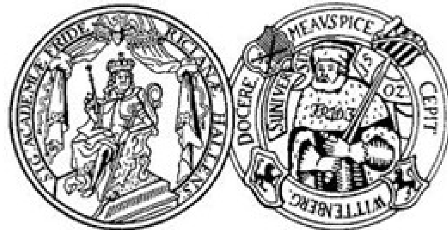
Analysis of sense transgene-induced gene silencing in introgression lines reveals the presence of silencing modulators in *Arabidopsis thaliana* accession genomes



Dissertation zur Erlangung
des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)
der Naturwissenschaftlichen Fakultät I
– Biowissenschaften –
der Martin-Luther-Universität Halle-Wittenberg,

vorgelegt

von Frau Le Phuong Dung
geboren am 16. Juli 1985 in Thai Nguyen, Vietnam

verteidigt am 25.04.2017, Halle (Saale)

Gutachter:

1. Prof. Dr. Thomas Altmann (IPK, Gatersleben, Martin-Luther-Universität, Germany)
2. Prof. Dr. Gunther Reuter (Martin-Luther-Universität, Halle-Wittenberg, Germany)
3. Prof. Dr. Daniel Schubert (Institut für Biologie, Freie Universität Berlin, Germany)

Acknowledgements

I express my sincere gratitude for the financial support of the Ministry of Education and Training (MOET) Vietnam. Receiving this scholarship helped me to fulfill my dream of studying abroad. I am also very grateful for all the chances that the scholarship granted by the German Academic Exchange Service (DAAD) provided for me.

I thank Prof. Dr. Thomas Altmann for the opportunity to conduct my work and studies at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben and the Martin Luther University of Halle-Wittenberg.

I express my gratitude to my supervisor Dr. Renate Schmidt for her support, encouragement, valuable suggestions, creative and constructive guidance. I could not have imagined having a better and supportive supervisor for my PhD study.

Likewise, I want to thank all my colleagues in the research group Genome Plasticity at the IPK, both present and past, for helpful discussions, suggestions, scientific advice as well as all the fun and memorable moment we had. In particular I want to acknowledge Dr. Hieu Xuan Cao and Loan Thanh Le and the helpful assistants Kristin Langanke, Helga Berthold and Christa Walter.

I would like to offer my special thanks to Dr. Michael Florian Mette for many valuable discussions and suggestions. I am grateful to Dr. Yusheng Zhao for his advice on the statistical analysis of data. My thanks also goes to Dr. Britt Leps for her kind and valuable support regarding administrative issues.

I wish to express my thanks to Assoc. Prof. Dr. Pham Hong Quang, Assoc. Prof. Dr. Le Ngoc Cong, Assoc. Prof. Dr. Nguyen Thi Tam, and Dr. Nguyen Huu Cuong for their encouragement and support.

I would like to express my gratitude to all my friends, who directly or indirectly, have given their hand whenever I needed it. My special thanks go to my Vietnamese friends living in Gatersleben, especially Dr. Hoang Trong Phan, Dr. Ha Minh Pham, Dr. Trung Duc Tran and my beloved little friends Minh Ha Phan (Nị) and Anna Phan for the warm care, sympathies and the important and memorable moments they shared with me.

Last but not least; I would like to thank my parents, Lê Ngọc Công and Bùi Thị Dậu, and my parents-in law for their love and support. Their faith in my abilities and encouragement throughout the journey of my PhD study has been invaluable. A special note of gratitude to all my relatives for their concern and encouragement. This acknowledgement cannot be complete without thanking my husband Dr. Thanh Nguyen Tien for his love, being very patient with me through the good and bad times, for his understanding, support and encouragement during all these years.

TABLE OF CONTENTS

List of figures i
List of tables iii
Abbreviations iv

1 INTRODUCTION 1
1.1 Transgene expression and silencing in plants 1
1.2 Sense-transgene induced post-transcriptional gene silencing in plants 3
1.3 Silencing spread 12
1.4 Impact of environmental conditions on gene silencing 13
1.5 Analysis of natural variation in *Arabidopsis thaliana* 14
1.6 Aims of the study 18

2 MATERIALS AND METHODS 19
2.1 Materials 19
2.1.1 Laboratory equipment 19
2.1.2 Chemicals, enzymes, kits and materials for plant cultivation 20
2.1.3 Buffers and solutions 20
2.1.4 *Arabidopsis thaliana* accessions and transgenic lines 21
2.1.5 Softwares 23
2.2 Methods 23
2.2.1 Plant growth conditions 23
2.2.2 Crossing of *Arabidopsis thaliana* accessions to *GFP* transgenic lines in Col-0 background 24
2.2.3 Isolation of DNA from plant leaves of *Arabidopsis thaliana* 24
2.2.4 Isolation of total DNA from aerial seedling tissues of *Arabidopsis thaliana* 25
2.2.5 Amplicon design 25
2.2.5.1 Amplicons for allelic diversity studies 26
2.2.5.2 Amplicons for RT-PCR and qRT-PCR 26
2.2.6 Polymerase chain reaction (PCR) 27
2.2.7 Agarose gel electrophoresis 27
2.2.8 Purification of PCR products for direct sequencing 28
2.2.9 Sequence analysis 28
2.2.9.1 Sequence alignments and comparisons 28
2.2.9.2 Polymorphism analysis 29
2.2.9.3 Identification of microsatellites 29
2.2.10 Generation of introgression lines (ILs) 29
2.2.11 Detection and imaging of *GFP* fluorescence 30
2.2.12 Analysis of *GFP* gene silencing 31
2.2.13 qRT-PCR experiments 32
2.2.13.1 Isolation of RNA from *Arabidopsis thaliana* aerial seedling tissues 32
2.2.13.2 DNAse treatment of total RNA 32
2.2.13.3 cDNA synthesis and RT-PCR 33
2.2.13.4 qRT-PCR experiment set up 33

3 RESULTS **35**
3.1 Sequence diversity in *Arabidopsis thaliana* genes that are involved in sense-transgene induced post-transcriptional gene silencing 35
3.1.1 Analysis of sequence variation in candidate genes – *NRPE1* as an example 35
3.1.2 Survey of sequence variation in *Arabidopsis thaliana* accessions revealed highly diverged allelic variants for several candidate genes in subsets of the accessions 38
3.1.3 Pairwise comparisons of selected allelic variants 44
3.2 Functional analysis of selected allelic variants 49
3.2.1 Gene expression analysis of selected allelic variants 50
3.2.2 Generation of introgression lines 53
3.2.3 Evaluation of molecular markers for indel polymorphisms in *Arabidopsis thaliana* accessions 56
3.2.4 Characterisation of the introgression lines with respect to number, length and position of introgressed segments 58
3.2.5 Analysis of *GFP* gene silencing 63
3.2.6 Analysis of introgression lines carrying Sq-8 allelic variants of the *HEN1* gene 65
3.2.7 Subpopulations of lines show a similar behaviour with respect to gene silencing 68
3.2.8 Comparative analysis of 6x*GFP* lines carrying different T-DNA locus combinations in the Col-0 genetic background 69
3.2.9 Several introgression lines show significantly more or less silencing than reference line 6x*GFP-F8/R127* 72
3.2.10 Analysis of introgression lines carrying Gie-0 alleles for the *AGO7* and *NRPD1* genes 75
3.2.11 Analysis of introgression lines carrying Sq-8 allelic variants of the *WEX* gene 76
3.2.12 Analysis of introgression lines carrying allelic variants of the *NRPE1* gene 82
3.2.13 Identification of genome regions in the Shahdara and Cvi-0 introgression lines which enhance post-transcriptional gene silencing 85

4 DISCUSSION **89**
4.1 Choice of candidate genes 89
4.2 Polymorphism patterns of twelve genes associated with PTGS in 25 *Arabidopsis thaliana* accessions 89
4.3 Expression analysis of selected alleles 93
4.4 Analysis of introgression lines with Indel markers 93
4.5 The study of gene silencing in the introgression lines 97
4.6 Comparisons between Col-0 transgenic lines carrying six *GFP* copies each 99
4.7 Assessing introgression lines for an impact on gene silencing 99
4.8 Analysis of lines showing a pronounced effect on gene silencing 101
5 SUMMARY **107**
6 ZUSAMMENFASSUNG **108**
7 REFERENCES **110**
8 SUPPLEMENTARY DATA **121**
Curriculum vitae **166**
Declarations **168**

List of Figures

Main Figures

Figure 1. Model for sense-PTGS pathway in *Arabidopsis thaliana* 4
Figure 2. Determining the presence and zygosity of a particular T-DNA locus in a transgenic line .. 30
Figure 3. Photographic documentation of a plant showing *GFP*-silencing 31
Figure 4. Amplicons developed for the *NRPE1* gene 35
Figure 5. Multiple alignment of sequences derived from *A. thaliana* accessions for a region of amplicon 9 of the *NRPE1* gene 36
Figure 6. Alignment of *WEX* gene sequences obtained for 26 *A. thaliana* accessions reveals a highly polymorphic region 41
Figure 7. RT-PCR experiments reveal expression of selected candidate genes in aerial seedling tissues………. 51
Figure 8. Expression analysis of *HEN1, SDE3, AGO7*, *NRPE1* and *WEX* genes in selected accessions .. 52
Figure 9. Map position of the candidate genes and *GFP* loci on the five chromosomes of *Arabidopsis thaliana* and crossing scheme for the generation of introgression lines 54
Figure 10. Evaluation of introgression lines for the presence and zygosity of T-DNA loci and alleles of interest...... 55
Figure 11. Characterisation of introgression lines containing allelic variants of the *WEX* gene........ 60
Figure 12. *GFP* expression and silencing in plants of introgression lines............ 63
Figure 13. Comparisons to determine significant differences between subpopulations of a particular line or between an introgression line and the reference line 6x*GFP-F8/R127*............ 65
Figure 14. Introgression lines carrying the Sq-8 allelic variant of the *HEN1* gene differ with respect to silencing............ 67
Figure 15. Comparison of 6x*GFP* lines carrying different T-DNA locus combinations with respect to *GFP* silencing 70
Figure 16. Comparison of the frequency of silencing of introgression lines carrying Gie-0 allelic variants of the *AGO7* and *NRPD1* genes 75
Figure 17. Comparison of *GFP* silencing between introgression lines carrying Sq-8 allelic variants of the *WEX* gene and the reference line 6x*GFP-F8/R127* 77
Figure 18. Position and extent of introgressed segments in introgression lines carrying Sq-8 allelic variants of the *HEN1* and/or *WEX* genes 80
Figure 19. Introgression lines with contrasting genotypes in regions of *Arabidopsis thaliana* chromosomes 2, 4 and 5 show differences with respect to gene silencing 81
Figure 20. IL_Shahdara_10 showed more silencing than IL_Shahdara_6............ 83
Figure 21. Significantly increased silencing in one of two introgression lines carrying the Cvi-0 allelic variant of the *NRPE1* gene 84
Figure 22. Position and extent of introgressed segments in introgression lines carrying allelic variants of the *NRPE1* gene............ 86
Figure 23. Introgression lines with contrasting genotypes in a region of *Arabidopsis thaliana* chromosome 2 show differences with respect to gene silencing............ 88

Supplementary figures

Supplementary figure 1. Pairwise genetic distances of 360 *A. thaliana* accessions using 149 SNPs 161
Supplementary figure 2. Characterisation of introgression lines that carry allelic variants of the *HEN1* gene with Indel markers 162
Supplementary figure 3. Chromosome maps of introgression lines containing allelic variants of the *SDE3* gene 163
Supplementary figure 4. Graphical genotypes of introgression lines carrying allelic variants of the *AGO7* and/or *NRPD1* genes 164
Supplementary figure 5. Chromosomal location and sizes of introgressed segments for introgression lines containing allelic variants of the *NRPE1* gene 165

List of Tables

Main tables

Table 1. List of *Arabidopsis thaliana* accessions used in this study 22
Table 2. Growth conditions of *Arabidopsis thaliana* plants 24
Table 3. Standard PCR reaction mixture and amplification conditions 27
Table 4. Sequence diversity of the *NRPE1* gene in 26 *Arabidopsis thaliana* accessions 37
Table 5. Sequence regions analysed for the different candidate genes with respect to allelic diversity.......... 39
Table 6. Alleles of several candidate genes show high SNP frequencies when compared to the corresponding Col-0 gene sequences 40
Table 7. Summary of the SNPs detected in 25 accessions for 12 candidate genes 42
Table 8. Indel variation of candidate genes in 25 *Arabidopsis thaliana* accessions 43
Table 9. Allelic variants selected for functional analysis 45
Table 10. Pairwise sequence identity levels of selected *NRPE1* alleles 46
Table 11. Pairwise identity levels of selected *WEX* alleles 47
Table 12. Screening of Indel markers 57
Table 13. Number of polymorphic Indel markers identified for selected accessions 58
Table 14. Characterisation of introgressed segments 62
Table 15. Comparison of the number of silenced and non-silenced plants in introgression lines carrying the Sq-8 allelic variant of the *HEN1* gene in different experiments 68
Table 16. Comparison of gene silencing revealed few significant differences between 6x*GFP* lines carrying different T-DNA locus combinations in the Col-0 genetic background 71
Table 17. Silencing frequencies observed for 6x*GFP* lines carrying different T-DNA locus combinations in the Col-0 genetic background 71
Table 18. Summary of significant differences with respect to gene silencing between introgression lines and the reference line 6x*GFP-F8/R127* 73

Supplementary tables

Supplementary table 1. Amplicons used in allelic diversity studies in 26 *Arabidopsis thaliana* accessions…… 121
Supplementary table 2. Amplicons used for amplification of specific regions of candidate genes in selected accessions 123
Supplementary table 3. Oligonucleotide pairs for semi-quantitative RT-PCR and/or qRT-PCR of reference and candidate genes 123
Supplementary table 4. Indel markers used for the analysis of introgression lines 124
Supplementary table 5. Indel markers and allele-specific oligonucleotides for selected accessions and candidate genes 127
Supplementary table 6. Primer sequences for the analysis of *GFP* T-DNA lines 127
Supplementary table 7. Regions of candidate genes and ORFs sequenced in all 26 accessions 128
Supplementary table 8. Compilation of SNPs and Indels detected in 26 accessions for 12 candidate genes 129
Supplementary table 9. cDNA information of twelve candidate genes 158
Supplementary table 10. Screening for polymorphic Indel markers 159

Abbreviations

Abbreviation	Definition
^{o}C	Degree centigrade
6x*GFP*	Six copies of the *GFP* gene
A	Adenine
AFLP	Amplified fragment length polymorphism
A. thaliana	*Arabidopsis thaliana*
AGO	Argonaute
BC	Back cross
bp	Base pair
C	Cytosine
C. elegans	*Caenorhabditis elegans*
CaMV	*Cauliflower Mosaic Virus*
cDNA	Complementary DNA
CTAB	Cetyltrimethyl ammonium bromide
DCL4	Dicer-like 4
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
EDTA	Ethylenediamine tetraacetic acid
EDS	Empty donor site
ERI	Enhancer of RNA interference
EST	Expressed sequence tag
G	Guanine
GFP	Green fluorescence protein
GWAS	Genome-wide association study
h	hour(s)
HEN1	Hua enhancer1
IL	Introgression line
Indels	Insertions/deletions
LB	Left border
Mbp	Mega base pair
miRNA	microRNA
mRNA	messenger RNA
N. benthamiana	*Nicotiana benthamiana*
nt	Nucleotide
NRPD1	Nuclear RNA polymerase D1
NRPE1	Nuclear RNA polymerase E1
ORF	Open reading frame
PCR	Polymerase chain reaction
Pol	DNA-dependent RNA polymerases
PTGS	Post-transcriptional gene silencing
qRT-PCR	Quantitative RT-PCR
RB	Right border
RdDM	RNA-directed DNA methylation
RDR	RNA-dependent RNA polymerase
RFLP	Restriction fragment length polymorphism
RIL	Recombinant inbred line
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
RT-PCR	Reverse transcription polymerase chain reaction
QTL	Quantitative trait locus
qRT-PCR	Quantitative real-time PCR
SDE3	Silencing defective 3
SDE5	Silencing defective 5
SDS	Sodium dodecyl sulfate
sec	Second(s)
SGS3	Suppressor of gene silencing 3
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
ssRNA	Single-stranded RNA
T	Thymine
TBE	Tris-borate-EDTA
T-DNA	Transfer DNA
TAIR	The Arabidopsis Information Resource
ta-siRNA	*trans*-acting siRNA
Tris	Tris (hydroxymethyl)-amino-methane
Tris-HCl	Tris (hydroxymethyl)-amino-methane hydrochloric acid
UTR	Untranslated region
XRN4	Exoribonuclease 4
WEX	Werner syndrome-like exonuclease

1 INTRODUCTION

1.1 Transgene expression and silencing in plants

Genetic transformation of plants has become a widely used technology that serves multiple purposes in plant biotechnology and research. For instance, transgene technology was used to engineer certain plant traits including disease resistance, stress tolerance, increased nutritional value and male sterility through the stable expression of transgenes (Daniell, 2002; Lanfranco, 2002). For the use of genetically modified crops high and stable expression of transgenes is in many cases an indispensable prerequisite, thus it is important to understand the factors which play a role not only in model organisms but also in crop plants (Kohli *et al*., 2006). Even more so as transgenic plants are also used in many studies as a tool to study gene function by over-expressing the genes of interest (Lloyd, 2003).

Transgenes, often delivered by *Agrobacterium tumefaciens* as part of the T-DNA, are integrated into different positions of a plant nuclear genome. In transgenic lines repeat arrangement of T-DNAs are frequently observed, likewise truncated and/or rearranged T-DNAs are readily found. Independent transgenic lines differ therefore with respect to number, arrangement and position of transgene copies in the genome (Feldmann, 1991; Tinland, 1996; Rios *et al*., 2002; Forsbach *et al*., 2003; Lechtenberg *et al*., 2003). Moreover, among the lines transformed with a particular transgene large variation with respect to transcript level of the introduced gene is seen (Holtorf *et al*., 1995), a subset can fail to express the introduced gene as a result of gene silencing (Matzke *et al*., 1989; Scheid *et al*., 1991). Gene silencing phenomena include all cases in which the inactivation of gene expression is not explained by an alteration or loss of DNA sequences. Two different types of gene silencing can be distinguished, transcriptional and post-transcriptional gene silencing (TGS, PTGS) (Meyer and Saedler, 1996; Vaucheret *et al*., 1998). Transgene expression can be inhibited at the level of transcription, thus a particular mRNA species is not synthesised any longer (Scheid *et al*., 1991). If transgenes are still transcribed but the transcript is not stable due to degradation one refers to post-transcriptional gene silencing (Napoli *et al*., 1990; Smith *et al*., 1990; Van der Krol *et al*., 1990). TGS and PTGS have the formation of double-stranded RNA (dsRNA) in common which is processed into short dsRNA fragments by an RNaseIII-type nuclease, Dicer. The small RNAs are then loaded into the RISC (RNA-induced

silencing complex) and target complementary RNA or DNA, resulting in RNA cleavage or translational inhibition in the case of PTGS or DNA methylation or chromatin modification in case of TGS (Baulcombe, 2004; Moazed, 2009). It should be noted that the phenomenon of RNA silencing is not limited to plants but some of the key components are evolutionarily conserved in other eukaryotes, such as animals, fungi, algae and protists (Waterhouse, 2001; Ghildiyal and Zamore, 2009).

TGS is typically associated with small interfering RNAs homologous to the promoter sequence, often DNA methylation of the promoter sequences is observed (Meyer, 1995; Mette *et al*., 2000; Vaucheret and Fagard, 2001). In PTGS, the accumulation of small interfering RNAs corresponding to the transcribed sequence of the transgene is observed (Hamilton and Baulcombe, 1999). If DNA methylation is found it is confined to transcribed regions of the transgene. Whereas TGS is usually mitotically and meiotically stable, PTGS is established during plant development and may spread throughout the plant, in each generation the process starts anew after resetting (Vaucheret *et al*., 1998).

Various factors are thought to affect the variation of transgene expression in independent transgenic lines. For instance, the choice of promoters influences transgene expression levels and also affects the magnitude of expression variability among individual transformants (Holtorf *et al*., 1995; De Bolle *et al*., 2003). Factors which have been implicated in the inactivation of transgenes included the transgene insertion site and copy number of introduced transgenes (Matzke and Matzke, 1998; Fagard and Vaucheret, 2000). A systematic study of transgene expression in *Arabidopsis thaliana* (Forsbach *et al*., 2003; Lechtenberg *et al*., 2003; Schubert *et al*., 2004) revealed that neither the position of transgene insertion in the genome nor the different repeat configurations of T-DNAs were sufficient to trigger gene silencing in lines carrying transgenes under the control of the strong CaMV 35S promoter. In contrast, the transcript level of different *A. thaliana* transgenic lines that carried the *GUS*, *GFP* or *SPT* transgenes under control of the CaMV 35S promoter depended on the copy number of a particular transgene. Transgene expression was positively correlated with the number of transgene copies and stable over all generations analysed unless the number of copies under the control of the CaMV 35S promoter exceeded a gene-specific threshold. However, not the transgene copy number as such triggered transgene silencing, rather silencing was elicited if the transcript level of a

transgene surpassed a gene-specific threshold. Variation in transgene copy number provided a suitable explanation for the pronounced variability of transgene expression among independent transformants. Based on molecular and phenotypic hallmarks in the silenced lines the mechanism was categorised as post-transcriptional gene silencing (Schubert *et al*., 2004).

1.2 Sense-transgene induced post-transcriptional gene silencing in plants

Since the discovery of RNA silencing in transgenic plants it has become clear that it represents an important layer in gene regulation (Meyer, 2013). Small noncoding RNAs play a role in many biological processes such as development, response to stress and the protection of the genome against viruses and transposable elements, more recently its role in plant-microbe interactions has been elucidated (Baulcombe, 2004; Voinnet, 2005; Peláez and Sanchez, 2013; Pumplin and Voinnet, 2013).

In plants, small RNAs can be classified into two major types; microRNAs (miRNAs) and small interfering RNAs (siRNAs). The majority of miRNAs are excised from DNA-dependent RNA polymerase II (Pol II) transcripts with stem-loop structures. In contrast, siRNAs always occur in populations of 21-24 nucleotides (nt) long duplexes and are produced from dsRNA. Fold-back structures of inverted-repeat transcripts as well as dsRNA generated through overlapping convergent transcription serve as precursors for siRNAs, but RNA-dependent RNA polymerases (RDRs) can also generate dsRNA from single-stranded RNA (Ruiz-Ferrer and Voinnet, 2009; Parent *et al.,* 2012; Meyer, 2013). The siRNA duplexes can be derived from viruses or transgenes (Vaucheret *et al.*, 2001), but endogenous genes also give rise to the so-called natural-antisense-transcript-siRNAs (nat-siRNAs; Borsani *et al*., 2005; Katiyar Agarwal *et al*., 2006) and trans-acting-siRNAs (ta-siRNAs; Peragine *et al*., 2004; Vazquez *et al*., 2004). In a transcriptional silencing process known as RNA-directed DNA methylation (RdDM) transcripts produced by the plant-specific DNA-dependent RNA polymerase IV (Pol IV) can be copied into long dsRNAs and processed to siRNAs (Matzke and Mosher, 2014).

Different silencing pathways have been elucidated, nevertheless all of them have several features in common, such as the formation of dsRNA and its processing into small RNAs (Brodersen and Voinnet, 2006; Mallory and Vaucheret, 2010). Sense transgene-induced post-transcriptional gene silencing (S-PTGS) is a process in which the transcripts from a highly transcribed transgene locus trigger PTGS. The initial observations of this phenomenon

were made in *Petunia*. When genes involved in flower pigmentation were introduced not only silencing of the transgenes was observed but also of endogenous genes that were sequence-related to the introduced genes. The phenomenon of coordinated suppression of homologous genes was termed cosuppression (Napoli *et al.,* 1990; Van der Krol *et al.,* 1990). S-PTGS was also observed in other plant species such as *A. thaliana*, tomato, tobacco and rice and yielded important insights into this process (Smith *et al*., 1990; Tanzer *et al*., 1997; Han and Grieson, 2002; Schubert *et al*., 2004; Luo and Chen, 2007; Kawakatsu *et al*., 2012; Shin *et al.,* 2014; Parent *et al.,* 2015).

Many factors of importance for S-PTGS have been identified, these studies that entailed forward genetic screens but also reverse genetic approaches were predominantly carried out in *A. thaliana*. Important classes of mutants are the suppressor of gene silencing (*sgs*) and silencing-defective (*sde*) mutants (Vaucheret *et al*., 2001; Brodersen and Voinnet, 2006). Figure 1 depicts the S-PTGS pathway as proposed by Mallory and Vaucheret (2010).

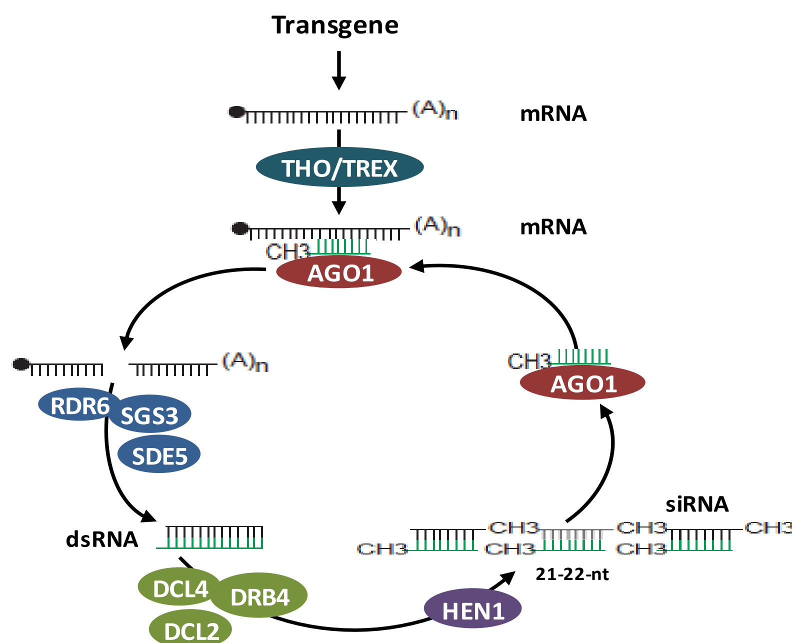


Figure 1. Model for Sense-PTGS pathway in *Arabidopsis thaliana* (modified after Mallory and Vaucheret, 2010).

Studies of transgenic lines showed that S-PTGS was triggered if transcription levels surpassed a gene-specific threshold (Schubert *et al*., 2004). The requirement of high transcript levels for the elicitation of silencing was corroborated by the characterisation of the *sgs8* mutant. In *sgs8* plants reduced transgene transcription was observed and transgenes silenced by

PTGS were reactivated. Importantly, SGS8 was required for high levels of transgene expression in a PTGS-independent manner. The gene affected in *sgs8* plants encodes the Histone3 Lysine4 di/trimethyl demthylase Jumonji-C domain-containg protein 14 (JMJ14) (Le Masson *et al*., 2012).

Transcripts from highly transcribed transgene loci are believed to include aberrant ones without poly(A) tail or 5'-cap structure. The 5'-3' exonuclease XRN4 degrades uncapped mRNAs, in *xrn4-1* plants uncapped mRNAs accumulated and gene silencing was triggered (Gazzani *et al.,* 2004). Mutations in the *XRN4* gene also affected the decay of miRNA target transcripts (Souret *et al*., 2004); *xrn4* plants were insensitive to ethylene and showed an enhanced heat stress tolerance (Olmedo *et al*., 2006; Nguyen *et al*., 2015). The study of small RNA populations in a loss-of-function mutant of the *XRN4* gene revealed that decapped transcripts of endogenous genes can become substrates for the biogenesis of small RNAs, in particular those of 21 nucleotides in length (Gregory *et al*., 2008). The *A. thaliana XRN* gene family consists of three genes, *XRN2*, *XRN3* and *XRN4*, all of which function as 5'-3' exoribonucleases, but only XRN4 exhibits activity in the cytoplasm whereas the other two proteins function in the nucleus (Kastenmayer and Green, 2000). As shown for XRN4 (Gazzani *et al*., 2004), XRN2 and XRN3 are endogenous suppressors of PTGS, as is their regulator FIERY1 (FRY1) (Gy *et al*., 2007).

Consistent with the finding that improperly terminated transcripts are more prone to S-PTGS (Luo and Chen, 2007), the study of enhanced silencing phenotype (*esp*) mutants revealed the impact of proteins that are involved in RNA processing and 3'-end formation on gene silencing (Herr *et al*., 2006). RNA quality control mechanisms are in place in eukaryotic cells in order to ensure that defective mRNAs are eliminated by degradation. If components of nonsense-mediated decay, deadenylation or exosome activity were impaired, enhanced S-PTGS was found, this implied that aberrant transgene RNAs are partitioned between RNA quality control and PTGS (Moreno *et al*., 2013; Yu *et al*., 2015). Characterisation of the *sgs14* mutant in which the gene coding for the nuclear ribonucleoprotein SmD1 was deleted showed that SmD1 facilitates PTGS, it was proposed that this protein protects the aberrant transgene RNAs from elimination by RNA quality control (Elvira-Matelot *et al*., 2016).

Several proteins are of importance for the conversion of aberrant RNA molecules into double stranded RNAs (dsRNAs) (Figure 1). These include SUPPRESSOR OF GENE SILENCING 2 (*SGS2*/*SDE1*/*RDR6* – Dalmay *et al.,* 2000; Morrain *et al*., 2000), SGS3 (SGS3 – Dalmay *et al*., 2000; Mourrain *et al.,* 2000), SDE5 (Hernandez-Pinzon *et al*., 2007; Jauvion *et al*., 2010) and possibly WERNER SYNDROME-LIKE EXONUCLEASE (WEX – Glazov *et al.,* 2003).

RNA-dependent-RNA polymerases use RNA templates for the synthesis of complementary RNAs. In the *A. thaliana* genome six *RNA-DEPENDENT-RNA POLYMERASE* (*RDR*) genes are found, RDR1, RDR2 and RDR6 share the C-terminal canonical catalytic DLDGD motif of eukaryotic RDRs while in the three *RDR* genes which form a cluster on chromosome 2, *RDR3*, *RDR4* and *RDR5*, the atypical motif DFDGD is found in the catalytic domain (Wassenegger and Krczal, 2006). The analysis of mutants in the *RDR6* gene (*sgs2*/*sde1* – Dalmay *et al.,* 2000; Morrain *et al*., 2000) showed its requirement for PTGS. In plants homozygous for both *xrn4-1* and *sde1-1* the level of decapped transcripts increased. It was therefore reasoned that decapped transcripts may serve as template for RDR6 so that silencing can be initiated and/or maintained (Gazzani *et al*., 2004). RDR2 is primarily involved in the RdDM pathway. However, it is likely that RDR2 and RDR6 compete for RNA templates, since siRNAs corresponding to transgenes that are subjected to S-PTGS are less abundant in *rdr2* plants than in plants carrying RDR2. Interestingly, S-PTGS is triggered earlier and/or is more efficient if RDR2 is impaired (Jauvion *et al*., 2012). Analysis of purified recombinant RDR2 and RDR6 proteins revealed that dsRNAs can be generated by using siRNAs as primers or by elongation of self-primed RNA templates (Devert *et al*., 2015).

SGS3 is also required for PTGS, it appears to function together with RDR6 in converting single-stranded RNA transcripts of sense transgenes and transcripts of DNA viruses into double-stranded RNA (Mourrain *et al.,* 2000; Muangsan *et al*., 2004). SGS3 is a plant-specific protein containing three protein domains: the rice gene X Homology (XH) domain, the rice gene X and SGS3 (XS) domain and the zinc finger-XS domain (Bateman, 2002). Of these, the XS domain acts as an RNA recognition motif (Zhang and Trudeau, 2008; Fukunaga and Doudna, 2009). It was demonstrated that SGS3 binds double stranded RNAs with a 5'-overhang (Fukunaga and Doudna, 2009). Loss of function mutations in the *SGS3* gene were found to have a phenotype similar to that of mutants in the *SGS2/SDE1/RDR6* gene, PTGS was abolished and methylation in the transgene coding sequences, an important hallmark of

S-PTGS, was severely reduced in *rdr6* and *sgs3* plants (Mourrain *et al.,* 2000). Consistent with the role of SGS3 and RDR6 in the same step of the PTGS pathway the proteins RDR6 and SGS3 were shown to interact and to colocalise in cytoplasmic granules (Kumakura *et al*., 2009). Both proteins have a central role for the production of nat-siRNAs (Borsani *et al*., 2005) and are also important for the regulation of the vegetative phase change and floral development since they are essential components for the biogenesis of ta-siRNAs (Peragine *et al*., 2004; Yoshikawa *et al*., 2005).

Like RDR6 and SGS3, SDE5 is neither involved in silencing triggered by inverted repeat transgenes nor for the biogenesis of miRNAs and DCL3-dependent 24 nt chromatin siRNAs, but it is required for S-PTGS and the production of trans-acting siRNAs. Whether it targets mRNAs or siRNAs remains to be elucidated but the presence of TAPC and PAM2 domains imply that SDE5 may play a role in RNA processing and/or trafficking (Hernandez-Pinzon *et al*., 2007; Jauvion *et al*., 2010).

The dsRNAs produced by the combined activities of RDR6, SDE5 and SGS3 are processed into 21-nt siRNAs by DICER-LIKE 4 (DCL4) in the S-PTGS pathway (Dunoyer *et al*., 2005). Then the siRNAs are methylated by HEN1 (Figure 1; Boutet *et al*., 2003; Li *et al*., 2005).

Dicer or dicer-like (DCL) proteins are known to play an important role in small RNA biogenesis pathways by processing long double-stranded RNAs into small RNAs with distinct products sizes (Park *et al*., 2002; Reinhart *et al*., 2002; Xie *et al*., 2004; Dunoyer *et al.,* 2005; Gasciolli *et al*., 2005; Xie *et al*., 2005; Yoshikawa *et al.,* 2005). In mammals, plants and insects, six domains are typically present in Dicer proteins; DExD-helicase, helicase-C, Duf283, PAZ, RNaseIII, and double stranded RNA-binding domains dsRBD whereas in lower eukaryotes, one or more of these domains appear to be absent (Margis *et al.,* 2006). In *A. thaliana* four DCLs have been identified (Schauer *et al*., 2002). All four Dicer like enzymes DCL1, DCL2, DCL3 and DCL4 have RNaseIII activity and can cleave double-stranded RNAs into short double-stranded RNA fragments of 21-nt in case of DCL1 and DCL4. DCL2 is important for the production 22-nt and 23-nt small RNAs and DCL3 generates 24-nt small RNAs. The majority of miRNAs are excised by DCL1, whereas DCL2, DCL3 and DCL4 are involved in the biogenesis of siRNAs (Xie *et al*., 2004; Xie *et al*., 2005; Parent *et al*., 2012).

DCL4 was shown to be responsible for the synthesis of trans-acting siRNAs (ta-siRNAs) (Xie *et al*., 2005; Yoshikawa *et al*., 2005), whereas both DCL4 and DCL2 produce siRNAs from viral substrates and transgenes (Blevins *et al*., 2006; Deleris *et al*., 2006; Fusaro *et al*., 2006; Henderson *et al*., 2006; Mallory and Vaucheret, 2009). In *dcl4* plants S-PTGS is initiated earlier and the amounts of transgene derived siRNAs are increased compared to plants containing DCL4, implying that the production of siRNAs by DCL2 alone is more efficient than in plants in which both DCL2 and DCL4 are present. In contrast, in *dcl2* plants silencing was delayed and the amount of transgene siRNAs was reduced, moreover the plants showed a mosaic pattern of silenced and unsilenced tissues (Parent *et al*., 2015). Dicers are associated with double-stranded RNA binding proteins (dsRPBs) that are encoded by five genes in *A. thaliana*. DCL4 and DCL1 were shown to interact with DRB4 and HYL1 (DRB1), respectively (Hiraguri *et al*., 2005).

HEN1 was shown to be critical for miRNA stability in *A. thaliana* (Park et *al*., 2002), but is also important for the accumulation of siRNAs in S-PTGS and virus-induced gene silencing (Boutet *et al*., 2003; Zhang *et al*., 2012). The *HEN1* gene encodes a methyltransferase (Li *et al*., 2005) that adds a methyl group on the ribose of the nucleotide at the 3'-end of miRNAs (Yu *et al*., 2005). This modification protects the small RNAs from degradation (Li *et al*., 2005). Interestingly, it was discovered that HEN1 has a stronger activity in *A. thaliana* accession Landsberg *erecta* (L*er*) than in Columbia-0 (Col-0), most likely due to the presence of a negative modulator of *HEN1* in the Col-0 genome, showing that the biogenesis of small RNAs is modulated by natural genetic variants (Yu *et al*., 2010). Elucidation of the structure of the *A. thaliana* HEN1 protein revealed that four domains directly interact with the small RNA substrate, whereas the structure of the fifth one, a PPIase-like domain, shows similarity to FK506-binding proteins. The four domains which are in direct contact with the small RNA consist of two dsRNA-specific binding domains, a domain with a La-type motif and one which harbours the methyltransferase activity that is dependent on Mg^{2+} (Huang *et al*., 2009).

The 21-nt small RNAs generated in the S-PTGS pathway can be bound by the AGO1 protein to form an AGO1-21nt-siRNA complex which guides the sequence-specific cleavage of homologous RNA (Baumberger and Baulcombe, 2005; Qi *et al*., 2005). There are ten AGO proteins in *A. thaliana*, which are divided into three major groups based on both their phylogenetic relationships; AGO1, AGO5 and AGO10 are belonging to group 1; group 2 is

made up of AGO2, AGO3 and AGO7 and AGO4; AGO6, AGO8 and AGO9 form the third group (Vaucheret, 2008; Mallory and Vaucheret, 2010). AGO1, AGO2, AGO7 and AGO10 are effector proteins for post-transcriptional RNA silencing processes, these proteins associate with 21 to 22-nt small RNAs (Fagard *et al.,* 2000; Morel *et al.,* 2002; Mallory *et al*., 2009; Carbonell *et al.,* 2012). In contrast, AGO4, AGO6 and AGO9 mostly associate with 24-nt small RNAs and are involved in transcriptional RNA silencing (Zilberman *et al.,* 2003; Zheng *et al.,* 2007; Havecker *et al.,* 2010). All AGO proteins contain four main domains; a variable N-terminal domain as well as PAZ, MID and PIWI domains. Crystal structure and biochemical analyses showed that the MID and PAZ domains bind to the 5'- and 3'-end of a small RNA, respectively. The PIWI domain shows similarity to the ribonuclease-H family of enzymes and exhibits endonuclease activity, the active site usually carries an Asp–Asp–His (DDH) motif (Hutvagner and Simard, 2008; Meister, 2013). The identity of the 5'-terminal nucleotide has an important role for the recruitment of small RNAs into distinct AGO complexes. For example, 21-nt small RNAs with an U at the 5'-end are sorted preferentially into AGO1 complexes (Mi *et al*., 2008). However, the duplex structure of miRNAs is also of importance for selective miRNA recruitment by AGOs (Zhang *et al*., 2014).

