *bru1* (Baghereh-Najjar *et al.*, 2005; Endo *et al.*, 2006; Gherbi *et al.*, 2001; Heitzeberg *et al.*, 2004; Molinier *et al.*, 2004; Takeda *et al.*, 2004) mostly display very severe developmental abnormalities in addition to increased HR frequency, *RPS5A::VP16-HRU* plants only exhibit mild phenotypic changes. Other than resistance to the DSB-inducing chemical bleomycin, which could be caused by an activated HR system, only a weak leaf phenotype and a small delay in flowering time were observed (Lindhout *et al.*, 2006). In the present study we corroborated and extended initial morphological observations (Figure 1). Furthermore, increased anthocyanin accumulation was found under high light intensity conditions.

The most dramatic phenotype of *RPS5A::VP16-HRU* plants, being the rise in somatic HR events, was analysed in further detail during early seedling development (Figure 2). Undoubtedly the HRU phenotype must be linked to the presence of a *RPS5A::VP16-HRU* transgene, but a simple straight forward connection of the spatio-temporal expression pattern of VP16-HRU as it might be expected to occur in transgenic seedlings and the extent of somatic HR events was lacking. The large majority of HR events occurred later than 7 days after seedling germination, so very much lagging behind the likely peaks of *RPS5A::VP16-HRU* expression during early phases of embryogenesis or meristematic activity (Weijers *et al.*, 2001) and in fact when VP16-HRU expression had already declined. Experiments with a tamoxifen induction system provided additional evidence that the relation between VP16-HRU expression and HR events must be complicated, as we were unable to find short or long term

VP16-HRU induction conditions that resulted in an increase in somatic HR. Based upon our experimental data thus far, it can be postulated that a dramatic increase in HR might only occur when VP16-HRU has been present during a critical stage of plant cell development, as if setting the stage for the recombination events. As a consequence, it must be realized that genes that are differentially regulated at the time when HR events do finally occur might only be very indirectly regulated by VP16-HRU. Just as well, genes apparently under more direct VP16-HRU control as they are found shortly after tamoxifen induction of VP16-HRU might not be causally linked to future HR events.

HR-inducing candidate genes

As described above, transcriptome analyses were performed in order to discover genes for which the level of mRNA depended upon the expression of the transcriptional regulator VP16-HRU. A rather stringent set of selection criteria, combined with some clemency for intriguing genes that just failed to meet all requirements, resulted in the assembly of a list of 13 genes (see Table 2) for which further experimental efforts were considered to be most appropriate regarding their potential role in the mitotic HR process. Since for some of these genes no suitable cDNA fragment could be obtained, the HR-inducing potential of 9 genes was assessed by ectopically expressing the corresponding mRNAs via the *RPS5A* promoter. This analysis revealed that several of these expression constructs led to enhanced frequencies of HR in transgenic primary transformants. The effect of two of these constructs was most conspicuous at 10 days after germination. Both *RPS5A* promoter-mediated expression of the cDNAs of genes At1g66300, corresponding to a gene encoding an unknown F-box protein, and At5g60250, corresponding to a gene encoding a zinc finger protein, stimulated HR-frequency about 4-fold. In contrast to plants expressing the At5g60250 construct, plants expressing the At1g66300 construct (F-box protein) continued to accumulate HR events in high frequency after 10 days of germination (Figure 6). F-box proteins are involved in substrate recognition for protein degradation through the ubiquitin proteasome system. Arabidopsis contains nearly 700 predicted F-Box proteins which are involved in a wide variety of processes, such as hormone response pathways, lateral root formation and phytochrome A-dependent light signalling (Lechner *et al.*, 2006). Besides its F-box, no other conserved domains or expression data were reported for At1g66300 that could disclose any leads to its binding partners or its precise function. At5g60250 belongs to a large group of genes as well, encoding about 300 zinc finger proteins with a so-called RING domain for which a large variety of functions have been proposed, such as transcriptional and translational

regulation, cell cycle regulation, and ubiquitin ligase activity (Kosarev *et al.*, 2002; Ren *et al.*, 2007). A precise function for At5g60250 can not yet be inferred, but it has been reported to be genotoxic responsive (http:// www.genevestigator.ethz.ch/). Although it is tempting to speculate that the F-box protein encoded by At1g66300 might somehow be connected with ubiquitin ligase activity of the At5g60250 protein, there is currently no experimental evidence confirming such a link.

Later in plant development (after 20 days, Figure 6B) several other candidate HR-inducing gene constructs started to stimulate HR frequency as well: *PARP2* (At4g02390), *SEP1* (At5g15800), At4g48720, encoding an unknown protein, and to a lesser extent, the calcineurin-like phosphoesterase family protein encoded by At3g10150. Considering these genes, a role for *PARP2* in regulation of HR seems rather plausible. In mammalian systems it was found that expression of PARP1 and PARP2 is DNA-damage dependent and from experiments in mice knocked out for *PARP1* it was evident that PARP1 functions in surveillance and maintenance of genome integrity (Ménissier de Murcia *et al.*, 2003; Schreiber *et al.*, 2002). Several studies are supportive for similar roles of PARP1 and PARP2 in plants. Both are found to be induced by DNA damaging irradiation (Doucet-Chabeaud *et al.*, 2001) and a PARP2 reporter construct was elevated in tobacco cells that were targeted to cell death due to DNA ligase I deficiency (Babiychuk *et al.*, 1998). Transient expression of Arabidopsis *PARP2* cDNA in soybean cells promoted DNA repair and inhibited cell death after mild oxidative stress, but promoted cell death upon severe oxidative stress (Amor *et al.*, 1998). Remarkably, while *PARP2* mRNA was significantly induced in 20-day old *RPS5A::VPP16:HRU* seedlings, *PARP1* (A2g31320) mRNA was not differentially expressed, this indicates that if any of the PARP proteins plays a role in the dramatically high HR frequency in the HRU mutant, this will be PARP2 rather than PARP1.

Ectopic *RPS5A*-mediated expression of either *SEP1* (At5g15800) or At4g48720 mRNA resulted in a 3-fold higher HR frequency in 20-day old seedlings. SEP1, annotated to be a MADS box transcription factor involved in flower and ovule development, is functionally redundant with SEP2 and SEP3. Flowers of *sep1/2/3* triple mutants show a conversion of petals and stamens to sepals (Pelaz *et al.*, 2000). Data mining in the public domain did not yet disclose any transcriptional response upon genotoxic treatment for these genes. However, in our data set *SEP1* mRNA is highly induced in all four arrays probed with cDNA from *RPS5A::VP16-HRU* seedlings and *SEP3* mRNA shows significant differential expression in arrays #3 and #4 representing the biological replicate and in array #1, hybridized with 10-day old *RPS5A::VP16-HRU* seedlings. *SEP2*

mRNA was not differentially regulated and seemed to be expressed at a lower level than *SEP1* and *SEP3* mRNA. As mentioned above, ectopic expression of *SEP1* mRNA in *RPS5A::SEP1* primary transformants frequently resulted in an extremely fast development, causing seedlings to bolt just after forming a few leaves. Since an increase in HR events was also observed in the wild-type parental line at more mature stages of development (data not shown), the higher HR frequency due to the *RPS5A::SEP1* construct can very well be explained as being caused by an accelerated maturation rather than by a more direct influence on HR rates. In this respect, the fact that *RPS5A::VP16-HRU* containing seedlings display delayed bolting makes the extremely high HR frequencies in these seedlings even more striking. Concerning At5g48720, data mining revealed that its expression is induced about 26-fold by genotoxic tress (http:://www.genevestigator.ethz.ch/), but further *in silico* analysis did not reveal any putative function.