AGO1 mediates miRNA- as well as siRNA-directed PTGS (Baumberger and Baulcombe, 2005). Many *ago1* mutants show severe developmental abnormalities and sterility, but fertile hypomorphic mutants were also described. Even the hypomorphic fertile mutants were impaired with respect to S-PTGS, revealing that this process is more sensitive to AGO1 defects than development (Morel *et al.,* 2002). During early embryo development AGO1 and AGO10 share a set of redundant functions. In *ago10* mutants AGO1 protein level is increased whereas its mRNA level was not affected, indicating that AGO10 acts as a negative regulator of the AGO1 protein level. The loss of AGO10 function in weak *ago1* mutants restores defects in leaf development as well as siRNA and miRNA pathways (Mallory *et al*., 2009).

AGO7 primarily functions in the regulation of developmental timing since it is important for the biogenesis of TAS3-derived trans-acting siRNAs (Adenot *et al*., 2006; Fahlgren *et al*., 2006; Hunter *et al*., 2006), for S-PTGS only a small effect was found in the AGO7-defective *zip-1* mutant (Hunter *et al*., 2003). Not all AGO proteins cleave their target RNAs in the region which shows complementarity to the miRNA or siRNA sequences, regulation of mRNA targets via translational repression has also been described (Brodersen *et al*., 2008).

However, in the case of AGO1 and AGO7 the slicer activities are required for normal plant development and ta-siRNA RNA biogenesis, since complementation of the *zip-1* and the *ago1-25* mutants depended on the catalytic residues (Carbonell *et al*., 2012).

Once the primary siRNAs cleave the transgene mRNA an amplification loop can be established, since the small RNAs and/or mRNA cleavage products may also serve as primers to promote further production of dsRNAs and secondary siRNAs, resulting in an amplified reaction. This phenomenon is referred to as transitivity (Brodersen and Voinnet, 2006).

SDE3 encodes a RNA helicase-like protein. Mutants in this gene impair PTGS, however, SDE3 is only required if PTGS is triggered by weak inducers, it is dispensable for strong ones. In contrast to RDR6, SGS3 and SDE5 it is not required for the ta-siRNA pathway (Dalmay *et al.,* 2000; Dalmay *et al.,* 2001; Jauvion *et al.*, 2010). The SDE3 protein is present with AGO1 and/or AGO2 in higher order complexes and genetically acts downstream of RDR6. It was proposed that the helicase function helps to unwind dsRNAs so that RDR6 could act on these single-stranded molecules repeatedly. A complex of SDE3 and siRNA-loaded AGO1 would furthermore be capable of the production of aberrant RNAs via endonucleolytic cleavage of the unwound dsRNAs. In this manner silencing amplification would be achieved (Garcia *et al.,* 2012).

The role of the *A. thaliana* Werner Syndrome-like exonuclease (WEX) in PTGS was elucidated by the study of Glazov *et al*. (2003). A T-DNA insertion mutant which showed strongly reduced WEX gene expression when compared to wild-type plants also revealed strong inhibition of *GFP* transgene silencing. Nonetheless, to date, where and how WEX acts in the PTGS pathway is not known. *WEX* is related to the *Caenorhabditis elegans MUT-7* gene, which has been demonstrated to be necessary for RNA interference (RNAi), PTGS and transposon silencing (Ketting *et al*., 1999). *WEX* was shown to encode an RNase D domain with similarity to that in MUT-7 and in human Werner Syndrome protein (WRN) (Plchova *et al*., 2003), but in contrast to WRN, WEX and MUT-7 lack the RecQ helicase domain.

HYPER RECOMBINATION1 (*HPR1/THO1*, *SGS9*) is homologous to one member of the THO/TREX complex which is involved in RNA trafficking. In *hpr* plants S-PTGS is suppressed but not abolished as in *sgs3*, the ta-siRNA pathway is affected in a similar fashion (Jauvion *et al*., 2010). *TEX1/THO3* and *THO6* encode other components of the THO/TREX complex, mutants in these genes also impair the ta-siRNA pathway (Jauvion *et al*., 2010; Yelina *et al*.,

2010). In *tho2* mutants not only the levels of siRNAs but also miRNAs were reduced. It has not been clarified where in the S-PTGS pathway the THO/TREX complex acts, however it was shown that THO2 interacts with miRNA precursors, this interaction may be of importance to recruit the precursors to the processing complex. Since the levels of other small RNA molecules were reduced in *tho2* mutants it was suggested that the complex has a rather broad affinity (Francisco-Mangilet *et al*., 2015).

A screen for *C. elegans* mutants with an enhanced sensitivity to dsRNAs in the nervous system revealed that a mutant in the *ERI-1* gene, *Enhancer of RNAi*, accumulated more siRNAs than wild-type animals (Kennedy *et al*., 2004). ERI-1 was found to degrade siRNAs with 2-nucleotide long 3'-overhangs *in vitro*, the nuclease activity is consistent with the fact that the protein belongs to the DEDDh family of 3'->5' exonucleases (Zuo and Deutscher, 2001). The *ERI-1* gene encodes an evolutionary conserved protein, and its role as a negative regulator of RNAi was not only documented in *C. elegans* but also in *Schizosaccharomyces pombe*. In the latter organism loss of *ERI-1* resulted in increased amounts of siRNAs that corresponded to centromeric repeats (Gabel and Ruvkun, 2008). In *A. thaliana*, the coding region of *At3g15140* was found to be most similar to *ERI-1* (Ramachandran and Chen, 2008). Overexpression of *At3g15140* caused reduction of 21-nucleotide long siRNAs, supporting the notion that *ERI* acts as a nuclease with specificity to siRNAs (Meyer *et al*., 2015).

Apart from the three essential DNA-dependent RNA polymerases, Pol I, Pol II and Pol III plants also contain two dispensable polymerases, Pol IV and Pol V. Similar to other DNA-dependent RNA polymerases, Pol IV and Pol V are also large protein complexes containing multiple subunits. Pol IV and V contain two different largest subunits, NRPD1 (NRPD1a/SDE4) and NRPE1 (NRPD1b), respectively, but share the same second-largest subunit (NRPD2) (Herr *et al*., 2005; Kanno *et al*., 2005; Onodera *et al*., 2005; Pontier *et al*., 2005). A comparison of amino acid regions in the conserved regions of NRPD1, NRPE1 and NRPD2 to the corresponding subunits of Pol II in *A. thaliana* and *Oryza sativa* revealed 10-20 times higher substitution rates in the Pol IV and Pol V subunits (Luo and Hall, 2007). Pol IV and V are involved in RdDM. This phenomenon was first discovered in tobacco plants in which potato spindle tuber viroid cDNAs had been introduced *via Agrobacterium*-mediated transformation (Wassenegger *et al.*, 1994). RdDM is enriched in heterochromatin, in euchromatic regions it is typically associated with transposable elements and other

dispersed repeats. It is assumed that Pol IV produces single-stranded RNAs which are used as templates by RDR2 for the generation of dsRNAs. Processing by DCL3 leads to the formation of 24-nt siRNAs which are loaded onto AGO4. Pol V transcripts are believed to pair with the AGO4-bound siRNAs, *de novo* DNA methylation is then carried out by the recruitment of DOMAINS REARRANGED METHYLTRANSFERASE (DRM2) (Matzke *et al*., 2009; Zhang and Zhu, 2011; Matzke and Mosher, 2014).

Key components of the RdDM pathway such as NRPD1, NRPE1 and RDR2 also play a role in the S-PTGS pathway (Herr *et al*., 2005). It was shown that NRPD1 and NRPE1 are neither needed for the iniation of S-PTGS nor for the production of secondary siRNAs, rather they are required for the maintenance of silencing (Eamens *et al*., 2008). Interplay between different silencing pathways also became apparent by studying the double mutants *hen1-2 nrpd1*, *hen1-2 nprd2* and *hen1-2* and *rdr2*. In these plants a competition between endogenous siRNAs and miRNAs for methylation was revealed, such a competition may also occur in wild-type situations (Yu *et al*., 2010).

1.3 Silencing spread

One of the remarkable hallmarks of RNA silencing is that it can spread. In plants, movement of the silencing signal can encompass both short-distance cell-to-cell movement most likely through plasmodesmata (Himber *et al*., 2003; Dunoyer *et al*., 2005; Kalantidis *et al*., 2006; Dunoyer *et al*., 2010) and long-distance transport via the vascular system (systemic silencing) (Palauqui *et al*., 1997; Voinnet and Baulcombe, 1997; Kalantidis *et al*., 2008).

Short-distance movement from cells in which silencing was initiated extends to 10-15 cells (Himber *et al*., 2003; Kalantidis et al., 2006), however, signal movement from root to shoot in *A. thaliana* can also occur in a similar manner (Liang *et al*., 2012). Cell-to-cell movement requires the presence of 21-nt siRNAs that are generated by DCL4 (Himber *et al*., 2003; Dunoyer *et al*., 2005). Studies employing *A. thaliana* mutants implicated also other proteins, such as AGO1, DCL1 and HEN1 (Dunoyer *et al*., 2007), whereas RDR6 was dispensable (Himber *et al*., 2003). Notably, certain components of the RdDM pathway, for example NRPD1 and RDR2, also affect this process (Dunoyer *et al*., 2007; Smith *et al*., 2007).

Silencing initiated in localised regions of a plant can be transmitted to other plant organs as shown by grafting and *Agrobacterium* infiltration experiments. The progression of silencing was dependent on a sequence-specific signaling mechanism and involved movement through plasmodesmata and the phloem (Palauqui *et al*., 1997; Voinnet *et al*., 1998). Based on the long-range mechanism silencing can spread across tissues, movement typically occurs from photosynthetic source to sink tissues. Analysis of phloem sap revealed miRNAs as well as siRNAs of different sizes, whereas RNAs were not found in xylem sap (Yoo *et al*., 2004; Buhtz *et al*., 2008). Grafting experiments provided evidence for movement of 22-nt and 24-nt siRNAs, but the approach used did not allow to assess 21-nt siRNAs (Molnar *et al*., 2010).

1.4 Impact of environmental conditions on gene silencing

In several studies the effect on environmental conditions, such as light and temperature, on gene silencing were reported (Szittya *et al*., 2003; Chellappan *et al*., 2005; Kotakis *et al*., 2010; Patil and Fauquet, 2015). For example, virus and transgene triggered silencing was studied in *Nicotiana benthamiana* plants that were grown at temperatures between 15 and 24°C. Silencing and siRNA accumulation were drastically reduced at the lower temperature (Szittya *et al*., 2003). Similar results were found for transgene-induced gene silencing in *A. thaliana* and potato, whereas miRNA accumulation was not affected. Cassava geminivirus-induced RNA silencing was more pronounced in *N. benthamiana* and cassava plants when the plants were cultivated at 30°C rather than at 25°C. The accumulation of siRNAs was higher at the elevated temperature, too (Chellappan *et al*., 2005). Kotakis *et al*. (2010) reported that *N. benthamiana* plants grown under higher light intensity showed more short-range and systemic silencing than under lower light conditions. They also showed that *DCL4* was upregulated by light, in case silencing had been initiated *DCL1*, *DCL2*, *DCL3* and *RDR6* were also expressed more highly under these conditions. Agroinfiltration studies in *N. benthamiana* revealed localised gene silencing at temperatures above 30°C as well as at light intensities higher than 450 $\mu E/m^2/s$, systemic spread of silencing was not observed under these conditions (Patil and Fauquet, 2015). In contrast, at light intensities of approximately 300 $\mu E/m^2/s$ and a cultivation temparature of 25°C strong movement of the systemic silencing signal was found. The observed differences were attributed to changes in the sink-source status of the leaves.

The impact of temperature and light on transgene silencing was also demonstrated for *A. thaliana* transgenic lines expressing *GFP* transgenes under the control of the CaMV 35S promoter by using fluorescence stereomicroscopy to study the initiation and spread of silencing in populations of transgenic plants (Arlt and Schmidt, 2006; Arlt, 2007).

1.5 Analysis of natural variation in *Arabidopsis thaliana*

A. thaliana, a small, self-pollinating cruciferous plant was discovered by Johannes Thal (hence, thaliana) in the Harz mountains of Northern Germany in the sixteenth century. It is a member of the mustard family (Brassicaceae), in contrast to the important crop plants of this family such as cabbage, broccoli and oilseed rape, it has no economic value (Meyerowitz, 1987). Friedrich Laibach was the first to recognize the versatility of *A. thaliana* for plant genetics (Laibach, 1943). Due to the small size of the plant, the ease of cultivation in limited space, the short generation time and the prolific seed production the plant lends itself well for genetical studies (Meinke *et al.,* 1998; Somerville and Koornneef, 2002). Many mutants affected in various biological processes were generated, characterised and mapped (Koornneef and Meinke, 2010). The discovery of the small genome size (Meyerowitz and Pruitt, 1985) together with the first report of *Agrobacterium tumefaciens* mediated transformation (Lloyd *et al.,* 1986) were important milestones in *A. thaliana* research. With about 125 to 150 Mbp distributed over five chromosomes, *A. thaliana* possesses a particular small genome among higher plants. In 2000, the annotated genome sequence of accession Col-0 was published by the Arabidopsis Genome Initiative (2000), it represented the first completed genome sequence of a higher plant. The generation of sequence-indexed collections of mutants in which genes were disrupted by transposon or T-DNA insertion in conjunction with the availability of the genome sequence enable systematic reverse genetics approaches (Alonso and Ecker, 2006). Alternatively, small RNA-based gene silencing can be used to systematically down-regulate single genes or multiple sequence-related genes (Ossowski *et al*., 2008). Due to the features described above *A. thaliana* has become an important model organism and molecular-genetic approaches provided important insights into various plant processes (Meinke *et al.,* 1998; Somerville and Koornneef, 2002; Koornneef and Meinke, 2010). Since many large-scale data sets have been generated it is of central importance that databases dedicated to data storage, distribution and analysis have been established (Graham and May, 2011).

A. thaliana is native to Europe and central Asia and is found in many diverse environments, especially in the northern hemisphere. However, it is now naturalised to many other places worldwide such as North America, Africa, Australia and Japan. The species has been found from sea level up to 4520 m, mostly on sandy or loamy soils in open or disturbed habitats (Al-Shehbaz and O'Kane, 2002; Hoffmann, 2002; Koornneef *et al.,* 2004). *A. thaliana* natural accessions collected from wild populations by Friedrich Laibach and mutants from Röbbelen and Kranz were first maintained by the Arabidopsis Information Service (AIS) seed stock center (Somerville and Koornneef, 2002). Nowadays, several seed stock centers propagate and distribute seeds of mutant lines and/or accessions (Scholl *et al.,* 2000; Knee *et al*., 2011).

A. thaliana shows impressive phenotypic and genetic diversity and the study of natural diversity is becoming more and more important. The analysis of gene variants that are found in nature cannot only be exploited to reveal insights into important processes in plants, but also offers the opportunity to unravel which allelic variants are important for adaptation to local environments (Koornneef *et al.,* 2004; Weigel and Nordborg, 2005; Benfey and Mitchell-Olds, 2008).

In many instances phenotypic differences between accessions are caused by allelic variation at more than one locus, furthermore an individual locus may contribute only little to the overall variation. Thus, statistical methods are needed to identify the regions in the genome which contribute to the variation of the trait of interest and to estimate the size of their effects. To unravel which genetic loci underlie the phenotypic diversity between accessions two different approaches can be used, quantitative trait locus mapping (QTL) or genome-wide association studies (GWAS). QTL studies require segregating populations that are derived from a cross of two accessions that ideally show variation for the trait of interest. Recombinant inbred line (RIL) populations lend themselves particularly well for QTL studies, since the essentially homozygous genotypes can be propagated indefinitely and repeatedly analysed with respect to phenotypic traits in many replicates and environments whereas the genotype information has to be established only once, consequently many different RIL populations were generated (Koornneef *et al*., 2011; Weigel, 2012). Initially, restriction fragment length polymorphism (RFLP) markers were used to establish molecular marker maps for RIL populations (Lister and Dean, 1993), but these have been substituted by polymerase chain reaction (PCR)-based marker systems such as amplified fragment length polymorphisms (AFLP) (Alonso-Blanco *et al*., 1998), microsatellite (Loudet *et al*., 2002),

insertion/deletion (Indel) (Salathia *et al.,* 2007; Hou *et al.,* 2010) and single nucleotide polymorphism (SNP) markers (El-Lithy *et al*., 2006; Törjék *et al*., 2006).

Variation has been found for many morphological and physiological characters and QTL for many traits were mapped and characterised in *A. thaliana* RIL populations. Flowering time was analysed particularly intensively (Alonso-Blanco *et al*., 2009; Koornneef *et al*., 2011), but traits such as rosette size, plant height, seed size, seed production (Alonso-Blanco *et al*., 1999; Simon *et al*., 2008), leaf shape and size (Juenger *et al*., 2005), root growth and architecture (Loudet *et al*., 2005) seed dormancy, seed longevity (Clerkx *et al*., 2004), plant biomass and early stage heterosis for biomass were also studied (Lisec *et al*., 2008; Meyer *et al*., 2010). Traits playing a role in the response to abiotic and biotic factors have also been analysed extensively. Notably, numerous genes and functional polymorphisms underlying different traits have been identified (Alonso-Blanco *et al*., 2009; Koornneef *et al*., 2011). The integrated analysis of a RIL population derived from the accessions L*er* and Cvi with respect to 139 phenotypic traits as well as transcript, protein and metabolite abundance revealed six QTL hot spots. Thus, despite the fact that the two accessions differ by more than 500,000 SNPs and that expression QTL were found for approximately 20% of the analysed seedling transcripts, the vast majority of molecular variants do not cause phenotypic variation across a range of environmental conditions (Fu *et al.,* 2009).

In genome-wide association studies the whole breadth of natural variation is assessed and the accessions are analysed for genotype-phenotype associations. This approach requires large collections of accessions and most importantly extensive genotype information in order to identify associations between the trait of interest and sequence variants (Weigel, 2012). Many *A. thaliana* accessions have been collected and panels for GWAS have been compiled (Atwell *et al*., 2010; Baxter *et al*., 2010; Platt *et al*., 2010; Horton *et al*., 2012). In a particular extensive study 107 traits were analysed in 96 to 192 accessions, genotyping of the accessions was performed using a custom Affymetrix SNP chip that contained 250,000 SNPs (Atwell *et al*., 2010).

Genotyping at such a large scale is only possible because considerable efforts have been made to study the genetic diversity of *A. thaliana* accessions. One of the first large-scale surveys of polymorphisms in many different accessions involved the sequence analysis of 876 short fragments, accounting together for almost 0.5 Mbp of the genome. In total, 96

plant samples that represented accessions as well as pairs of individuals of 25 selected populations were analysed (Nordborg *et al*., 2005). Analysis of 27 disease resistance genes using the same methodology and panel of plant samples revealed that this particular class of genes was characterised by a generally higher nucleotide diversity and more recombination when compared to the findings obtained for the 876 fragments (Bakker *et al*., 2006). High-density array sequencing of 20 diverse strains revealed even more polymorphism information (Clark *et al*., 2007). The results of this study provided the basis for the development of the custom Affymetrix SNP chip with 250,000 SNPs (Kim *et al*., 2007). However, array sequencing also revealed that certain gene families show exceptional polymorphism levels and that a considerable proportion of the different accession genomes were either highly dissimilar or even deleted relative to the reference accession Col-0 (Clark *et al*., 2007). The latter finding also implied that in accession genomes many regions may be present that are absent in the reference genome, therefore it was an important goal to access also regions in the accession genomes that are not present in the reference genome. With this motivation in mind the 1001 Genomes Project was initiated in 2007. It aims at sequencing the genome of *A. thaliana* accessions from various geographic regions as well as several individuals of selected populations. Different technologies and depths of sequence coverage are used to produce genome sequences of accessions at different levels of accuracy and completeness (Nordborg and Weigel, 2008; Weigel and Mott, 2009; 1001 Genomes Consortium, 2016). Importantly, for selected accessions additional reference sequences were established (Gan *et al*., 2011; Schneeberger *et al*., 2011). Whole-genome sequencing of 80 strains revealed almost 5,000,000 SNPs and more than 800,000 Indels smaller than 20 bp in 80 strains when compared to the reference sequence. Furthermore, many examples of larger deletions and copy number variation of coding sequences were found. Interestingly, in more than 6,000 genes SNPs were detected that altered the coding sequence; start codons were altered, premature stop codons were introduced, open reading frames were extended, splice donor or acceptor sites were affected. In addition, more than 27,000 indels were identified potentially causing frame shifts. Considering only premature stop codons, 4,263 genes were affected in at least one of the accessions. Genes of the NB-LRR, F-box, RLK and RING families were particularly prone to such changes. It is important to note that approximately 10% of the *MIRNA* loci were missing in one or more of the strains (Cao *et al*., 2011). Gan *et al.* (2011) generated genome sequences and transcriptomes for 18

different accessions and based on these data they reannotated the genomes of the different accessions. When the genome sequences were compared to the reference sequence of Col-0 the coding region of many genes appeared to be disrupted, reannotation of the different genomes revealed alternative gene models with restored coding potential (Gan *et al*., 2011). The study of Long *et al.* (2013) reported high levels of genetic variation in lines from a single geographic region. Notably, large differences in genome size were found among the Swedish accessions which could be attributed to copy number variation at the 45S ribosomal DNA loci.

1.6 Aim of study

This study aimed at a first insight into the role of natural variation in the process of sense transgene-induced post-transcriptional gene silencing in *Arabidopsis thaliana*. To address this it was intended to survey sequence variation in *A. thaliana* genes involved in S-PTGS and/or other RNA silencing pathways. A particularly important goal of this work was the identification of genome regions which modulate S-PTGS in *Arabidopsis thaliana*.

To study allelic diversity in genes involved in the S-PTGS pathway, amplicon sequencing would be performed for selected *A. thaliana* accessions. Alignments of the accession sequences to the *A. thaliana* Col-0 reference gene and open reading frame sequences would identify single nucleotide polymorphisms and Indels and also establish which polymorphisms would affect the amino acid sequences. Of particular interest in this context would be alleles with high levels of sequence divergence and/or differential expression in comparison to reference accession Col-0.

In order to evaluate whether particular genome regions of *A. thaliana* accessions modulate S-PTGS, it was planned to introgress them into Col-0 transgenic lines carrying *GFP* transgenes, since multiple *GFP* transgene copies under the control of the strong CaMV 35S promoter in the Col-0 genome are readily subjected to S-PTGS and represent a sensitive monitoring system for transgene silencing. Silencing of the *GFP* transgenes in the different introgression lines would be analysed at several stages of development and compared to the performance of the *GFP* transgenes in the Col-0 genetic background in order to reveal whether *GFP* silencing would be altered in certain introgression lines. Using suitable molecular markers the introgression lines would be characterised in detail with respect to number, position and length of introgressed regions.

2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Laboratory equipment

Manufacturer/supplier, City, Country	**Material(s)**
Biofrontier Technology Pte Ltd., Prestige Centre, Singapore	Heating block HLC
Biometra GmbH, Goettingen, Germany**; Bio-rad**, München, Germany	Thermocycler
Carl Roth GmbH & Co. KG, Karlsruhe, Germany	Forceps
Cell Biosciences Inc., Santa Clara, CA	AlphaImager HP
Duran Group GmbH, Wertheim/Main, Germany; **Schott AG**, Mainz, Germany	Glass ware
Gilson Inc, Middleton, USA	Pipettes (P2, P10, P20, P200, P1000)
Heidolph Instruments GmbH & Co. KG, Schwabach, Germany	Shaker
Heraeus Instruments GmbH, Wiesloch, Germany; **Eppendorf AG**, Hamburg, Germany	Centrifuges
Julabo GmbH, Seelbach, Germany	Water bath
Kern & Sohn GmbH, Balingen-Frommern, Germany	Balances
Leica Microsystems GmbH,Wetzlar, Germany	Leica Schott KL1500 LCD, Leica MZ16 F, Leica DFC 345FX
Peqlab Biotechnology GmbH, Erlangen, Germany	Centrifuge Perfect Spin P, Electrophoresis power supply, Electrophoresis chambers, NanoDrop ND1000 Spectrophotometer, Thermocycler
Retsch GmbH, Haan, Germany	Mixer Mill MM 400
Sanyo Electric Co., Ltd., Japan	Freezer -80°C
Bosch, UK	Freezer -20°C
Liebherr, Germany	Freezer -20°C, Fridge 4°C
Sartorius, Madrid, Spain	Balances
Scientific Industries Inc., Bohemia, USA	Vortex-Genie 2
Sharp, USA	Microwave
Starlab GmbH, Hamburg, Germany	Electrophoresis chambers
Thermo Fisher Scientific Inc., Waltham, USA	ABI PRISM® 7900 HT real-time PCR System

2.1.2 Chemicals, enzymes, kits and materials for plant cultivation

Manufacturer/supplier, City, Country	**Material(s)**
Biozym Scientific GmbH, Hessisch Oldendorf, Germany	LE agarose, NuSieve 3:1 agarose, Metaphor agarose
Carl Roth GmbH & Co. KG, Karlsruhe, Germany	ß-mercaptoethanol, boric acid, bromophenol blue, chloroform, diethylpyrocarbonate (DEPC), EDTA, ethanol, ethidium bromide, formaldehyde (37%), formamide, glycerol, HCl, isopropanol, NaCl, NaOH, phenol/chloroform/isoamyl alcohol, RNAse, SDS, Tris.
Eurofins Genomics, Ebersberg, Germany	custom oligonucleotides
Klasmann-Deilmann GmbH, Geeste, Germany	Substrate 1
Merck, Darmstadt, Germany	sodium acetate, sodium disulfite
Öre Protect Biologischer Pflanzenschutz GmbH, Schwentinental, Germany	Novo Nem® F
Peqlab Biotechnology GmbH, Erlangen, Germany	PeqGOLD RNAPure™
Serva Feinbiochemica GmbH & Co., Heidelberg, Germany	CTAB
Sigma Aldrich Chemie GmbH, Steinheim, Germany	MOPS, bromophenol blue
Thermo Fisher Scientific Inc., MA, USA	dNTPs, GeneRuler™ 100 bp DNA Ladder Plus, GeneRuler™ 1 kb DNA Ladder, Dream *Taq* DNA-Polymerase and 10x Dream *Taq* buffer, Exonuclease I (Exo I), FastAP™ Thermosensitive Alkaline Phosphatase, Maxima H Minus First Strand cDNA Synthesis kit, Maxima SYBR Green/Fluorescein qPCR Master Mix (2x), Quant-iT PicoGreen dsDNA Assay Kit, "TURBO DNA-free™" DNAse Kit

2.1.3 Buffers and solutions

10x TBE buffer	0.9 M Tris base 0.9 M boric acid 20 mM EDTA pH 8.0
10x DNA loading buffer	0.25% bromophenol blue 1 mM EDTA, pH 8.0 50% glycerol

DNA extraction buffer A	200 mM Tris-HCl, pH 7.5
	250 mM NaCl
	25 mM EDTA
	0.5% SDS
DNA extraction buffer B	100 mM Tris-HCl, pH 8.0
	1.4 M NaCl
	20 mM EDTA, pH 8.0
	2% CTAB
	13.15 mM sodium disulfite
	0.1% ß-mercapto ethanol (only added prior to use)
100x RNAse solution	10 mM Tris, pH 7.5
	15 mM sodium chloride
	10 mg/ml RNAse A
	(solution boiled for 15 min and diluted 100-fold before use)
Staining solution	125 µl of a 1% ethidium bromide solution
	500 ml 1x TBE buffer
10x RNA running buffer	0.2 M MOPS
	80 mM sodium acetate (NaAc)
	10 mM EDTA
	(adjusted to pH 7.0 with 10 M NaOH)
RNA agarose gel solution (100 ml)	1.2 g agarose
	72 ml DEPC-treated dH_2O
	10 ml 10x RNA running buffer
	18 ml formaldehyde
	(for a 11x 14 cm gel)
RNA sample buffer (1 ml)	100 µl 10x RNA running buffer
	500 µl formamide
	178 µl formaldehyde
	222 µl DEPC-treated dH_2O
RNA loading buffer	0.25% bromophenol blue
	1 mM EDTA, pH 8.0
	50% glycerol

2.1.4 *Arabidopsis thaliana* accessions and transgenic lines

Twenty-six *A. thaliana* accessions were selected from a set of 360 accessions (Baxter *et al.*, 2010; Platt *et al.*, 2010). Genetic distances between the 360 accessions are shown in Supplementary figure 1. Seeds for accessions Col-0 and C24 had been ordered from The

European Arabidopsis Stock Center (NASC). The remaning accessions were acquired from Prof. Dr. Marcel Quint (Martin Luther University Halle-Wittenberg). Information about the accessions used is given in Table 1.

Table 1. List of *Arabidopsis thaliana* accessions used in this study

Abbreviated name	Location	Stock number	Country of origin
Amel-1	Ameland	CS28014	Netherlands
Ang-0	Angleur	CS28018	Belgium
Baa-1	Baarlo	CS28054	Netherlands
Bor-4	Borky	CS76100	Czech Republic
C24		CS76106	
Col-0		CS76113	
Cvi-0	Cape Verde Islands	CS76116	Cape Verde
Gie-0	Gießen	CS28280	Germany
Kas-1	Kashmir	CS76150	India
Kin-0	Kindalville	CS76153	USA
Kl-5	Köln	CS28394	Germany
Kno-18	Knox	CS76154	USA
L*er*-1	Landsberg	CS76164	Poland
Ll-0	Llagostera	CS76172	Spain
Lp2-2	Lipovec	CS76176	Czech Republic
Lz-0	Lezoux	CS76179	France
Mt-0	Martuba	CS76192	Lybia
Pu2-23	Prudka	CS76215	Czech Republic
Ra-0	Randan	CS76216	France
RRS-7	North Liberty	CS28713	USA
Sapporo-0	Sapporo	CS28724	Japan
Shahdara	Pamiro-Alay	CS76227	Tadjikistan
Sq-8	Ascot	CS76230	UK
Tscha-1	Tsagguns	CS28779	Austria
Tsu-0	Tsushima	CS28780	Japan
Ws-0	Wassilewskija	CS28824	Belarus

A. thaliana Col-0 T-DNA insertion lines harbouring six or eight copies of the *GFP* (green fluorescent protein) reporter gene under the control of the CaMV 35S promoter conferring high constitutive expression were established previously. In all lines locus *R127* was present which carries two T-DNA copies in an inverted repeat orientiation (Lechtenberg *et al*., 2003). In addition, the different lines carried one or two single-copy T-DNA loci, *F8*, *F18* and *F128* (Forsbach *et al*., 2003; Schubert *et al*., 2004). Five transgenic lines were used in total. Two lines, 8x*GFP-F8/F18/R127* and 8x*GFP-F8/F128/R127* carried eight copies of the *GFP* transgene and three lines six, 6x*GFP-F8/R127,* 6x*GFP-F18/R127* and 6x*GFP-F128/R127* (Arlt, 2007; Thanh Loan Le, unpublished results).

2.1.5 Softwares

BioEdit	Ibis Biosciences, Carlsbad, USA http://www.mbio.ncsu.edu/bioedit/bioedit.html
AlphaImager HP camera	Cell Biosciences Inc., Santa Clara, USA
BLAST	NCBI, Bethesda, USA http://www.ncbi.nlm.nih.gov
CodonCode Aligner	CodonCode Corporation, Dedham, USA
Electronic PCR	http://www.ncbi.nlm.nih.gov/projects/e-pcr/reverse.cgi
GCG	Wisconsin Package (version 10.0-UNIX; Genetics Computer Group, Madison, USA)
Leica Application Suite Version 3.7.0	Leica Microsystems GmbH, Wetzlar, Germany
Microsatellite Repeats Finder	http://insilico.ehu.es/mini_tools/microsatellites/
Primer3	http://bioinfo.ut.ee/primer3
R	https://www.r-project.org/
TAIR 10	https://www.arabidopsis.org
SDS 2.2.2	Thermo Fisher Scientific Inc., Waltham, USA

2.2 METHODS

2.2.1 Plant growth conditions

A. thaliana seeds were sown in pots with substrate 1 (Klasmann-Deilmann GmbH) and placed at 4°C for 3 days for stratification, then the cultivation of the plants was carried out in long-day conditions (16 h light/8 h dark). During the first ten days pots were covered with suitable plastic lids. Ten days after sowing single seedlings were transferred into individual pots containing substrate 1. The pots were placed in small or large trays accommodating 35 or 54 pots each, respectively. The trays were covered by plastic lids for one more week. Plants were either cultivated in a growth room or in a growth cabinet (Table 2). Every two weeks the plants were treated with nematodes (Novo Nem® F, ÖRE Bio-Protect Biologischer Pflanzenschutz GmbH). Cultivation took place in the growth cabinet if plants were evaluated with respect to *GFP* transgene silencing or gene expression, all other experiments were performed in the growth room. Leaf material from individual six to eight weeks old plants was harvested for DNA isolation, selected plants were kept for seed production.

Seeds of late flowering accessions, Amel-1, Cit-0, Kno-18 and Ws-0, were sown out on soil and after the stratification treatment vernalised for 6 weeks at 4°C under short-day conditions (10 h light/14 h dark). During the vernalisation period trays were covered with plastic lids. After vernalisation the plants were transferred to the growth room and cultivated as described above.

Table 2. Growth conditions of *Arabidopsis thaliana* plants. n.d: not determined

	Day/light period				**Night period**			
Condition	**Time (h)**	**Temparature (°C)**	**Humidity (%)**	**Light intensity (µE $m^{-2} s^{-1}$) (µmol/m^2)**	**Time (h)**	**Temparature (°C)**	**Humidity (%)**	**Light intensity (µE $m^{-2} s^{-1}$) (µmol/m^2)**
Growth room	16	21	n.d	70-120	8	21	n.d	0
Growth cabinet	16	20	70	120	8	16	60	0
4°C	10	4	n.d	10-22	14	4	n.d	0

2.2.2 Crossing of *Arabidopsis thaliana* accessions to *GFP* transgenic lines in Col-0 background

The crossing of *A. thaliana* accessions was carried out as described (Koornneef *et al*., 2006). Yellowish siliques were placed in small paper bags until the siliques opened.

2.2.3 Isolation of DNA from plant leaves of *Arabidopsis thaliana*

DNA was isolated from leaves of single plants following the protocol of Edwards *et al*. (1991). Two to three leaves of six to eight weeks old *A. thaliana* plants were placed in 2 ml safe-lock microtubes and frozen in liquid nitrogen. Immediately after grinding by a mixer mill (27 Hz, 45 s), 400 µl DNA extraction buffer A were added to each sample. After vortexting for 30 sec the samples were centrifuged for 2 min at 20800 g. The supernatants were subsequently transferred to a new 1.5 ml microtube and 350 µl isopropanol were added to each sample. The microtubes were inverted two to three times, left at room temperature for two min and then centrifuged for five min at 20800 g. The DNA pellets were rinsed with 350 µl of 70% ethanol and centrifuged for 3 min at 20800 g. The supernatants were discarded and the

pellets were dried at room temperature for 15 min before they were resuspended in 50 µl of 1x RNAse solution each and incubated at 37°C for 30 min. Samples were stored at -20°C until further use.

2.2.4 Isolation of total DNA from aerial seedling tissues of *Arabidopsis thaliana*

Total DNA from aerial seedling tissues was isolated using a protocol modified after Dellaporta *et al.* (1983). Twohundred milligrams of aerial tissues of 17 days old seedlings or leaves of adult plants were harvested into 2 ml safe-lock microtubes, frozen with liquid nitrogen and ground by a mixer mill (27 Hz, 45 s). After addition of 790 µl of DNA extraction buffer B to each sample the preparations were vortexed for 30 sec. The samples were subsequently incubated at 65°C for 15 min, every 5 min the tubes were inverted. After addition of 276 µl of 5 M potassium acetate, the samples were vortexed, placed on ice for 15 min and then centrifuged for 10 min at 20800 g. The upper aqueous phases were then transferred to new 2 ml microtubes and 900 µl of phenol/chloroform/isoamyl alcohol were added to each preparation. The samples were inverted several times and then centrifuged at 6800 g for 5 min. The aqueous phases were transferred to new 2 ml microtubes, before 1 ml of chloroform was added. The samples were then inverted again and centrifuged at 6800 g for 5 min. The extraction of the supernatants with 1 ml of Chloroform was repeated once more. Then the upper aqueous phases were transferred to new 1.5 ml microtubes and 500 µl of isopropanol were added. The samples were gently inverted and centrifuged for 3 min at 20800 g. The supernatants were removed and the DNA pellets were rinsed with 1 ml of ethanol (70% v/v) followed by a centrifugation at 20800 g for 2 min. The ethanol was removed carefully and the pellets were dried at room temperature for 15 min before they were resuspended in 50 µl of 1x RNAse solution and incubated at 37°C for 2 hours. Concentration of total DNA was measured by NanoDrop-ND 1000 using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific Inc.) according to the manufacturer's instructions. After dilution the samples were kept at -20°C until further use.

2.2.5 Amplicon design

PRIMER3 software (http://bioinfo.ut.ee/primer3) was used to select oligonucleotides for amplicon design. The Electronic PCR web server (http://www.ncbi.nlm.nih.gov/projects/e-

pcr/reverse.cgi) was used to assess whether oligonucleotide sequences were present uniquely in the *A. thaliana* genome (Schuler, 1997; Rotmistrovsky *et al*., 2004)

2.2.5.1 Amplicons for allelic diversity studies

Col-0 sequences of the candidate genes were retrieved from the TAIR database (http://www.arabidopsis.org) and used to design gene-specific oligonucleotide pairs. For PCR amplification and DNA sequencing, primers were designed according to the following criteria, if possible; the GC content of a particular primer should range from 40% to 60%, its melting temperature should be approximately 60°C and its length between 18 and 26 nucleotides. To cover the entire coding region of a gene including introns, each gene was divided into several amplicons spanning about 1 kbp in length each (Supplementary table 1). Adjacent amplicons overlapped by 100 to 200 bp. In case amplification products for a subset of accessions were repeatedly not observed, alternative amplicons were designed (Supplementary table 2).

2.2.5.2 Amplicons for RT-PCR and qRT-PCR

To trace amplification of contaminating total DNA in RT-PCR and/or qRT-PCR experiments, the forward and reverse oligonucleotides were placed such that they matched sequences either side of an exon/exon junction. Oligonucleotide pairs for qRT-PCR were designed using the software PRIMER3 with the following criteria, if possible; melting temperature should be 60 ± 1°C, the GC content should be larger than 45% and an oligonucleotide should be between 19 and 23 nucleotides in length, amplicon sizes should range from 70 to 154 bp. Oligonucleotides were placed in regions that did not show polymorphisms in the accessions to be analysed for gene expression to prevent that sequence variants may have an impact on amplification efficiency. For each candidate and reference gene, at least one amplicon was designed. The established amplicons are given in Supplementary table 3.