The At3g10150 protein, stimulating HR frequency between 2- and 3-fold at 20 days after germination when expressed via the *RPS5A* promoter, was predicted to encode a calcineurin-like phosphoesterase. Proteins belonging to this family include a wide variety of protein phosphoserine phosphatases, nucleotidases as well as nucleases, such as bacterial SbcD or yeast MRE11 (Koonin, 1994; Sharples and Leach, 1995). Additionally, it was found that many archaeal DNA polymerases II enzymes contain an apparently active calcineurin-like phosphatase superfamily domain and that an inactive calcineurin-like phosphatase superfamily protein domain is present in the small subunits of eukaryotic DNA polymerases α, δ and ε (Aravind and Koonin, 1998). Taken together, it might very well be that the calcineurin-like phosphoesterase encoded by At3g10150 is involved in DNA repair but detailed functional analysis will be necessary to demonstrate such a role. Since At3g10150 has recently also been annotated as *AtPAP16*, a purple acid phosphate belonging to a family of at least 28 members in Arabidopsis (Zhu *et al.,* 2005), a careful analysis is warranted.

As mentioned above, At1g43780 mRNA levels were most dramatically induced in VP16-HRU expressing seedlings, from 20- to 100-fold in the microarrays and about 600-fold according to Q-RT-PCR data (Figure 6). However, *RPS5A*-mediated expression of the coding sequence only resulted in about 2-fold induction of HR frequency after 20 days of growth (Figure 7). Supposing that further verification would lead to similar data, a 2-fold raise in HR frequency can not be disregarded; several HR-mutants described in literature only have approximately 2-fold altered recombination frequencies (either up or down) at a similar type of reporter locus (Hartung *et al.*, 2007; Heitzeberg *et al.*, 2004; Li *et al.*, 2006). When only specific cell types contribute to the 20- to 100-

fold elevated levels of At1g43780 mRNA measured in the population of transcripts, it can very well be imagined that *RPS5A*-driven expression does not reach to those expression levels required for a strong effect upon HR. Obviously, such a consideration applies to all *RPS5A*-driven constructs and future experiments should address this possibility. Nevertheless, a near to complete lack of HR-induction by the *RPS5A*-mediated expression approach, such as displayed for At5g07610 (encoding another F-box family protein) and At5g20850 (encoding RAD51), should be taken as strong evidence that the particular protein is irrelevant for the high HR frequencies in VP16-HRU expressing seedlings.

Since At5g07610 mRNA accumulation could not be verified by Q-RT-PCR (Figure 6), lack of HR-inducing capacity might be expected. For *Rad51*, higher mRNA levels were corroborated, but ectopic expression of *Rad51* could not induce more HR events. Although connections between RAD51 and HR have been described, it has been shown in yeast that there are two different HR pathways, namely the single strand annealing (SSA) pathway, which is independent of RAD51, and the double-strand break repair pathway (DSBR), being RAD51-dependent. Intra-chromosomal HR on homologous sequences arranged in tandem, is thought to occur mainly through the RAD51-independent SSA pathway (reviewed in Puchta, 2005). This might very well explain our finding that ectopic expression of *Rad51* did not increase the number of HR events between the partial GUS repeats in line 1406.

Albeit none of the *RPS5A*-driven constructs did not even approach the 200- to 300-fold HR-inducing capacity of the *RPS5A::VPP16:HRU* construct, an increment of 2- to 4-fold in HR-frequency is certainly not marginal when comparing these values to those reported for other hyper recombinogenic mutants (Table 3), assayed on the same or very similar HR reporter loci.

Table 3: Somatic HR mutants

mutant gene	type	reference	fold change in somatic HR
recql4a	T-DNA insertion line	Bagherieh-Najjar *et al.*, 2005	7.5
caf-1 (mutants fas1 and fas2)	T-DNA insertion line	Endo *et al.*, 2006	40
rad50	T-DNA insertion line	Gherbi *et al.*, 2001	8 to 10
rad17	T-DNA insertion line	Heitzeberg *et al.*, 2004	2.5
rad9	T-DNA insertion line	Heitzeberg *et al.*, 2004	3.8
bru1	T-DNA insertion line	Takeda *et al.*, 2004	4
centrin2	Activation-tagged line (downregulated Centrin2)	Molinier *et al.*, 2004	36 in initial line, 3 -15 in RNAi lines
At1g66300 (F-Box protein)	RPS5A promoter mediated expression	This work	4-5
At4g60250 (zinc finger protein)	RPS5A promoter mediated expression	This work	4
At4g02390 (*Parp2*)	RPS5A promoter mediated expression	This work	4-5
At5g15800 (SEP1)	RPS5A promoter mediated expression	This work	3-4
At4g48720 (unknown protein)	RPS5A promoter mediated expression	This work	3-4
RPS5A::VP16-HRU	ZF-ATF line	Lindhout *et al.*, 2006; this work	200-300

Data comparison to previous publications

All microarray experiments performed in the present study demonstrated that substantial numbers of genes were differentially expressed as a consequence of VP16-HRU presence. Four of the microarrays were probed with cDNA samples derived from seedlings containing a *RPS5A::VP16-HRU* transgene, harvested at a time that HR events were taking place. At least a fraction of the combined data derived from these experiments should thus reflect changes in the transcriptome that might be relevant for hyper recombination. We therefore compared our combined data with available data regarding other transcriptome analyses during changes in DNA metabolism after DNA damaging treatments. Using the bio-array resource for Arabidopsis functional genomics (http://bar.utoronto.ca/, Toufighi *et al.*, 2005) and genevestigator (https://www.genevestigator.ethz.ch, Zimmermann et al 2004) we found that our combined dataset, regarding differentially expressed genes or locus IDs putatively up- or downregulated in plant tissues with enhanced HR frequency, hardly had any overlap with supposedly similar datasets derived from other experiments. This, however, seemed to be a rather general phenomenon since pair wise comparisons between single transcriptome studies (Chen *et al.*, 2003 ; Molinier *et al.*, 2005 Nagata *et al.*, 2005) other than our own also showed a near lack of overlap, even though in many cases comparable numbers of differentially expressed genes were reported. When all available datasets were combined and normal selection criteria (at least 2-fold up- or downregulated, $p \leq 0.05$) were applied, we did not find any overlap at all (data not shown). This indicated that overlap between datasets from different experimental investigations might at least partly have been caused by chance, a logical consequence when the transcriptional

activity of large numbers of genes is compared. Taking this into consideration, we decided to investigate the behaviour of genes that are more generally acknowledged as being involved in recombination and DNA repair in more detail in our microarray data. Complementary to that, and when possible, we also set out to check the behaviour of genes that were at least 2-fold up- or downregulated in the strongly correlating series of *RPS5A* arrays from 20-day old seedlings (arrays #3, #4, and #5) in the datasets from other experiments.

The *Rad51* gene, known to be involved in HR but probably irrelevant for recombination between repetitive sequences as mentioned above, was induced (2.4- to 4.1-fold) in 20-day old *RPS5A::VP16-HRU* seedlings and 1.6-fold in 10-day old *RPS5A::VP16-HRU* seedlings. Also in a dataset generated using bleomycin-treated plant material (Molinier *et al.*, 2005) and a dataset obtained after γ-irradiation of Arabidopsis seedlings (Nagata *et al.*, 2005) *Rad51* gene expression was found to be induced. More extensive analysis of publicly available microarray data revealed that *Rad51* appeared to be specifically induced after genotoxic stress, but not after other kinds of abiotic stress, like temperature -, salt -, or osmotic stress (http://bar.utoronto.ca/). It is interesting to note that the HRU mutant was found to be more resistant to bleomycin-induced DNA damage (Lindhout et al., 2006), a feature that might thus possibly be attributed to enhanced *Rad51* expression.

Further comparison with Molinier's work (2005) revealed that also At4g34510 (a fatty acid elongase annotated to be involved in wax biosynthesis) was strongly induced (4.6-fold to 7.3-fold) in the transcription profiles derived from 20-day old *RPS5A::VP16-HRU* seedlings. No upregulation in 10-day old *RPS5A::VP16-HRU* seedlings was found for this gene (array #1). Further examination of this gene in the public domain (http://bar.utoronto.ca/) indicated that it can also be induced by a combination of mitomycin C and bleomycin. Mitomycin C is a potent DNA cross linker which can result in DNA DSBs comparable with stalled replication-forks (Tomasz, 1995). However, since At4g34510 was reported to encode a very-long-chain fatty acid elongase involved in wax biosynthesis (James *et al.*, 1995), a role in DNA metabolism seems to be very unlikely.

Comparing our microarray data with high-density colony array transcription profiles obtained after combined treatment with both bleomycin and mitomycin C (Chen *et al.*, 2003), we found three genes, At3g27630, At3g27060, and At5g49480, that were also upregulated in 20-day old *RPS5A::VP16-HRU* seedlings. At3g27630, annotated as a very small hypothetical protein without any predicted structural domains, was found to be 3-fold induced in Chen's dataset whereas we found inductions ranging from 7.2- to 20.9-fold for 20-day old

seedlings and 2.2-fold using 10-day old seedlings. At3g27060, encoding the small chain of ribonucleotide reductase, was induced 10-fold by bleomycin/mitomycin (Chen *et al*., 2003) and between 2.2- and 4.6-fold in the 20-day-old *RPS5A::VP16-HRU* seedlings, but not significantly in array #1. In other studies with DNA damaging treatments no differential expression for both genes was reported. At5g49480, a sodium-inducible calcium binding protein found to be 5-fold upregulated by Chen and co-workers, was 2- to 3-fold induced only in 20-day old *RPS5A::VP16-HRU* seedlings. At5g49480 was annotated as being bleomycin/mitomycin-inducible in microarray experiments as well (http://bar.utoronto.ca/).

We have found a significant induction of *PARP2* (At4g02390) in the 20-day old *RPS5A::VP16-HRU* expressing seedlings (ranging from 3.6-fold to 8.7-fold), but not of *PARP1*. Induction of *PARP2* was also found in response to a short treatment with γ-irradiation (Nagata *et al*., 2005) and also a response to genotoxic treatment was annotated (http://bar.utoronto.ca/; https://www.genevestigator.ethz.ch). In Arabidopsis, both *PARP1* and *PARP2* have been described as genes that are induced by DNA breaks (Doucet-Chabeaud *et al*., 2001). As mentioned above, ectopic expression of *PARP2* did result in increased somatic HR events in our studies.

Some genes that were significantly repressed in 20-day old *RPS5A::VP16-HRU* seedlings were also found to be downregulated in other studies. At2g41800 and At2g41810, adjacent to each other and both encoding similar unknown expressed proteins, were found to be downregulated after bleomycin treatment (Molinier *et al*., 2005). These genes were not significantly downregulated in 10-day old *RPS5A::VP16-HRU* seedlings. Three other genes were downregulated upon γ-irradiation: At1g18710, encoding MYB transcription factor 47, At5g25390, encoding an AP2 domain transcription factor, and At5g54190, encoding protochlorophyllide oxidoreductase A (Nagata *et al*., 2005). At1g18710 and At5g25390, also showed downregulation in 10-day old *RPS5A::VP16-HRU* seedlings (array #1), whereas no downregulation was found for At5g54190 in this array.

Consistent with the data of Molinier *et al.* (2005) and Nagata *et al.* (2005), we did not find differential expression of several genes with a presumed role in DNA metabolism and repair. These include MRX constituents *Rad50*, *Mre11*, and *Nbs1/Xrs2* (Gallego *et al*., 2001; Bundock and Hooykaas, 2002; Akutsu *et al*., 2007), the *Rad51* paralogs (*Rad51B, Rad51C, Rad51D, XRCC2* and *XRCC3*) and members from the *RecQ* helicase family, except for two *RecQ* family members (At1g10930 and At4g35740) which were found differentially expressed upon bleomycin treatment (Molinier *et al.,* 2005). Also *Ku70*, *Ku80*, *XRCC4* and

Lig4, active in the NHEJ pathway (Bleuyard *et al.*, 2006) did not show differential expression in any of the here compared datasets, including our own.

Concluding remarks

Altogether, the results in this report illustrate that plants expressing VP16-HRU have a remarkably mild phenotype in all aspects, except for their extremely high frequency of somatic HR. By transcriptome analysis we have shown that many genes are differentially regulated due to VP16-HRU expression, among them genes known to be involved in DNA repair but also a variety of genes with different or unknown functions. We identified novel genes that can act as a positive modulator of HR, although none of them were yet able to meet VP16-HRU in terms of orchestrating a 200- to 300-fold increase in somatic HR frequency. Whether or not an endogenous plant gene can be discovered that fulfils such a role, the very existence of a dominant HR-inducing factor like VP16-HRU offers exciting possibilities to gain fundamental knowledge about HR, possibly providing means to bring HR under experimental control once we gain further understanding of its mode of action. Future studies are required to address these issues.

MATERIALS AND METHODS

Bacterial strains

For cloning procedures *E.coli* strain DH5α was used. For transformation of plants, constructs (see below) were mobilized to *Agrobacterium* strain AGL1 (Lazo *et al.,* 1991) via tri-parental mating (Ditta *et al.,* 1980).

Plant material and transformation

Arabidopsis ecotype Columbia plant lines *1406* and *1415* (Gherbi *et al.*, 2001), containing a HR substrate as a direct repeat and inverted repeat, respectively, were used for transformation via floral dip (Clough and Bent, 1998)

Constructs.