For candidate genes, for which only parts of the ORF region were confirmed by EST and/or cDNA sequences, amplicons were designed in order to verify experimentally those parts of the gene structures that only relied on predictions (Supplementary table 3).

2.2.6 Polymerase chain reaction (PCR)

PCR amplifications were for example carried out to determine the presence and zygosity of certain T-DNA loci in transgenic lines, to establish particular fragments of candidate genes for sequence analysis and for the analysis of Indel polymorphisms. The reactions were performed in a final volume of 20 µl. The details for a PCR reation mixture and the standard program used for amplification are presented in Table 3.

Table 3. Standard PCR reaction mixture and amplification conditions. The asterisk indicates that the annealing temperature needs to be adjusted for certain amplicons, for example for particular Indel markers (Supplementary table 4).

PCR reaction mixture

Components	Volume
10x Dream *Taq* buffer	2 µl
dNTPs (10 mM)	2 µl
Dream *Taq* DNA polymerase (5 U/µl)	0.2 µl
DNA template (10 ng/µl)	2 µl
Forward primer (10 pmol/µl)	1 µl
Reverse primer (10 pmol/µl)	1 µl
ddH_20	11.8 µl
Total volume	**20 µl**

Standard program

Steps	Duration	Temperature
Initial denaturation	10 min	95°C
Denaturation	45 sec	95°C
Annealing	45 sec	60°C*
Extension	1 min	72°C
Final extension	10 min	72°C
Hold	∞	15°C
Steps 2-4 were repeated 34 times		

2.2.7 Agarose gel electrophoresis

To analyse PCR products and polymorphic patterns of Indel markers, DNA was separated by agarose gel electrophoresis in 1x TBE buffer. Depending on the size of the DNA fragments to be analysed, 0.8%, 1.2%, 2% (w/v) LE agarose or 3% NuSieve 3:1 agarose gels were used. Prior to gel electrophoresis, PCR products were mixed with an appropriate volume of 10x DNA loading buffer. Electrophoresis was performed in a chamber containing 1x TBE buffer with an applied voltage of 8 - 10 V/cm. After electrophoresis, the gels were stained with an ethidium bromide staining solution for 15 min. If needed, gels were destained in an aequeous solution. DNA fragments were visualised using UV and documented as images. The sizes of separated DNA fragments were analysed relative to GeneRuler DNA ladders (Thermo Fisher Scientific Inc.).

2.2.8 Purification of PCR products for direct sequencing

Purification of PCR products was carried out using a protocol modified from Werle *et al.* (1994). Applying this method, unincorporated oligonucleotides, which would interfere with

direct DNA sequencing, were removed from the PCR reactions. An aliquot of 5 µl of a PCR reaction was mixed with 10 U of Exonuclease I and 1 U of FastAP™ (Thermo Fisher Scientific Inc.). After an incubation at 37°C for 15 min, the mixture was heated to 85°C and kept at this temperature for another 15 min.

For DNA sequencing, 1 µl of purified PCR product was added to 4 µl ddH_2O and 1 µl of forward or reverse primer (5 pmol/µl). All samples were sequenced at The Plant Genome Resource Center (PGRC) sequencing service at the IPK using ABI 3730 XL automatic sequencers.

2.2.9 Sequence analysis

2.2.9.1 Sequence alignments and comparisons

For allelic diversity studies PCR amplifications were performed using gene-specific amplicons (Supplementary table 1) and DNA of twenty-six accessions as templates. Most PCR products were directly sequenced using oligonucleotides of the forward as well as the reverse orientations. The sequences of all accessions were manually edited to remove low quality stretches at the 5'- and 3'-ends of the reads. If necessary, miscalled bases were identified and changed. In cases in which discrepancies were noted between two sequences of the same accession for a particular amplicon, additional sequences were established. The accession sequences for each amplicon were aligned to the Col-0 reference sequence using Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html).

Sequence assembly and analysis were performed using the Wisconsin Package (version 10.0-UNIX; Genetic Computer Group, Madison, WI). ORFs of accession gene sequences were predicted based on alignments with Col-0 cDNA sequences. Sequence alignments were performed with the Bestfit program. For alignment of nucleotide and amino acid sequences gap creation penalties of 50 and 8 as well as gap extension penalties 1 and 2 were used, respectively. Unless otherwise stated, the comparisons were restricted to the regions from start to stop codons.

2.2.9.2 Polymorphism analysis

All SNPs and Indels that were present in the aligned amplicon sequences of the different accessions were determined using the *A. thaliana* Col-0 sequence as reference, then it was assessed which of the polymorphisms were present in coding regions. It was recorded whether SNPs corresponded to transitions, A → G, G → A, T → C or C → T, or to transversions, A → T, T → A, A → C, C → A, G → T, T → G, G → C or C → G. For SNPs which were located in coding regions it was also determined whether they were present at the 1st, 2nd or the 3rd position of a codon and whether they caused synonymous or nonsynonymous substitutions. The length of Indels was determined, if possible. Furthermore, it was assessed whether Indels caused amino acid insertions/deletions, affected exon/intron borders or caused a frame shift.

SNP and Indel frequencies of a particular candidate gene were determined for each accession in gene and coding regions relative to the total length (bp) of sequenced regions in all accessions.

2.2.9.3 Identification of microsatellites

The Col-0 sequences of the twelve candidate genes that were used to design gene-specific amplicons were also used for the identification of microsatellites using Microsatellite repeats finder (http://insilico.ehu.es/mini_tools/microsatellites/). Repeat units of a length between 2 and 10 bp were identified in the Col-0 sequence, if at least three perfect repeats were present. Mononucleotide stretches were considered that spanned at least 6 bp.

2.2.10 Generation of introgression lines (ILs)

To generate an introgression line, a selected accession harbouring an allelic variant of interest and a transgenic line carrying six or eight copies of the *GFP* transgene were used as female and male parents in a backcrossing strategy, respectively. The resulting F_1 plants were backcrossed four times to *GFP*-containing transgenic lines until the BC_4 generations were obtained. In each generation it was assessed with PCR-based assays which plants carried the allelic variant of interest (Supplementary table 5). Furthermore, the presence and zygosity of the T-DNA loci was determined (Supplementary table 6). Selected plants of the

BC_4 generation were carried on after self-pollination up to the BC_4F_2 generation in order to obtain introgression lines that harboured the allelic variant of interest and two *GFP* loci homozygously. The principle how the presence and zygosity of a particular T-DNA locus is assessed with the help of two different oligonucleotide combinations in individual lines is outlined in Figure 2.

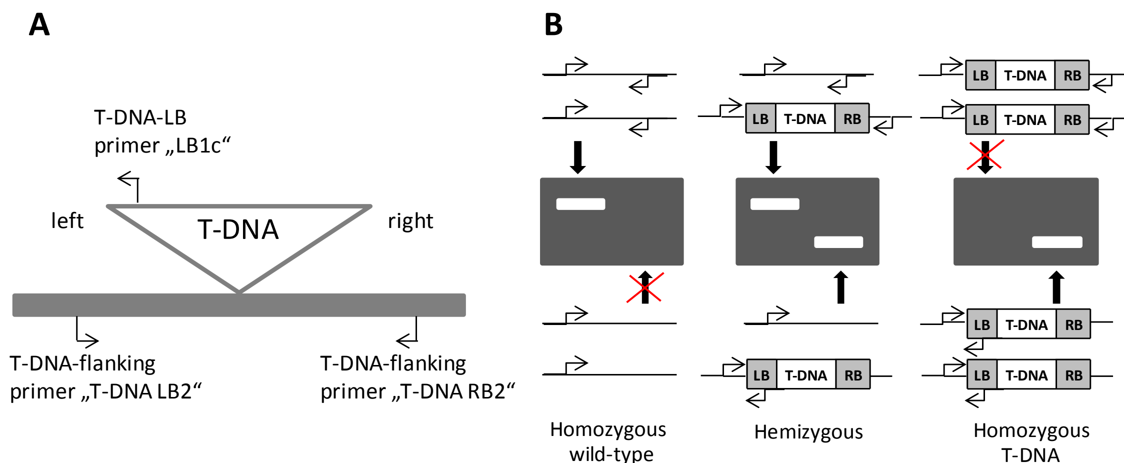


Figure 2. Determining the presence and zygosity of a particular T-DNA locus in a transgenic line. (A) Locations of the oligonucleotides that are used to identify the presence and zygosity of a T-DNA insertion. The flanking primer "T-DNA LB2" is on the one hand combined with primer "LB1c" and on the other hand with the other flanking primer "T-DNA RB2". The oligonucleotide sequences for "LB1c", "T-DNA LB2" and "T-DNA RB2" are given in Supplementary table 6. (B) Depending on the outcome of the PCR amplifications with the two different amplicons it can be deduced whether the analysed T-DNA locus is present hemizygously or homozygously in a particular plant or not at all. The standard PCR conditions used did not permit amplification across the T-DNA locus.

2.2.11 Detection and imaging of *GFP* fluorescence

Visual detection of *GFP* fluorescence was performed using the fluorescence stereomicroscope Leica MZ16 F. The fluorescence stereomicroscope was equipped with three sets of filters:

GFP 2	Excitation filter: 480/40 nm	Barrier filter: 510 LP
GFP 3	Excitation filter: 470/40 nm	Barrier filter: 525/50 nm
CY5	Excitation filter: 620/60 nm	Barrier filter: 700/75 nm

Photographic documentation was carried out using the digital camera "Leica DFC 345FX" and the software "Leica Application Suite Version 3.7.0". Figure 3 shows an example of a silenced plant, which was documented using filter sets GFP3 (A), CY5 (B) and the overlay image (C).

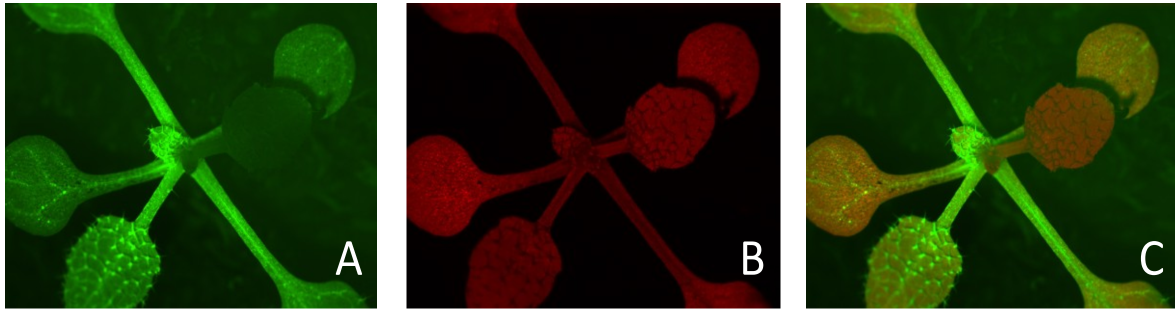


Figure 3. Photographic documentation of a plant showing *GFP* silencing. A seventeen days old plant was documented using two filter sets. (A) *GFP* fluorescence is seen as bright green with filter set GFP3, whereas tissues exhibiting *GFP* silencing appear dark green. (B) Due to chlorophyll fluorescence aerial tissues are bright red when evaluated with filter CY5. (C) Overlay of the images shown in panels A and B.

2.2.12 Analysis of *GFP* gene silencing

Silencing behaviour of *GFP* transgene silencing in BC_4F_3 introgression lines was evaluated by fluorescence stereomicroscopy using the scoring system that had been developed for transgenic *GFP* lines in the Col-0 genetic background (Arlt, 2007). Seventy plants were analysed for each line in a particular experiment. The plants were distributed to two small trays containing 35 plants each and cultivated in a growth cabinet under a long day regime (Table 2). To minimise position effects of the trays in growth cabinet, all trays were rotated daily and shifted every alternate day; the positions of the plants in a tray were also randomised twice a week, but plants of one tray were never transferred to the other tray.

GFP reporter gene activity was monitored from day 17 after sowing onwards. For a period of five weeks, it was recorded twice a week for all plants whether silencing had occurred and which proportion of the aerial tissues of a particular plant exhibited *GFP* silencing. Based on the estimated percentage of aerial tissue that showed *GFP*-silencing, the plants were divided into six different categories. Plants exhibiting no silencing at all or silencing in the entire aerial tissues were grouped into categories 0 and 5, respectively. Plants showing silencing in less than 10% and more than 90% of the aerial tissues belonged to categories 1 and 4, respectively. Plants of categories 2 and 3 had between 10% and 50% and between 50% and 90% silenced aerial tissues, respectively. In order to obtain a simple and quantitative description of the *GFP* transgene expression, “Frequency of silencing” (F) was used. Frequency of silencing describes the proportion of silenced plants out of the total number of

all plants analysed. For each line and time point of a particular experiment the frequency of silencing was calculated.

The numbers of silenced and non-silenced plants as well as the number of plants classified into the six different categories were compiled for each introgression line and compared to the values of line 6x*GFP-F8/*R127. Data obtained for each time point of a particular experiment were used for statistical analysis using Fisher's exact test in the R environment. If the obtained p-values were smaller than 0.05, the results were considered to be significant.

2.2.13 qRT-PCR experiments

2.2.13.1 Isolation of RNA from *A. thaliana* aerial seedling tissues

RNA was isolated from 80 mg of aerial seedling tissues of 10 days old *A. thaliana* plants. The tissue samples were frozen in liquid nitrogen and ground by a mixer mill at 27 Hz for 45 s. RNA extraction was performed using "PeqGOLD RNA Pure™" (Peqlab) according to the manufacturer's instructions. Total RNA was dissolved in 40 – 50 µl DEPC-treated water and stored at -80°C until further use.

The RNA concentration of samples was determined by using a NanoDrop ND-1000 spectrophotometer (Peqlab). The ratio of the absorbance values at 260 nm and 280 nm (A260/A280) was furthermore used to assess the purity of the RNA in 1 µl aliquots. Only RNA samples showing A260/A280 ratios higher than 1.8 were used for qRT-PCR experiments.

2.2.13.2 DNAse treatment of total RNA

RNA was purified using the "TURBO DNA-free™" DNAse Kit (Thermo Fisher Scientific Inc.) to remove DNA. Total RNA of a particular sample, 10 µg, was mixed gently with 5 µl TURBO DNase buffer and 1 µl TURBO DNase, then DEPC-treated water was added in order to obtain a 50 µl reaction volume. The mix was incubated for 30 min at 37°C. After the enzymatic reaction 5 µl of DNA Inactivation Reagent were added and the sample was incubated for 5 min at room temperature. After centrifugation at 10600 g at 4°C for 90 s, the RNA solution was transferred to a new 1.5 ml microtube and stored at -80°C until further use.

RNA quality and quantity were also assessed via formaldehyde-agarose gel electrophoresis. Electrophoresis equipment was soaked in DEPC-treated water prior to the experiments. A 1.2% agarose gel was prepared by dissolving 1.2 g agarose in 72 ml DEPC-treated water and 10 ml of 10x RNA running buffer. After the solution cooled down to 65°C, 18 ml of formaldehyde were added and the gel (11x14 cm) was casted in a fume hood. The agarose gel was left for 1 hour at room temperature and then placed into the electrophoresis chamber and covered with 1x RNA running buffer. Prior to the separation of the RNA samples electrophoresis was performed for 30 min at 120 V. RNA samples containing between 5-10 µg RNA in a volume of 4 µl were mixed with 4 µl of RNA sample buffer. The samples were incubated at 65°C for 10 min and chilled on ice. Subsequently, samples were mixed with 4 µl of RNA loading buffer and loaded onto the gel. The gel was run at 3.6 V/cm for 3 hours and then soaked twice in DEPC-treated water for 10 to 15 min to remove formaldehyde. The gel was stained in an aequeous solution containing 1 µg/ml ethidium bromide for 10 min and destained in DEPC-treated water for 3 hours before placing it under UV to visualise the RNA for documentation.

2.2.13.3 cDNA synthesis and RT-PCR

Reverse transcription reactions were performed using the “Maxima H Minus First Strand cDNA Synthesis Kit” (Thermo Fisher Scientific Inc.). First-strand cDNA was synthesised using 1 µg total RNA and oligo (dT) primer as described in the manufacturer’s instructions. The cDNA synthesis was assessed by RT-PCR amplification using oligonucleotides flanking an intron sequence. For each cDNA sample 2 ng were used as template in RT-PCR reactions with a final volume of 20 µl. PCR reactions were set up in the same way as described for DNA samples, furthermore, the same amplification conditions were applied (section 2.2.6). The oligonucleotide sequences used for RT-PCR amplifications are listed in Supplementary table 3. After PCR amplification, the products were separated using high resolution agarose gel electrophoresis.

2.2.13.4 qRT-PCR experiment set up

Several genes which exhibited expression stability in the study of Czechowski *et al*. (2005) were considered as reference genes. From these genes *At4g34270* was selected to serve as

reference gene since its expression level was in a similar range as that of the candidate genes to be analysed.

For each qRT-PCR amplicon PCR efficiency was calculated based on several template dilutions using formula $E = (10^{(-1/slope)}-1)*100$ which was previously described by Pfaffl (2001). To create a reliable standard curve for the relative quantification, serial dilutions of 1:5, 1:10, 1:20, 1:40, 1:80, and 1:160 of the same cDNA template were used for qRT-PCR reactions. Each dilution was prepared three times, for each of the resulting samples three technical replicates were analysed as described below.

The qRT-PCR reactions were performed in 384-well plates with the ABI PRISM® 7900 HT real-time PCR System, SYBR green was used to monitor PCR amplifications. Reactions contained 7.5 µl of 2x Maxima SYBR Green/ Fluorescein Master Mix reagent (Thermo Fisher Scientific Inc.), 0.3 µM of each forward and reverse oligonucleotides and 2.0 ng cDNA template in a final volume of 15 µl. A cocktail consisting of all reagents as well as the cDNA template and the oligonucleotides was prepared and dispensed into the relevants wells. The following program was used for all qRT-PCR reactions:

1. Initial denaturation at 95°C for 10 min
2. Denaturation at 95°C for 15 s
3. Annealing at 60°C for 30 s
4. Extension at 72°C for 30 s

Steps 2 to 4 were repeated 39 times

Amplicon dissociation curves were recorded after cycle 40 by heating from 60 to 95°C to assess the specificity of the qRT-PCR products. The qRT-PCR experiments were carried out for three independent biological replicates with four technical replicates each.

Data were analysed using the SDS 2.2.2 software. The cycle threshold (C_t) values for product detection with baseline set to cycles 3-15 were used to calculate the relative expression levels. The expression of candidate genes in selected accessions compared to the expression in Col-0 was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001), an amplicon of *At4g34270* served as reference.

Amount of target gene= $2^{-\Delta\Delta Ct}$ where $\Delta\Delta Ct = \Delta Ct(accession) - \Delta Ct(Col\text{-}0)$

3 RESULTS

3.1 Sequence diversity in *Arabidopsis thaliana* genes that are involved in sense-transgene induced post-transcriptional gene silencing

This study aims at a survey of sequence variation in coding regions of genes which are known or suspected to play a role in sense-transgene induced post-transcriptional gene silencing in *A. thaliana*. *AGO1* (*At1g48410*), *AGO7* (*At1g69440*), *DCL4* (*At5g20320*), *ERI* (*At3g15140*), *HEN1* (*At4g20910*), *NRPD1* (*At1g63020*), *NRPE1* (*At2g40030*), *SDE3* (*At1g05460*), *SDE5* (*At3g15390*), *SGS3* (*At5g23570*), *WEX* (*At4g13870*) and *XRN4* (*At1g54490*) were chosen as candidate genes. For the analysis of the 12 genes with respect to allelic diversity 26 *A. thaliana* accessions including Col-0 were used which capture portions of the genetic diversity found in this species (Baxter *et al*., 2010; Platt *et al*., 2010).

3.1.1 Analysis of sequence variation in candidate genes – *NRPE1* as an example

To investigate sequence variation in the *NRPE1* gene, the Col-0 reference sequence of the *NRPE1* gene was retrieved from The Arabidopsis Information Resource (TAIR, https://www.arabidopsis.org) and used to design gene-specific oligonucleotide pairs (Supplementary table 1). It was the aim to cover for this gene almost the entire coding region including introns therefore the gene was divided into 10 amplicons spanning about 1 kbp in length each. Adjacent amplicons overlapped by 93 to 223 bp (Supplementary table 7). Figure 4 shows the location of the ten oligonucleotide pairs relative to a schematic drawing of the *NRPE1* exon/intron structure. The amplicons were denoted NRPE1-1 to NRPE1-10.

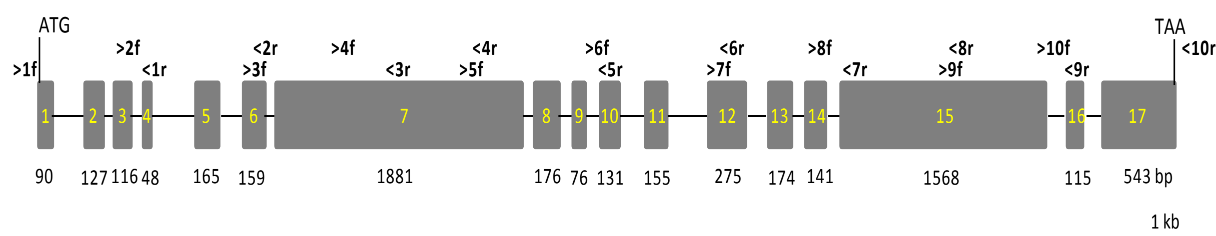


Figure 4. Amplicons developed for the *NRPE1* gene. Schematic drawing of the *NRPE1* exon/intron structure, the region from the ATG to the stop codon is shown. The grey boxes and the black bars indicate exon and intron regions, respectively. The numbers below the grey boxes indicate the lengths of the different exons in bp. The locations of the ten oligonucleotide pairs are shown above the gene structure. Arrowheads indicate the orientation of forward and reverse primers.

DNA samples of the 26 accessions were used as templates for PCR amplification and all resulting amplification products were sequenced. The sequences for the different amplicons were derived from bi-directional sequencing with the exception of amplicons 5 and 7, which were only sequenced with the oligonucleotides in forward orientation. The sequences of all accessions were manually edited and aligned to the Col-0 reference sequence using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Based on these alignments, SNPs and/or Indel polymorphisms were determined (Figure 5, Supplementary table 8). The alignments were manually curated, if necessary. The combined regions that were covered by the established sequences of all 26 accessions spanned together 8050 bp, this corresponded to 88.6% of the region from start to stop codon in the Col-0 reference gene sequence.

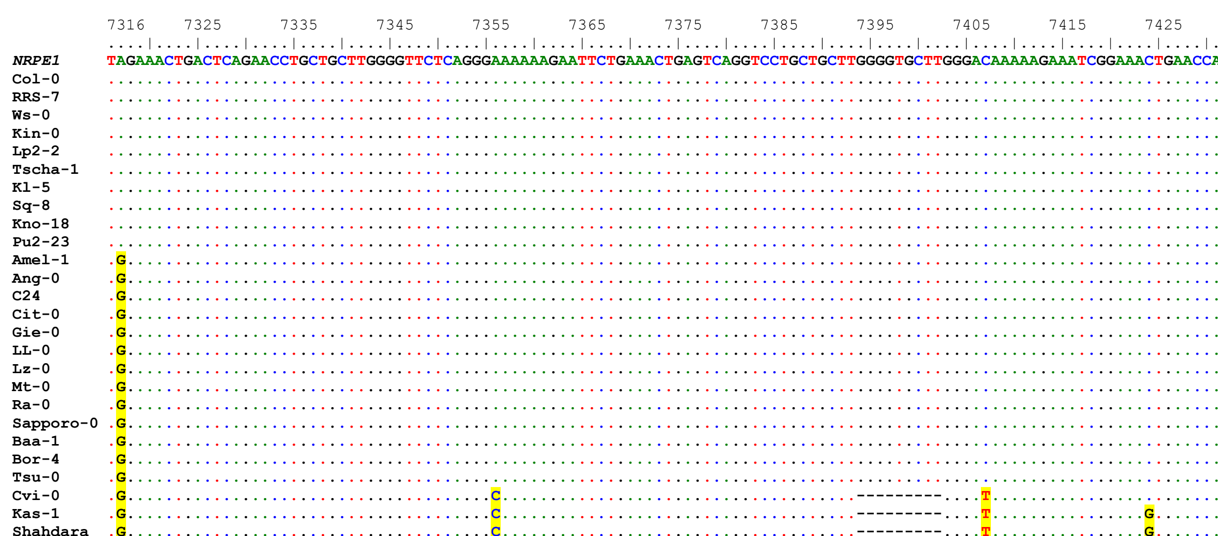


Figure 5. Multiple alignment of sequences derived from *A. thaliana* accessions for a region of amplicon 9 of the *NRPE1* gene. The *NRPE1* gene sequence retrieved from TAIR served as reference. Dots indicate bases that are identical to the reference sequence, the symbol (–) indicates a deleted base. SNPs are highlighted in yellow. Accessions were ordered according to the polymorphism patterns.

The sequences established for accession Col-0 did not show any differences to the reference sequence. In the other 25 accessions the analysis of edited sequences for the 10 amplicons of the *NRPE1* gene revealed at 144 positions SNPs, moreover 23 different Indels were found in those regions of the gene that were covered by sequences in all accessions (Supplementary table 8). DNA polymorphisms were found in all regions of the gene, in both exons and introns. SNP frequencies were established for the different accessions by dividing the total number of SNPs detected for a particular accession by the combined length of the gene regions analysed in all accessions, the Col-0 gene sequence served as reference in all

cases. Five accessions Kin-0, Lp2-2, RRS-7, Tscha-1, and Ws-0 did not reveal any SNP. In approximately two-thirds of the accessions sequence differences were found, but SNP frequencies were lower than 0.2%. For three accessions, Cvi-0, Kas-1 and Shahdara, SNP frequencies higher than 1% were observed (Table 4). With respect to the Col-0 gene sequence four insertion and 19 deletion polymorphisms were found. The majority of Indels, 14 out of 23, were shorter than 10 bp, the lengths of seven indels varied from 11-50 bp, and two deletions spanned 72 and 114 bp (Table 4).

Haplotypes represent distinct sets of linked single nucleotide and/or indel polymorphisms. Based on the number, identity and position of SNPs and/or Indels the *NRPE1* sequences of the different accessions were classified into different haplotypes. Twenty different haplotypes were found among the 26 accessions studied. Sixteen haplotypes were found only once, two haplotypes twice and and another two haplotypes were observed in three accessions (Table 4).

Table 4. Sequence diversity of the *NRPE1* gene in 26 *Arabidopsis thaliana* accessions. Accessions were ordered according to haplotypes. The symbols (-) or (+) indicate deletions or insertions relative to the Col-0 reference sequence, respectively.

Accession	SNP frequency (%)	No. of SNP(s)	Length(s) of Indel(s) in bp	Haplotype
Col-0	0	0	0	1
RRS-7	0	0	0	1
Ws-0	0	0	-6	2
Kin-0	0	0	-6	2
Lp2-2	0	0	-12	3
Tscha-1	0	0	-18	4
Sq-8	0.012	1	-12	5
Kno-18	0.025	2	-24	6
Bor-4	0.112	9	-1;-30;+6	7
LL-0	0.124	10	+6;+6	8
Sapporo-0	0.124	10	+6;+6	8
Tsu-0	0.124	10	+6;+6	8
Amel-1	0.124	10	-1;-36;+6	9
Gie-0	0.124	10	-1;-36;+6	9
Baa-1	0.124	10	-1;-36;+6	9
C24	0.137	11	+12;+6	10
Cit-0	0.137	11	+12;+6	11
Pu2-23	0.137	11	-1;-18;+6	12
Kl-5	0.137	11	-1;-24;+6	13
Mt-0	0.137	11	+6;+6	14
Lz-0	0.137	11	+12;+6	15
Ra-0	0.161	13	-18;+6	16
Ang-0	0.161	13	-24;+6	17
Kas-1	1.143	92	-1;-1;-7;-1;-3;-1;-9;-1;-114;+6	18
Cvi-0	1.168	94	-1;-1;-7;-1;-3;-1;-9;-1;-6;+12;-12;-6	19
Shahdara	1.217	98	-1;-1;-7;-1;-3;-1;-2;-9;-1;-6;+12;-72;-6	20

It was determined which and how many of the SNPs and Indels were present in the sequence of the open reading frame (ORF). Moreover, it was analysed whether and how many of the sequence polymorphisms caused changes in the amino acid sequence. Eighty-one of the SNPs (56.3%) were located in the 5698 bp of ORF sequences that were present in the sequences established for all 26 accessions. SNPs which were located in the first, second and third codon position were found in 14, 19 and 48 cases, respectively. Thirty-four of the 81 SNPs (42.0%) caused amino acid replacements. Two SNPs located in the last exon of the gene led to premature stops. The premature stop in the Cvi-0 accession removed 51 amino acids, whereas the one in the Shahdara accessions truncated the protein by 2 amino acids. Sixteen (70.0%) out of the 23 Indels were found in exon sequences, the 4 insertions and 12 deletions were in-frame (Supplementary table 8).

3.1.2 Survey of sequence variation in *Arabidopsis thaliana* accessions revealed highly diverged allelic variants for several candidate genes in subsets of the accessions

Candidate genes *AGO1, AGO7, DCL4, ERI, HEN1, NRPD1, SDE3, SDE5, SGS3, WEX* and *XRN4* were analysed in the same manner as described for the *NRPE1* gene. Specific oligonucleotide pairs for 60 gene-specific amplicons were designed in total for the eleven candidate genes. For the longest gene, *DCL4,* 11 amplicons were developed and for the two shortest genes, *WEX* and *ERI*, three amplicons each. The other candidate genes that spanned approximately 3000 to 7000 bp were divided into four to eight amplicons. The sequences of the oligonucleotide pairs are listed in Supplementary table 1.

Amlification products were observed in all accessions used in this study for 67 out of the 70 amplicons that were designed for the twelve candidate genes. For amplicons AGO7-1, HEN1-5 and WEX-3F repeatedly no product was obtained in a subset of accessions. Sequences of accessions Gie-0 and Bor-4 corresponding to *AGO7* showed many sequence differences in amplicons AGO7-2, AGO7-3 and AGO7-4 when compared to the gene sequence of the Col-0 reference accession. However, they were highly similar to whole genome shotgun sequences (WGS) that had been established for accession L*er*. Therefore, an alternative amplicon was designed to cover a similar region as AGO7-1 based on the WGS contig available for L*er* (AFMZ01005520). Using amplicon AGO7-1a sequences could be established for accessions Gie-0 and Bor-4. Accessions Lp2-2 and Sq-8 did not yield amplification products with

oligonucleotide pair HEN1-5. An oligonucleotide corresponding to the open reading frame of *At4g20900* in combination with oligonucleotide HEN1-5f was suitable for amplification of Lp2-2 and Sq-8 as templates DNAs, this oligonucleotide combination was used to establish sequences for the region corresponding to HEN1-5. For accessions Kin-0, Mt-0 and Sq-8 no amplification was observed using amplicon WEX-3. The combination of oligonucleotides WEX-2bf and WEX-3r enabled to amplify sequences for Mt-0 and Sq-8. Oligonucleotide WEX-2cs was developed based on WEX-2b amplicon sequences that had been established for Kin-0. This oligonucleotide in combination with WEX-3r was suitable to establish Kin-0 sequences for the region covered by amplicon WEX-3.

The edited sequences established for the 26 accessions were aligned to the Col-0 reference sequence of each candidate gene in order to detect sequence variation. Without exception the amplicon sequences established for Col-0 did not differ from the gene sequences which were retrieved from the reference genome sequence (TAIR, https://www.arabidopsis.org; Supplementary table 8).

Table 5. Sequence regions analysed for the different candidate genes with respect to allelic diversity. The Col-0 sequences of the candidate genes were retrieved from TAIR. In case more than one splicing variant was available, the ORF corresponding to splicing variant 1 was used.

	Gene		ORF	
	Length of Col-0 sequence	Length covered in all accessions (%)	Length of Col-0 sequence	Length covered in all accessions (%)
AGO1	6489 bp	5540 bp (85.4%)	3147bp	2986 bp (94.9%)
AGO7	3640 bp	3275 bp (90.0%)	2973 bp	2796 bp (94.0%)
DCL4	10049 bp	8721 bp (86.8%)	5109 bp	4681 bp (91.6%)
ERI	2250 bp	2083 bp (92.6%)	1014 bp	1014 bp (100%)
HEN1	4608 bp	3582 bp (77.7%)	2829 bp	2393 bp (84.6%)
NRPD1	7548 bp	5728 bp (75.9%)	4362 bp	4243 bp (97.3%)
NRPE1	9090 bp	8050 bp (88.6%)	5931 bp	5698 bp (96.1%)
SDE3	3880 bp	3014 bp (77.7%)	3009 bp	2721 bp (90.4%)
SDE5	3731 bp	2845 bp (76.3%)	1473 bp	1383 bp (93.9%)
SGS3	2927 bp	2324 bp (79.4%)	1878 bp	1773 bp (94.4%)
WEX	2264 bp	1929 bp (85.2%)	858 bp	769 bp (89.6%)
XRN4	6726 bp	6480 bp (96.3%)	2844 bp	2798 bp (98.4%)
Sum	63202 bp	53571 bp (84.8%)	35427 bp	33255 bp (93.9%)

For each of the genes it was calculated which areas of the gene sequences were covered by amplicon sequences in all accessions analysed. For genes *AGO7*, *ERI* and *XRN4* the combined sequences established for the accessions accounted for 90%, 93% and 96% of the gene sequences, respectively. Between 76 and 79% of the *HEN1, NRPD1, SDE3, SDE5* and *SGS3*

gene sequences were covered by amplicon sequences in all accessions. For all other genes between 85 and 89% of the gene regions were represented in all accessions. With respect to the ORF region, an even better coverage had been achieved, for ten out the twelve genes more than 90% of the ORF sequences were covered by the established amplicon sequences (Table 5; Supplementary table 7).

Table 6. Alleles of several candidate genes show high SNP frequencies when compared to the corresponding Col-0 gene sequences. SNP frequencies are given in percent. The light and dark green coloured boxes mark SNP frequencies higher than 0.5% and 1%, respectively. Asterisks indicate that Indels larger than 50 bp were present in exon regions. Only regions that were covered by amplicon sequences in all accessions were analysed.

Accession	*AGO1*	*AGO7*	*DCL4*	*ERI*	*HEN1*	*NRPD1*	*NRPE1*	*SDE3*	*SDE5*	*SGS3*	*WEX*	*XRN4*
Col-0	0	0	0	0	0	0	0	0	0	0	0	0
Amel-1	0.289	0.061	0.424	0.240	0.223	0.663	0.124	0	0.211	0.086	0.207	0.077
Ang-0	0.325	0.092	0.046	0.384	0.112	0.628	0.161	0.199	0.211	0.215	3.110	0.062
Baa-1	0	0.122	0.046	0.336	0.112	0.611	0.124	0.199*	0.176	0.129	3.162	0.062
Bor-4	0.235	3.206	0.138	0	0.140	0.576	0.112	0	0	0.129	2.385	0.062
C24	0.036	0.031	0.092	0.240	0.112	0.663	0.137	0.232	0.176	0.215	0.259	0.077
Cit-0	0.307	0.031	0.424	0.192	0.112	0.663	0.137	0.265	0.176	0.043	1.244	0.062
Cvi-0	0.181	0.153	0.413	0.288	0.084	0.576	1.168	0.531	0.105	0.129	2.229	0.108
Gie-0	0.235	3.206	0.057	0.144	0.112	0.698	0.124	0.133	0.176	0.086	1.244	0.062
Kas-1	0.271	0.061	0.183	0.240	0.084	0.628	1.143*	0.232	0.211	0.043	2.851	0.031
Kin-0	0.289	0.061	0.424	0.192	0.056	0.663	0	0.166	0.246	0.172	4.355	0.062
Kl-5	0.235	0	0	0.144	0.056	0.594	0.137	0	0.035	0.172	2.540	0.077
Kno-18	0.235	0.031	0.115	0.288	0.195	0.628	0.025	0.033	0.141	0.172	0.415	0
LL-0	0	0.122	0.046	0.336	0.112	0.611	0.124	0.133	0.176	0.215	2.696	0.062
Lp2-2	0.018	0.122	0.172	0.048	0.921	0.681	0	0	0	0.172	2.799	0.062
Lz-0	0.307	0.153	0.011	0.288	0.028	0.646	0.137	0.531	0.211	0.086	2.696	0
Mt-0	0	0.061	0.172	0.240	0.084	0.576	0.137	0.199	0.176	0	2.592	0.077
Pu2-23	0.235	0.122	0.183	0.336	0.056	0.576	0.137	0.265	0	0.129	2.799	0.077
Ra-0	0.090	0.153	0.046	0.288	0.028	0.559	0.161	0.531	0.176	0.086	2.696	0
RRS-7	0.090	0.092	0.046	0.336	0.195	0.646	0	0.199	0.211	0.172	0.207	0.062
Sapporo-0	0	0.183	0.138	0.144	0.056	0.628	0.124	0.265	0.070	0.172	0.311	0.046
Shahdara	0.271	0.061	0.195	0.144	0.084	0.628	1.217*	0.299	0.035	0.043	2.851	0.031
Sq-8	0	0.092	0.138	0.336	1.089	0.646	0.012	0.033	0.176	0.086	2.592	0.077
Tscha-1	0.235	0.183	0.034	0	0.112	0.628	0	0	0.141	0.086	2.592	0.015
Tsu-0	0.018	0	0.057	0.288	0.112	0.594	0.124	0.166	0.070	0.215	0.259	0.062
Ws-0	0.307	0.061	0.183	0.336	0.195	0.611	0	0.199*	0.176	0.043	2.488	0.031

As the next step, SNP frequencies were calculated for each of the candidate genes, the Col-0 gene sequences served as reference in all cases (Table 6). SNP frequencies higher than 1% were found for the *AGO7*, *HEN1*, *NRPE1* and *WEX* genes. In case of the *AGO7* gene, two accessions, Bor-4 and Gie-0, showed 3.2% nucleotide differences, whereas less than 0.2%

were found for the remaining accessions. Nineteen accessions showed SNP frequencies higher than 1% when the *WEX* gene was evaluated. Among them, Kin-0 was the accession that showed the highest value with 4.4%, for accessions Ang-0 and Baa-1 3.1% and 3.2% nucleotide differences were observed, respectively. Fourteen of the nineteen accessions revealed values between 2.2% and 2.9% and two accessions showed 1.2% nucleotide differences. The sequence alignment shown in Figure 6 reveals the high degree of polymorphism that was found in many accessions for the *WEX* gene (Supplementary table 8).