Constructs expressing *VP16* and *VP16-HRU* from the *RPS5A* promoter were described earlier (Lindhout *et al.*, 2006). For inducible expression of *VP16, VP16-HRU* and *GFP*, a system consisting of a tamoxifen-activated GVT protein under control of the *CaMV35S* promoter, allowing tamoxifen-activation of a flanking

gene cassette under control of a yeast GAL4 UAS in virtually all cells of a plant, was taken from the vector pINTAM3 (Friml *et al.*, 2004) and further adapted. Since the pCAMBIA backbone of pINTAM3 contains multiple *Not*I sites, a 4 kb *Cla*I-*Fsp*I fragment harboring the GVT gene and the flanking UAS-driven cassette was isolated from this vector and cloned into pRF (Lindhout *et al.*, 2006), which was digested with *Cla*I and *Pme*I, thereby replacing the *RPS5A* gene cassette of pRF. The resulting plasmid was digested with *Spe*I for ligation of an *Xba*I-digested PCR product obtained by PCR on pRF with primers gff9807F (5' CGTGTCTAGACGCTCTGTTTCTCTCACCACAG) and gff10017R (5' GCACTCTAGAGCTCTATAAGCGGGTGACGAAC). This fragment provided the expression vector with a translational start site, a FLAG-tag, a unique *Not*I site, and a sequence encoding the C-terminal 36 amino acids of the VirF protein of *Agrobacterium tumefaciens*. The resulting plasmid was verified by sequence analysis and designated pTAM35S, referring to its tamoxifen-inducible *CaMV35S* promoter. After cloning a *Not*I fragment encoding a GFP reporter protein within this vector, its ability to confer tamoxifen-inducible GFP expression was verified in transgenic plants by fluorescence microscopy and western blot (Figure 3). Subsequently, *VP16* and *VP16-HRU* encoding sequences were cloned into the *Not*I site of pTAM35S.

For *RPS5A* promoter-mediated expression of HR candidate genes, the plant expression vector pGPTV-KAN (Becker *et al.*, 1992) was modified in order to obtain vector pRN. Briefly, *Not*I and *Sfi*I sites within the vector backbone were removed via sequential digestion and treatment with Klenow enzyme and T4 DNA polymerase, respectively. These modifications did not hamper the frequency of plant transformation. The promoterless GUS coding sequence was replaced by a 1.7 kb XmaI–SacI fragment containing the RPS5A promoter (Weijers *et al.*, 2001). The resulting plasmid was digested with *Sac*I to be ligated with a fragment composed of two annealed oligos pRN-*Not*-fw (5' GACGCTCTGTTTCTCTCACCACAGCGGCCGCCAGCT3') and pRN-*Not*-rv (5'-GGCGGCCGCTGTGGTGAGAGAAACAGAGCGTCAGCT3') providing the vector with a *Not*I site for subcloning cDNAs of candidate genes. The resulting pRN vector, with a unique *Not*I site, was sequence verified.

HR-candidate genes were amplified using *PfuIUltra®* DNA polymerase (Stratagene) from cDNA using primers listed in supplementary Table 4, PCR products were tailed with dATP prior to cloning into pGEM-T Easy® (Promega) according to manufacturer's instructions. HR candidate genes were sequence verified and subloned as *Not*I fragments into pRN.

Western blot analysis

TAM35S::GFP seedlings were treated with tamoxifen (1µM) for varying lengths of time in liquid B5 medium (0, 0.5, 1, 2, 4, 8, and 24 hrs). Total protein was extracted, separated by SDS-PAAGE and blotted to Immobilon P membranes (Millipore) following standard procedures (Sambrook *et al*., 1989). GFP protein was detected via N-terminal FLAG epitope using ANTI-FLAG® M2 antibody (Sigma), conjugated to alkaline phosphatise according to manufacturer's instructions.

Transcriptome analysis.

For array #1, seedlings were grown at 21 ºC for 10 days on 1/2MS plates containing 100 mg/l timentin. For array 2, involving tamoxifen induced samples, T2 seedlings originating from 50 primary transformants were grown on 1/2MS plates containing timentin and kanamycin at 100 mg/l and 30 mg/l, respectively. After 8 days of growth at 21 ºC, the seedlings were sub-cultured in liquid B5 medium for 48 hours and subsequently induced for 7 hours by adding tamoxifen to a concentration of 1 µM. For arrays #3, #4 and #5 plant material was grown in standard greenhouse conditions for 20 days. All plant material was harvested in liquid nitrogen. RNA was isolated according to manufacturer's instructions (Qiagen's RNeasy plant kit, including a DNase treatment on the column). Cy3/5 labelling was done using MessageAmpII (Ambion). After hybridization on Agilent® ArabidopsisIII arrays conform manufacturer's instruction, data were extracted using Agilent feature extraction software (version 8.1) and analysed using Rosetta Resolver System 5.1 (Rosetta Biosoftware), hosted by the Netherland Bioinformatics Centre (NBIC). For comparison with other transcriptome studies, our array data were imported into one Microsoft Office Excel file together with available transcriptome data from other studies (Chen *et al*., 2003; Molinier *et al*., 2005; Nagata *et al*., 2005) and compared using the auto filter option. Probe IDs from other studies using a different array platform were converted into AGI codes to facilitate comparisons.

Quantitative reverse transcriptase PCR

For Q-RT-PCR analysis RNA was isolated from 20-day old plants grown in the greenhouse (comparable with material used for arrays #3, #4 and #5). cDNA was synthesised using 1 µg RNA in a total reaction volume of 20 µl with a mixture of oligo-dT primers and random hexamers (iScript™cDNA Synthesis Kit, Bio-RAD). Subsequently 0.25 µl was used as template for real-time PCR (iQ™ SYBR® Green Supermix BIO-RAD) in a total volume of 25 µl with 0.4 µM of each

primer. PCRs were run on a DNA Engine Thermal Cycler (MJ Research) equipped with a Chromo4 Real-Time PCR Detection system (BIO-RAD). Normalization of relative gene expression was based upon at least 3 out of 5 housekeeping genes, according to the GeNorm normalization procedure (Vandesompele *et al.*, 2002). These reference genes included *actin2* (At3g18780), *β-6-tubilin* (At5g12250), *elongation factor 1α* (At5g60390), *adenosyl-phosphoribosyltransferase* (At1g27450), and *Roc1* (At4g38740).

Primers for reference genes (Czechowski *et al.*, 2004) and HR candidate genes are listed in supplementary Table 3. The following thermal profile was used: 95ºC for 3 min; 40 cycles of 95 ºC for 15 sec and 61 ºC for 30 sec, followed by a melting curve from 50 ºC to 95 ºC with 0.5 ºC interval read out. Data were analysed using Opticon Monitor software 3.10 (BIO-RAD).

Homologous recombination assay in plants expressing HR candidate genes

GUS staining was performed as earlier described (Lindhout *et al.*, 2006). HR events were measured in primary transformants as the average number of GUS positive spots per seedling at 10 and 20 days of growth on selective medium.

ACKNOWLEDGEMENTS

This work was supported by the Netherlands Organization for Scientific Research (NWO Programma Genomics, grant 050-10-123). We would like to thank Rob Verbeek for technical assistance and Maartje van Kregten for critically reading the manuscript

REFERENCES

Akutsu, N., Iijima, K., Hinata, T. and Tauchi, H. (2007) Characterization of the plant homolog of Nijmegen breakage syndrome 1: Involvement in DNA repair and recombination. *Biochem. Biophys. Res. Commun.*, **353**, 394-398.

Amor, Y., Babiychuk, E., Inzé, D. and Levine, A. (1998) The involvement of poly(ADP-ribose) polymerase in the oxidative stress responses in plants. *FEBS Lett.*, **440**, 1-7

Aravind, L. and Koonin, E.V. (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res.*, **26**, 3746-3752.

Babiychuk, E., Cottril, P.B., Storozhenko, S., Fuangthong, M., Chen, Y., O'Farrell, M.K., Van Montagu, M., Inzé, D. and Kushnir, S. (1998) *Plant J.,* **15**, 635-645

Bagherieh-Najjar, M.B., de Vries, O.M., Hille, J. and Dijkwel, P.P. (2005) Arabidopsis RecQI4A suppresses homologous recombination and modulates DNA damage responses. *Plant J.*, **43**, 789-798.

Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol. Biol.*, **20**, 1195-1197.

Beltran, A., Liu, Y., Parikh, S., Temple, B. and Blancafort, P. (2006) Interrogating genomes with combinatorial artificial transcription factor libraries: asking zinc finger questions. *Assay Drug Dev. Technol.*, **4**, 317-331.

Bleuyard, J.Y., Gallego, M.E. and White, C.I. (2006) Recent advances in understanding of the DNA double-strand break repair machinery of plants*. DNA Repair (Amst)*, **5**, 1-12.

Bundock, P. and Hooykaas, P.J.J (2002) Severe developmental defects, hypersensitivity to DNA-damaging agents, and lengthened telomeres in Arabidopsis MRE11 mutants. *Plant Cell*, **14**, 2451-2462.

Chen, I.P., Haehnel, U., Altschmied, L., Schubert, I. and Puchta, H. (2003) The transcriptional response of Arabidopsis to genotoxic stress - a high-density colony array study (HDCA). *Plant J.*, **35**, 771-786.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana. Plant J.*, **16**, 735-743.

Czechowski, T., Bari, R.P., Stitt, M., Scheible, W.R. and Udvardi, M.K. (2004) Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J.*, **38**, 366-379.

De Block, M., Verduyn, C., De Brouwer, D. and Cornelissen, M. (2005) Poly(ADP-ribose) polymerase in plants affects energy homeostasis, cell death and stress tolerance. *Plant J.*, **41**, 95-106.

Ditta, G., Stanfield, S., Corbin, D. and Helinski, D.R. (1980) Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl Acad. Sci. USA*, **77**, 7347-7351.

Doucet-Chabeaud, G., Godon, C., Brutesco, C., de Murcia, G. and Kazmaier, M. (2001) Ionising radiation induces the expression of PARP-1 and PARP-2 genes in Arabidopsis. *Mol. Genet. Genomics*, **265**, 954-963.

Endo, M., Ishikawa, Y., Osakabe, K., Nakayama, S., Kaya, H., Araki, T., Shibahara, K., Abe, K., Ichikawa, H., Valentine, L., Hohn, B. and Toki, S. (2006) Increased frequency of homologous recombination and T-DNA integration in Arabidopsis CAF-1 mutants. *EMBO J.*, **25**, 5579-5590.

Friml, J., Yang, X., Michniewicz, M., Weijers, D., Quint, A., Tietz, O., Benjamins, R., Ouwerkerk, P.B.F, Ljung, K., Sandberg, G., Hooykaas, P.J.J, Palme, K. and Offringa, R. (2004) A PINOID-dependent binary switch in apical-basal PIN polar targeting directs auxin efflux. *Science*, **306**, 862-865.

Gallego, M.E., Jeanneau, M., Granier, F., Bouchez, D., Bechtold, N. and White, I. (2001) Disruption of the Arabidopsis RAD50 gene leads to plant sterility and MMS sensitivity. *Plant J.*, **25**, 31-41.

Gherbi, H., Gallego, M.E., Jalut, N., Lucht, J.M., Hohn, B. and White, C.I. (2001) Homologous recombination in planta is stimulated in the absence of Rad50. *EMBO Rep.*, **2**, 287-291.

Hartung, F., Suer, S. and Puchta, H. (2007) Two closely related RecQ helicases have antagonistic roles in homologous recombination and DNA repair in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **104**, 18836-18841.

Heitzeberg, F., Chen, I.P., Hartung, F., Orel, N., Angelis, K.J. and Puchta, H. (2004) The Rad17 homologue of Arabidopsis is involved in the regulation of DNA damage repair and homologous recombination. *Plant J.*, **38**, 954-968.

Honma, T. and Goto, K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, **409**, 525-529.

James, D.W., Jr., Lim, E., Keller, J., Plooy, I., Ralston, E. and Dooner, H.K. (1995) Directed tagging of the Arabidopsis FATTY ACID ELONGATION1 (FAE1) gene with the maize transposon activator. *Plant Cell*, **7**, 309-319.

Koonin, E.V. (1994) Conserved sequence pattern in a wide variety of phosphoesterases. *Protein Sci.*, **3**, 356-358.

Kosarev, P., Mayer, K.F. and Hardtke, C.S. (2002) Evaluation and classification of RING-finger domains encoded by the Arabidopsis genome. *Genome Biol.*, **3**, RESEARCH0016.

Lazo, G.R., Stein, P.A. and Ludwig, R.A. (1991) A DNA transformation-competent Arabidopsis genomic library in Agrobacterium. *Biotechnology (N Y)*, **9**, 963-967.

Lechner, E., Achard, P., Vansir, A., Potuschak, T. and Genschik, P. (2006) F-box proteins everywhere. *Curr. Opin. Plant Biol.,* **9**, 631-638

Li, W., Chen, C., Markmann-Mulisch, U., Timofejeva, L., Schmelzer, E., Ma, H. and Reiss, B. (2004) The Arabidopsis AtRAD51 gene is dispensable for vegetative development but required for meiosis. *Proc. Natl Acad. Sci. USA*, **101**, 10596-10601.

Li, L., Jean, M. and Belzile, F. (2006) The impact of sequence divergence and DNA mismatch repair on homeologous recombination in Arabidopsis. *Plant J.*, **45**, 908-916.

Lindhout, B.I., Pinas, J.E., Hooykaas, P.J.J. and van der Zaal, B.J. (2006) Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in Arabidopsis. *Plant J.*, **48**, 475-483.

Ménissier de Murcia, J., Riboul, M., Tartier, L., Niedergang, C., Huber, A., Dantzer, F., Schreiber, V., Amé, J-C., Dierich, A., Lemeur, M., Sabatier, L., Chambon, P. and De Murcia, G. (2003) Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse. *EMBO J.,* **22**, 2255-2263

Molinier, J., Ramos, C., Fritsch, O. and Hohn, B. (2004) CENTRIN2 modulates homologous recombination and nucleotide excision repair in Arabidopsis. *Plant Cell*, **16**, 1633-1643.