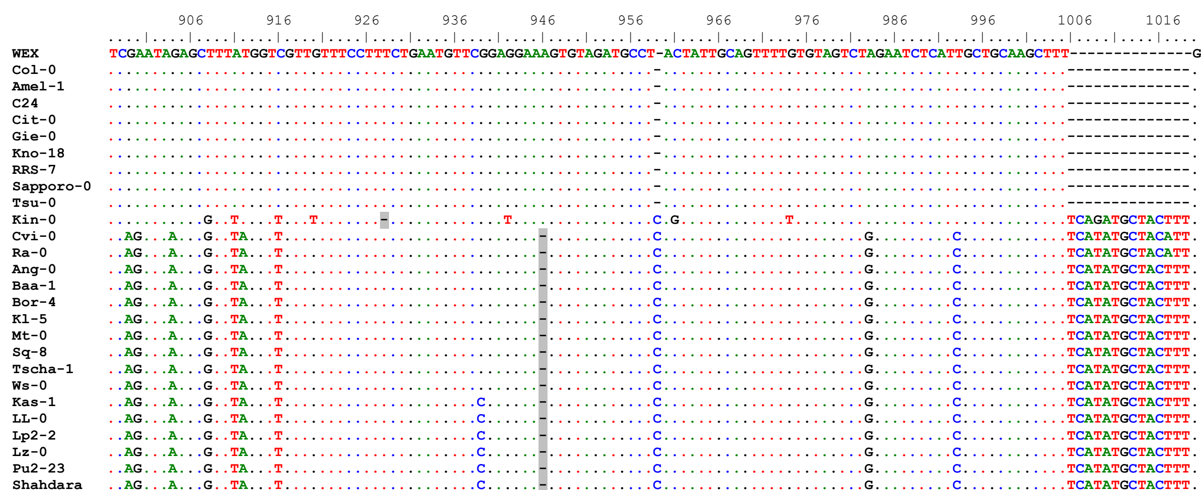


Figure 6. Alignment of *WEX* gene sequences obtained for 26 *A. thaliana* accessions reveals a highly polymorphic region. The *WEX* gene sequence was retrieved from TAIR and served as reference. Only the nucleotides that differ from the reference sequence of accession Col-0 are shown for the different accessions whereas bases identical to the references sequence are represented as dots. Indels are marked with "–". Accessions were ordered based on their polymorphism patterns.

Analysis of the *NRPE1* sequences revealed that accessions Cvi-0, Kas-1 and Shahdara showed approximately tenfold higher SNP frequencies than the seventeen accessions for which SNPs were observed. In case of the *HEN1* gene, Lp2-2 and Sq-8 showed fourfold to twentyfold higher SNP frequencies than the remaining 23 accessions which were analysed. Three accessions, Cvi-0, Lz-0 and Ra-0, showed SNP frequencies of 0.53% when the *SDE3* gene was evaluated, whereas in the other accessions values were found that were lower than 0.3%. In case of the *NRPD1* gene all 25 accessions showed sequence differences between 0.56% and 0.68% when compared to the Col-0 gene sequences. For six of the genes, *AGO1, DCL4, ERI, SDE5, SGS3* and *XRN4*, values below 0.5% were noted for all accessions analysed. In case of the *XRN4*, *SGS3* and *SDE5* genes SNP frequencies established for the different accessions did not exceed 0.11%, 0.22% and 0.25%, respectively. For the *AGO1*, *ERI* and *DCL4* genes overall

higher values were noted in the 25 accessions, but in none of the accessions values of 0.33%, 0.38% and 0.42% were surpassed, respectively.

Table 7. Summary of the SNPs detected in 25 accessions for 12 candidate genes. Gene and ORF regions refer to the combined areas that are covered by amplicon sequences in all accessions analysed (Supplementary table 7). SNP frequency refers to the number of positions in which SNPs were detected in any of the 25 accessions. ([2]) and ([3]) indicate that two and three nucleotide changes are present in the same accession(s) and affect the same codon, respectively (Supplementary table 8).

	Gene region				ORF region								
Candidate gene	**SNP frequency (%)**	**Total**	**Transitions**	**Transversions**	**SNP frequency (%)**	**Total**	**Transitions**	**Transversions**	**1st codon position**	**2nd codon position**	**3rd codon position**	**Synonymous substitutions**	**Nonsynonymous substitutions**
AGO1	1.101	61	31	30	0.469	14	9	5	2	1	11	12	2
AGO7	3.878	127	70	57	3.720	104	63	41	14	11	79	72[2]	30[2]
DCL4	1.101	96	59	37	0.833	39	27	12	12	10	17	16	23
ERI	0.720	15	5	10	0.690	7	3	4	2	0	5	3	4
HEN1	2.010	72	42	30	1.755	42	24	18	11	7	24	20	22
NRPD1	1.187	68	43	25	1.013	43	28	15	10	11	22	18	23[2,2]
NRPE1	1,789	144	79	65	1,422	81	47	34	14	19	48	45	36
SDE3	1.194	36	22	14	1.103	30	19	11	6	5	19	17[2]	11[2]
SDE5	0.578	16	9	7	0.578	8	4	4	3	2	3	1	6[2]
SGS3	0.645	15	10	5	0.677	12	8	4	2	2	8	7	5
WEX	8.450	163	81	82	5.462	42	21	21	9	10	23	19[3]	21
XRN4	0.231	15	5	10	0.071	2	0	2	1	0	1	0	2
Sum		828	456	372		424	253	171	86	78	260	230	185

It was determined which of the SNPs were present in coding regions and whether they caused changes in the amino acid sequences. None of the SNPs identified in this study affected the exon/intron borders of the candidate genes as present in the Col-0 reference sequences. Approximately half of the SNPs were located in protein-coding exons. Transitions were with 55% and 60% more frequent than transversions in the analysed gene and ORF regions, respectively. For the *ERI* and *XRN4* genes, transversions were more frequent than transitions, but only few SNPs were observed for these two genes. Out of the 424 SNPs that were located in ORFs 260 (61%) represented the third codon position, whereas 86 (20%) and 78 (18%) were present in the first and second codon position, respectively. Synonymous substitutions were with 55% more prevalent than nonsysnoymous substitutions, however, half of the genes that were analysed, *DCL4*, *ERI*, *HEN1*, *SDE5*, *WEX* and *XRN4*, showed more nonsynonymous substitutions than synonymous substitutions (Table 7; Supplementary table 8).

Indel polymorphisms were found in the analysed regions of all candidate genes except *SGS3* (Table 8 ; Supplementary table 8). Among the eleven genes for which Indels were detected, four genes, *HEN1*, *NRPD1*, *SDE5* and *XRN4*, did not show any Indels in the coding regions. All but one of the length polymorphisms identified in exon regions were in-frame indels, only a 1 bp deletion found in the Shahdara allele of the *SDE3* gene caused a frame shift. This deletion was located in the last exon at position 2710 in the Col-0 ORF, 258 bp upstream of the stop codon. In the Shahdara allele the frame shift caused a change in the 86 amino acids at the carboxy terminus of the deduced protein sequence. In the last exon of the *SDE3* gene a particularly large deletion of 51 bp was found in accessions Baa-1 and Ws-0.

Table 8. Indel variation of candidate genes in 25 *Arabidopsis thaliana* accessions. For all analyses the combined regions that were covered by amplicon sequences in all accessions analysed were taken into account.

	AGO1	*AGO7*	*DCL4*	*ERI*	*HEN1*	*NRPD1*	*NRPE1*	*SDE3*	*SDE5*	*SGS3*	*WEX*	*XRN4*
No. of indels in genes	11	10	12	11	17	9	23	3	3	0	24	7
Frequencies of Indels in genes (%)	0.199	0.305	0.138	0.528	0.475	0.157	0.286	0.100	0.105	0	1.244	0.108
No. of indels in exons	1	3	1	9	0	0	16	2	0	0	2	0
Frequencies of Indels in exons (%)	0.033	0.107	0.021	0.888	0	0	0.281	0.074	0	0	0.260	0
No. of microsatellites in Col-0 sequences	21	26	47	10	19	31	26	19	12	12	12	36
Microsatellite frequency (%) in Col-0	0.379	0.794	0.539	0.480	0.530	0.541	0.323	0.630	0.422	0.516	0.662	0.556
No. of Indels affecting microsatellites	4	2	3	9	15	1	11	1	0	0	3	4

Those Col-0 sequences of the twelve candidate genes that had been analysed with respect to SNPs and Indels were also screened for the presence of microsatellites using Microsatellite repeats finder (http://insilico.ehu.es/mini_tools/microsatellites/). The following parameters were used for the identification of microsatellites; for repeat units that encompassed between two and ten bases three perfectly repeated units had to be present, mononucleotide stretches were considered if they spanned at least 6 bp. Table 8 summarises the number of microsatellite loci observed for the different genes. In total, 271 microsatellite loci were observed in Col-0. Dinucleotide motifs were with 51.3% most abundant, followed by trinucleotide motifs with 20.7% and hexanucleotide motifs with 0.4%. Mononucleotide motifs represented 27.7%. It was also analysed how many of the microsatellite loci identified in Col-0 were affected by Indels that had been detected in the

25 accessions. Therefore it was evaluated whether Indels had an impact on the length of microsatellites or whether entire microsatellites were deleted. In total, 53 indels affecting microsatellite loci were identified, this corresponded to 41% of the 130 detected Indels. For seven microsatellites at least two different size variants were observed in the 25 accessions. Particularly polymorphic microsatellites were found in the *ERI*, *HEN1* and *NRPE1* genes. In the first *HEN1* intron 16 different size variants were found for a TA repeat motif. For a trinucleotide repeat motif in the first *ERI* exon 10 different size variants were observed. In the last *NRPE1* exon a "CAGTCT" repeat motif was found 17 times in Col-0, for this hexanucleotide motif ten other size variants were found in the 25 accessions (Supplementary table 8).

3.1.3 Pairwise comparisons of selected allelic variants

A subset of the accessions analysed with respect to sequence diversity showed for genes *AGO7*, *HEN1, NRPE1, SDE3* and *WEX* SNP frequencies higher than 0.5% (Table 6) and/or large Indels in exon regions when compared to Col-0 gene sequences, in case of the *NRPD1* gene this was true for all 25 accessions. At least two accessions which showed large indels in exon regions and/or which showed SNP frequencies higher than 0.5% when compared to Col-0 were selected for further analysis in order to determine if diverged alleles can affect the process of PTGS. For the *AGO7* and *HEN1* genes two accessions each met these criteria. In case of *NRPE1*, the three accessions Cvi-0, Kas-1 and Shahdara that showed SNP frequencies higher than 1% were selected, moreover for the Shahdara and Kas-1 alleles 72 and 114 bp long Indels in the last exon had been observed, respectively (Table 4). Two accessions that showed 51 bp long Indels in the last exon of the *SDE3* gene (Supplementary table 8) were chosen and in addition two accessions that showed SNP frequencies of 0.53%. For genes *NRPD1* and *WEX* only a subset of the accessions showing high SNP frequencies were considered because 25 and 19 accessions, respectively showed high SNP frequencies when compared to Col-0. In case of *NRPD1*, the same accessions were chosen for further analysis as for *AGO7* because the *NRPD1* gene maps within less than 3 Mbp of the *AGO7* gene on chromosome 1. For the *WEX* gene, accessions Kin-0, Baa-1, Ang-0 and Shahdara were selected because these accessions showed with 4.4%, 3.2%, 3.1% and 2.9%, respectively the highest SNP frequencies when compared to Col-0. Accessions Lp2-2 and Sq-8 were also included since these accessions had been chosen for the analysis of the *HEN1* gene which is located approximately 3.2 Mbp apart from the *WEX* gene on chromosome 4. In total, 13

accessions were analysed further. For six accessions, Baa-1, Bor-4, Gie-0, Lp2-2, Shahdara and Sq-8 two candidate genes each were studied (Table 9).

Table 9. Allelic variants selected for functional analysis. Alleles from accessions were chosen based on SNP frequencies (S), large Indels in exon sequences (Ind) or both (S/Ind).

Accessions	*AGO7*	*HEN1*	*NRPD1*	*NRPE1*	*SDE3*	*WEX*
Ang-0						S
Baa-1					Ind	S
Bor-4	S		S			
Cvi-0				S		
Gie-0	S		S			
Kin-0						S
Kas-1				S/Ind		
Lp2-2		S				S
Lz-0					S	
Ra-0					S	
Shahdara				S/Ind		S
Sq-8		S				S
Ws-0					Ind	

In order to assess sequence variation in the selected accessions for the entire candidate gene regions from ATG to stop codon, it was necessary to design additional oligonucleotide pairs (Supplementary table 2) for all regions in which sequences established for amplicons did not overlap. For example, two more amplicons were developed for the *NRPE1* gene, since the sequences obtained for amplicons 4, 5 and 6 did not show overlaps. The amplicon sequences that were originally designed for the *NRPE1* and *NRPD1* genes spanned the regions around the ATG and stop codons, whereas this was not the case for the *AGO7*, *HEN1*, *SDE3* and *WEX* genes. For the latter four genes it was assessed whether an ORF mapped within 1500 bp of the ORF of a particular candidate gene. In such cases amplicons were developed (Supplementary table 2) between the flanking gene and the candidate gene in order to obtain sequences around the ATG and stop codons of the candidate gene. In this manner the regions around the start and stop codons could be analysed for the *WEX* gene, for the *HEN1* and *SDE3* genes amplicons were developed that included stop codons. Sequences were established for the selected accessions with these newly developed amplicons and manually edited, if necessary. All sequences corresponding to a particular accession and candidate gene were then assembled.

Based on the alignments of the allele sequences to the Col-0 ORF sequences, the ORFs were deduced for all variants. All exon/intron borders were found to be conserved. Amino acid

sequences were deduced for all alleles by translation of the ORF sequences. The gene, ORF and amino acid sequences that had been established for the selected allelic variants were then compared to the Col-0 sequences of the candidate gene and among each other. For the comparison of the gene sequences only the regions spanning from the start to the stop codons were considered. In those cases in which the sequence assemblies of the allelic variants lacked the 5'- and/or the 3'-parts of a particular gene, the comparisons were restricted to the region which was represented in all allelic variants analysed.

Table 10. Pairwise sequence identity levels of selected *NRPE1* alleles. For the gene sequences only the regions from ATG to stop codons were taken into account. All values are given in %.

(a) Pairwise comparisons of gene sequences

Accession	Col-0	Cvi-0	Kas-1	Shahdara
Col-0	100	98.89	98.89	98.75
Cvi-0	98.89	100	99.46	99.63
Kas-1	98.89	99.46	100	99.57
Shahdara	98.75	99.63	99.57	100

(b) Pairwise comparisons of ORF sequences

Accession	Col-0	Cvi-0	Kas-1	Shahdara
Col-0	100	99.17	99.21	99.02
Cvi-0	99.17	100	99.39	99.62
Kas-1	99.21	99.39	100	99.62
Shahdara	99.02	99.62	99.62	100

(c) Pairwise comparisons of deduced amino acid sequences

Accession	Col-0	Cvi-0	Kas-1	Shahdara
Col-0	100	99.22	98.45	98.72
Cvi-0	99.22	100	99.12	99.64
Kas-1	98.45	99.12	100	99.12
Shahdara	98.72	99.64	99.12	100

Comparison of the *NRPE1* gene sequences (Table 10 (a)) revealed that the Cvi-0, Kas-1 and Shahdara sequences were approximately 99% identical when compared to Col-0. When Cvi-0, Kas-1, and Shahdara allele sequences were compared to each values ranging from 99.5 to 99.6%, were found, indicating that all three alleles were more similar to each other than to Col-0. Nucleotide identities were also analysed for the ORF and the deduced amino acid level (Table 10 (b) and (c)). For the ORF, nucleotide identity values between accessions Cvi-0, Kas-1 and Shahdara varied from 99.4% to 99.6%. Comparisons at the amino acid level showed sequence identities between 99.1% and 99.6%. As observed for the comparisons of the gene

sequences, the ORF and amino acid sequences of the selected accessions were more similar to each other than to Col-0.

Table 11. Pairwise identity levels of selected *WEX* alleles. For the gene sequences only the regions from ATG to stop codons were taken into account. All values are given in %.

(a) Pairwise comparisons of gene sequences

Accession	Col-0	Ang-0	Baa-1	Kin-0	Lp2-2	Shahdara	Sq-8
Col-0	100	96.72	96.67	95.73	97.17	97.11	97.22
Ang-0	96.72	100	99.95	94.86	98.54	98.48	98.37
Baa-1	96.67	99.95	100	94.80	98.48	98.43	98.32
Kin-0	95.73	94.86	94.80	100	95.36	95.31	95.14
Lp2-2	97.17	98.54	98.48	95.36	100	99.84	98.27
Shahdara	97.11	98.48	98.43	95.31	99.84	100	98.21
Sq-8	97.22	98.37	98.32	95.14	98.27	98.21	100

(b) Pairwise comparisons of ORF sequences

Accession	Col-0	Ang-0	Baa-1	Kin-0	Lp2-2	Shahdara	Sq-8
Col-0	100	97.54	97.54	98.02	98.25	98.14	97.67
Ang-0	97.54	100	100	97.54	98.83	98.71	98.13
Baa-1	97.54	100	100	97.54	98.83	98.71	98.13
Kin-0	98.02	97.54	97.54	100	98.02	97.90	97.32
Lp2-2	98.25	98.83	98.83	98.02	100	99.88	98.02
Shahdara	98.14	98.71	98.71	97.90	99.88	100	97.90
Sq-8	97.67	98.13	98.13	97.32	98.02	97.90	100

(c) Pairwise comparisons of the deduced amino acid sequences

Accession	Col-0	Ang-0	Baa-1	Kin-0	Lp2-2	Shahdara	Sq-8
Col-0	100	96.14	96.14	96.50	96.50	96.15	97.55
Ang-0	96.14	100	100	96.14	98.60	98.25	97.19
Baa-1	96.14	100	100	96.14	98.60	98.25	97.19
Kin-0	96.50	96.14	96.14	100	96.15	95.80	96.50
Lp2-2	96.50	98.60	98.60	96.15	100	99.65	97.20
Shahdara	96.15	98.25	98.25	95.80	99.65	100	96.85
Sq-8	97.55	97.19	97.19	96.50	97.20	96.85	100

Table 11 shows the pairwise identity levels for selected alleles of the *WEX* gene. At the gene level, the Kin-0 allele shows most sequence differences when compared to Col-0. Nucleotide identity is with 95.7% approximately 1% lower than those values that were observed when the allele sequences of the other five selected accessions were compared to Col-0. The pairwise comparisons revealed that the *WEX* alleles of accessions Ang-0 and Baa-1 were almost identical to each other, likewise those of Lp2-2 and Shahdara. Most other pairwise comparisons between the alleles of the selected accessions revealed nucleotide identity values of around 98%, however, the pairwise comparisons with the Kin-0 allele showed lower values of approximately 95%. At the ORF level, the pairwise comparisons displayed nucleotide identity levels close to 98% when the alleles of the six selected accessions were

compared to Col-0. When the Kin-0 and Sq-8 alleles were compared to those of Ang-0, Baa-1, Lp2-2 and Shahdara values of approximately 98% were observed. With the exception of the identical Ang-0 and Baa-1 ORFs and the almost identical ORFs of Lp2-2 and Shahdara (99.9%), all remaining pairwise comparisons between the alleles of the selected accessions yielded identity levels close to 99%. At the amino acid level, these numbers decreased to approximately 96% when the sequences of accessions Ang-0, Baa-1, Kin-0, Lp2-2 and Shahdara were compared to Col-0, the deduced amino acid sequence of Sq-8 was with 97.6% more similar to that of Col-0. Consistent with the data for the ORF sequences the pairwise comparisons of the amino acid sequences of the selected accessions revealed values of 96% and 97% when the Kin-0 and Sq-8 sequences were analysed, respectively. Comparisons involving the other accessions revealed values of approximately 98%, exceptions were noted for the identical sequences of Ang-0 and Baa-1 and the very similar amino acid sequences of Lp2-2 and Shahdara.

In case of *AGO7*, the assemblies of the gene sequences for the different accessions were missing the sequences at the 5'- and 3'-end of the coding region, therefore the comparisons were restricted to the regions which were represented in the sequences established for Bor-4 and Gie-0. The gene, ORF and amino acid sequences of the Bor-4 and Gie-0 alleles were identical to each other and showed 96.8%, 96.9% and 98.0% sequence identity, respectively when compared to Col-0.

The amplicon sequences established for the *NRPD1* gene included the ATG and the stop codon, hence the region from start to stop codon was used for the sequence alignments. The Bor-4 and Gie-0 alleles of *NRPD1* were between 99.8 and 99.9% identical when analysed with respect to gene, ORF and amino acid sequences. However, when compared to Col-0, lower values were obtained, ranging from 99.3% to 99.6%.

The sequences assembled for the selected *SDE3* alleles did not contain the 5'-end of the coding sequence but all of them harboured the region around the stop codon, hence comparisons were confined to the region which was represented in all alleles up to the stop codon. Comparisons of the *SDE3* gene sequences established for accessions Baa-1, Lz-0, Ra-0, and Ws-0 revealed two groups. On the one hand alleles of Baa-1 and Ws-0 were identical to each other at sequence level, on the other hand identical *SDE3* sequences were also

found for accessions Lz-0 and Ra-0. In comparison to the Col-0 gene sequences, sequence identity values of 99.8% and 99.5% indicated that the accessions in the first group represented by Baa-1 and Ws-0 were more similar to Col-0 than those of the second one comprising Lz-0 and Ra-0, respectively. Pairwise comparisons of allele sequences that belonged to the different groups yielded sequence identity values of 99.3%. Very similar values were obtained when the ORF sequences were compared. At the amino acid level sequences belonging to the first and second group were 99.9% and 99.6% identical when compared to Col-0, respectively. If sequences of accessions were analysed that represented the two different groups a value of 99.7% was found.

The sequence assemblies which were established for the *HEN1* alleles of the Lp2-2 and Sq-8 accessions missed the 5'-end of the ORF, thus the comparisons were confined to the region which was covered by amplicon sequences of these accessions up to the stop codon. In case of the Sq-8 allele it was not possible to determine the exact length of the TA microsatellite that was present in the first intron. The sequence identity values of Sq-8 and Lp2-2 alleles were close to 99% when compared to Col-0 at the level of the gene and ORF sequences. Sequence identity of 99.8% between the Sq-8 and Lp2-2 alleles indicated that these alleles were more similar to each other than to the Col-0 allele, regardless whether gene or ORF sequences were compared. The Sq-8 and Lp2-2 amino acid sequences were 99.6% identical, whereas values of 98.5% and 98.7% resulted when the sequences of Sq-8 and Lp2-2 were used in the comparisons, respectively.

3.2 Functional analysis of selected allelic variants

Accessions in which large indels were found in exon regions of the candidate genes and/or which had substantially diverged sequence variants of candidate genes compared to the reference accession Col-0 were selected for functional analysis (Table 9). The steady-state transcript levels of the chosen 19 allelic variants were investigated by RT-PCR and/or qRT-PCR experiments. Furthermore, the 19 allelic variants were introgressed into Col-0 transgenic lines carrying *GFP* transgenes. For all established introgression lines number, length and position of introgressions were determined. The introgression lines were then analysed with respect to *GFP* silencing alongside transgenic lines that carried *GFP* transgenes in the Col-0

genetic background in order to evaluate whether and to which extent allelic variants and/or chromosome regions originating from selected accessions may affect *GFP* silencing.

3.2.1 Gene expression analysis of selected allelic variants

It was evaluated for all candidate genes used in this study whether and which of the exon/intron structures available in TAIR (https://www.arabidopsis.org) were supported by cDNA sequences. BLAST analyses (NCBI, http://www.ncbi.nlm.nih.gov; TAIR BLAST (version 2.2.8)) revealed that all but one of the exon/intron structures were supported by full-length cDNA sequences. For eight out of the twelve genes of interest *AGO1*, *AGO7*, *DCL4*, *ERI*, *NRPD1, NRPE1*, *SDE3* and *SGS3* one full-length cDNA sequence each was found. Two genes, *HEN1* and *WEX,* were represented by two full-length cDNA sequences each, and for the *XRN4* gene four full-length cDNA sequences were available. Furthermore, cDNAs which only confirmed parts of the ORF region or which lacked 5'-UTR and/or 3'-UTR sequences were also found (Supplementary table 9).

For candidate gene *SDE5* only the 5'-UTR together with exons 1 and 2 was covered by a cDNA sequence. Amplicons were developed (Supplementary table 3) based on the predicted exon/intron structure of the *SDE5* gene in TAIR (https://www.arabidopsis.org) in order to validate the remainder of predicted gene structure experimentally. RNA extracted from flower tissues of reference accession Col-0 served as template for RT-PCR amplifications. All resulting RT-PCR products were sequenced, manually edited if necessary, and aligned to the Col-0 sequences of the *SDE5* gene and the predicted ORF. In this manner all exon/intron boundaries predicted for the *SDE5* gene were confirmed (data not shown).

In order to assess the expression of those candidate genes for which allelic variants had been chosen for further analysis, RT-PCR amplifications were carried out with cDNAs generated from RNA extracted from seedlings. Aerial tissues of plants were harvested ten days after sowing for the 13 accessions that had been selected for further analysis (Table 9) as well as for reference accession Col-0. Since some of the allelic variants showed substantial sequence divergence, it was taken care that gene-specific oligonucleotide pairs suitable for RT-PCR and qRT-PCR were placed in monomorphic regions. The sequences of all oligonucleotides corresponding to the different RT-PCR and/or qRT-PCR amplicons are listed in Supplementary table 3. Amplifications products obtained for the *ACTIN2* (*At3g18780*) gene served as

reference. For each accession except Ra-0, two independent RNA preparations were analysed. Expression in seedling tissues was observed for all genes of interest, regardless which accession was analysed (Figure 7).

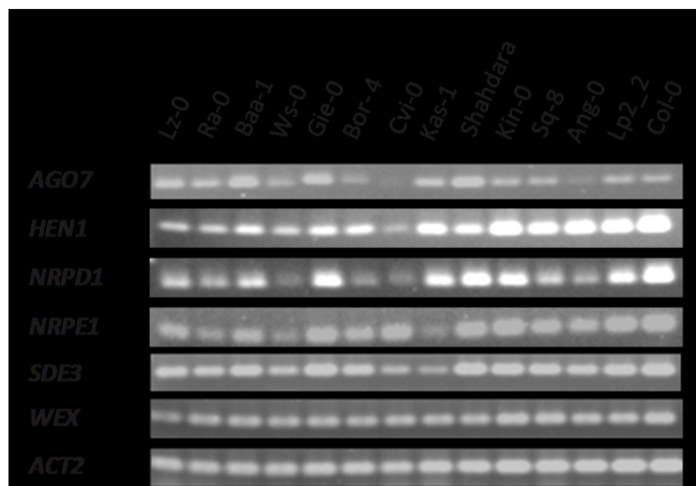


Figure 7. RT-PCR experiments reveal expression of selected candidate genes in aerial seedling tissues. PCR products obtained after 35 amplification cycles were analysed using gel electrophoretic separation. *ACTIN-2* served as reference gene.

Quantitative real-time PCR (qRT-PCR) was performed to refine the RT-PCR results for candidate genes *AGO7*, *HEN1*, *NRPE1*, *SDE3* and *WEX*. It was assessed whether any of the allelic variants that had been selected for functional studies showed an altered expression level in aerial seedling tissues when compared to that of the corresponding alleles in the reference accession Col-0. Several genes for which expression stability had been demonstrated in a study of Czechowski *et al*. (2005) were considered as reference genes (Supplementary table 3). From this set, *At4g34270* was selected since its expression level was in a similar range as that of the candidate genes. Amplification efficiencies of all oligonucleotide pairs were evaluated and only those amplicons, which showed efficiencies between 90% and 110% were chosen for qRT-PCR experiments. The $2^{-\Delta\Delta Ct}$ method was used to compare expression of the candidate genes in different accessions (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008; Bolha *et al*., 2012). For each of the accessions three biological replicates were analysed for the expression of a particular candidate gene and *At4g34270*. The resulting values were normalised to the level of the reference gene and the median normalised expression value of accession Col-0 was set to 1, median values of the three biological replicates were used in all comparisons unless indicated otherwise.

For the analysis of each candidate gene apart from the *WEX* gene, one accession, which carried an allele that was identical or very similar at sequence level to the Col-0 allele was also included (Supplementary table 8). Accessions Bor-4, Kas-1, Kin-0 and Sapporo-0 were selected for the analysis of candidate genes *SDE3, AGO7*, *NRPE1* and *HEN1*, respectively.

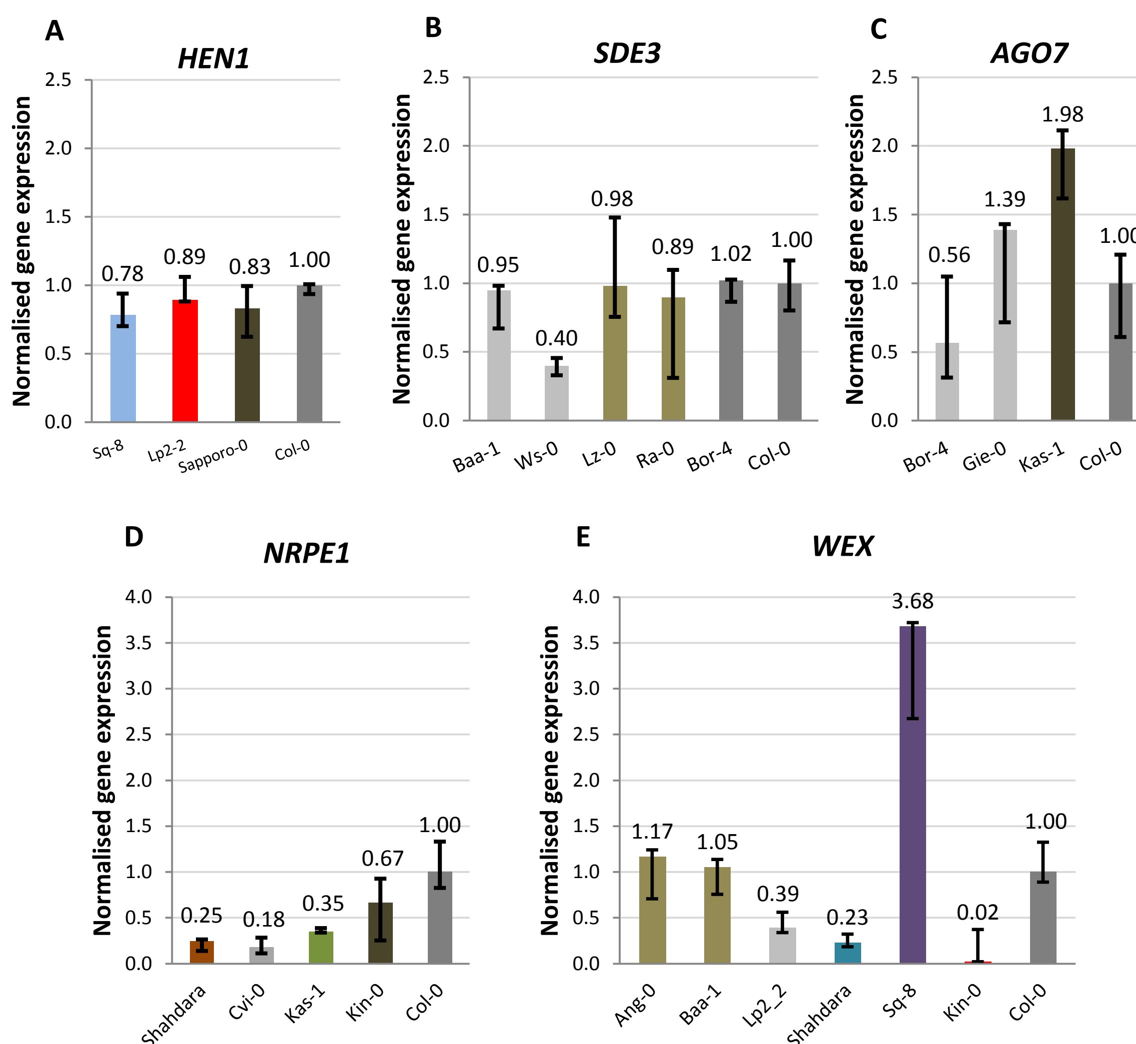


Figure 8. Expression analysis of *HEN1, SDE3, AGO7, NRPE1* and *WEX* genes in selected accessions. All expression values were normalised to those determined for the *At4g34270* gene. The bars indicate the deviation of two biological replicates from the median value. Expression values obtained for reference accession Col-0 were set to 1 in all comparisons. The bars of accessions that have identical sequences in the coding region are represented in the same colour.

The results of the expression analysis for candidate genes *HEN1*, *SDE3*, *AGO7, WEX* and *NRPE1* are shown in Figure 8. In case of the *HEN1* gene, very similar expression values were observed for accessions Col-0 and Sapporo-0 and those carrying the selected allelic variants, Lp2-2 and Sq-8 (Figure 8A). The normalised median expression values did not differ by more

than a factor of 1.3. In case of the *SDE3* and *AGO7* genes identical sequences had been observed for some of the alleles, nonetheless in some cases more than twofold differences with respect to expression were observed. For example, accessions Baa-1 and Ws-0 had the same haplotype in the *SDE3* coding region, but the median expression level of the alleles differed by a factor of 2.4. In case of the *AGO7* gene, a 2.5-fold expression difference was observed for accessions Bor-4 and Gie-0 that did not differ in the coding region. The Kas-1 allele showed only one SNP in the *AGO7* coding region when compared to Col-0, nonetheless expression of this gene differed by a factor of two in these two accessions (Figure 8C). The normalised expression values of the Baa-1, Lz-0 and Ra-0 allelic variants of the *SDE3* gene did not differ by more than a factor of 1.1 when compared to the values observed for reference accession Col-0 and Bor-4 which is identical in the coding region to Col-0.

All three accessions which carried allelic variants of the *NRPE1* gene showed lower expression values for this gene when compared to Col-0 (Figure 8D). *NRPE1* expression in the Cvi-0, Kas-1 and Shahdara accessions was approximately 5.6-fold, 4-fold and 2.9-fold lower than in Col-0, respectively. Analysis of the *WEX* gene revealed similar expression values for accessions Ang-0, Baa-1 and Col-0. *WEX* expression values of these three accessions did not differ more than 1.2-fold (Figure 8E). In contrast, in accessions Lp2-2, Shahdara and Kin-0 the *WEX* gene was not expressed as highly as in Col-0. Col-0 showed 2.6-fold and 4.3-fold higher *WEX* expression than Lp2-2 and Shahdara, respectively. For the Kin-0 allelic variant 50-fold lower expression of the *WEX* gene than in the reference accession Col-0 was observed if the median values were considered, however one of the biological replicates showed a less drastic reduction of 2.7-fold. The only accession in which considerable more expression of the *WEX* gene was found than in Col-0 was Sq-8, normalised expression in this accession differed by a factor of 3.7 from that of the reference accession.

3.2.2 Generation of introgression lines

Five Col-0 transgenic lines that had been established previously were used in this study; these lines carried six or eight *GFP* reporter gene copies under the control of the CaMV 35S promoter. Loci *F8*, *F18* and *F128* represent single-copy T-DNA insertions and carry one *GFP* transgene each (Schubert *et al*., 2004) whereas locus *R127* contains two T-DNA copies in an inverted repeat orientiation, it harbours two copies of the *GFP* transgene (Lechtenberg *et al*.,

2003). These loci map to chromosomes 1 and 3 (Forsbach *et al*., 2003, Lechtenberg *et al*., 2003, Figure 9A). The lines with eight *GFP* copies contained the *R127* locus homozygously and additionally either *F8* and *F18* or *F8* and *F128* in the homozygous fashion. The lines harbouring six copies were also homozygous for *R127* and contained either the *F8*, the *F18* or the *F128* locus homozygously (Arlt *et al*., 2007; Thanh Loan Le, unpublished results).

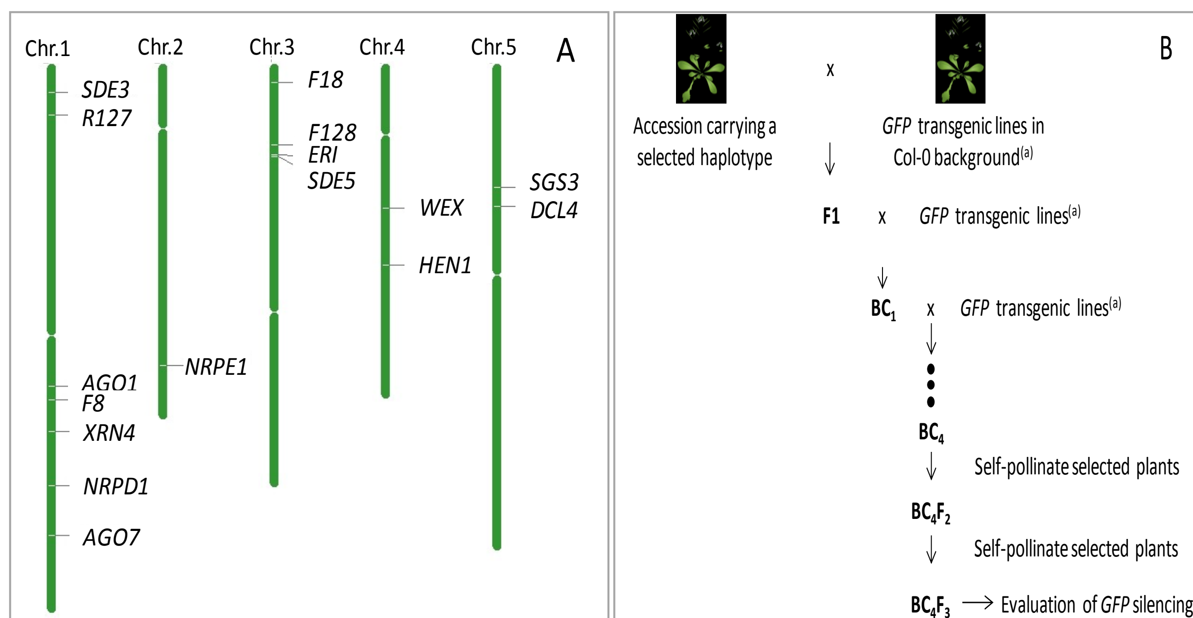


Figure 9. Map position of the candidate genes and *GFP* loci on the five chromosomes of *Arabidopsis thaliana* and crossing scheme for the generation of introgression lines. (A) The chromosome sequence maps show the map positions of the T-DNA-loci and of the candidate genes used in this study. (B) Plants derived from a cross between an accession carrying a selected allelic variant and the *GFP* transgenic lines were backcrossed four times to *GFP* containing transgenic lines. Selected BC_4 plants were self-pollinated. In the BC_4F_2 generation plants were selected which carried the allelic variants of interest as well as two *GFP* loci in the homozygous fashion. Plants of the BC_4F_3 generation were evaluated for the initiation and spread of silencing. [(a)]: Transgenic lines contained six or eight copies of *GFP*.

To generate introgression lines (ILs), the F_1 plants, derived from a cross between a selected *A. thaliana* accession and a transgenic line carrying six or eight copies of the *GFP* transgene in the Col-0 background, were backcrossed four times consecutively to *GFP*-containing transgenic lines until plants of the BC_4 generations were obtained (Figure 9B). The transgenic lines served as pollinators. If possible, the transgenic lines were selected such that the map positions of the *GFP* loci and of the candidate genes were genetically not closely linked.