Molinier, J., Oakeley, E.J., Niederhauser, O., Kovalchuk, I. and Hohn, B. (2005) Dynamic response of plant genome to ultraviolet radiation and other genotoxic stresses. *Mutat. Res.* , **571**, 235-247.

Nagata, T., Yamada, H., Du, Z., Todoriki, S. and Kikuchi, S. (2005) Microarray analysis of genes that respond to gamma-irradiation in Arabidopsis. *J. Agric. Food Chem.*, **53**, 1022-1030.

Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, **252**, 809-817.

Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M.F. (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, **405**, 200-203.

Puchta, H. (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.*, **56**, 1-14.

Ren, H., Santner, A., Del Pozo, J.C., Murray, J.A. and Estelle M. (2007) Degradation of the cyclin-dependent kinase inhibitor KRP1 is regulated by two different ubiquitin E3 ligases. *Plant J.*

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: A laboratory manual 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Schreiber, V., Ame, J.C., Dolle, P., Schultz, I., Rinaldi, B., Fraulob, V., Menissier-de Murcia, J. and de Murcia, G. (2002) Poly(ADP-ribose) polymerase-2 (PARP-2) is required for efficient base excision DNA repair in association with PARP-1 and XRCC1. *J. Biol. Chem.*, **277**, 23028-23036.

Sharples, G.J. and Leach, D.R. (1995) Structural and functional similarities between the SbcCD proteins of *Escherichia coli* and the RAD50 and MRE11 (RAD32) recombination and repair proteins of yeast. *Mol. Microbiol.*, **17**, 1215-1217.

Takeda, S., Tadele, Z., Hofmann, I., Probst, A.V., Angelis, K.J., Kaya, H., Araki, T., Mengiste, T., Mittelsten Scheid, O., Shibahara, K., Scheel, D. and Paszkowski, J. (2004) BRU1, a novel link between responses to DNA damage and epigenetic gene silencing in Arabidopsis. *Genes Dev.*, **18**, 782-793.

Tomasz, M. (1995) Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.*, **2**, 575-579.

Toufighi, K., Brady, S.M., Austin, R., Ly, E. and Provart, N.J. (2005) The Botany Array Resource: e-northerns, expression angling, and promoter analyses. *Plant J.*, **43**, 153-163.

Tsukaya, H. (2002) The leaf index: heteroblasty, natural variation, and the genetic control of polar processes of leaf expansion. *Plant Cell Physiol.*, **43**, 372-378.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**, RESEARCH0034.

Weijers, D., Franke-van Dijk, M., Vencken, R.J., Quint, A., Hooykaas, P.J.J. and Offringa, R. (2001) An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene. *Development*, **128**, 4289-4299.

Wright, D.A., Thibodeau-Beganny, S., Sander, J.D., Winfrey, R.J., Hirsh, A.S., Eichtinger, M., Fu, F., Porteus, M.H., Dobbs, D., Voytas, D.F. and Joung, J.K. (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat. Protoc.*, **1**, 1637-1652.

Zhu, H., Qian, W., Lu, X., Li, D., Liu, X., Liu, K. and Wang, D. (2005) Expression patterns of purple acid phosphatase genes in Arabidopsis organs and functional analysis of AtPAP23 predominantly transcribed in flower. *Plant Mol. Biol.*, **59**, 581-594.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W. (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiol.*, **136**, 2621-2632.

SUPPLEMENTARY TABLES

(S) Table 1: Genes significantly upregulated (at least 2-fold) in all four *'RPS5A'* arrays. For each of the arrays fluorescence intensity values and fold regulations are represented.

		20 days greenhouse array #3			20 days greenhouse array #4			20 days greenhouse array #5			10 days tissue culture array #1			7 hours induction array #2		
Primary sequence name	Annotation	1406	1406-VP16-HRU	Fold change	1415	1415- VP16-HRU	Fold change	1406-VP16	1406-VP16-HRU	Fold change	1406	1406-VP16-HRU	Fold change	1406-VP16	1406-VP16-HRU	Fold change
At1g43780	serine carboxypeptidase S10 family protein	34	2447	71.0	41	819	20.0	33	2495	76.7	36	3742	104.2	230	111	-2.1
At5g66150	glycosyl hydrolase family 38 protein	34	879	26.0	49	465	9.5	118	1297	10.9	61	1963	32.1	291	294	1.0
At1g35730	pumilio/Puf RNA-binding domain-containing protein	58	1404	24.2	38	340	8.9	270	1540	5.7	69	1458	21.0	1316	856	-1.5
At5g13825	hypothetical protein	31	647	21.0	39	180	4.6	37	644	17.5	28	377	13.6	23	62	2.7
At5g03860	malate synthase	38	683	17.8	38	339	8.9	26	659	25.3	505	3192	6.3	1054	497	-2.1
At3g27630	hypothetical protein	315	5533	17.6	199	1424	7.2	245	5131	20.9	590	1317	2.2	852	786	-1.1
At2g27880	argonaute protein, putative / AGOPF02171: PIwi domain	71	852	12.0	41	194	4.8	206	926	4.5	171	3244	19.0	394	179	-2.2
At4g16050	expressed protein	30	350	11.6	37	168	4.6	35	382	10.9	44	517	11.8	296	227	-1.3
At5g60250	zinc finger (C3HC4-type RING finger) family protein	105	1180	11.2	74	426	5.8	112	1324	11.9	202	561	2.8	523	670	1.3
At4g05470	F-box family protein (FBL21) contains similarity to N7 protein GI:3273101 from [Medicago truncatula]	41	457	11.2	36	194	5.4	310	655	2.1	1311	4563	3.5	64	72	1.1
At3g27250	expressed protein	169	1475	8.8	185	767	4.1	524	1632	3.1	176	561	3.2	147	176	1.2
At3g11402	DC1 domain-containing protein	278	2002	7.2	253	1338	5.3	444	1812	4.1	258	2977	11.5	7411	3520	-2.1
At3g55890	yippee family protein similar to mdgl-1 [Mus musculus] GI:10441648	1022	7306	7.1	619	2267	3.7	802	6510	8.1	382	1997	5.2	976	646	-1.5
At5g32161	hypothetical protein	64	449	7.0	69	280	4.1	100	490	4.9	103	570	5.5	127	111	-1.1
At5g20710	beta-galactosidase, putative / lactase,	112	769	6.9	93	439	4.7	57	832	14.5	429	3017	7.0	602	454	-1.3
At3g10150	calcineurin-like phosphoesterase family protein SEC14 cytosolic factor, putative /	192	1313	6.9	119	605	5.1	244	1179	4.8	155	882	5.7	200	92	-2.2
At4g34580	phosphoglyceride transfer protein,	243	1627	6.7	223	1031	4.6	435	1603	3.7	2576	5275	2.0	2404	1494	-1.6
At3g55910	expressed protein PA26	528	3369	6.4	431	1108	2.6	440	2854	6.5	471	1463	3.1	1462	964	-1.5
CHR4v012120U	Unknown	2449	15214	6.2	1888	13946	7.4	2736	12627	4.6	137	1149	8.4	211	156	-1.4
AT5G55410	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	88	550	6.2	82	255	3.1	125	582	4.6	78	810	12.1	152	97	-1.5
At1g59660	nucleoporin family protein	691	4248	6.1	778	2945	3.8	1051	3903	3.7	1135	5193	4.6	5756	4558	-1.3
At5g15800	developmental protein SEPALLATA1 / floral homeotic protein (AGL2) (SEP1)	266	1578	5.9	169	903	5.3	293	1893	6.5	52	384	7.3	88	106	1.2
At1g59920	hypothetical protein	545	3204	5.9	470	3022	6.4	632	2720	4.3	198	1264	6.4	295	285	-1.0
At1g59930	hypothetical protein	629	3614	5.7	494	3270	6.6	612	3008	4.9	213	1187	5.6	265	235	-1.1
At5g16360	NC domain-containing protein expressed protein similar to auxin down-regulated protein ARG10 [Vigna radiata]	709	4005	5.6	525	1354	3.5	399	3544	8.9	1508	6962	4.6	5735	3211	-1.8
At3g22850	Gi:29700510	4257	23182	5.4	3286	13878	4.2	3841	19186	5.0	22811	55796	2.4	40854	31346	-1.3
At4g05640	hypothetical protein	47	242	5.2	65	222	3.4	85	350	4.1	74	434	5.9	55	85	1.5
At3g26320	cytochrome P450 71B36, putative (CYP71B36)	883	4148	4.7	692	2285	3.3	1006	3756	3.7	477	3605	7.6	817	670	-1.2
At2g17040	no apical meristem (NAM) family protein	8269	37832	4.6	10649	35281	3.3	18528	37311	2.0	2031	4533	2.2	1523	1260	-1.2
At4g25000	alpha-amylase	616	2789	4.5	722	1726	2.4	859	2512	2.9	82	432	5.2	177	162	-1.1
At5g58770	dehydrodolichyl diphosphate synthase,	3334	14962	4.5	2196	6573	3.0	2437	12656	5.2	955	9292	9.7	7368	3590	-2.1