Plants were screened for the presence and zygosity of the different *GFP* loci in each generation. To determine the presence of the different T-DNA loci, amplicons were used that consisted of one oligonucleotide specific for T-DNA sequences and another one

matching the *A. thaliana* sequences flanking the T-DNA at a particular locus (Figure 2, Supplementary table 6). Using these amplicons it was not possible to discriminate plants that were hemi- or homozygous for the different T-DNA loci. Zygosity of a particular T-DNA locus was established with the help of a second amplicon that spanned across the insertion site or empty donor site (EDS) of the T-DNA in the *A. thaliana* genome. Using standard PCR conditions amplification products were only obtained if plants did not contain the T-DNA locus at all or if the plants were hemizygous for it. The results obtained for the amplicons that were specific for a particular T-DNA locus in combination with those for the amplicons specific for the empty donor site permitted to determine unambiguously whether a plant was hemi- or homozygous for a particular T-DNA locus or whether it did not contain it at all (Figure 10A).

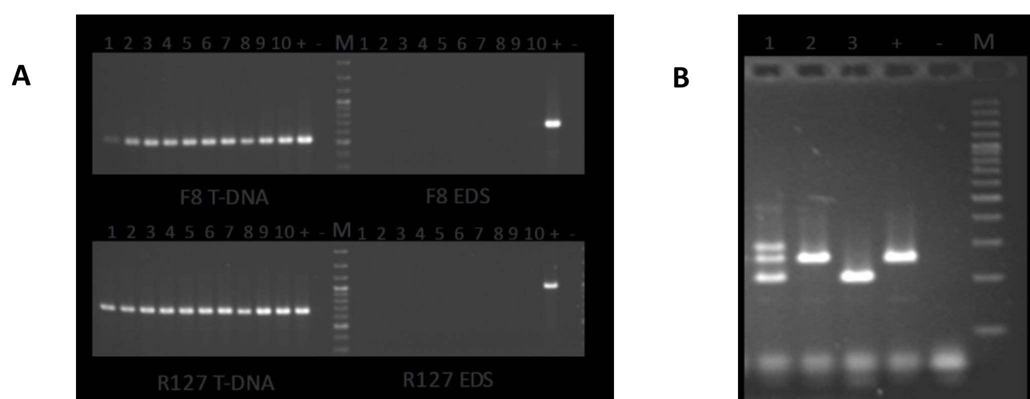


Figure 10. Evaluation of introgression lines for the presence and zygosity of T-DNA loci and alleles of interest. (A) The presence of T-DNA loci *F8* and *R127* in plants was analysed with oligonucleotide pairs *F8* T-DNA and *R127* T-DNA. These consisted of an oligonucleotide matching T-DNA sequences (LB1c) and an oligonucleotide specific for *A. thaliana* sequences flanking the T-DNA at loci *F8* and *R127*, respectively (Figure 2, Supplementary table 6). The presence of amplification products revealed that plants 1-10 contained T-DNA loci *F8* and *R127*. Using oligonucleotide pairs which matched *A. thaliana* sequences that flank a particular T-DNA locus on both sides, F8 EDS and R127 EDS, products were only obtained with standard PCR amplification conditions if the plants did not contain this particular T-DNA locus or if plants were hemizygous for the T-DNA locus. For plants 1-10 products were not observed using these oligonucleotide pairs, this indicates that plants 1-10 contained loci *F8* and *R127* homozygously. In the lanes labeled (+) and (-) the results for appropriate control templates are shown. (B) Lanes 1-3 show the amplification results for an Indel marker specific for the *SDE3* gene using template DNA of plants that carried the Col-0 allele homozygously (2), the Ws-0 *SDE3* allele containing a 51 bp deletion homozygously (3) and both alleles (1). DNA of Col-0 wild-type plants was used as control (+).

The presence of allelic variants of interest was determined in each generation either with the help of allele-specific amplicons or with amplicons that were suitable to discriminate allelic variants of interest from the Col-0 reference allele based on size differences (Supplementary table 5). For plants of the BC_4F_2 generation it was not only necessary to analyse the presence of a particular variant, but also its zygosity. This was readily possible

with the Indel markers that had been developed to discriminate allelic variants of interest from the Col-0 reference allele (Figure 10B). In cases in which the presence of allelic variants was analysed with allele-specific amplicons, zygosity was determined with the help of a second amplicon that specifically amplified the Col-0 allele. The results for both amplicons taken together indicated whether a particular plant carried the allelic variant of interest or the Col-0 allele homozygously, or both alleles (data not shown).

It was the aim to identify in the BC_4F_2 generation plants that contained the allelic variant of interest homozygously and in addition six copies of the *GFP* gene. In order to avoid segregation with respect to the T-DNA loci in the BC_4F_3 generation it was necessary to identify plants that contained both locus *R127* and one of the single-copy T-DNA loci homozygously.To reduce the screening effort that was necessary to obtain plants that were homozygous for three different loci it was the aim to identify plants homozygous for at least one of the T-DNA loci already in the BC_4 generation.

Using the scheme described above, plants which carried the allele of interest and two *GFP* loci homozyously were in most of the cases readily identified in the BC_4F_2 generation. However, in the case of the *SDE3* gene which maps about 1.3 Mbp apart from the *R127* locus on chromosome 1 (Figure 9A) many plants had to be screened in order to find introgression lines which carried the candidate allele of interest and the *R127* locus in the homozygous fashion. Candidate genes *NRPD1* and *AGO7* map about 2.8 Mbp apart on chromosome 1 (Figure 9A). In order to find lines, which have only one of the two candidate alleles introgressed into Col-0 large plant populations had to be analysed as well. In few instances it was only possible to find suitable lines in later generations. At least two introgression lines each were established for each of the 19 allelic variants of interest (Table 9).

3.2.3 Evaluation of molecular markers for indel polymorphisms in *Arabidopsis thaliana* accessions

Molecular markers were used to evaluate the different introgression lines for the relative contributions of the Col-0 genome and that of the accession carrying the allelic variant of interest. Large Indel marker collections are available for *A. thaliana* (Loudet *et al*., 2002; Salathia *et al.,* 2007; Hou *et al.,* 2010; Păcurar *et al*., 2012), but for many markers polymorphism information was only available for few genotypes.

To assess whether markers showed insertion or deletion polymorphisms in the accessions carrying the allelic variants of interest when compared to Col-0, PCR amplifications with 146 Indel markers were performed using DNA of Col-0 and of the thirteen accessions of interest as templates. The sizes of the resulting products were evaluated after agarose gel electrophoresis. For each marker it was evaluated whether the different accessions showed fragment sizes larger or smaller than Col-0 (Table 12) or whether one or several accessions repeatedly failed to yield amplification products. Only size differences that were clearly discernible on 2% agarose or alternatively on 3% Nusieve 3:1 gels were taken into account. The oligonucleotide sequences of the 146 Indel markers used in this study are listed in Supplementary table 4 together with their map positions.

Table 12. Screening of Indel markers. The letters "a", "b", and "c" indicate that the size of amplification products is very similar, longer or shorter when compared to Col-0, respectively. The letter "n" indicates that amplification products were repeatedly not obtained.

Accession	IND I_21	F11P17-4615	1-7539	IND I_24	F5I14-IND	1-8645	UPSC 1-26627	IND I_27	ATHATPASE	UPSC 1-29617
Col-0	a	a	a	a	a	a	a	a	a	a
Ang-0	c	c	c	a	b	c	b	a	c	a
Gie-0	a	c	c	a	b	c	b	a	c	b
Shahdara	c	c	a	c	b	a	b	a	c	b
Sq-8	c	c	c	n	b	c	b	a	c	b
Kin-0	a	c	c	c	b	c	b	a	c	a
Bor-4	a	a	c	a	b	a	b	a	c	b
Kas-1	c	c	c	a	b	a	b	a	c	b
Cvi-0	a	b	c	c	b	a	b	a	c	b
Lp2-2	a	c	c	c	b	c	b	b	c	b
Baa-1	c	c	c	n	a	c	b	a	c	a
Lz-0	c	c	c	a	b	c	b	a	c	a
Ra-0	c	c	c	a	b	c	b	a	c	a
Ws-0	c	c	c	c	b	c	b	b	c	a

Only five out of the 146 markers tested were monomorphic in all accessions tested, the remainder of 141 markers showed for one or more accessions clearly discernible size polymorphisms when compared to Col-0 (Supplementary table 10). The accession Shahdara revealed with 112 (77%) the highest number of polymorphic indel markers, whereas the lowest numbers were observed for accessions Ra-0 and Baa-1 with 88 (60%) and 89 (61%) markers, respectively (Table 13).

Table 13. Number of polymorphic Indel markers identified for selected accessions. The total number is given together with the data for the individual chromosomes.

Accessions	Chr. 1	Chr. 2	Chr. 3	Chr. 4	Chr. 5	Total
Ang-0	29	22	17	21	16	105
Baa-1	25	20	12	20	12	89
Bor-4	24	21	14	21	17	97
Cvi-0	26	22	14	22	14	98
Gie-0	26	26	15	18	14	99
Kas-1	29	25	13	23	17	107
Kin-0	31	24	16	21	13	105
Lp-2_2	27	22	17	19	17	102
Lz-0	27	22	12	19	13	93
Shahdara	30	25	17	23	17	112
Sq-8	30	24	16	16	16	102
Ra-0	26	21	12	17	12	88
Ws-0	27	22	15	21	14	99
Number of analysed Indel markers	**40**	**31**	**25**	**29**	**21**	**146**

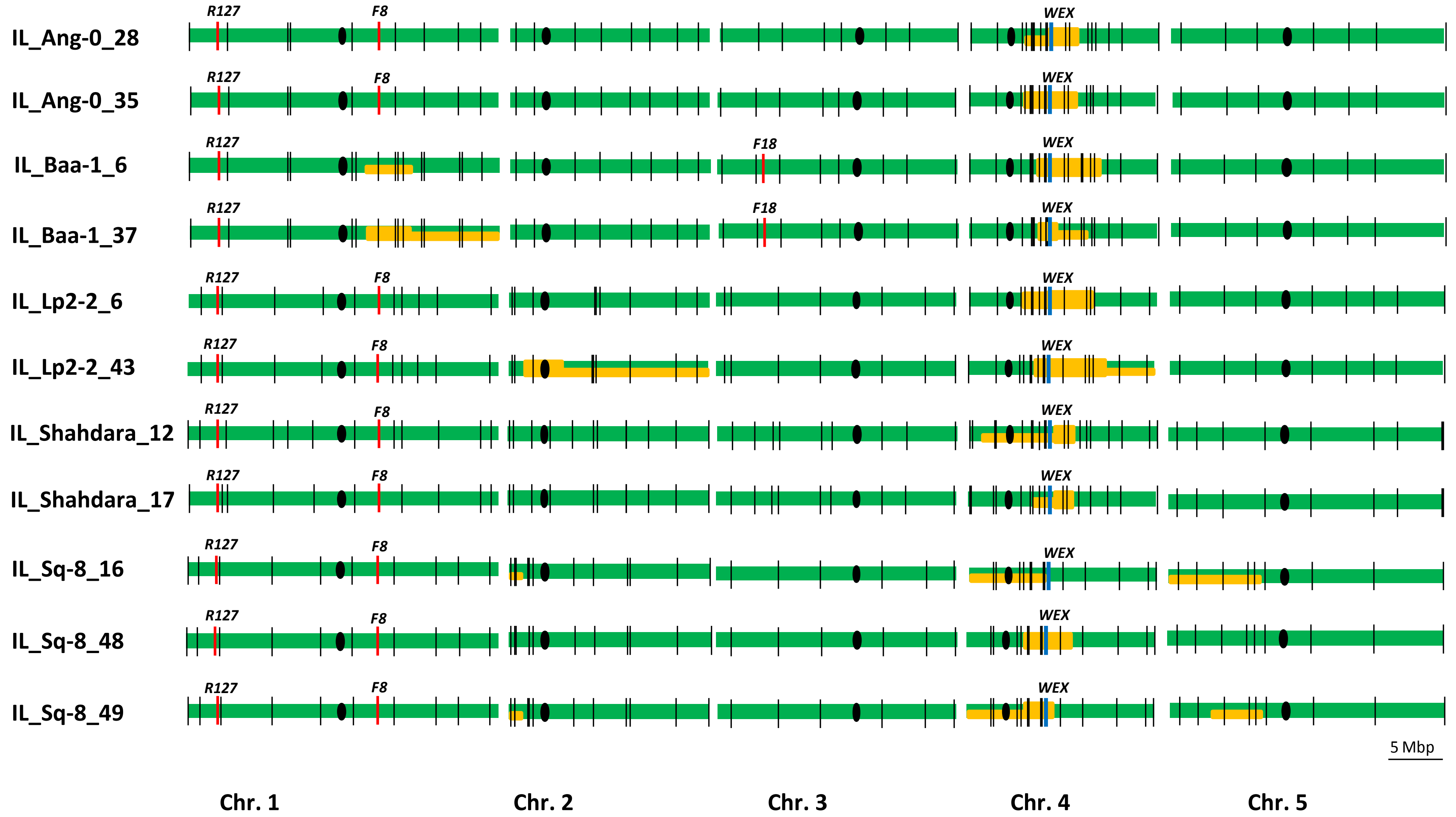
3.2.4 Characterisation of the introgression lines with respect to number, length and position of introgressed segments

The analysis of the 51 introgression lines with allele-specific markers or Indel markers corresponding to the candidate genes showed that the allelic variants of interest had been successfully introgressed into the Col-0 background (Section 3.2.2). Each introgression line was then characterised with Indel markers to identify number, size and position of the introgressed segments. Indel markers spaced every 4 to 5 Mbp were used, if possible. Since the number of polymorphic indel markers was for selected accessions rather low in some regions of the *A. thaliana* genome the spacing was more sparse in few regions of the genome, for example in the distal region of the long arm of chromosome 5 it was nearly 7 Mbp for 10 out of the 13 accession that were analysed (Figure 11, Supplementary figures 2-5). A higher density of markers was used, if available, for the genomic regions containing the candidate gene locus. For the regions for which breakpoints between introgressed segments and regions corresponding to the Col-0 genome had been observed, all available polymorphic markers were used for the analysis in order to determine the breakpoints as precisely as possible. In case the locus of the candidate gene itself was flanking a breakpoint amplicons corresponding to the candidate gene were sequenced to confirm that the breakpoint was not residing in the candidate gene.

Between 44 and 65 Indel markers were used to analyse the different lines. For 22 (43%) introgression lines a single introgression segment was observed, for the remaining 29 (57%) lines between two and four segments were detected. In total, 94 introgressed segments were found in the 51 ILs with Indel markers. Based on the analysis with Indel markers, 42 (45%) and 24 (26%) segments were entirely homozygous and hemizygous with respect to the accession genome, respectively. For the remaining 28 (30%) cases homozygous segments and heterozygous regions were found next to each other. Estimations of the sizes of the introgression segments were based on the assumption that the breakpoints were present midway between the position of markers immediately flanking the breakpoints. The sizes of the detected introgressed segments ranged from a minimum of 0.4 Mbp to a maximum 18.1 Mbp (Table 14). Estimated sizes smaller than 5 Mbp were found for 45 (48%) introgression segments. For 25 (27%) segments sizes ranged between 5 and 10 Mbp and 24 (26%) segments were larger than 10 Mbp.

Based on the position of the markers on the *A. thaliana* chromosome sequence maps and the genotype scores for the Indel markers used, the positions and sizes of the introgressed segments were indicated on the five chromosomes for all introgression lines analysed (Figure 11, Supplementary figures 2-5). The graphical genotypes are displayed such that the breakpoints are indicated approximately midway between the markers immediately flanking the breakpoint. The chromosome maps for the introgression lines revealed for 48 (94%) out of the 51 lines an introgression segment in the chromosome region where the candidate gene resides in the Col-0 genome. Exceptions were only noted for three introgression lines carrying the Kin-0 allele of the *WEX*-gene, IL_Kin-0_3, IL_Kin-0_12 and IL_Kin-0_14 (Figure 11).

The introgression lines carry different combinations of the *GFP* T-DNA loci. The ILs established for the allelic variants of the *NRPE1* and *HEN1* genes contained loci *F8* and *R127*, whereas all introgression lines carrying allelic variants of the *NRPD1* and/or *AGO7* genes harboured loci *F128* and *R127*. Out of the nineteen lines which possess allelic variants of the *WEX* gene 17 share the locus combination *F8/R127*, the two exceptions derived from accession Baa-1 carry loci *F18* and *R127*. Among the introgression lines containing allelic variants of the *SDE3* gene, five, three and two lines carry locus combinations *F8/R127*, *F128/R127* and *F18/R127*, respectively (Figure 11, Supplementary figures 2-5).


IL_Ang-0_28
IL_Ang-0_35
IL_Baa-1_6
IL_Baa-1_37
IL_Lp2-2_6
IL_Lp2-2_43
IL_Shahdara_12
IL_Shahdara_17
IL_Sq-8_16
IL_Sq-8_48
IL_Sq-8_49
R127
F8
F18
WEX
5 Mbp
Chr. 1
Chr. 2
Chr. 3
Chr. 4
Chr. 5

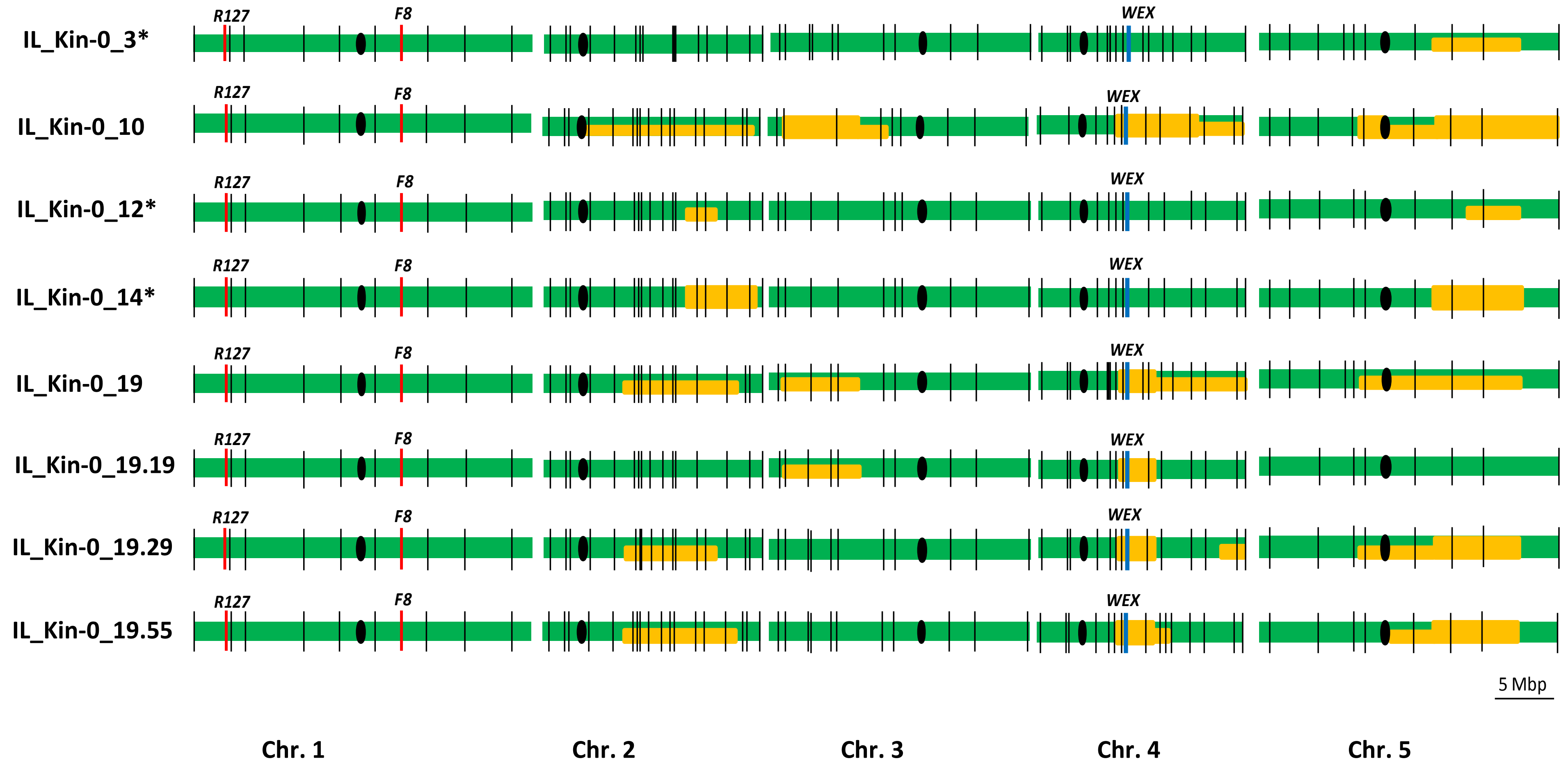


Figure 11. Characterisation of introgression lines containing allelic variants of the *WEX* gene. Black circles indicate the centromeres. The bold blue, red and black bars mark the locations of the candidate genes, the *GFP* loci and the Indel markers, respectively. Columbia-0 genome regions are shown in green colour and yellow colour indicates segments of the other accessions. For the introgression lines labelled with an asterisk the homozygous presence of the allelic variant of the *WEX* gene had been confirmed but Indel markers flanking the candidate gene did not reveal an introgression segment. For IL_Sq-8_16 a heterozygous introgression segment was found in the area of the genome where the *WEX* gene maps in Col-0, but homozygosity of the Sq-8 allelic variant had been proven using an Indel marker.

Table 14. Characterisation of introgressed segments. Only segments detected with Indel markers are shown. Asterisks indicate that markers specific for the allelic variant confirm the presence of the allelic variant, but that Indel markers which map in the vicinity in the candidate gene in the Col-0 genome do not reveal an introgressed segment. The sizes of the introgressed segments were estimated by assuming that breakpoints occurred midway between markers flanking the breakpoints.

Candidate gene(s)	Introgression line	Number of Indel markers used for analysis	Number of introgression segments detected with Indel markers	Length of segments (Mbp)
HEN1	IL_Lp2-2_27	51	1	10.2
	IL_Lp2-2_30	51	1	8
	IL_Sq-8_6	55	2	11.7/3.1
	IL_Sq-8_7	55	2	10.2/2.4
	IL_Sq-8_8	55	2	10.2/2.4
NRPD1	IL_Bor-4_35.33.3	49	3	5.1/4.7/0.5
	IL_Bor-4_35.2.9	49	1	6.5
	IL_Bor-4_35.33.10	49	3	5.1/4.7/0.5
	IL_Bor-4_35.2.11	49	1	6.5
NRPD1/AGO7	IL_Bor-4_35.41.24.12	48	3	9/2.6/0.5
	IL_Bor-4_35.41.24.50	48	2	9/0.5
	IL_Gie-0_3a	46	2	4/3.3
	IL_Gie-0_6(6.18)	46	2	4/3.3
NRPE1	IL_Cvi-0_6(6.25)	48	1	17.5
	IL_Cvi-0_19.27	48	1	14.4
	IL_Cvi-0_39	48	1	17.5
	IL_Kas-1_18	46	1	4.3
	IL_Kas-1_32	44	1	4.3
	IL_Kas-1_39	46	1	4.3
	IL_Shahdara_6	58	2	16.7/12.2
	IL_Shahdara_10	59	2	17.9/12.2
	IL_Shahdara_19.30	57	2	13.3/9.5
SDE3	IL_Baa-1_9	56	1	2.3
	IL_Baa-1_21	56	1	2.3
	IL_Lz-0_15	55	1	2.5
	IL_Lz-0_20	54	2	3.5/2.5
	IL_Lz-0_38	59	2	2.5/0.4
	IL_Ra-0_26	45	1	1.7
	IL_Ra-0_51	45	1	1.7
	IL_Ws-0_11	65	2	10.3/2.6
	IL_Ws-0_23	57	3	2.6/2.5/2
	IL_Ws-0_24	64	3	10.3/3.5/2.6
WEX	IL_Ang-0_28	49	1	5.2
	IL_Ang-0_35	49	1	5.2
	IL_Baa-1_6	55	2	6.3/4.5
	IL_Baa-1_37	55	2	13.1/4.9
	IL_Kin-0_3	65	1*	7.9
	IL_Kin-0_10	59	4	18.1/15.7/11.4/9.6
	IL_Kin-0_12	60	2*	4.8/2.6
	IL_Kin-0_14	60	2*	7.9/6.2
	IL_Kin-0_19	65	4	14.6/11.4/9.9/7
	IL_Kin-0_19.19	58	2	7/3.2
	IL_Kin-0_19.29	61	4	14.6/8.2/3.2/2.2
	IL_Kin-0_19.55	61	3	11.8/9.9/4.5
	IL_Lp2-2_6	48	1	8.2
	IL_Lp2-2_43	48	2	18.1/12.3
	IL_Shahdara_12	62	1	9
	IL_Shahdara_17	56	1	3.8
	IL_Sq-8_49	50	3	8.5/4.9/1.3
	IL_Sq-8_16	49	3	9/8.5/1.3
	IL_Sq-8_48	49	1	4.5

3.2.5 Analysis of *GFP* silencing

In order to assess whether any of the allelic variants and/or chromosomal regions that were introgressed from selected accessions into the Col-0 genetic background had an impact on post-transcriptional gene silencing, all 51 introgression lines that had been established were evaluated with respect to the silencing of the *GFP* transgene. It was exploited that silencing of *GFP* transgenic lines can be readily observed using fluorescence stereomicroscopy. Even if isogenic plants are grown in controlled environment conditions, the onset of silencing is different in individual plants of a particular population (Arlt and Schmidt, 2006; Arlt, 2007). Silencing is often first observed as a small or large sector and frequently but not always spreads to other parts of the plant during plant development (Figure 12), silencing may be observed in the entire aerial tissue of a particular plant.

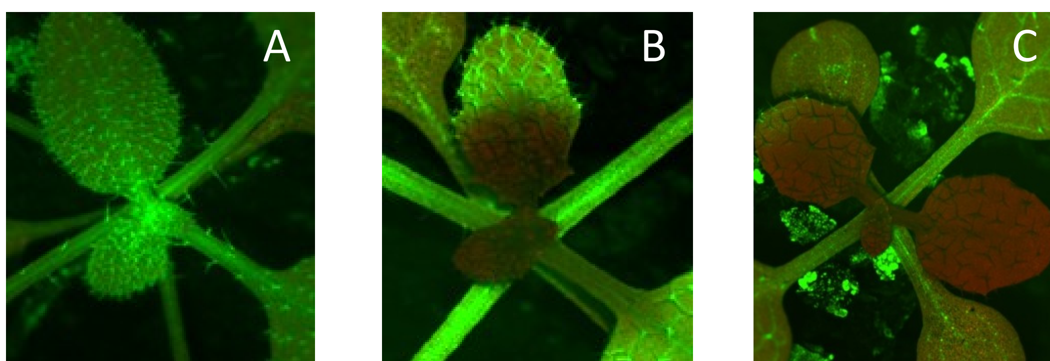


Figure 12. *GFP* expression and silencing in plants of introgression lines. The plants were analysed 20 days after sowing with a fluorescence microscope with filter sets GFP3 and CY5; in the overlay images bright green indicates GFP fluorescence, whereas *GFP* silenced tissues are shown in red due to chlorophyll fluorescence. (A) *GFP* expression is seen in all parts of the plant. (B,C) Specimen reveal *GFP* silencing in leaves or parts thereof.

Owing to the variable onset of silencing in individual plants, populations rather than individual plants have to be analysed at different stages of development (Arlt, 2007). For each line 70 plants were evaluated. Line 6x*GFP-F8/R127* in the Col-0 genetic background served as reference in all experiments. The 70 plants of each plant population to be analysed were distributed to two small trays of 35 plants each. The positions of the plants in the individual trays were randomised twice a week, but plants of one tray were never transferred to the other tray, thus the 35 plants in each of the trays could be treated as subpopulations. All plants were grown in controlled environment conditions using a long-day regime (16 h light at 20°C/8 h dark at 16°C, Table 2). At day 17 after sowing *GFP* fluorescence

was evaluated for the first time. In each experiment *GFP* fluorescence was analysed over a period of five weeks at ten different time points.

The genotypes of all lines were confirmed every time the plants were evaluated with respect to silencing. The identity and zygosity of the *GFP* loci was checked to ensure that all lines to be analysed had six copies of the *GFP* transgene. It was also tested whether the allele of the candidate gene of interest was present homozygously. For each introgression line polymorphic Indel markers immediately flanking the breakpoints were used to validate the identity and zygosity of the introgression segments.

In order to obtain a simple and quantitative description of the silencing behaviour in a particular plant population the proportion of silenced plants out of the total number of all plants was analysed. This is called hereafter "Frequency of silencing" or "F" (Arlt, 2007). Frequency of silencing was calculated for all lines and time points of a particular experiment. In order to record whether a particular plant showed silencing in small or large areas of the plant a previously developed scoring system was adopted (Arlt, 2007). Based on the estimated percentage of aerial tissues that showed silencing of the *GFP* gene, the plants were grouped into six different categories. Plants of category 0 did not exhibit any silencing and in plants of category 1 less than 10% of the aerial tissues of a particular plant revealed silencing. In plants of categories 2, 3 and 4 between 10% and 50%, between 50% and 90% and more than 90% of the total above-gorund tissues showed *GFP* silencing, respectively. The entire aerial tissues of an individual plant were affected by *GFP* silencing in category 5 plants.

Fisher's exact test (Section 2.2.12) was used to analyse whether the two subpopulations comprising 35 plants each of a particular line showed significant differences with respect to silencing behaviour and whether a particular introgression line showed significantly more or less silencing than line 6x*GFP-F8/R127* in the Col-0 genetic background that served as reference (Figure 13). For the comparisons between lines seventy plants were taken into account for each of the lines. On the one hand, the number of non-silenced plants, plants of category 0, and the plants showing silencing, all plants belonging to categories 1 to 5, were compared for subpopulations or different lines in the statistical analysis, this is termed "FN" hereafter. On the other hand, the number of plants that had been grouped in the six different categories provided the basis for the comparisons of two different subpopulations or lines, this evaluation is abbreviated "C". These analyses were carried out for all

experiments and all different time points which had been analysed. Comparisons were only made between lines that had been evaluated in the same experiment. If the resulting p-values were smaller than 0.05, the results were considered to be significant. All introgression lines were analysed following the experimental design and statistical analysis described above.

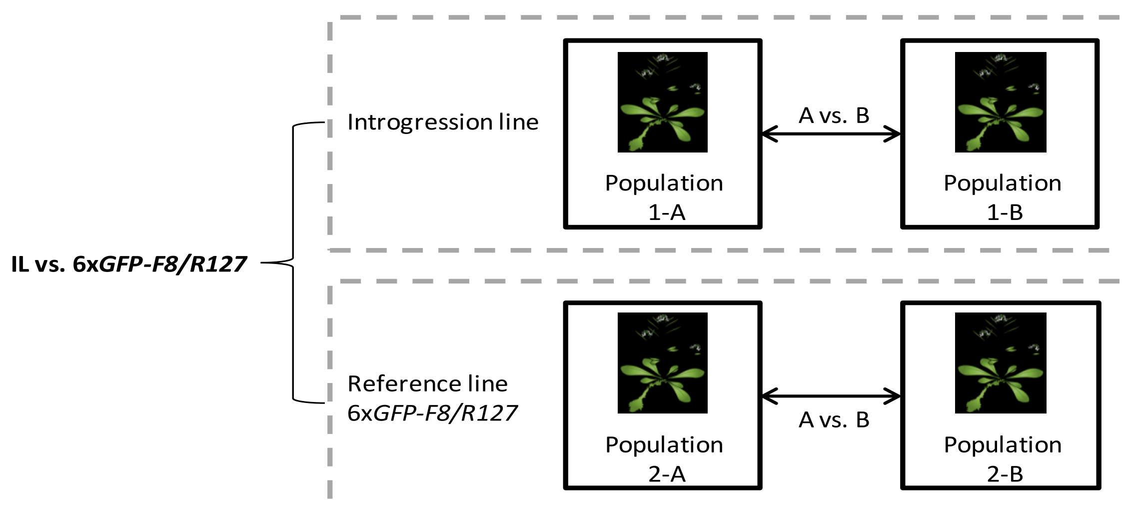


Figure 13. Comparisons to determine significant differences between subpopulations of a particular line or between an introgression line and the reference line 6x*GFP-F8/R127*. The populations of lines were divided into two subpopulations comprising 35 plants each (A and B). Subpopulations of the same line were compared to each other (A *versus* B). Comparisons between an introgression line and the reference line took into account 70 plants each for both lines (IL *versus* 6x*GFP-F8/R127)*.

3.2.6 Analysis of introgression lines carrying the Sq-8 allelic variant of the *HEN1* gene

Three introgression lines were established that carried the Sq-8 allelic variant of the *HEN1* gene, IL_Sq-8_6, IL_Sq-8_7 and IL_Sq-8_8. Figure 14 shows the data obtained in experiment 08-14 in which all three lines were analysed alongside with reference line 6x*GFP-F8/R127*. In all four plant populations, the number of plants showing silencing increased with the age of the plants. The percentage of plants showing *GFP* silencing in the three introgression lines was between 1.4% and 5.7% lower when compared to the reference line 6x*GFP-F8/R127* for the first four time points. At time points 5 and 6, the values were approximately 3% higher or up to 4% lower, significant differences with respect to FN were not found. However, at the last three and four time points IL_Sq-8_7 and IL_Sq-8_8 showed significantly more silenced plants than the reference line, respectively, whereas the third line, IL_Sq-8_6, showed values that were similar to that of the reference line. At the end of the experiment the percentage of silenced plants in line 6x*GFP-F8/R127* and IL_Sq-8_6 was approximately

65%, whereas about 90% of the plants in the populations of IL_Sq-8_7 and IL_Sq-8_8 showed silencing.

Differences between the introgression lines as well as between them and the reference line 6x*GFP-F8/R127* were also observed with respect to C. Plants of categories 0 and 1 were observed for IL_Sq-8_6, in contrast, plants belonging to categories 0, 1, 2 and 3 were found for IL_Sq-8_7. In the reference line and IL_Sq-8_8, plants of all six different categories were observed. Significant differences with respect to the C values were observed for ILs Sq-8_7 and Sq-8_8 when compared to the reference line, but not for IL_Sq-8_6 (Figure 14).

Additional experiments in which one or two introgression lines were evaluated in comparison to the reference line corroborated these results. Table 15 shows the number of silenced and non-silenced plants for each introgression line as well as for the reference line, for all time points of the different experiments. In experiment 12-13 the number of silenced plants in IL_Sq-8_6 is similar to that of the reference line for the first three time points but then smaller differing by up to 11 plants at time point 9. Significant differences were not observed between IL_Sq-8_6 and line 6x*GFP-F8/R127*. In contrast, IL_Sq-8_8 always shows more silenced plants than the reference line. Significant differences between those two lines were observed for the last three time points of the analysis.

The number of silenced and non-silenced plants observed for IL_Sq-8_6 in experiment 03-14 also did not differ significantly from the data established for the reference line, regardless which time point was analysed. Plants of lines 6x*GFP-F8/R127* and IL_Sq-8_6 started silencing at days 34 and 38 after sowing, respectively. The number of silenced plants was always smaller than that of the non-silenced plants, even at the later time points. IL_Sq-8_7 shows at least twice the number of silenced plants compared to the other lines from days 31 to 45 after sowing (Table 15). Significant differences ($p<0.05$) were observed for the last three time points when IL_Sq-8_7 was compared to the reference line. Significantly enhanced silencing in IL_Sq-8_7 was also seen in the last experiment listed in Table 15.

Silenced plants were already found at the first time point in this experiment both for IL_Sq-8_7 and line 6x*GFP-F8/R127*. The number of plants showing silencing increased more rapidly in IL_Sq-8_7 than in the reference line. Significant differences between both lines were detected from the fifth time point onwards. During this period the number of silenced plants increased from 14 to 67 in IL_Sq-8_7, in the reference line the values changed from 3 to 45.

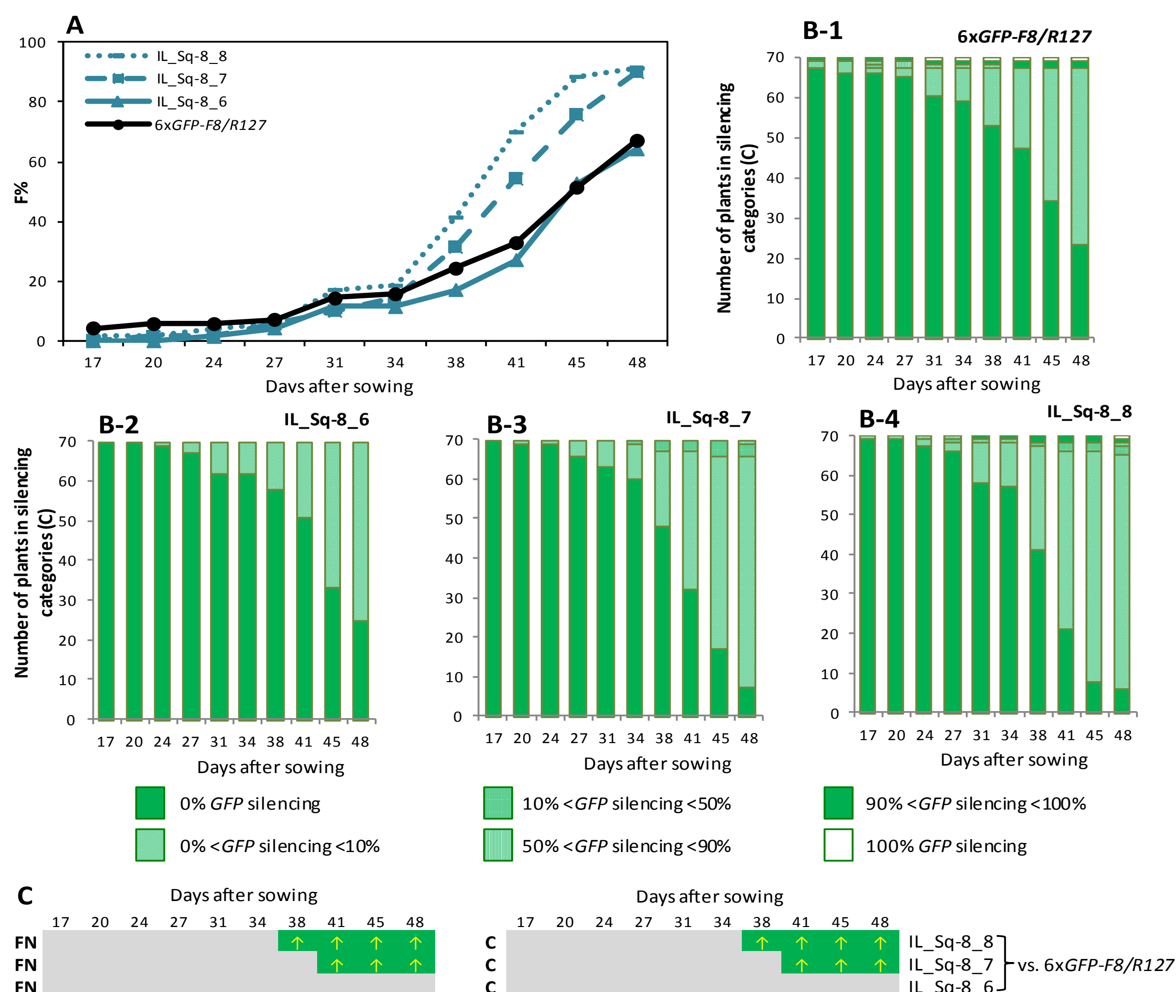


Figure 14. Introgression lines carrying the Sq-8 allelic variant of the *HEN1* gene differ with respect to silencing. (A) Comparison of frequency of silencing between introgression lines and reference line 6x*GFP-F8/R127*, the data were established in experiment 08-14. (B-1 – B-4) Numbers of plants in the different categories are displayed, the categories are explained in the key below the bar charts. Panel C summarises the results of the statistical analysis with respect to FN and C. The dark green and grey coloured boxes represent significant and non-significant values, respectively. The arrows (↑) illustrate significantly higher number of silenced plants in the introgression line than in reference line 6x*GFP-F8/R127*.

The different experiments confirmed that IL_Sq-8_7 and IL_Sq-8_8 show significantly more silencing than line 6x*GFP-F8/R127*. Likewise, it was corroborated that IL_Sq-8_6 did not show enhanced silencing when compared to the reference line. However, the data shown in Table 15 also reveal that considerable differences were seen with respect to the number of plants showing silencing in individual experiments. For example, in IL_Sq-8_7 three and 27 non-silenced plants were observed in experiments 10-14 and 03-14 at the last time point of the experiment (Table 15), respectively, in the experiment shown in Figure 14 seven non-silenced plants were observed 48 days after sowing. Similarly, the numbers of non-silenced

plants in the reference line differed at time point 10 in experiments 03-14 and 10-14 by a factor of approximately two (Table 15).

Table 15. Comparison of the number of silenced and non-silenced plants in introgression lines carrying the Sq-8 allelic variant of the *HEN1* gene in different experiments. Si. and N-si. represent silenced plants and non-silenced plants, respectively. For each of the 10 time points of the analysis, the silenced and non-silenced plants in each introgression line compared to the values of reference line 6x*GFP-F8/R127* are listed. Green boxes indicate significant differences.

Exp	Lines	Days after sowing																			
		17		20		24		27		31		34		38		41		45		48	
		Si.	N-si.	Si.	N-si.	Si.	N-si.	Si.	N-si.	Si.	N-si.	Si.	N-si.	Si.	N-si.	Si.	N-si.	Si.	N-si.	Si.	N-si.
12-13	**IL_Sq-8_6**	0	70	0	70	0	70	0	70	0	70	1	69	4	66	12	58	22	48	32	38
	p-value	1		1		1		1		0.244604		0.11556		0.15712		0.668908		0.08311		0.61211	
	IL_Sq-8_8	1	69	1	69	1	69	3	67	7	63	8	62	17	53	32	38	50	20	59	11
	p-value	1		1		1		0.6195450		0.3255190		0.7792400		0.1981540		0.0039480		0.0056940		0.0000530	
	6x*GFP-F8/R127*	0	70	0	70	0	70	1	69	3	67	6	64	10	60	15	55	33	37	36	34
03-14	**IL_Sq-8_6**	0	70	0	70	0	70	0	70	0	70	0	70	4	66	5	65	18	52	27	43
	p-value	1		1		1		1		1		1		1		1		0.415845		0.207605	
	IL_Sq-8_7	0	70	0	70	0	70	0	70	2	68	2	68	9	61	22	48	36	34	43	27
	p-value	1		1		1		1		0.4964029		1		0.243487		0.000442		0.000079		0.0000782	
	6x*GFP-F8/R127*	0	70	0	70	0	70	0	70	0	70	1	69	4	66	5	65	13	57	19	51
10-14	**IL_Sq-8_7**	3	67	5	65	9	61	10	60	14	56	20	50	35	35	47	23	64	6	67	3
	p-value	0.6195450		0.2085600		0.0551620		0.0772956		0.0080248		0.0040820		0.0027728		0.0003483		0.00000003		0.0000034	
	6x*GFP-F8/R127*	1	69	1	69	2	68	3	67	3	67	6	64	17	53	25	45	34	36	45	25

In experiment 08-14 23 non-silenced plants were observed at 48 days after sowing (Figure 14), this value is very similar to that observed in experiment 10-14. Differences were also noted with respect to the onset of silencing, for example in experiments 08-14 and 10-14 the first silenced plants were observed 17 days after sowing for two of the lines, whereas in experiment 03-14 this was only the case at 31 days after sowing for one line.

3.2.7 Subpopulations of lines show a similar behaviour with respect to gene silencing

In any of the experiments which had been carried out in the context of this study, the results for the two subpopulations of all lines were evaluated with respect to gene silencing for the ten time points which had been analysed. In total, 1390 comparisons each were conducted for the FN and C data. In 64 (4.6%) cases significant differences were observed with respect to FN and/or C. For 34 pairwise comparisons of subpopulations, significant differences were found with respect to FN and C. In nine cases significant differences were only observed if

the FN data were taken into account. Twenty-one comparisons revealed significant differences only if the number of plants in the different silencing categories were analysed. For ten lines, IL_Ang-0_35, IL_Bor-4_35.2.9, IL_Bor-4_35.2.11, IL_Kas-1_32, Cvi-0_6, IL_Kin-0_14, IL_Lz-0_15, IL_Ra-0_51, IL_Sq-8_7 and 6x*GFP-F8/R127*, significant differences with respect to FN were observed for at least two of the later time points, day 34 until day 48 after sowing. For line 6x*GFP-F8/R127* significant differences at later stages of the experiment were found in three out of 22 experiments. For seven of the lines only a single experiment was affected, in case of IL_Ang-0_35 and Sq-8_7 this was seen in two experiments.

3.2.8 Comparative analysis of 6x*GFP* lines carrying different T-DNA locus combinations in the Col-0 genetic background

Line 6x*GFP-F8/R127* was used as reference in all experiments. Most of the introgression lines, 36 out of 51, also contained T-DNA loci *F8* and *R127* homozygously. However, 4 and 11 introgression lines carried locus combinations *F18/R127* and *F128/R127* instead (Figure 11, Supplementary 2-5). If introgression lines carrying *GFP* loci *F18* and/or *F128* were analysed in a particular experiment, lines 6x*GFP-F18/R127* and/or 6x*GFP-F128/R127* were in most of the cases analysed alongside with line 6x*GFP-F8/R127*.

Examples of comparisons with respect to F between 6x*GFP* lines containing different T-DNA locus combinations in the Col-0 genetic background are seen in Figure 15. Panel A shows the proportion of silenced plants, which were observed for lines 6x*GFP-F8/R127* and 6x*GFP-F128/R127*. The F values differed at most of the time points but the differences only ranged from 0.5% to 6.4%, significant differences were not found. Panel B shows a comparison between line 6x*GFP-F8/R127* and 6x*GFP-F18/R127*. The percentages of silenced plants in 6x*GFP-F8/R127* were always higher than for 6x*GFP-F18/R127* with the exception of the two last time points. However, a significant difference was only observed at day 34 after sowing. Panel C of Figure 15 shows the results of an experiment in which all three 6x*GFP* lines were included. The percentages of silenced plants differed by up to 2.9% between the three 6x*GFP* lines during the first six time points, but at later stages differences increased up to 27%. The frequency of silencing in lines 6x*GFP-F18/R127* and 6x*GFP-F128/R127* increased comparatively slowly, from 14.3% at the seventh time point to 40.0% at the last time point, in contrast, F increased during this period from 18.6% to 67.1% in line 6x*GFP-F8/R127*.

Significant differences between the reference line and 6x*GFP-F128/R127* were observed for the last three time points with respect to FN and C. During this period significant differences were also observed between lines 6x*GFP-F8/R127*and 6x*GFP-F18/R127*.

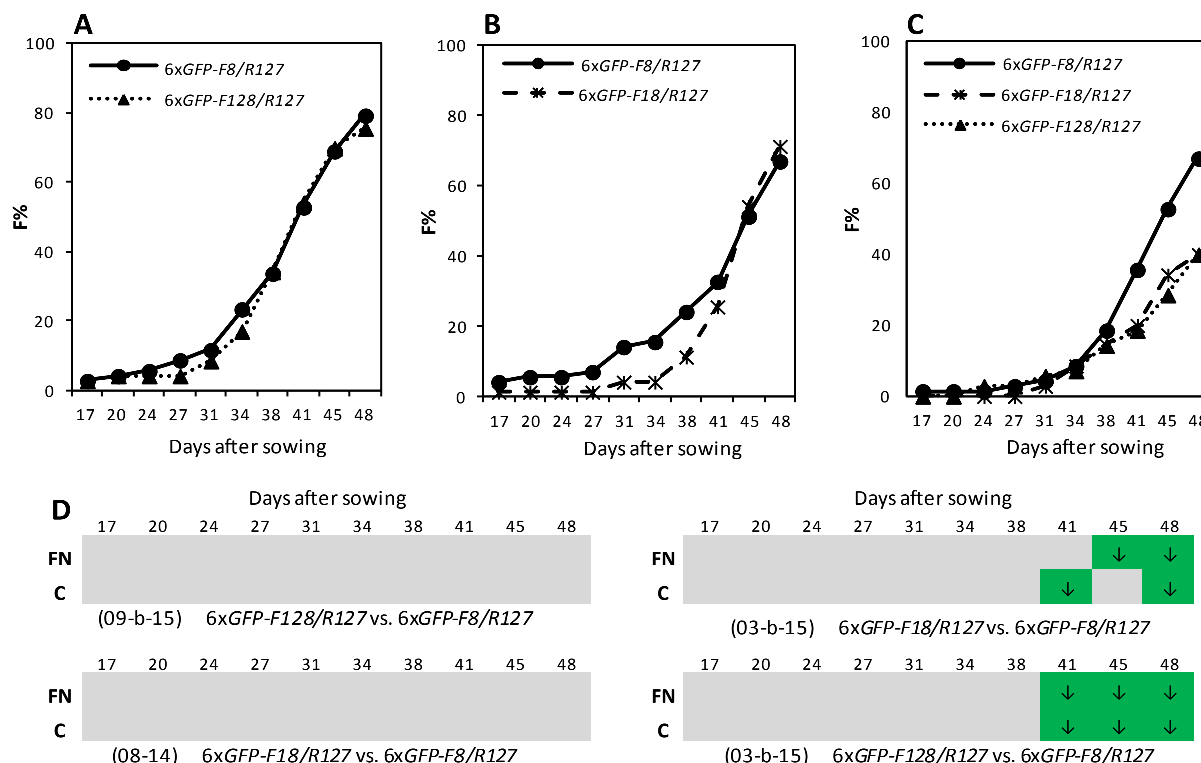


Figure 15. Comparison of 6x*GFP* lines carrying different T-DNA locus combinations with respect to *GFP* silencing. Panels A, B and C show data from three different experiments, 09-b-15, 08-14 and 03-b-15, respectively. The results of the statistical analyses are displayed in panel D. Arrows (↓) illustrate significantly lower numbers of silenced plants in lines 6x*GFP-F18/R127* or 6x*GFP-F128/R127* than in reference line 6x*GFP-F8/R127*.

Lines 6x*GFP-F8/R127* and 6x*GFP-F18/R127* were analysed alongside of each other in three experiments, lines 6x*GFP-F8/R127* and 6x*GFP-F128/R127* in seven experiments. The results are summarised in Table 16. Out of a total of 100 comparisons, 13 significant differences each with respect to FN and C were found. Significant differences with respect to FN or C were observed in 8 out of 10 experiments. In four experiments significant differences were only found at single time points. Significant differences for consecutive time points with respect to FN were seen in three instances, in all cases for later time points of the analyses.

As reported in the section in which Sq-8 introgression lines carrying the Sq-8 allelic variant of *HEN1* were described, differences with respect to silencing frequency were observed when data of different experiments were compared. The different 6x*GFP* lines were analysed

multiple times, hence it was possible to look at this point in more detail. Line 6x*GFP-F8/R127* was analysed in 22 different experiments. The silencing frequency at day 48 after sowing ranged from 27.1 to 79.4% in the 22 experiments (Table 17), mean and median values were 58.5% and 63.6%, respectively. For lines 6x*GFP-F18/R127* and 6x*GFP-F128/R127* the frequencies of silencing at the last time points of the different experiments ranged from 40.0% to 71.4 % and from 27.1% to 75.7%, respectively.

Table 16. Comparison of gene silencing revealed few significant differences between 6x*GFP* lines carrying different T-DNA locus combinations in the Col-0 genetic background

6x*GFP*	Expermiment	Number of time points showing significant differences (p<0.05) between different 6x*GFP* lines			
		FN		C	
		No.	Time point	No.	Time point
F8/R127	03-14	1	8	2	8, 10
vs.	08-14	1	6	0	
F18/R127	03-b-15	2	9, 10	2	8, 10
	06-14	1	6	1	6
	09-b-14	0		0	
F8/R127	12-14	0		1	5
vs.	01-15	4	7, 8, 9, 10	4	7, 8, 9, 10
F128/R127	03-b-15	3	8, 9, 10	3	8, 9, 10
	04-15	0		0	
	07-15	1	8	0	

Table 17. Silencing frequencies observed for 6x*GFP* lines carrying different T-DNA locus combinations in the Col-0 genetic background. Only the data for plants that were scored 48 days after sowing are shown.

Experiment	Frequency of silencing (F%)			Experiment	Frequency of silencing (F%)		
	F8/R127	***F18/R127***	***F128/R127***		***F8/R127***	***F18/R127***	***F128/R127***
08-13	54.3			10-14	64.3		
12-13	51.4			11-14	67.1		
02-b-14	37.1			11-b-14	67.1		
03-14	27.1	44.3		12-14	41.4		40.0
04-14	41.4			12-b-14	62.9		
06-14	74.3		61.4	01-15	70.0		27.1
06-b-14	48.6			02-15	40.0		
07-14	45.7			03-b-15	67.1	40.0	40.0
08-14	67.1	71.4		04-15	55.7		51.4
09-14	75.7			06-15	71.4		
09-b-14	79.4		75.7	07-15	78.6		65.7

Two or more significant differences were observed between lines 6x*GFP-F8/R127* and 6x*GFP-F18/R127* in two experiments, in experiment 03-14 silencing was more pronounced in 6x*GFP-F18/R127*, whereas in experiment 03-b-15 more plants showed silencing in line 6x*GFP-F8/R127* (Tables 16 and 17).

3.2.9 Several introgression lines show significantly more or less silencing than reference line 6x*GFP-F8/R127*

Out of the 51 established introgression lines five, ten, four, nine and 19 lines contained allelic variants of the *HEN1, SDE3, NRPD1/AGO7, NRPE1* and *WEX* genes, respectively, four additional lines harboured the Bor-4 allelic variant of the *NRPD1* gene only. The data provided in Table 18 summarise all comparisons that were performed between introgression lines and the reference line 6x*GFP-F8/R127.* For each allelic variant of a particular accession at least two different introgression lines were analysed, in most of the cases twice. Only the two introgression lines established for the Ra-0 alleles of *SDE3* were evaluated once. The same was true for IL_Shahdara_6 and IL_Shahdara_30, but IL_Shahdara_10 was investigated in two experiments. It was evaluated for all the different time points of the different experiments whether an introgression line showed significantly more or less silencing than the reference line or not according to the FN and the C data. In comparison to reference line 6x*GFP-F8/R127*, ten out of 51 ILs did not reveal any significant differences with respect to FN and C. Nineteen and 13 ILs showed significantly higher and lower silencing than the reference line for FN or C, respectively. For nine lines significantly higher values were observed than for 6x*GFP-F8/R127* in at least one experiment, whereas another experiment revealed significantly lower silencing than the reference line. Significant differences could be observed at any time point and were seen for FN or C, but in most cases for FN and C. Seven ILs, IL_Cvi-0_6/ 6.25, IL_Shahdara_10, IL_Gie-0_3a, IL_Gie-0_6/ 6.18, IL_Sq-8_49, IL_Sq-8_7 and IL_Sq-8_8, revealed significant differences compared to the reference line in two or more experiments with respect to FN and C. All of these lines showed significant results in at least two or three consecutive time points towards the end of the experiments for FN and C. The results for lines IL_Sq-8_7 and IL_Sq-8_8 were described in detail in a previous section, a detailed analysis of the data obtained for the other five lines is presented in the following sections.

As can be seen in Table 18, introgression lines that were established for the same allelic variant of a particular accession may show differences with respect to gene silencing. This was described in a previous section for the lines that carry Sq-8 allelic variants of the *HEN1* gene, but also holds true for the Sq-8 allelic variant of the *WEX* gene. Independent introgression lines carrying the Shahdara or Cvi-0 allelic variants of the *NRPE1* gene also differ with respect to their silencing behaviour.

Table 18. Summary of significant differences with respect to gene silencing between introgression lines and the reference line 6x*GFP-F8/R127*. The dark green and grey coloured boxes represent significant and non-significant values, respectively (p<0.05). The arrows, (↑) or (↓), illustrate significantly higher or lower silencing in the introgression line than in reference line 6x*GFP-F8/R127*, respectively. The different experiments are listed below the names of the introgression lines.

		Days after sowing										Days after sowing										Days after sowing									
HEN1		17	20	24	27	31	34	38	41	45	48	17	20	24	27	31	34	38	41	45	48	17	20	24	27	31	34	38	41	45	48
IL_Lp2-2_27	FN								↑	↑	↑																				
03-14, 08-14, 10-14	C								↑	↑	↑																				
IL_Lp2-2_30	FN								↑		↑					↓		↓													
03-14, 08-14, 10-14	C										↑					↓		↓													
IL_Sq-8_6	FN																														
12-13, 03-14, 08-14	C																														
IL_Sq-8_7	FN								↑	↑	↑								↑	↑	↑					↑	↑	↑	↑	↑	↑
03-14, 08-14, 10-14	C								↑	↑	↑								↑	↑	↑						↑	↑	↑	↑	↑
IL_Sq-8_8	FN								↑	↑	↑							↑	↑	↑	↑										
12-13, 08-14	C								↑	↑	↑							↑	↑	↑	↑										
NRPE1		17	20	24	27	31	34	38	41	45	48	17	20	24	27	31	34	38	41	45	48	17	20	24	27	31	34	38	41	45	48
IL_Cvi-0_6	FN									↑	↑									↑	↑								↑	↑	↑
04-14, 06-b-14, 12-b-14	C									↑	↑								↑	↑	↑									↑	↑
IL_Cvi-0_6.25	FN								↑	↑	↑																				
02-15	C								↑	↑	↑																				
IL_Cvi-0_19.27	FN																														
12-b-14, 02-15	C																														
IL_Cvi-0_39	FN																														
12-b-14	C																														
IL_Kas-1_18	FN																														↓
02-b-14, 06-b-14, 12-b-14	C																														
IL_Kas-1_32	FN																														
02-b-14, 06-b-14	C																														
IL_Kas-1_39	FN																				↓										
06-b-14, 12-b-14	C																				↓										
IL_Shahdara_6	FN					↑	↑	↑	↑	↑	↑																				
06-b-14	C					↑	↑	↑	↑	↑	↑																				
IL_Shahdara_10	FN			↑	↑	↑	↑	↑	↑	↑	↑				↑	↑	↑	↑	↑	↑	↑										
04-14, 06-b-14	C			↑	↑	↑	↑	↑	↑	↑	↑				↑	↑	↑	↑	↑	↑	↑										
IL_Shahdara_30	FN																														
12-b-14	C																														
SDE3		17	20	24	27	31	34	38	41	45	48	17	20	24	27	31	34	38	41	45	48	17	20	24	27	31	34	38	41	45	48
IL_Baa-1_9	FN																														
12-13, 09-b-14	C																														
IL_Baa-1_21	FN								↑						↓		↓														
12-13, 09-b-14	C								↑						↓		↓	↓	↓												
IL_Lz-0_15	FN						↓			↓											↓										
09-b-14, 03-b-15	C						↓		↓	↓	↓																				
IL_Lz-0_20	FN				↓	↓	↓	↓	↓	↓	↓																				
09-b-14	C				↓		↓	↓	↓	↓	↓																				
IL_Lz-0_38	FN																													↓	↓
04-14, 09-b-14, 03-b-15	C						↑				↑																		↓		
IL_Ra-0_26	FN																														
03-b-15	C																														
IL_Ra-0_51	FN																														
03-b-15	C																														
IL_Ws-0_11	FN						↑	↑	↑									↑										↑			
08-13, 09-b-14, 03-b-15	C						↑	↑	↑	↑	↑							↑	↑	↑	↑							↑			
IL_Ws-0_24	FN																														
08-13	C									↑	↑																				
IL_Ws-0_23	FN																														
09-b-14, 03-b-15	C								↑	↑																					

		Days after sowing										Days after sowing										Days after sowing									
NRPD1 and/or AGO7		17	20	24	27	31	34	38	41	45	48	17	20	24	27	31	34	38	41	45	48	17	20	24	27	31	34	38	41	45	48
IL_Bor-4_24.12	FN											↑	↑	↑	↑	↑	↑	↑	↑	↑	↑										
01-15, 07-15	C											↑	↑	↑	↑	↑	↑	↑	↑	↑	↑										
IL_Bor-4_24.50	FN																↑	↑	↑	↑	↑										
01-15, 07-15	C															↑	↑	↑	↑	↑	↑										
IL_Bor-4_35.2.9	FN								↑	↑	↑										↓		↓								
12-14, 01-15, 07-15	C								↑	↑	↑																				
IL_Bor-4_35.2.11	FN				↑	↑	↑	↑	↑	↑	↑																				
12-14, 01-15, 07-15	C					↑	↑	↑	↑	↑	↑																			↓	↓
IL_Bor-4_35.33.3	FN																														
12-14	C																														
IL_Bor-4_35.33.10	FN										↑																				
12-14	C																														
IL_Gie-0_3a	FN							↓	↓	↓	↓																	↓	↓	↓	↓
06-14, 12-14, 01-15	C							↓	↓	↓	↓																	↓	↓	↓	↓
IL_Gie-0_6	FN								↓	↓	↓							↓									↓	↓	↓	↓	↓
02-b-14, 12-14, 01-15	C								↓	↓	↓							↓									↓	↓	↓	↓	↓
IL_Gie-0_6.18	FN								↓	↓	↓																				
06-14	C								↓	↓	↓																				
WEX		17	20	24	27	31	34	38	41	45	48	17	20	24	27	31	34	38	41	45	48	17	20	24	27	31	34	38	41	45	48
IL_Ang-0_28	FN									↑																					
07-14, 11-14	C																				↑										
IL_Ang-0_35	FN											↑	↑	↑	↑			↑	↑	↑	↑										
07-14, 11-14	C											↑					↑	↑	↑	↑	↑										
IL_Baa-1_6	FN																														
07-14, 11-14	C																		↑	↑											
IL_Baa-1_37	FN											↑	↑	↑																	
07-14, 11-14	C											↑			↑		↑	↑	↑												
IL_Kin-0_3	FN						↓																								
06-b-14, 09-14	C						↓																								
IL_Kin-0_10	FN							↓	↓	↓	↓																				
10-14	C							↓	↓	↓	↓																				
IL_Kin-0_12	FN			↑	↑	↑		↑		↑																					
10-14, 06-15	C			↑		↑		↑		↑	↑																				
IL_Kin-0_14	FN										↓																				
10-14, 06-15	C									↓	↓																				
IL_Kin-0_19	FN					↓																									
06-b-14, 09-14	C					↓																									
IL_Kin-0_19.19	FN									↓	↓																				
11-b-14, 06-15	C								↓	↓	↓										↑										
IL_Kin-0_19.29	FN								↓	↓	↓																				
11-b-14, 06-15	C							↓	↓	↓	↓																				
IL_Kin-0_19.55	FN									↓	↓																				
11-b-14, 06-15	C									↓	↓										↓										
IL_Lp2-2_6	FN							↑											↓	↓											
09-14, 11-b-14	C																		↓	↓	↓										
IL_Lp2-2_43	FN																	↑	↑												
09-14, 11-b-14	C																		↑		↑										
IL_Shahdara_12	FN													↑	↑	↑	↑				↑								↓	↓	↓
04-14, 07-14, 11-b-14	C													↑	↑	↑	↑				↑								↓	↓	↓
IL_Shahdara_17	FN			↑	↑															↓	↓										
07-14, 11-b-14	C			↑	↑														↓	↓	↓										
IL_Sq-8_16	FN						↑	↑	↑																						
11-14	C						↑		↑																						
IL_Sq-8_48	FN																														
11-14, 02-15	C																														
IL_Sq-8_49	FN	↑	↑	↑	↑	↑	↑	↑	↑		↑		↑	↑	↑	↑	↑	↑	↑	↑	↑				↑	↑	↑	↑	↑	↑	↑
09-14, 11-14, 02-15	C	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑		↑					↑	↑	↑	↑				↑	↑	↑	↑	↑	↑	↑

3.2.10 Analysis of introgression lines carrying Gie-0 alleles for the *AGO7* and *NRPD1* genes

All introgression lines which carried the Gie-0 allelic variants of candidate genes *NRPD1* and *AGO7* contained T-DNA loci *F128* and *R127* homozygously (Supplementary figure 4). In three experiments the two established introgression lines were therefore compared not only to reference line 6x*GFP-F8/R127* but also to line 6x*GFP-F128/R127* (Figure 16).

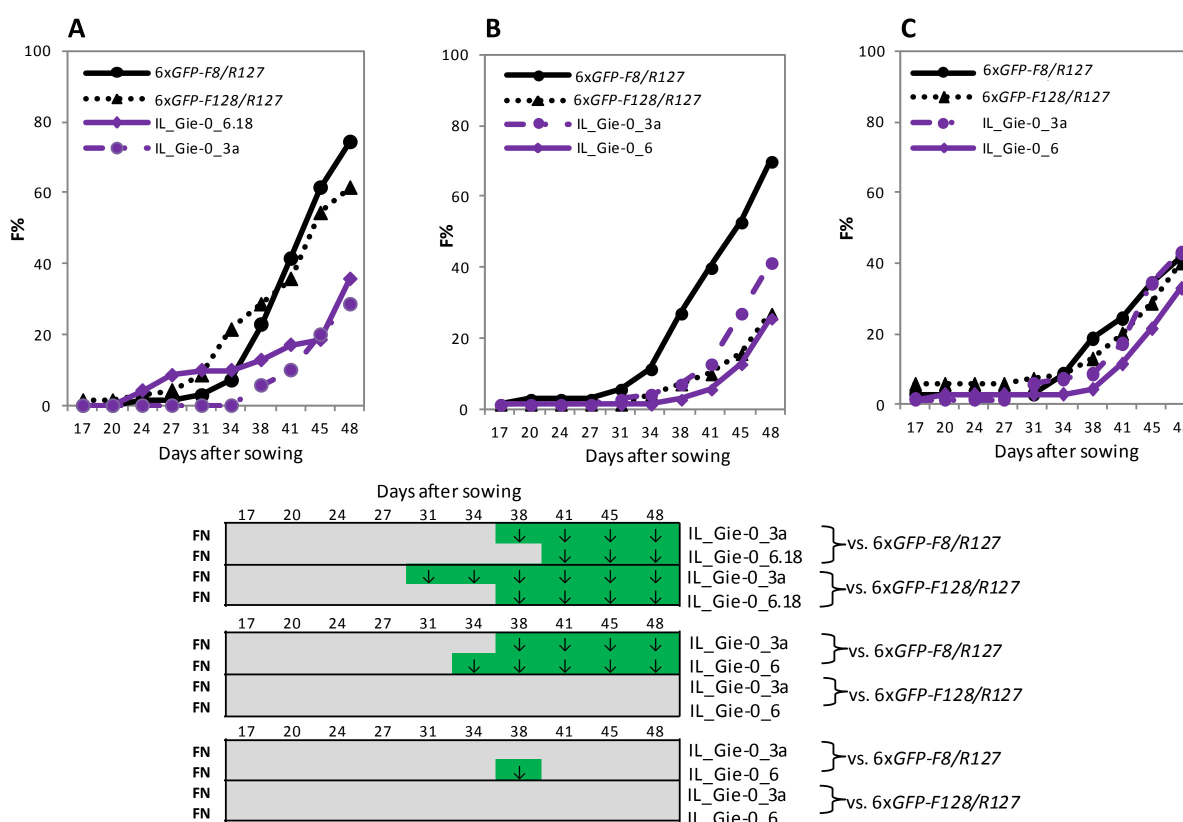


Figure 16. Comparison of the frequency of silencing of introgression lines carrying Gie-0 allelic variants of the *AGO7* and *NRPD1* genes. The data shown in panels (A), (B) and (C) were obtained in experiments 06-14, 01-15 and 12-14, respectively. Below the graphs the results of the statistical analysis of the FN data are shown. Dark green and grey coloured boxes indicate significant and non-significant values, respectively. The arrows (↓) illustrate significantly fewer silenced plants in the introgression line than in reference line 6x*GFP-F8/R127* or 6x*GFP-F128/R127*.

In experiment 06-14, overall more silenced plants in the populations of both 6x*GFP* lines are found from time point 7 onwards when compared to the two introgression lines. Significant differences between the different lines were not detected for the first four time points. However, at time point 5, a significant difference was found between 6x*GFP-F128/R127* and IL_Gie-0_3a, the former line showed 8.6% more silencing than the latter one. At the last time point the population of line 6x*GFP-F128/R127* showed 61% silenced plants, 74% were

observed for reference line 6x*GFP-F8/R127*. Significantly lower values of 29% and 36% were established for IL_Gie-0_3a and IL_Gie-0_6.18, respectively.

In experiment 01-15 the proportion of silenced plants was comparable in all four lines analysed for the first four time points, the lines did not differ from each other by more than 1.4%. From time point 6 onwards significant differences were seen between the reference line 6x*GFP-F8/R127* and IL_Gie-0_6. IL_Gie-0_3a also showed significant differences when compared to line 6x*GFP-F8/*R127 for time points 7 to 10. In contrast to that, line 6x*GFP-F128/R127* did not show significantly more silencing than the two introgression lines, it exhibited significantly less silencing than the reference line from day 38 after sowing onwards (Table 17). At the end of the experiment, 70% of silenced plants were found in population 6x*GFP-F8/R127*. For IL_Gie-0_3a, 6x*GFP-F128/R127* and IL_Gie-0_6.18 values of 41%, 27% and 26% were observed, respectively.

During the first four time points, the highest silencing frequency in experiment 12-14 was with 5.7% observed in line 6x*GFP-F128/R127*. The percentage of *GFP* silenced plants in the remaining lines 6x*GFP-F8/R127,* IL_Gie-0_3a and IL_Gie-0_6 were 2.9%, 1.4% and 2.9%, respectively. The value for IL_Gie-0_6 stayed the same at time points 5 and 6, but higher values of 7.1% and 8.6% were found in the other lines at day 34 after sowing. From the seventh time point onwards, the proportion of silenced plants in all four lines continuously increased. However, the percentage of silenced plants at day 48 after sowing observed for reference line 6x*GFP-F8/R127* was with 41% not as high as in experiments 06-14 and 01-15. In experiment 12-14, the silencing frequency of 6x*GFP-F128/R127* was 21% lower than in experiment 06-14 but 13% higher than in experiment 01-15. At day 38 after sowing a significant difference with respect to FN was observed between line 6x*GFP-F8/R127* and IL_Gie-0_6.

3.2.11 Analysis of introgression lines carrying the Sq-8 allelic variant of the *WEX* gene

Three introgression lines contained the Sq-8 allelic variant of the *WEX* gene. Selected data for lines IL_Sq-8_48 and IL_Sq-8_49, which were analysed two and three times, respectively, are shown in Figure 17. The proportion of silenced plants of IL_Sq-8_49 was higher than that of the reference line regardless at which time point the plants were evaluated in experiment 09-14.

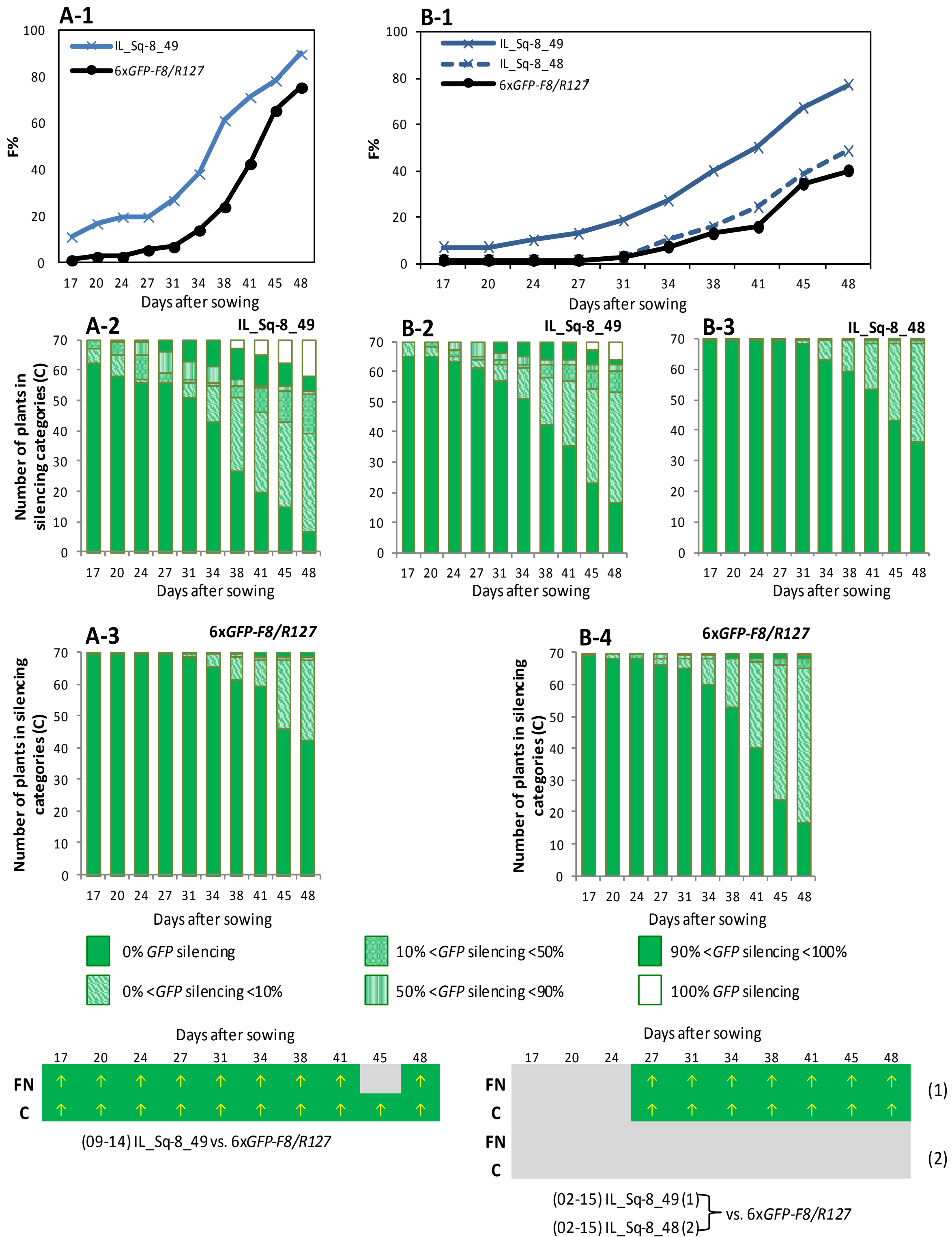


Figure 17. Comparison of *GFP* silencing between introgression lines carrying the Sq-8 allelic variant of the *WEX* gene and the reference line 6x*GFP-F8/R127*. (A-1 and B-1) Proportions of silenced plants are shown for different plant populations. (A-2 – B-4) The panels indicate how many of the analysed plants were found in the six different scoring categories. The key for the categories is found below the panels. The data shown in the B panels were derived from experiment 02-15 whereas the ones shown in the A panels were obtained in experiment 09-14. Below the bar charts the results of the statistical analysis are displayed, dark green and grey coloured boxes represent significant ($p<0.05$) and non-significant values, respectively. The arrows (↑) illustrate significantly more silencing in the introgression line than in reference line 6x*GFP-F8/R127*.

The silencing frequency of IL_Sq-8_49 was 10% higher than that in the 6x*GFP* reference line at the first time point. For the remainder of the analysis, the percentage of plants that showed silencing was always at least 12% higher than that of the reference line. The highest difference of 37% was found at the age of 38 days after sowing between the two lines. At the end of the experiment, 90% of the plants in the population of IL_Sq-8_49 showed silencing in parts of the different plants, but only 76% in the reference line. The statistical analysis revealed that the lines were significantly different for nine out of ten time points analysed.

The grouping of plants in the six different categories showed that silencing in IL_Sq-8_49 was not confined to small sectors but also spread throughout the plants, some plants even showed silencing in the entire aerial tissues (Figure 17A-2). At the first time point, eight out of 70 plants showed silencing, among them, five plants belonged to category 1 and three to category 2 whereas only one plant of category 1 was seen in the population of the reference line (Figure 17A-3). At the seventh time point, 43 out of 70 plants of IL_Sq-8_49 had been placed in categories 1 to 5. In the case of the reference line only 17 plants showed silencing at this stage, 15 of the plants belonged to category 1 and two plants to category 4. At the end of the experiment 32 plants were found in category 1, 13 plants in category 2, one plant in category 3, five plants in category 4 and 12 plants in category 5 for population IL_Sq-8_49. For the reference line 6x*GFP-F8/R127* 48 and three plants belonged to categories 1 and 2, respectively. One plant each were classified to categories 4 and 5. Significant differences with respect to C were found between the two lines for all ten time points analysed.

A second experiment was performed for IL_Sq-8_49 and the 6x*GFP* reference line to reveal whether the pronounced differences with respect to silencing between the lines were reproducible. Figure 17B-1 confirms that the proportion of plants showing silencing increases with plant age. Like in the first experiment, the percentage of plants displaying silencing in IL_Sq-8_49 was higher during the entire analysis than in the reference line. The introgression line revealed up to 10% silenced plants during the three first time points, but the values were not significantly different from the 1.4% found for line 6x*GFP-F8/R127*. However, from the fourth time point onwards, the FN values in the IL_Sq-8_49 population always showed significant differences when compared to the reference line. The lowest and highest differences between the lines during this time period were 11.4% and 37.1% at days 27 and 48 after sowing, respectively. The results obtained for the different categories of the two

populations also showed significant differences between the two lines for all except the three first time points. In all three experiments in which Sq-8_49 was analysed alongside 6x*GFP-F8/R127* the former line showed enhanced silencing in comparison to the reference line (Table 18). When compared to the reference line, line IL_Sq-8_48 did not show significant differences, neither with respect to FN nor to C (Figure 17). Silencing frequency of IL_Sq-8_48 was 8.6% higher than for the reference line at days 41 and 48 after sowing. For all other time points differences of up to 4.3% were found. During the first half of the experiment each line had only one or two silenced plants each. The proportion of plants displaying silenced sectors was at the end of the experiment lower than 50% for both lines (Figure 17).