Continuation of (S) Table 1: Genes significantly upregulated (at least 2-fold) in all four *'RPS5A'* arrays. For each of the arrays fluorescence intensity values and fold regulations are represented.

		20 days greenhouse array #3			20 days greenhouse array #4			20 days greenhouse array #5			10 days tissue culture array #1			7 hours induction array #2		
Primary sequence name	Annotation	1406	1406-VP16-HRU	Fold change	1415	1415- VP16-HRU	Fold change	1406-VP16	1406-VP16-HRU	Fold change	1406	1406-VP16-HRU	Fold change	1406-VP16	1406-VP16-HRU	Fold change
At5g45095	expressed protein	119	525	4.4	65	376	5.8	30	474	16.0	211	909	4.3	512	210	-2.4
At5g01870	lipid transfer protein	374	1405	3.8	250	694	2.8	99	1416	14.4	11869	29217	2.5	10712	10365	-1.0
At1g58150	hypothetical protein	882	3244	3.7	821	2031	2.5	1156	3130	2.7	1499	5634	3.8	8927	4417	-2.0
At1g35310	Bet v I allergen family protein	114	419	3.7	126	317	2.5	176	554	3.1	1093	3295	3.0	1240	1732	1.4
At1g55380	DC1 domain-containing protein	172	620	3.6	108	404	3.8	339	800	2.4	251	673	2.7	423	327	-1.3
At5g50130	short-chain dehydrogenase/reductase (SDR) family protein	1843	6150	3.3	1674	3538	2.1	1792	5378	3.0	1654	3806	2.3	1191	1008	-1.2
At3g28910	myb family transcription factor (MYB30)	5496	17684	3.2	4683	10032	2.1	4801	14808	3.1	8044	26978	3.4	13610	9795	-1.4
At2g24440	expressed protein	634	2033	3.2	635	1487	2.3	564	2069	3.7	1956	5222	2.7	4093	3143	-1.3
At4g39510	cytochrome P450 family protein	2069	6482	3.1	1335	3036	2.3	975	5088	5.2	6138	13357	2.2	5561	5796	1.0
At1g60110	jacalin lectin family protein	201	624	3.1	147	509	3.5	124	567	4.6	203	568	2.8	360	216	-1.7
At3g12510	expressed protein	823	2458	3.0	669	1544	2.3	801	2328	2.9	1186	4273	3.6	17963	9893	-1.8
At1g53480	expressed protein	232	687	3.0	152	526	3.5	230	769	3.3	155	818	5.3	229	352	1.5
At3g59900	expressed protein	1211	3562	2.9	1119	3783	3.4	768	3342	4.4	1470	5606	3.8	4155	2905	-1.4
At1g77590	long-chain-fatty-acid--CoA ligase family protein / long-chain acyl-CoA synthetase family protein (LACS9) 3 [SP|Q63151] from Rattus norvegicus;	2646	7762	2.9	2091	5429	2.6	1968	6725	3.4	4223	16541	3.9	7349	5001	-1.5
At3g08570	phototropic-responsive protein,	662	1853	2.8	636	1661	2.6	839	2042	2.4	1278	2704	2.1	1710	1390	-1.2
At4g12700	expressed protein	871	2409	2.8	804	1693	2.1	793	2562	3.2	1790	4107	2.3	2864	2083	-1.4
At1g06420	expressed protein	357	973	2.7	180	501	2.8	312	878	2.8	649	1933	3.0	989	700	-1.4
At5g35120	hypothetical protein	1107	3011	2.7	898	2959	3.3	863	2482	2.9	545	1546	2.8	605	521	-1.2
At1g02690	importin alpha-2 subunit	561	1513	2.7	468	1008	2.2	606	1560	2.6	4655	10357	2.2	6577	5783	-1.1
At4g34650	farnesyl-diphosphate farnesyltransferase 2 / squalene synthase 2 (SQS2)	596	1574	2.6	453	1070	2.4	364	1572	4.3	623	1485	2.4	1023	930	-1.1
At3g05600	epoxide hydrolase	1842	4841	2.6	668	1868	2.8	1060	4010	3.8	3763	12909	3.4	4364	4392	1.0
At1g09890	expressed protein	140	368	2.6	177	412	2.3	230	469	2.0	3669	17881	4.6	663	490	-1.4
At1g20870	expressed protein	143	368	2.6	110	225	2.0	163	449	2.8	206	743	3.6	791	678	-1.2
At5g57780	expressed protein similar to unknown protein (emb CAB79759.1)	7234	18454	2.6	5680	15626	2.8	3694	14864	4.0	2518	5747	2.3	2620	1964	-1.3
At4g05440	temperature sensing protein-related	424	1057	2.5	313	721	2.3	487	1055	2.2	469	1122	2.4	1392	1333	-1.0
At1g34315	expressed protein	838	2071	2.5	531	1075	2.0	647	1791	2.8	295	1103	3.7	630	378	-1.7
At5g35490	expressed protein (MRU1)	6835	16610	2.4	4951	14813	3.0	4384	13540	3.1	1124	3732	3.3	1053	968	-1.1
At5g35480	expressed protein	32812	72845	2.2	20147	75309	3.7	21631	65920	3.0	1891	6743	3.6	1782	1457	-1.2
CHR1v012120	Unknown	78	171	2.2	62	170	2.7	98	247	2.5	128	473	3.7	161	229	1.4
At1g21050	expressed protein	13244	28641	2.2	9855	22072	2.2	11178	24207	2.2	11308	25052	2.2	8970	6416	-1.4
At1g78440	gibberellin 2-oxidase / GA2-oxidase (GA2OX1)	133	271	2.0	105	360	3.4	121	391	3.2	184	513	2.8	183	72	-2.5
At1g32900	starch synthase	2966	5956	2.0	1439	3718	2.6	1749	5845	3.3	3251	7533	2.3	4717	3543	-1.3