The contrasting results obtained for lines of IL_Sq-8_48 and IL_Sq-8_49 suggest that the observed enhanced silencing may not be due to the allelic variant of the *WEX* gene itself. The characterisation of the introgression lines with Indel markers revealed differences between the lines with respect to number, position and size of the introgressed segments. For example, IL_Sq-8_48 carried only one homozygous segment which spanned 4.5 Mbp on chromosome 4, whereas IL_Sq-8_49 carried a 8.5 Mbp long introgression on this chromosome. The latter line also harboured two heterozygous segments with lengths of 1.3 Mbp and 4.9 Mbp on chromosomes 2 and 5, respectively (Table 14; Figures 11 and 18). It is important to note, that introgression lines IL_Sq-8_7 and IL_Sq-8_8 which carried the Sq-8 allelic variant of the *HEN1* gene and showed enhanced silencing when compared to the reference line also harboured an introgression segment on chromosome 5 which partly overlapped the introgression segments present in lines IL_Sq-8_16 and IL_Sq-8_49. In contrast, line IL_Sq-8_6, neither showed significant differences with respect to silencing when compared to the reference line nor contained an introgression in this region (Table 14, Figure 18, Supplementary figure 2).

As a next step, it was evaluated whether some of the introgressed regions that differed between IL_Sq-8_48 and IL_Sq-8_49 caused increased gene silencing. Plants were therefore selected among the progeny of IL_Sq-8_49 that carried the introgression segment on chromosome 5 homozygously for Col-0 or Sq-8. Likewise, plants were selected that either carried the segment delineated by markers FRI-IND and UPSC_4-6222 homozygously for Col-0 or Sq-8.

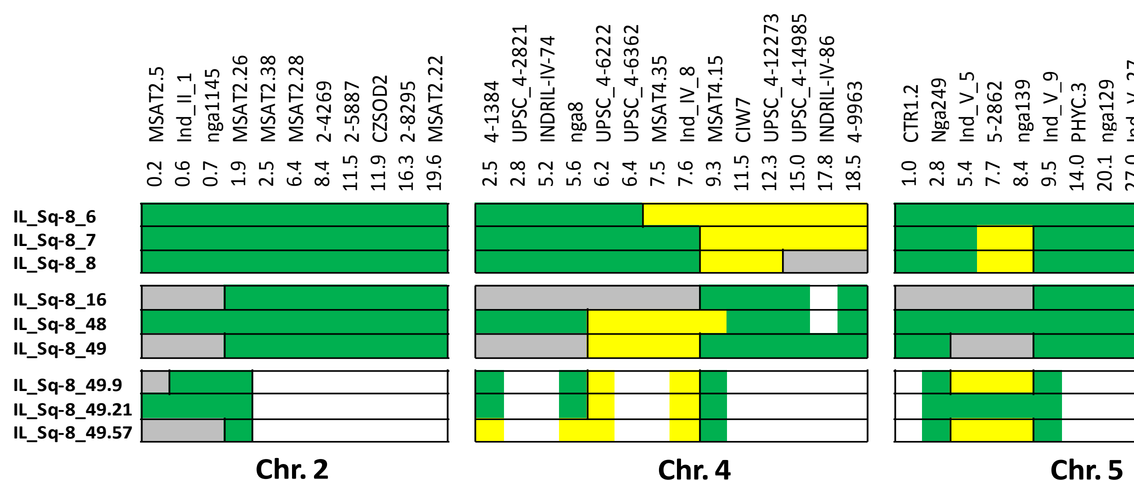


Figure 18. Position and extent of introgressed segments in introgression lines carrying Sq-8 allelic variants of the *HEN1* and/or *WEX* genes. At the top alias names of Indel markers and their position on the sequence map in Mbp are shown. Yellow and dark green boxes represent regions homozygous for the Sq-8 accession genome and Col-0, respectively. Heterozygous areas are indicated in grey. Missing data points are displayed as white boxes. IL_Sq-8_49.9, IL_Sq-8_49.21 and IL_Sq-8_49.57 were established from progeny plants of IL_Sq-8_49.

The newly established derivatives of IL_Sq-8_49, IL_Sq-8_49.9, IL_Sq-8_49.21 and IL_Sq-8_49.57 were analysed with respect to gene silencing (Figure 19). Chromosomes 2 and 5 were homozygous for the Col-0 genotype in IL_Sq-8_49.21. IL_Sq-8_49.9 and IL_Sq-8_49.57 carried heterozygous introgression segments of different size on chromosome 2 and were homozygous for a Sq-8 introgression on chromosome 5. The chromosome region flanked by markers FRI-IND and UPSC_4-6222 was homozygous for Sq-8 in IL_Sq-8_49.57, instead IL_Sq-8_49.9 and IL_Sq-8_49.21 showed the Col-0 genotype for this segment (Figure 18).

IL_Sq-8_49.21 showed less silencing compared to the other two lines, regardless which time point was analysed. At the first time point silenced plants were observed in all three lines; 4 silenced plants (5.7%) were observed for IL_Sq-8_49.21, whereas 6 (8.6%) and 8 plants (11.4%) were found for IL_Sq-8_49.9 and IL_Sq-8_49.57, respectively. From time points 2 to 5 the number of silenced plants slowly increased in IL_Sq-8_49.21 to 6 plants, in IL_Sq-8_49.9 and IL_Sq-8_49.57 more pronounced increases to 16 and 19 plants were seen, respectively. At the end of the experiment 97% and 96% of the plants were silenced in IL_Sq-8_49.9 and IL_Sq-8_49.57, respectively, in IL_Sq-8_49.21 only 59% silenced plants were observed. Significant differences with respect to FN were found from time points 2 and 5 onwards when IL_Sq-8_49.21 was compared to IL_Sq-8_49.57 and IL_Sq-8_49.9,

respectively. Significant differences between IL_Sq-8_49.57 and IL_Sq-8_49.9 were not found (Figure 19C).

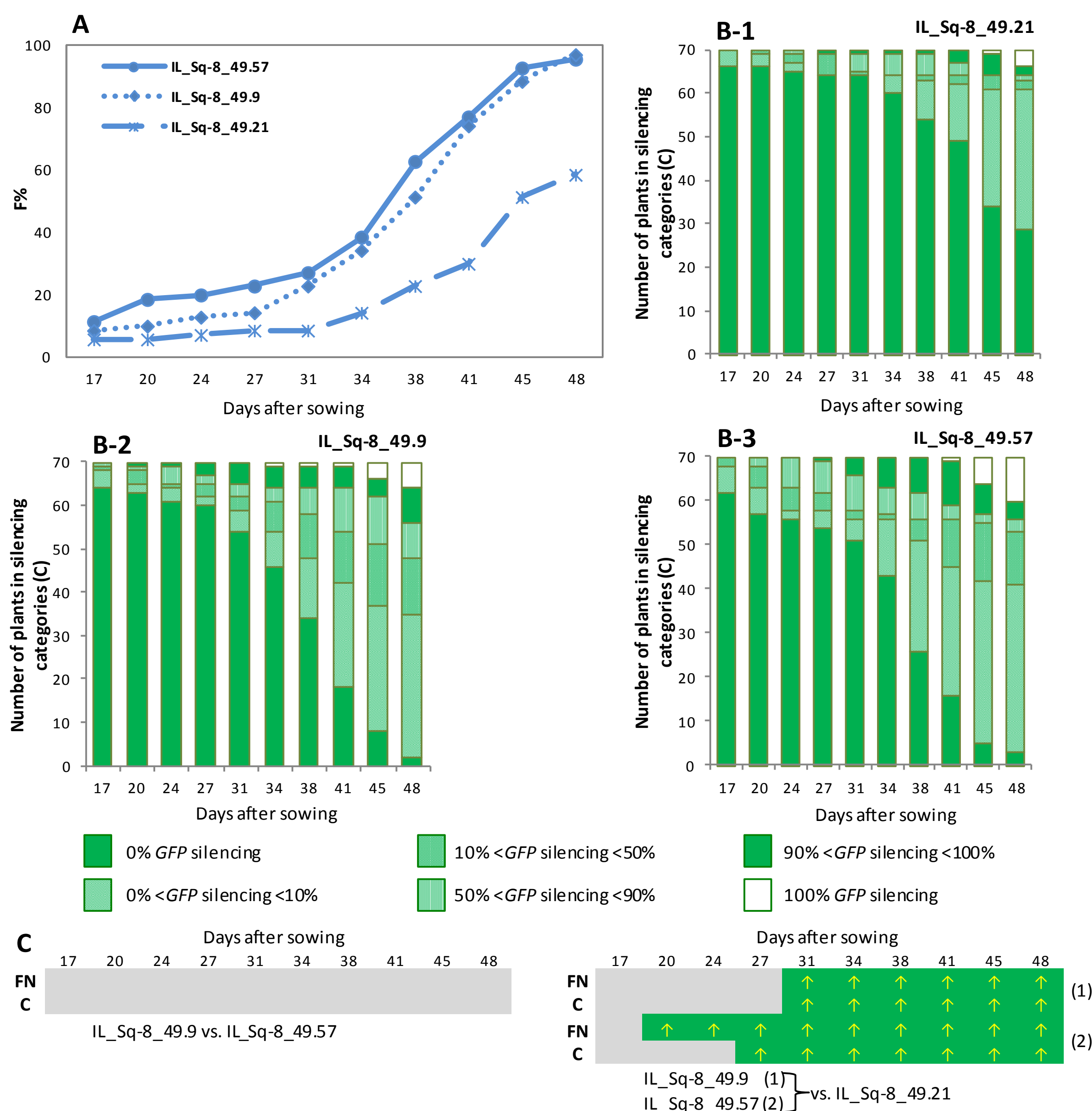


Figure 19. Introgression lines with contrasting genotypes in regions of *Arabidopsis thaliana* chromosomes 2, 4 and 5 show differences with respect to gene silencing. (A) Comparison of frequency of silencing between introgression lines which either carried the Col-0 haplotype or Sq-8 haplotype, respectively, in regions of *A. thaliana* chromosomes 2, 4 and 5 (Figure 18). (B-1 – B-3) Number of silenced plants in different *GFP* silencing categories of IL populations. The silencing categories are marked as shown in the key below the graphs. (C) The tables summarise the results of the statistical analysis. The dark green and grey coloured boxes represent significant and non-significant values, respectively.

Figure 19B shows the differences with respect to *GFP* silencing categories in populations of the three newly established introgression lines for all ten time points. At the first time point, silenced plants were observed in all three lines, however in IL_Sq-8_49.21 only plants of

category 1 were found whereas in the other lines silenced plants of categories 2 and/or 3 were also found. Plants that showed complete *GFP* silencing were found in IL_Sq-8-49.9, IL_Sq-8_49.57 and IL_Sq-8_49.21 from time points 6, 8 and 9 onwards, respectively. Significant differences in terms of C were found from time points 4 and 5 onwards when IL_Sq-8_49.21 was compared to IL_Sq-8_49.57 and IL_Sq-8_49.9, respectively.

3.2.12 Analysis of introgression lines carrying allelic variants of the *NRPE1* gene

Nine introgression lines were established that carried Cvi-0, Kas-1 or Shahdara alleles of the *NRPE1* gene. However, only two out of three lines with introgressions of the accession Shahdara, IL_Shahdara_6 and IL_Shahdara_10, and one out of three lines carrying introgressions of the Cvi-0 accession, IL_Cvi-0_6/6.25, showed significantly more silencing when compared to the reference line. The remaining six lines, among them all three lines with introgressions from the accession Kas-1, showed a silencing behaviour that was not significantly different to that of line 6x*GFP-F8/R127* (Table 18).

The data that were obtained in experiment 06-b-14 for IL_Shahdara_6 and IL_Shahdara_10 are displayed in Figure 20. Both lines showed significantly more silencing than the reference line. However, on average silencing in the IL_Shahdara_10 population was initiated earlier and in more plants than in IL_Shahdara_6, regardless which time point was analysed. For the first three time points, frequency of silencing in IL_Shahdara_10 was between 2.9% and 10% higher than in IL_Shahdara_6 or the reference line. During the following four time points, the percentage of silenced plants in IL_Shahdara_10 raised quickly from 24% to 97%. In contrast, during this phase in IL_Shahdara_6 and in line 6x*GFP* the percentages of plants showing silencing increased from approximately 13% to 54% and 4.3% to 20%, respectively. At the end of the experiment, all plants in the IL_Shahdara_10 population, 90% of the IL_Shahdara_6 population and 49% of plants in the reference line population showed silencing.

Comparison of the data for the different silencing categories at the ten time points also revealed differences between the lines (Figure 20). For example, plants in all six categories were found in the population of IL_Shahdara_10. In the other two lines only plants of categories 1 to 3 were observed, but the proportion of plants belonging to category 1 in IL_Shahdara_6 was for time points 2 to 9 always higher than in the reference line.

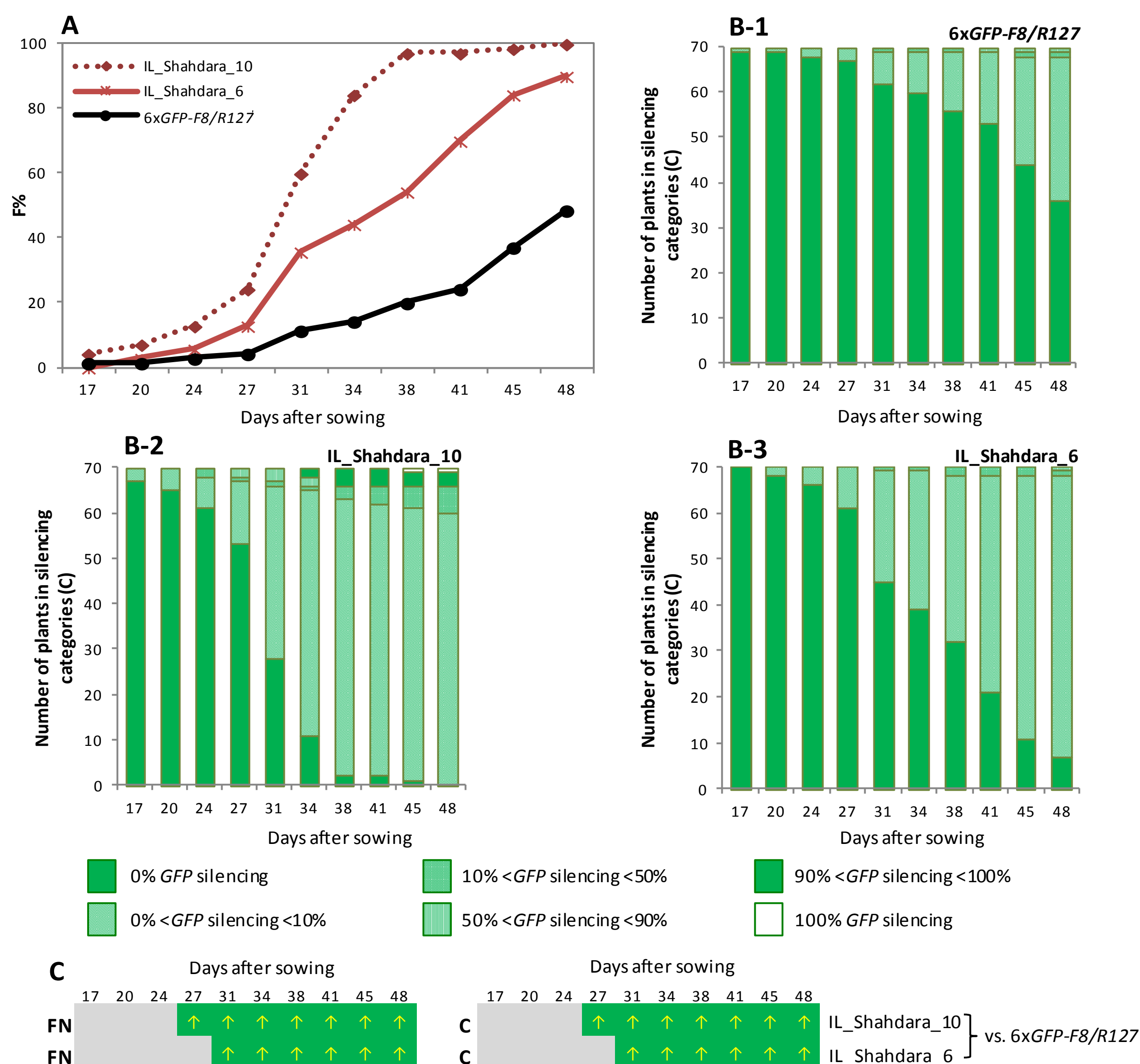


Figure 20. IL_Shahdara_10 showed more silencing than IL_Shahdara_6. (A) Frequency of silencing is displayed for introgression lines and reference line 6x*GFP-F8/R127*, the data were established in experiment 06-b-14. (B-1 – B-3) The number of silenced plants in the different categories are shown for the different plant populations. The categories are indicated according to the key shown below the graphs. Panel C summarises the statistical results with respect to FN and C. The dark green and grey coloured boxes indicate significant ($p<0.05$) and non-significant values, respectively. The arrows (↑) illustrate significantly more silencing in the introgression line than in reference line 6x*GFP-F8/R127*.

Statistical analysis of the FN and C data revealed that the plants of IL_Shahdara_10 showed significantly more silencing than the reference line for time points 4 to 10, in case of the IL_Shahdara_6 population significantly more silencing was found for time points 5 to 10 (Figure 20C). Significantly increased silencing of line IL_Shahdara_10 was also seen in experiment 04-14 when compared to 6x*GFP-F8/R127*, in this case significant differences with respect to FN and C were found for the last eight time points. In contrast, line

IL_Shahdara_30, that was analysed once did not show significant differences in comparison to the reference line (Table 18).

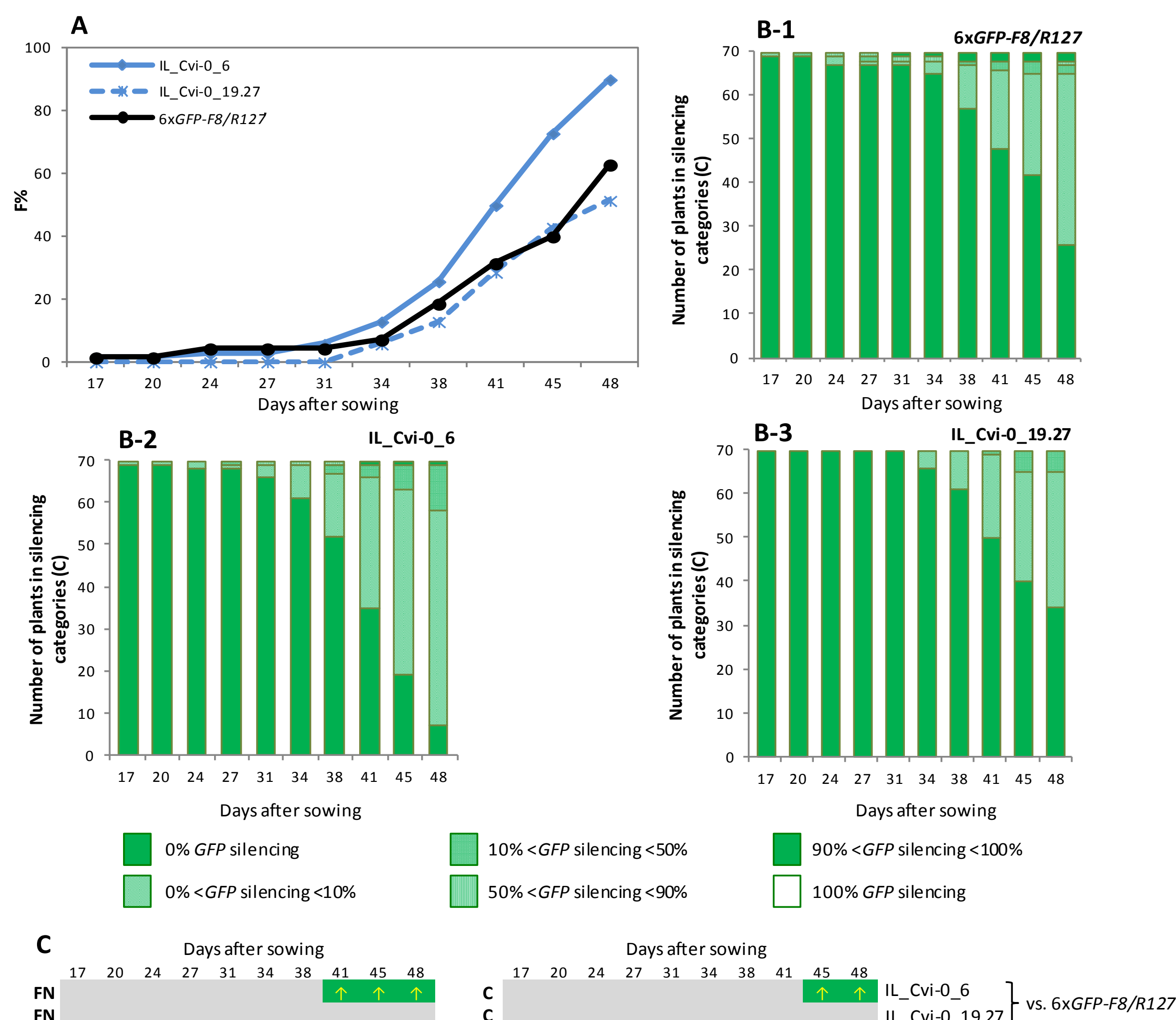


Figure 21. Significantly increased silencing in one of two introgression lines carrying the Cvi-0 allelic variant of the *NRPE1* gene. (A) The proportion of plants showing silencing at ten different time points is shown for two introgression lines and the reference line 6x*GFP-F8/R127*. The data were established in experiment 12-b-14. (B-1 – B-3) Number of silenced plants in different *GFP* silencing categories of two IL populations and a reference line. The different categories are displayed as indicated in the key below the graphs. Results of the statistical analysis of the FN and C values are shown below the graphs. The dark green and grey coloured boxes represent significant and non-significant values, respectively. The arrows (↑) illustrate significantly higher number of silenced plants in the introgression line than in reference line 6x*GFP-F8/R127*.

Analysis of two introgression lines established for the Cvi-0 allelic variant, IL_Cvi-0_6 and IL_Cvi-0_19.27, was carried out in experiment 12-b-14 (Figure 21). IL_Cvi-0_6 displays a higher proportion of silenced plants for time points 6 to 10 when compared to the reference line 6xGFP-*F8/R127*. The values differ from approximately 1.4% to 27%. For the last three time points of the experiment, the results for these lines are significantly different. The proportion of plants showing silencing in line Cvi-0_19.27 differed at the different time

points by 1.4% to 11% from the values observed for the reference line. None of the differences between these two lines were significant. The data for the silencing categories are illustrated in Figure 21B. Taking the results of all time points into consideration, plants of categories 0 to 4 were found in populations of the reference line and IL_Cvi-0_6, whereas in IL_Cvi-0_19.27 only plants of categories 0 to 2 were observed. The total number of silenced plants classified in categories 1 and 2 was 41 for the reference line at the end of the experiment. This value was significantly lower than that determined for IL_Cvi-0_6 with 62 but similar to that of IL_Cvi-0_19.27 with 36.

The results obtained in experiment 12-b-14 (Figure 21) were corroborated for line IL_Cvi_6/6.25 in three additional experiments, 04-14, 06-b-14 and 02-15, and for line IL_Cvi-0_19.27 in experiment 02-15. A third introgression line carrying the Cvi-0 allelic variant of the *NRPE1* gene, IL_Cvi-0_39, was analysed only once and did not show significant differences when compared to the reference line (Table 18).

3.2.13 Identification of genome regions in the Shahdara and Cvi-0 introgression lines which enhance post-transcriptional gene silencing

Out of the six introgression lines established for the Cvi-0 and Shahdara alleles of the *NRPE1* gene only lines IL_Cvi_6/6.25, IL_Shahdara_6 and IL_Shahdara_10 showed significantly more silencing than the reference line towards the end of the experiments (Table 18). These results suggested that the effect on gene silencing is most probably not due to the presence of the allelic variants of the candidate gene *NRPE1* itself. The characterisation of the introgression lines by Indel markers revealed differences with respect to the position and size of the introgressed segments (Supplementary figure 5; Table 14). The three introgression lines which carried the Shahdara allelic variant showed apart from introgressions located on chromosome 2 also another large one on chromosome 3 each. For the three introgression lines carrying the Cvi-0 haplotype only one introgression each on chromosome 2 was detected. Notably, all three introgression lines that showed enhanced silencing when compared to the reference line had introgressions in overlapping regions of the genome that correlated with the enhanced silencing. In case of the Shahdara introgression lines markers 2-3475 and MSAT2.17 flank the part of the introgressions, which is present in the introgression lines showing enhanced silencing, IL_Shahdara_6 and

IL_Shahdara_10, but in IL_Shahdara_19.30 with a similar silencing behaviour as the reference line this area was homozygous for Col-0. IL_Cvi_6/6.25 that showed more silencing than the reference line carried a region flanked by markers MSAT2.28 and MSAT2.17 homozygously for the Cvi-0 genotype (Figure 22) whereas in the lines that showed a similar silencing behaviour as the reference line, IL_Cvi-0_19.27 and IL_Cvi-0_39, this segment was heterozygous.

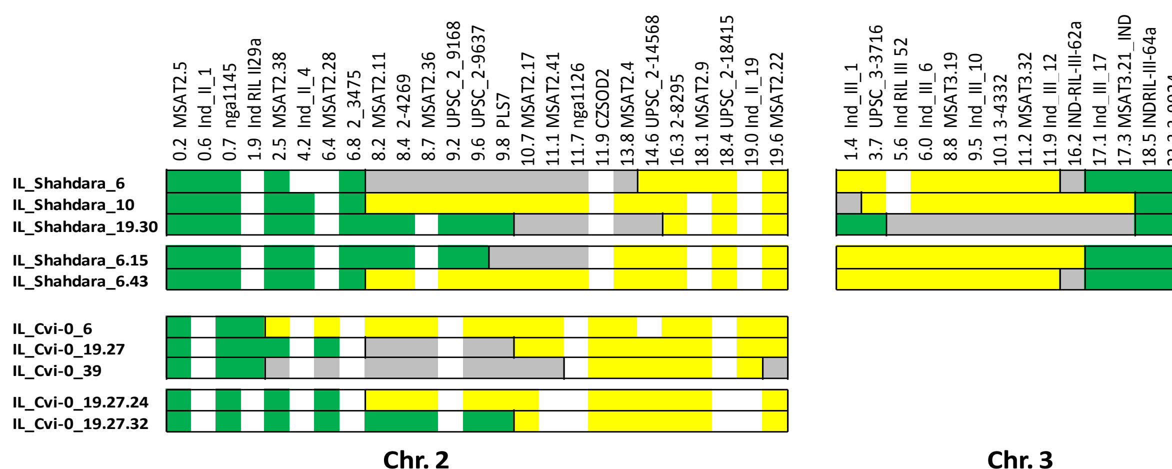


Figure 22. Position and extent of introgressed segments in introgression lines carrying allelic variants of the *NRPE1* gene. Alias names of Indel markers and their position on the sequence map in Mbp are shown at the top. Regions homozygous for the Cvi-0 or Shahdara accession genomes and Col-0 are shown as yellow and dark green boxes, respectively. Grey boxes indicate heterozygous areas. Missing data points are displayed as white boxes. IL_Shahdara_6.15 and IL_Shahdara_6.43 as well as IL_Cvi-0_19.27.24 and IL_Cvi-0_19.27.32 were established from IL_Shahdara_6 and IL_Cvi-0_19.27, respectively.

In order to assess whether the chromosome 2 segment which is located upstream of the *NRPE1* gene caused the enhanced silencing, IL_Shahdara_6 and IL_Cvi-0_19.27, which carried the segment of interest heterozygously were selected for further analysis. Molecular markers polymorphic for the Shahdara or Cvi-0 accession in comparison to Col-0 were used to screen among the progeny of these lines for plants that carried the segment of interest either homozygously for the Col-0 genotype or for the Shahdara and Cvi-0 genotypes, respectively.

These newly established lines were analysed with respect to gene silencing in a pairwise fashion. IL_Shahdara_6.43 and IL_Cvi-0_19.27.24 carried the segment of interest homozygously for Shahdara and Cvi-0, respectively, they showed enhanced silencing when compared to IL_Shahdara_6.15 and IL_Cvi-0_19.27.32 that carried the Col-0 region of interest (Figure 23). In case of the two newly established introgression lines derived from IL_Shahdara_6, significantly more silencing was seen in plants of the IL_Shahdara_6.43

population already at day 27 after sowing when compared to the IL_Shahdara_6.15 population. The proportion of plants showing silencing increased rapidly in introgression line IL_Shahdara_6.43, 79% and 100% of the plants showed silencing at 38 and 48 days after sowing, respectively. In contrast, the number of plants exhibiting silencing in introgression line IL_Shahdara_6.15 increased more slowly, only 26% and 51% of the plants showed silencing at these time points. Significant differences with respect to FN were seen from the eighth time point onwards between the lines derived from IL_Cvi-0_19.27 (Figure 23). From time points 7 to 10 the percentage of silenced plants in populations of IL_Cvi-0_19.27.32 and IL_Cvi-0_19.27.24 increased from 11% to 74% and from 2.9% to 21%, respectively.

In IL_Shahdara_6.43, the first silenced plant was found at the second time point, during the scoring period plants belonging to categories 0 to 4 were observed. In contrast, in the population of IL_Shahdara_6.15 a silenced plant was first observed at the fifth time point, in the course of the experiment only plants of categories 0 to 2 were found. Significant differences between the lines with respect to C were observed from time point 4 onwards. In the population of IL_Cvi-0_19.27.24 the first silenced plant was seen at the second time point, during the entire experiments only plants of categories 0 to 2 were found. This was also the case for IL_Cvi-0_19.27.32, but in this case the first silenced plant was only observed at time point 5. For time points 8 to 10 significant differences between the two introgression lines with respect to C were observed.

It is important to note that IL_Cvi-0_19.27.32 showed significantly less silencing than the reference line at the last four time points of experiment 04-15, IL_Cvi-0_19.27.24 showed significantly more silencing than 6x*GFP-F8/R127*, but only at time point 10. Line IL_Shahdara_6.43 revealed significantly more silencing than the reference line 6x*GFP-F8/R127* at the last six time points both with respect to FN and C, whereas this was not the case for IL_Shahdara_6.15 in experiment 12-b-14 (data not shown). More silencing in IL_Shahdara_6.43 than in the reference line was also observed in experiment 02-15. From time points 3 and 4 onwards significant differences were observed for the C and FN data, respectively. In contrast, IL_Shahdara_6.15 showed a similar silencing behaviour as reference line 6x*GFP-F8/R12*. Significantly increased silencing of IL_Shahdara_6.43 when compared to IL_Shahdara_6.15 was also observed in experiment 02-15. Significant differences with respect to FN and C were seen for time points 3 to 10 (data not shown).

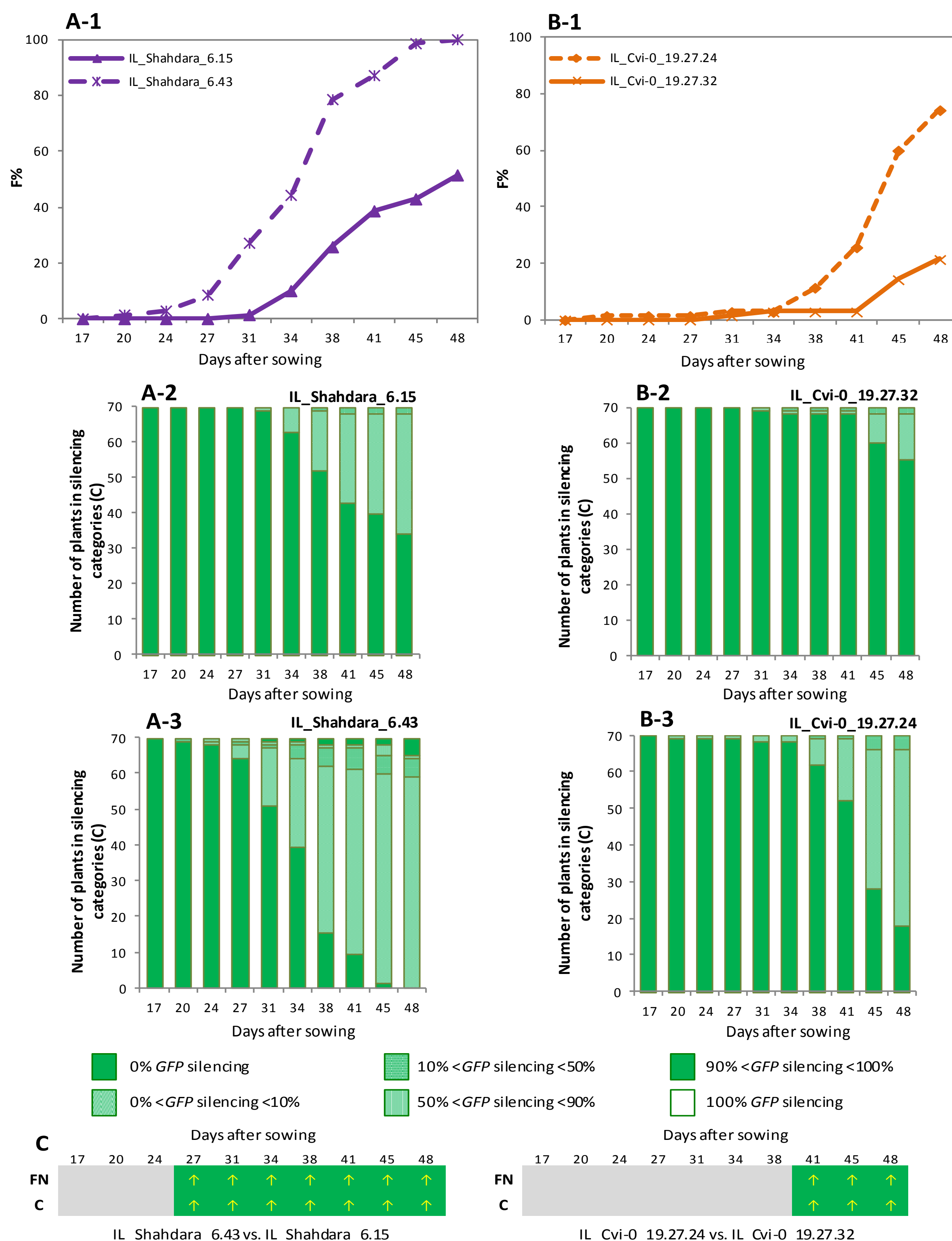


Figure 23. Introgression lines with contrasting genotypes in a region of *Arabidopsis thaliana* chromosome 2 show differences with respect to gene silencing. (A-1, B-1) Comparison of frequency of silencing between introgression lines which either carried the Col-0 haplotype or Shahdara and Cvi-0 haplotype, respectively, in a region of *A. thaliana* chromosome 2. The data for the Shahdara and Cvi-0 introgression lines were established in experiment 12-b-14 and 04-15, respectively. (A-2, A-3, B-2, B-3) Number of plants in different *GFP* silencing categories is displayed. The silencing categories are marked as shown in the key below the graphs. (C) The tables summarise the results of the statistical analysis. The dark green and grey coloured boxes represent significant (p<0.05) and non-significant values, respectively.

4 DISCUSSION

4.1 Choice of candidate genes

Many genes involved in RNA silencing pathways have been described (Brodersen and Voinnet, 2006; Mallory and Vaucheret, 2010; Matzke and Mosher, 2014). In order to select candidate genes for this study preference was given to genes for which an involvement in sense-transgene induced PTGS had been demonstrated, since it was intended to use *GFP* transgenes prone to S-PTGS for the functional analysis of alleles. Several components involved in the S-PTGS pathway are also playing a role in other silencing pathways, for example in the biogenesis of ta-siRNAs (Brodersen and Voinnet, 2006; Mallory and Vaucheret, 2010), hence allelic variants with an impact on S-PTGS may also affect other RNA silencing phenomena. The genes were selected such that different steps in the S-PTGS pathway were covered; SGS3 and SDE5 are for example involved in the formation of dsRNA, DCL4 and HEN1 perform processing of dsRNA and modification of siRNAs, respectively. Factors of importance for silencing spread such as NRPD1 and NRPE1 were also included (Brodersen and Voinnet, 2006; Mallory and Vaucheret, 2010; Mermigka *et al*., 2015).

4.2 Polymorphism patterns of twelve genes associated with PTGS in 25 *Arabidopsis thaliana* accessions

The number of studies describing sequence variation in *A. thaliana* accessions are increasing rapidly, including those that compare entire genomes of accessions (Nordborg *et al*., 2005; Clark *et al*., 2007; Ossowski *et al*., 2008; Zeller *et al*., 2008; Cao *et al*., 2011; Gan *et al*., 2011; Schneeberger *et al*., 2011; 1001 Genomes Consortium, 2016). These studies revealed not only the diversity of selected *A. thaliana* accessions (Clark *et al*., 2007; Zeller *et al*., 2008; Cao *et al*., 2011; Gan *et al*., 2011; Schneeberger *et al*., 2011) but also showed which classes of genes diverge particularly rapidly (Bakker *et al*., 2006; Clark *et al*., 2007; Zeller *et al*., 2008).

One goal of the work presented here was a description of allelic diversity of genes that are involved in the S-PTGS pathway. It was of particular interest whether any of the candidate genes chosen showed in one or more accessions a high level of sequence divergence when compared to the sequence of *A. thaliana* reference accession Col-0 (Arabidopsis genome initiative, 2000). Sequence variation was assessed for twelve genes in 26 *A. thaliana* accessions including Col-0 by amplicon sequencing, since at the time this project was started

sequence information provided by the 1001 Genome project (Weigel and Mott, 2009) was only available for few accessions; the Col-0 reference genome served as basis for amplicon design. The 26 accessions used capture portions of the genetic diversity found in *A. thaliana* (Baxter *et al*., 2010; Platt *et al*., 2010; Supplementary figure 1).

It was the aim to analyse the majority of the coding regions including intron sequences, due to their length the genes had to be divided into several amplicons each. Comparative analyses of promoter regions were not performed. PCR amplification was observed for 67 out of 70 amplicons in all accessions analysed, 95.7%. Only for amplicons AGO7-1, HEN1-5 and WEX-3 amplification products for selected accessions were repeatedly not obtained, however, the use of alternative amplicons that was facilitated by the availability of whole genome shotgun sequences of selected accessions (Schneeberger *et al*., 2011) resulted in amplification products for these accessions. Based on the PCR-based analyses the twelve candidate genes were therefore present in all accessions tested. In contrast, in the *RDR* gene cluster that is comprised of three genes in Col-0, deletions of the *RDR4* gene were noted in three out of the same 25 accessions used here (Thanh Loan Le, unpublished results). Unambiguous sequence information was obtained for all amplicons and accessions, thus indications for copy number variation were not observed.