(S) Table 2: Genes significantly downregulated (at least 2-fold) in all four *'RPS5A'* arrays. For each of the arrays fluorescence intensity values and fold regulations are represented.

		20 days greenhouse array #3			20 days greenhouse array #4			20 days greenhouse array #5			10 days tissue culture array #1			7 hours induction array #2		
Primary sequence name	**Annotation**	**1406**	**1406-VP16-HRU**	**Fold change**	**1415**	**1415- VP16-HRU**	**Fold change**	**1406-VP16**	**1406-VP16-HRU**	**Fold change**	**1406**	**1406-VP16-HRU**	**Fold change**	**1406-VP16**	**1406-VP16-HRU**	**Fold change**
At1g21460	nodulin MtN3 family protein	4158	1554	-2.7	4245	2022	-2.1	3219.1	1399.4	-2.3	2558	1028	-2.5	1.0	1922.5	2010.4
At1g11170	expressed protein	678	227	-3.0	731	314	-2.3	1108.4	458.7	-2.4	668	304	-2.2	1.0	663.6	680.0
At1g55320	acyl-activating enzyme 18 (AAE18)	372	122	-3.0	326	138	-2.4	359.1	113.0	-3.2	634	241	-2.6	-1.1	589.4	553.4
At1g18710	myb family transcription factor (MYB47)	1682	501	-3.4	2280	1000	-2.3	1399.9	509.8	-2.7	659	107	-6.2	-1.2	79.2	67.1
At3g46370	leucine-rich repeat protein kinase	2677	730	-3.7	2059	930	-2.2	1524.1	748.2	-2.0	570	74	-7.7	1.0	108.8	112.6
At2g18050	histone H1-3 (HIS1-3)	5641	1398	-4.0	6109	1764	-3.5	2921.8	1288.6	-2.3	8605	3292	-2.6	1.1	1555.6	1643.0
At2g43590	chitinase	2540	575	-4.4	1353	314	-4.3	2159.8	627.4	-3.4	15519	7618	-2.0	-1.3	6293.0	4867.9
At3g29370	expressed protein	1605	224	-7.2	1268	582	-2.2	814.3	227.6	-3.6	3466	1043	-3.3	-1.6	1572.5	1007.0

(S) Table 3: Q-RT-PCR primers (from 5' to 3')

gene label	forward primer	reverse primer
At3g18780	TCCCTCAGCACATTCCAGCAGAT	AACGATTCCTGGACCTGCCTCATC
At5g12250	ACCACTCCTAGCTTTGGTGATCTG	AGGTTCACTGCGAGCTTCCTCA
At5g60390	TGAGCACGCTCTTCTTGCTTTCA	GGTGGTGGCATCCATCTTGTTACA
At1g27450	GTTGCAGGTGTTGAAGCTAGAGGT	TGGCACCAATAGCCAACGCAATAG
At4g38740	GAACGGAACAGGCGGTGAGTC	CCACAGGCTTCGTCGGCTTTC
At1g43780	CTTATGGAGCTTGGTTCGAC	CGTGAAGGCTGCGAGTAC
At5g07610	AGACGAACCGCTACCTG	CTTCACGGACCAGCTTG
At1g66300	TTCTTGTCGGGATCTCTGGTG	ATTCGGGCAGCTCTCAAG
At5g60250	GGAGATGTGTTGAATGTCGTGG	TTCGTTCCAGGTTGGGCA
At4g02390	CTCGGCAAGATAAGCAAGTCC	GAGTGTCTATAACAAACTGGC
At3g10150	TGGGACAAAGCAATCTCTC	AACGAAGGGGAGGAATACC
At5g15800	AGCAAGTTCGGTCCATCAA	ATCCTCCTCCTCCTCCCATA
At5g20850	CCAACAACAAGACGATGAAG	GTGAACCCCAGAGGAACTA
At5g48720	GAGGACACGAACAGGAAA	GCAATCAGCATACCACTCTT

(S) Table 4: Primers (from 5' to 3') for amplification of cDNAs of putative HR-affecting genes.

gene label	forward primer	reverse primer
At1g43780	GAAAGAGAAATGGTGGGAGGTAAGTGGCGGT	CTTCTCCTTCAATCATGAAGAGCAGGAGGCGA
At1g66300	GAAGTTACATGGACGAAGACGGAGAGA	CACAGTATGTTCGTCAACGTGAATCAGAGAAGGATAACTT
At3g10150	GAAGTCGATGAAAAAACCGTCGTTGTTTCAGA	GATAAATAGACTAATCATATGTTAGGTTAACTTCACT
At3g55890	CAAAAGATGGGTAGGGTTTTTATGGTTGAT	GCCTCAATCAAGCATCATCTCCATCA
At3g55910	GATACAAATGAAGAAGGCGCTCACCACA	GTTGATGGAATTAATTTTCCCACAAATATATTTTCCT
At4g02390	GACGAAAATGGCGAACAAGCTCAAAGTCGACGA	CATAAGTTTTAGTGCTTGTAGTTGAATTTGACTTGGA
At5g07610	CCAGCGATGTCCTCTTGTTCAAGAACAAGAA	CTACTATTTAGACATTAGCAAGAGACTTGATGAACTGGA
At5g15800	AAAGAAATGGGAAGAGGAAGAGTAGAGCTGAA	CATGACTCTCAGAGCATCCACCCCGGGAT
At5g20850	GAGAGAAATGACGACGATGGAGCAGCGTAGA	CCAAAAACATCAATCCTTGCAATCTGTTACACCT
At5g35120	CAAAAAGATGGATCAAGAGGCTTTTTTGCGT	GTTTCAGACTAGTGGTGGCTACCATCCATTGA
At5g46915	GAAAGCGATGGCTACTACGAACCCTGGATA	TTCCTCAGTAATTTCTTCTCACCCTCACATCCA
At5g48720	GTCAAAATGGTGCTCTATGGTCAGAATCCA	CTTAGCTTTAGCCTCTAGTCCTCATTATCGTGATGCT
At5g60250	GCAAGTATGGATTTCGATTCCGAGTATGCTTTTCGGT	CCAAACCTCAGAAGTGAGGATAGTCAGGATGGAA