Gene models supported by cDNA sequences were available for the Col-0 reference sequence of all candidate genes (Supplementary table 9), based on this information sequence differences were assigned to exon and intron sequences. All Col-0 amplicon sequences that were generated for the twelve candidate genes were identical to the sequence of the reference genome, evidence for potential errors in the reference sequence was therefore not found. In seven accessions (28%) SNPs were identified in all twelve candidate genes; in 13 (52%), one (4%) and four (16%) accessions SNPs were revealed in 11, ten and nine genes, respectively (Table 6). In total, 828 SNPs were found, approximately half were located in exons (Table 7). Indel polymorphisms relative to the reference sequence Col-0 were also identified in the analysed regions of the candidate genes, except for *SGS3* (Table 8). In the analysed gene and exon sequences of the twelve genes 130 and 34 Indels were found in the 25 accessions, respectively. As noted previously, Indels occurred less frequently than SNPs (Nordborg *et al*., 2005; Zeller *et al*., 2008; Ganz *et al*., 2011). The candidate genes differed with respect to the frequencies of SNP sites. In the *AGO7* and *WEX* genes the frequency of

SNP sites in the 25 accessions amounted to 3.9 and 8.5%, respectively. For another eight candidate genes several-fold lower values of around 1% or even less were found and for the *HEN1* and *NRPE1* genes values of 2.0 and 1.8% were noted. Transitions were with 55% and 59% more prevalent than transversions in gene and ORF regions, respectively. Synonymous substitutions were with 55% found more often than nonsysnonymous substitutions (Table 7). On average, the frequency of nonsynonymous substitutions amounted to 0.56% for the twelve candidate genes in the 25 accessions analysed. The lowest values were found with 0.07% for *AGO1* and *XRN4*. Values above average were noted with 2.73%, 1.07%, 0.92% and 0.63% for the *WEX*, *AGO7*, *HEN1* and *NRPE1* genes, respectively (Tables 5 and 7). In a study of sequence variation that had been carried out in 30 accessions of *A. thaliana* for five starch synthase genes the frequencies of nonsynonymous substitutions in the analysed exon sequences ranged from 0.46% to 1.1%, with an average of 0.66% (Schwarte *et al*., 2015). Thus, the study of the candidate genes for PTGS revealed much larger differences with respect to the frequencies of nonsynonymous substitutions, despite that fewer accessions were studied. Evidence for the fast evolvement of some genes involved in RNA silencing had been documented previously. A comparison of the *A. thaliana* and rice *NRPD1* and *NRPE1* genes with the corresponding subunit in Pol II revealed that rates of amino acid substitutions in the Pol IV and Pol V subunits were approximately 20 times faster than in the Pol II subunit (Luo and Hall, 2007).

The study of sequence variation in 20 *A. thaliana* accessions had revealed that as much as 9% of the genes were affected by polymorphisms that altered the exon/intron structure in certain accessions (Clark *et al*., 2007). A case of a major effect polymorphism in a gene involved in RNA silencing had been described in *Nicotiana benthamiana*. A natural variant of a salicylic acid-inducible RNA-dependent RNA polymerase was found that contained an Indel with in-frame stop codons in the 5'-region of the open reading frame, the abolished function of this particular *RDR* rendered the plant hypersusceptible to viruses (Yang *et al.*, 2004). Approximately 85% and 94% of the gene and exon regions, respectively were evaluated on average in the candidate genes (Table 5), therefore it was not possible to analyse all exon/intron borders, ATG and stop codons of the candidate gene alleles in this study. The region around the ATG was for example sequenced in all accessions for six genes, *AGO1*, *ERI*, *NRPD1*, *NRPE1*, *SGS3* and *SDE5*, and the area encompassing the stop codon for four genes,

ERI, *NRPE1*, *WEX* and *XRN4*. In the established data set evidence for one Indel polymorphism was identified that resulted in a frame shift, a deletion in the coding region of the Shahdara *SDE3* allele changed the carboxy terminus of the deduced protein sequence. Two premature stop codons were also determined, they truncated the Cvi-0 and Shahdara alleles of the *NRPE1* gene by 51 and two amino acids, respectively (Supplementary table 8).

An analysis of the four candidate genes in which particularly high frequencies of SNP sites were found revealed substantial differences between individual accessions (Table 6). For example, accessions Bor-4 and Gie-0 showed SNP frequencies of 3.2% for the *AGO7* gene whereas for the other 23 accessions values smaller than 0.2% were established. In all cases in which accessions showed for particular candidate genes SNP frequencies of around 1% or higher these values differed at least threefold from the values observed for the remainder of accessions. Based on these results all accessions that showed SNP frequencies of around 1% or higher were considered for further studies. In case of the *AGO7*, *HEN1* and *NRPE1* genes between two and three accessions fulfilled this criterion, but different accessions were affected in the three genes (Table 6), all were included in additional analyses. The Cvi-0, Kas-1 and Shahdara *NRPE1* alleles were not only characterised by many SNPs, Kas-1 and Shahdara also contained the largest Indels that were observed in exon regions in this study (Table 4) and the Cvi-0 allele was truncated by a premature stop codon. Nineteen of the accessions revealed SNP frequencies of 1.2% or higher for the *WEX* gene, for this candidate gene only the six accessions with the highest values were taken into account for further studies (Table 9). When the *AGO1*, *DCL4*, *ERI*, *SDE5*, *SGS*3 and *XRN4* genes were analysed none of the individual accessions showed SNP frequencies larger than 0.5%. Three accessions exceeded this value when the *SDE3* gene was analysed, in case of the *NRPD1* gene this was observed for all accessions studied. Despite the fact that the differences between SNP frequencies of different accessions was not as pronounced as for the *AGO7*, *HEN1*, *NRPE1* and *WEX* genes several accessions were also studied further for *NRPD1* and *SDE3*. Due to the presence of a 51 bp long Indel in an exon of the *SDE3* gene, accessions Baa-1 and Ws-0 were also included in additional analyses (Table 9).

4.3 Expression analysis of selected alleles

Promoter regions had not been included in the sequence comparisons of the candidate genes in the different accessions. It was therefore deemed important to analyse the transcript levels of the candidate genes in the different accessions and to compare them to the expression in Col-0. Rather than analysing all accessions for expression of the candidate genes, the comparative analysis was focused on those alleles and accessions that had been selected for further studies (Table 9). Since the selected allelic variants showed considerable sequence variation when compared to the Col-0 reference sequence (Tables 6 and 8) it was mandatory that the oligonucleotides of the RT/qRT-PCR-amplicons were placed in monomorphic regions. In aerial seedling tissues all six selected genes were expressed in accession Col-0 as well as in the 13 accessions (Figure 7). In eight out of the 17 cases in which an allelic variant was compared to the Col-0 counterpart steady-state transcript levels differed less than twofold (Figure 8). All but one of the allelic variants of the *AGO7*, *HEN1* and *SDE3* genes belonged to this class, only the Ws-0 variant of the *SDE3* gene showed a 2.5 fold reduction when compared to Col-0. However, the Ws-0 and Baa-1 alleles that were identical in sequence in the analysed region of the *SDE3* gene differed with respect to steady-state transcript levels by a factor of 2.4, thus almost as much as the Ws-0 and Col-0 alleles. All three of the *NRPE1* allelic variants analysed showed reduced transcript levels when compared to Col-0, the reductions ranged from 2.9-fold for the Kas-1 allele to 5.6-fold for the Cvi-0 allele. The candidate gene for which the highest degree of sequence differences had been found also revealed pronounced differences with respect to transcript levels for some of the alleles analysed. The Sq-8 and Shahdara alleles showed approximately fourfold higher and lower steady-state transcript levels when compared to the Col-0 allele, respectively. Moreover, the allelic variant of the *WEX* gene in accession Kin-0 showed in two out of the three biological replicates a drastically reduced steady-state transcript level. Preferably, additional biological replicates should be analysed for this accession.

4.4 Analysis of introgression lines with Indel markers

Based on the comparative sequence analysis 19 allelic variants corresponding to six candidate genes were selected for functional analysis (Table 9). In total, 51 introgression lines were established, for each of the alleles at least two lines were generated. Indel

markers were used to assess the position and size of introgressed segments in plants of the BC_4F_2 generation that carried the allele of interest homozygously and in addition two homozygous T-DNA loci that together carried six copies of the *GFP* gene.

The majority of Indel markers used in this study had been described previously for *A. thaliana*, but data with respect to polymorphism information was in most cases only available for two accessions (Loudet *et al*., 2002; Salathia *et al*., 2007; Hou *et al*., 2010). Only in one of the studies polymorphism information had been compiled for seven accessions (Păcurar *et al*., 2012). Among the 146 markers tested, 31 showed polymorphisms between all thirteen accessions which had been used to establish introgression lines and Col-0, only five of the markers were monomorphic. More than half of the markers showed polymorphisms in more than 75% of the analysed accessions whereas only 14.4% revealed polymorphisms relative to Col-0 in less than 25% of the thirteen accessions (Supplementary table 10). Thus, many of the Indel markers that had been identified in other studies to be polymorphic between two or few accessions were suitable to detect polymorphisms between the thirteen accessions and Col-0. Nonetheless, differences with respect to the degree of polymorphism were observed between accessions, the average distance between polymorphic markers in the accessions analysed varied between 1.1 and 1.4 Mbp (Supplementary tables 4 and 8). The marker density was different on the five chromosomes since Indel marker screening had been focused on the three chromosomes that carried the allelic variants targeted for introgression mapped (Figure 9).

It should be noted that 22 out of the 146 Indel markers, 15.1%, were not amplified in some accessions. In total, 48 presence/absence polymorphisms were observed (Supplementary table 10). Such polymorphisms were only useful to distinguish between regions homozygous for Col-0 and a particular accession, but it was not possible to discriminate regions that were homozygous for Col-0 and heterozygous for both accessions. Few markers that showed clearly discernible polymorphisms between Col-0 and a particular accession were not suitable to identify heterozygous segments in a reliable manner. For example, preferential amplification of the Col-0 allele relative to the alleles in accessions Shahdara and Cvi-0 did not permit the identification of heterozygous genotypes in case of marker Ind_II_9.

In order to establish the number, position and size of introgressed segments in the different introgression lines only subsets of the polymorphic markers were used (Table 13). It was the aim to assess polymorphic markers every 4 to 5 Mbp, but in case an introgression segment was identified in a particular genomic region, additional markers were used in order to delimit the introgressed region as precisely as possible. Using between 44 and 65 polymorphic markers (Table 14) the average distances between Indel markers ranged from 1.9 in several Ws-0 and Kin-0 introgression lines to 2.9 Mbp in IL_Kas-1_39 (Figure 11, Supplementary figures 2-5). The preferential screening and use of Indel markers that were located on chromosomes which harboured the candidate alleles is reflected in the higher resolution of the graphical genotypes on chromosomes which carried a particular allelic variant of interest. The average distances between analysed markers ranged from 1.0 to 2.0 Mbp on those chromosomes (Figure 11, Supplementary figures 2-5). Intervals between adjacent polymorphic markers larger than 5.0 Mbp were part of the different maps, but their number was small, between one and five such segments were observed in the different introgression lines (Figure 11; Supplemental figures 2-5). In summary, apart from few areas in certain introgression lines, introgression larger than 5 Mbp were readily detectable using a subset of the assembled Indel marker set.

For 22 out of 51 (43.1%) introgression lines only one introgression segment was detected, whereas 18 (35.3%), eight (15.6%) and three lines (5.9%) contained two, three and four segments, respectively (Table 14). Similar results were observed in the study of Keurentjes *et al*. (2007), they found that 40 out of 92 (43.4%) introgression lines carried a single introgressed segment, the remainder contained between two and four segments. In contrast, for two reciprocal sets of introgression lines that together consisted of 140 ILs overall fewer segments per line were reported (Törjék *et al*., 2008).

After the fourth backcross, 96.9% of the genome are expected to be contributed by the recurrent parent. To estimate the sizes of the introgression segments it was assumed that the recombination breakpoints were present midway between the positions of two markers. In only seven out of the 51 ILs small introgressions were observed which represented between 1.4% and 3.2% of the donor genome. In 23 and 14 lines the recurrent parent contributed between 90.9% to 96.4% and between 80.9% to 89.8% of the genome, respectively. In five of the remaining seven lines, the introgression segments covered

between 22% and 25.5% of the genome. Two lines which carried the Kin-0 allelic variant of the *WEX* gene had with four introgressed segments not only the highest number of introgressed segments but with 64% and 54% also by far the smallest proportions of the recurrent parent genome. The mean size of the different introgressed segments in the 51 established lines was 6.9 Mbp, segment sizes ranged from 0.4 to 18.1 Mbp (Table 14).

In the study of Törjék *et al*. (2008) smaller average sizes of 4.5 and 5.0 Mbp had been established for two sets of reciprocal introgression lines. However, when these authors analysed BC_3F_1 lines 18.7% and 19.5% of the genome were represented as heterozygous segments, thus also in this study introgression segments larger than expected for this generation were found. They had selected lines especially suited for the generation of introgression lines, analysed several descendants in the BC_4F_2 generation and used a comparatively high density of markers. This explains why overall fewer segments and of smaller size were obtained than in the current work. In the study presented here the initial focus had to be on the identification of lines that carried the allelic variant of interest and two *GFP* loci homozygously, thus it was not possible to implement steps which would have resulted in fewer and smaller segments in the BC_4F_2 generation. Owing to the same fact for 31 out of the 51 established lines heterozygous segments were present (Figure 11, Supplementary figures 2-5), whereas in the study of Keurentjes *et al*. (2007) and Törjék *et al*. (2008) most and all of the lines carried homozygous introgression segments, respectively.

An introgressed segment was identified using markers which flanked the candidate gene on one or both sides in the Col-0 genome for 48 out of 51 introgression lines (Figure 11, Supplementary figures 2-5). Exceptions were noted for three lines which were descendants of the same BC_4 plant. These lines carried the Kin-0 allelic variant of the *WEX* gene, but the analysis with markers mapping 0.4 Mbp upstream and 1.3 Mbp downstream of the *WEX* gene in the Col-0 genome did not reveal an introgression (Figure 11). It is possible that the introgression segment was too small to be detected, but the results may also be explained by gene conversion. The study of a 170 kbp region on chromosome 4 revealed that gene conversion can make up a substantial proportion of the observed recombination (Haubold *et al*., 2002). It is also possible that the *WEX* gene is located in accession Kin-0 in another genome region as in Col-0, but a homozygous introgression segment in common to all lines containing the Kin-0 allele of the *WEX* gene homozygously was not detected (Figure 11).

4.5 The study of gene silencing in the introgression lines

Col-0 populations carrying six copies of the *GFP* gene under the control of the CaMV 35S promoter showed under the growing conditions used steadily increasing silencing frequencies during plant development. Even at the end of the experiments, at day 48 after sowing, not all plants of a particular population showed silencing. Six *GFP* gene copies were therefore deemed a suitable reporter system for the monitoring of gene silencing (Arlt, 2007; Figure 15). Introgression lines were therefore established that carried six copies of the *GFP* gene at homozygous loci, in order to avoid segregation of the transgenes in the BC_4F_3 generation that was analysed for *GFP* silencing. All introgression lines carried locus *R127* with two *GFP* transgenes (Lechtenberg *et al*., 2003), but differed with respect to a locus that contained one *GFP* copy (Schubert *et al.*, 2004). In each of the experiments introgression lines were analysed side by side with lines carrying six copies of the *GFP* reporter gene in the Col-0 genetic background. Reference line 6x*GFP-F8/R127* was included in all experiments, lines 6x*GFP-F18/R127* and 6x*GFP-F128/R127* only if introgression lines were analysed that carried these particular *GFP* loci.

It had been possible to detect the influence of environmental effects on gene silencing using populations of 70 transgenic *A. thaliana* plants each (Arlt and Schmidt, 2006; Arlt, 2007), hence the same population size was used for the study of the introgression lines. For each analysed line two subpopulations of 35 plants each were investigated in the different experiments for a period of five weeks. At each of the ten scoring time points the frequency of silencing in the population was recorded, moreover the individual plants were classified based on the proportion of aerial tissues that showed silencing. Comparisons either involved different lines or subpopulations of a line (Figure 13). For the statistical evaluation of lines and/or subpopulations number of silenced and non-silenced plants or number of plants in the different silencing categories were assessed with Fisher´s exact test by applying a significance threshold of 0.05, these evaluations were termed “FN” and “C”, respectively.

In accordance with a recent suggestion (Bergelson *et al*., 2016), the identity of the different lines had been validated in all experiments with suitable PCR-based assays. Nonetheless, the analysis of the same introgression line in different experiments usually resulted in different silencing frequencies as documented for Sq-8 introgression lines in Figure 14 and Table 15. Likewise, in line 6x*GFP-F8/R127* that had been analysed in 22 experiments the silencing

frequencies varied at the end of the different experiments from 27.1 to 79.4% (Table 17). This happened even though the analysed plants in the different experiments were in most cases derived from the same seed lot of a particular line. The results of several studies had revealed an effect of temperature and/or light conditions on gene silencing (Szittya *et al*., 2003; Chellappan *et al*., 2005; Arlt and Schmidt, 2006; Arlt, 2007; Kotakis *et al*., 2010; Patil and Fauquet, 2015). The evaluation of *GFP* silencing in the introgression lines was carried out under controlled environment conditions in order to minimise variations with respect to temperature and/or light conditions during the experimental period. But since the plant populations had to be removed from the controlled environment conditions for several hours each week in order to evaluate *GFP* silencing a certain impact of the repeated removal of the plants from the controlled environment conditions cannot be excluded.

On average, 58.5 +/- 15.0% of the plants showed silencing at the end of the experiments in case of line 6x*GFP-F8/R127*, the highest and lowest values observed in individual experiments differed 20.9 and 31.4% from the mean, respectively (Table 17). A comparison of the 6x*GFP-F8/R127* subpopulations revealed that silencing frequencies differed on average by 11.4 +/- 8.3% at the end of the different experiments, in only three cases the values differed by more than 20% (data not shown). Thus, differences between silencing frequencies were larger for plants grown in different experiments than for plants cultivated in the same experiment. It was therefore not meaningful to compare silencing frequencies of individual lines that had been established in different experiments.

Each of the 139 populations that were cultivated in a total of 22 experiments were analysed at ten time points each. For the majority of pairwise comparisons of subpopulations, 1326 (95.4%) out of 1390, significant differences with respect to FN and/or C were not observed (Section 3.2.7). Restricting the analysis to the second half of the experiments revealed significant differences between subpopulations in 55 cases. In 21 populations only a single time point was affected, but 14 populations differed significantly from each other at two to four time points. Significant differences at consecutive time points during the second half of the experiments were with 28 cases even rarer. Such data were found for eight introgression lines. IL_Sq-8_7 was affected in two out of three experiments, the other seven only in a single experiment. Only in the line which had been analysed 22 times, 6x*GFP-F8/R127*, significant differences at consecutive time points were seen in three experiments (08-14, 02-

15, 07-15). Since large differences with respect to silencing behaviour were occasionally even observed for subpopulations it was required to analyse individual lines more than once with respect to gene silencing in order to draw meaningful conclusions.

4.6 Comparisons between Col-0 transgenic lines carrying six *GFP* copies each

Three different lines carrying six *GFP* copies under the control of the CaMV 35S promoter in the Col-0 genetic background were analysed in this study with respect to gene silencing. It was shown previously that independent single-copy T-DNA loci confer comparable steady-state *GFP* transcript levels and GFP fluorescence (Schubert *et al*., 2004). Lines 6x*GFP-F18/R127* and 6x*GFP-F128/R127* were analysed alongside reference line 6x*GFP-F8/R127* in three and seven experiments, respectively (Tables 16 and 17). The silencing frequencies at the last time point of the different experiments ranged from 40.0 to 71.4% for line 6x*GFP-F18/R127* and from 27.1 to 75.7% for line 6x*GFP-F128/R127*. Rather similar values, 27.1 to 79.4%, had been determined for reference line 6x*GFP-F8/R127*. The mean values established for line 6x*GFP-F18/R127* and line 6x*GFP-F128/R127* amounted to 51.9% and 51.6%, respectively, also these values were comparable to the 58.5% that had been established for line 6x*GFP-F8/R127*. In only two out of the nine experiments in which at least two of the 6x*GFP* lines were grown side by side, silencing frequencies at time point 10 were significantly different to each other. Strikingly, in both of these experiments, 01-15 and 03-b-15, one of the 6x*GFP* lines each showed the lowest silencing frequency value out of all experiments. Overall, the results suggest a comparable silencing behaviour of the three different 6x*GFP* lines, consistent with the results of previous studies (Schubert *et al*., 2004). It was therefore justified to use line 6x*GFP-F8/R127* as reference in all experiments, even if introgression lines were analysed carrying loci *F18* or *F128* rather than locus *F8*.

4.7 Assessing introgression lines for an impact on gene silencing

The analysis of subpopulations and of the different 6x*GFP* lines in the Col-0 genetic background had revealed exceptional data in individual experiments. For this reason only the 40 introgression lines that were evaluated at least twice are discussed in the following section (Table 18). In order to assess whether individual introgression lines showed an effect on gene silencing or not it had to be taken into account that silencing frequencies obtained for the same line but in different experiments were hardly comparable. Hence, the results

for a particular introgression line were compared to the line that was evaluated in all experiments, 6x*GFP-F8/R127* in the Col-0 genetic background. Twelve out of the 40 lines that were analysed repeatedly showed in replicated experiments a consistent silencing behaviour in comparison to the reference line. Seven of the lines, 17.5%, did not show any significant difference with respect to FN when compared to the reference line, but for five lines, 12.5%, significantly more silencing was observed for at least two consecutive time points during the second half of the different experiments (Table 18).

For the majority of lines, 28 (70%), deviating results with respect to gene silencing were obtained in individual experiments. Six of the lines, 15%, revealed in one of the experiments significantly more silencing than the reference line and in another one significantly less silencing. An examination of the time points at which significant differences were observed revealed that in several cases only a single time point or non-consecutive time points were affected. Moreover, the time points at which significant differences were seen often did not coincide in the individual experiments. Twenty lines, 50%, showed significant results in at least one of the experiments, whereas in the remaining experiments significant differences were either not detected when compared to the reference line or at single time points only. It would be necessary to carry out additional experiments, preferably with larger population sizes, in order to determine whether some of the effects on gene silencing that were observed in individual experiments can be replicated.

Significant differences with respect to FN had been observed 43 times in the 139 comparisons of subpopulations that were analysed at 10 time points each, 3.1%. Even rarer were significant differences at consecutive time points during the second half of the experiments, such a pattern was only found for nine out of 139 populations. Based on these results, introgression lines were required to show significant differences when compared to the reference line for at least two consecutive time points during the second half of the experiments in order to be considered as candidates for further studies. Moreover, it was demanded that a particular introgression line should either show consistently or at least in the majority of experiments significantly higher or lower silencing frequencies than the reference line when analysed repeatedly. These criteria were met by seven out of the 40 lines, 17.5%. Five lines, IL_Cvi-0_6/6.25, IL_Shahdara_10, IL_Sq-8_7, IL_Sq-8_8 and IL_Sq-8_49, showed significantly more silencing when compared to 6x*GFP-F8/R127* in all experiments in

which they were analysed. Two lines, IL_Gie-0_3a and IL_Gie-0_6/6.18, showed significantly less silencing in all but one of the three and four experiments in which they were studied (Table 18). Thus, the use of six *GFP* gene copies premitted the identification of introgression lines showing significantly more silencing or less silencing than the reference line. Notably, for six of the seven lines with a pronounced effect on gene silencing, at least one other introgression line carrying the same allelic variant also revealed a comparable silencing behaviour.

In the 22 experiments 960 pairwise comparisons were performed between introgression lines that were analysed repeatedly and the reference line. In 218 cases, 22.7%, significant results with respect to FN and/or C were obtained. The different 6x*GFP* lines in the Col-0 genetic background were more similar to each other, 14.5% significant differences were observed for these lines. The comparisons of the subpopulations yielded the lowest value, only 4.6% of the comparisons showed significant differences.

Significant results for FN and C coincided in 159 out of 218 cases, 72.9%, in which the introgression lines were compared to the reference line. Thus, scoring of the lines with respect to the frequency of silencing and the different silencing categories yielded overall similar results. The seven lines for which consistently a pronounced effect on gene silencing had been observed accounted for 21.9% of the comparisons with the reference line, but in this data subset 41.7% of all significant differences were found. Out of the 91 significant differences that were observed in total for the seven lines 83 (91.2%) showed significant results both with respect to FN and C. A large proportion of the observed significant differences were also found by applying a stricter significant threshold of 0.01. In total, 138 significant differences were found among the 960 pairwise comparisons, 80 (58%) of which were accounted for by the seven lines for which effects were observed consistently (data not shown). Thus, the match between the FN and C evaluations and the proportion of significant differences which surpassed a more stringent threshold were both higher for lines for which effects were found consistently than for the remainder of introgression lines.

4.8 Analysis of lines showing a pronounced effect on gene silencing

Candidate genes *AGO7* and *NRPD1*, which were reported to play the role in post-transcriptional gene silencing, map less than 3 Mbp apart on chromosome 1 in *A. thaliana* (Dalmay *et al*., 2000; Hunter *et al*., 2003; Herr *et al*., 2005; Dunoyer *et al*., 2007; Smith *et al*., 2007; Eamens *et al*., 2008;). A pronounced effect on silencing was observed in lines Gie-0_3a

and Gie-0_6/6.18, but not in all experiments. Significantly less silencing for several consecutive time points was found during the second half of experiments 06-14 and 02-b-14, but in experiment 12-14 only line Gie-0_6 showed a significant difference at a single time point. The reference line showed with 41.4% a comparatively low silencing frequency at the end of this particular experiment. An additional repetition in experiment 01-15 again revealed significantly less silencing in both introgression lines when compared to the reference line (Figure 16; Table 18). Both lines showed two introgressions each of very similar or identical size, a homozygous one encompassing the two candidate genes and a heterozygous one which was also located on chromosome 1 (Supplementary figure 4). Pairwise comparisons of the lines among each other only revealed a single significant difference in one of the three different experiments in which they were analysed alongside of each other (data not shown). This result was consistent with the similar or identical structure of the introgressions in these lines that were derived from the same BC_4 plant.

Those parts of the *AGO7* gene that were analysed at sequence level in this study were identical in Gie-0 and Bor-4, the *NRPD1* genes of these two accessions shared between 99.8 and 99.9% identity at the gene, ORF and amino acid levels. Six lines with chromosome 1 introgressions from accession Bor-4 were established. Two lines carried Bor-4 alleles of *AGO7* and *NRPD1* and four lines contained the Bor-4 alleles of *NRPD1*, but the Col-0 alleles of *AGO7* (Supplementary figure 4). In none of the Bor-4 introgression lines consistently more silencing or less silencing than in the reference line was observed (Table 18). Taking into account the high similarity of the Bor-4 and Gie-0 alleles of the two genes it appears unlikely that the pronounced effect on gene silencing that was seen for the Gie-0 but not for the Bor-4 introgression lines was due to one or both of the candidate genes. In future, it will be important to clarify which of the two introgressions in the Gie-0 ILs caused the observed effect on gene silencing. Two genes which were implicated in PTGS pathways previously potentially map to the region which was present heterozygously in the Gie-0 introgression lines, *THO2* (Francisco-Mangilet *et al.*, 2015) and *UPF3* (Moreno *et al.*, 2013). In case the observed effect on S-PTGS in the Gie-0 ILs should be due to this introgression it will be important to clarify whether it carried the Gie-0 alleles of *UPF3* and/or *THO2*.

Three lines each had been established for the Sq-8 *HEN1* and *WEX* allelic variants as well as the Cvi-0 and Shahdara alleles of the *NRPE1* gene. In three of the four sets, two introgression

lines each had revealed a pronounced enhancing effect on gene silencing whereas the remaining line showed a silencing behaviour that was not significantly different to the reference line. In case of the Cvi-0 lines only one out of the three revealed significantly more silencing than the reference line (Table 18, Figure 21). The homozygous presence of the allelic variants had been confirmed for all lines, thus the deviating silencing behaviour in individual lines was not explained by the presence or absence of the allelic variant of the candidate gene. Graphical genotypes had been established for all lines carrying a particular allele thus it was possible to examine whether the differences with respect to gene silencing in individual lines correlated with the genotypes of particular regions of the genome.

In the lines carrying the Sq-8 *HEN1* or Sq-8 *WEX* allele a region derived from chromosome 5 of Sq-8 was present in all lines showing increased silencing, IL_Sq-8_7, IL_Sq-8_8, IL_Sq-8_16 and IL_Sq-8_49. In contrast, in lines with a silencing behaviour not significantly different to the reference lines, IL_Sq-8_6 and IL_Sq-8_48, the entire chromosome 5 was derived from Col-0. It is tempting to speculate that the same locus caused an increase in gene silencing in the subsets of the Sq-8 *HEN1* and Sq-8 *WEX* introgression lines (Figure 18). These results demonstrated that lines carrying multiple introgression segments offered the opportunity to assess segments of the accession genome which had not been targeted for introgression with respect to an effect on gene silencing.

All three lines carrying the Cvi-0 allelic variant of the *NRPE1* gene had introgressions of varying length on chromosome 2. The region flanked by markers MSAT2.28 and MSAT2.17 correlated with the deviating silencing response in the three lines. Enhanced silencing was found in IL_Cvi-0_6/6.25 in which the region was homozygous for Cvi-0, whereas it was heterozygous in the two lines which did not differ significantly from the reference line, IL_Cvi-0_19.27 and IL_Cvi-0_39 (Figure 22).

Two lines carrying Shahdara introgressions which had been established independently showed enhanced gene silencing when compared to the reference line using the same criteria as applied in the work presented here. IL_Shahdara_83 carried introgressions on chromosomes 1 and 2 and IL_Shahdara_112 on chromosomes 2 and 4. Only the region flanked by Indel markers MSAT2.28 and UPSC_2_9637 was in common in both lines, thus it most likely represented the Shahdara segment that caused the increase in silencing (Thanh Loan Le, personal communication). Interestingly, this chromosome 2 region overlapped with

a region in which the zygosity of the Shahdara introgressions analysed in the work presented here correlated with the frequency of gene silencing. IL_Shahdara_10 in which the region was homozygous for Shahdara and IL Shahdara_6 in which it was heterozygous both showed significantly more silencing when compared to the reference lines, but they also differed significantly when compared to each other from time point 5 onwards (Figure 20; data not shown). In contrast, IL_Shahdara_30 that carried the Col-0 version of this region did not differ significantly from the reference line. All three lines also carried introgressions of different lengths on chromosome 3, but a region in which presence and zygosity correlated with the frequency of silencing was not identified on chromosome 3 (Figure 22). It is tempting to speculate that the locus which caused an increase in silencing in IL_Shahdara_83 and IL_Shahdara_112 coincided with the one present in IL Shahdara_10 and IL Shahdara_6.

In summary, in all lines for which significantly enhanced silencing had been observed candidate regions were identified in which the genotype of the introgression correlated with the frequency of silencing. To substantiate these findings it was taken advantage of lines that were heterozygous for the three different candidate regions. Among the progeny of these lines plants were identified that either carried a particular introgressed segment of interest homozygously for Col-0 or for the accession that had been used to establish introgression lines.

The analysis of IL_Sq-8_49.9, IL_Sq-8_49.21 and IL_Sq-8_49.57 revealed that those lines in which the region flanked by markers Nga249 and Ind_V-9 was homozygous for Sq-8 showed significantly more silencing than line IL_Sq-8_49.21 in which chromosome 5 was entirely derived from Col-0 (Figures 18 and 19). IL_Sq-8_49.9 and IL_Sq-8_49.57 also showed introgression segments of different lengths on chromosomes 2 and 4, but significant differences with respect to FN were not observed. This makes it unlikely that these regions had a considerable impact on gene silencing. As a matter of fact, IL_ Sq-8_49.57.67, in which the Sq-8 introgression segment on chromosome 2 was not present also showed considerable more silencing than Sq-8_49.21.21, a descendant of IL_Sq-8_49.21 (Renate Schmidt, personal communication). Two genes for which a role in post-transcriptional gene silencing had previously been documented, *DCL4* (Dunoyer *et al*., 2005; Gasciolli *et al*., 2005; Xie *et al*., 2005) and *SGS3* (Mourrain *et al*., 2000) and which had been analysed with respect to allelic diversity in this study were mapping to the Sq-8 segment on chromosome 5 which was

implicated in gene silencing. In comparison to Col-0 only one and four amino acid substitutions had been identified for the Sq-8 alleles of *DCL4* and *SGS3*, respectively (Supplementary table 8). Sequencing of amplicons of these genes confirmed the presence of the Sq-8 alleles in IL_Sq-8_7, IL_Sq-8_8, IL_Sq-8_16 and IL_Sq-8_49, whereas in lines IL_Sq-8_6 and IL_Sq-8_48 the Col-0 alleles of both genes were found. Currently it is unclear whether the Sq-8 allelic variants of the *DCL4* and/or *SGS3* genes caused enhanced gene silencing or whether the effect observed in IL_Sq-8_7, IL_Sq-8_8, IL_Sq-8_16 and IL_Sq-8_49 has to be attributed to another locus. The *ESP5* gene which had been implicated in RNA silencing (Herr *et al.*, 2006) is also mapping to this area. Comparative analyses of steady-state transcript levels for accessions and Col-0 alongside the introgression lines will be informative in order to reveal the expression level of the alleles of the different candidates.

The study of additional introgression lines, IL_Cvi-0_19.27.24 and IL_Cvi-0_19.27.32, corroborated that the presence of the homozygous Cvi-0 region flanked by Indel markers MSAT2.28 and MSAT2.17 correlated with increased gene silencing (Figure 22 and 23). Additional Shahdara introgression lines derived fom IL_Shahdara_6 were analysed in the same experiment as IL_Shahdara_19.30. IL_Shahdara_6.43 showed significantly more silencing than IL_Shahdara_6.15 and IL_Shahdara_30 for time points 4 to 10 (Figure 23; data not shown). In contrast, significant differences were neither seen when IL_Shahdara_30 was compared to IL_Shahdara_6.15 nor to the reference line. These results in conjunction with the information obtained from the graphical genotypes delineated the region that caused an increase in gene silencing to the segment flanked by Indel markers 2_3475 and PLS7. An enhancing effect on gene silencing had also been observed for Shahdara introgressions flanked by Indel markers MSAT2.28 and UPSC_2_9637 (Thanh Loan Le, personal communication). If the effect on gene silencing was due to the same locus in both sets of Shahdara introgression lines, the Shahdara segment which caused an increase in gene silencing would be flanked by Indel markers 2-3475 and UPSC_2_637. Interestingly, the regions in the Cvi-0 and Shahdara introgression lines which showed significantly enhanced gene silencing largely overlap (Figure 22). Since some of the Cvi-0 and Shahdara introgression lines were analysed side by side it was possible to compare the lines among each other. For example, in experiments 12-b-14 and 02-15 IL_Cvi-0_6/6.25 showed significantly more silencing than IL_Shahdara_6.15 for time points 9 to 10 and 8 to 10,

respectively, but significantly less silencing than IL_Shahdara_6.43 for time points 5 to 10 and 3 to 10, respectively (data not shown). Thus, although IL_Cvi-0_6/6.25 and IL_Shahdara_6.43 showed repeatedly significantly more silencing than the reference line, the pairwise comparisons between the introgression lines revealed that the effect of the Shahdara introgression was more pronounced.

The introgressions which caused an increase in gene silencing in a subset of the Shahdara and Cvi-0 introgression lines harbour a cluster of RNA-dependent RNA polymerase genes consisting of the three tandemly arranged genes, *RDR3*, *RDR4* and *RDR5* (Wassenegger and Krczal, 2006; Willmann *et al*., 2011). Several T-DNA insertions that mapped to the *RDR* gene cluster were characterised in Col-0. For two T-DNA insertion lines significantly more silencing when compared to the reference line was documented (Thanh Loan Le, personal communication). When the cluster was studied with respect to allelic diversity, several accessions were noted showing more than 0.5% sequence divergence when compared to the Col-0 reference sequence in at least one of the three genes, among these were accessions Cvi-0 and Shahdara. Different haplotypes were observed for the gene cluster in the two accessions (Thanh Loan Le, unpublished results). Based on these results it is conceivable that allelic variants of the *RDR* gene cluster may have an impact on post-transcriptional gene silencing. However, other genes for which a role in PTGS had been reported are also mapping to this region, *RRP44A* (Moreno et al., 2013) and *THO6* (Yelina *et al*., 2010).

Previous studies indicated that *HEN1* has a stronger activity in Landsberg *erecta* than in Col-0, possibly due to the presence of a negative modulator of gene silencing in Col-0 (Yu *et al*., 2010). The analysis of the introgression lines presented here revealed at least three regions which also modulate gene silencing. The approach taken turned out to be useful to determine segments that increased gene silencing but was also suitable to detect regions that diminished gene silencing. The analysis of independent lines in conjunction with graphical genotype information was crucial to delimit the regions which showed an effect on gene silencing. In future, genetic dissection of these genome regions will be required in order to determine the causative loci and whether these correspond to genes for which a role in S-PTGS had been described previously. It will be particularly interesting to clarify whether alleles of the same locus caused the effect in the Shahdara and Cvi-0 introgression lines.

SUMMARY

Highly expressed transgenes are readily subjected to post-transcriptional gene silencing. Many genes involved in PTGS have been identified through the characterisation of *Arabidopsis thaliana* mutants, but comparatively little is known about the role of natural variation in this process. In order to address this question, allelic diversity was analysed in 12 genes playing a role PTGS. Amplicon sequencing of these candidate genes was performed for 25 *A. thaliana* accessions capturing portions of genetic diversity in this species. Accession Col-0 served as reference in all sequence comparisons. Differences with respect to SNP frequencies were noted between the 12 candidate genes as well as among the 25 accessions for individual candidate genes. For the six most conserved genes all accessions showed SNP frequencies lower than 0.5%. In case of the least conserved genes, *AGO7*, *HEN1*, *NRPE1* and *WEX*, SNP frequencies ranging from 1.1% to 4.4% were observed, but only in selected accessions. Based on the results of the diversity studies, 19 alleles characterised by high SNP frequencies and/or large Indels in exon regions were selected for further studies. Transcript analyses revealed that all 19 alleles were expressed, with transcript levels in aerial seedling tissues of the different accessions not differing more than threefold from the values established for the reference Col-0 in most cases. In order to evaluate whether and to which extent the selected alleles would affect PTGS, the alleles were introgressed into Col-0 transgenic lines carrying six *GFP* transgene copies under the control of the CaMV 35S promoter. *GFP* silencing was monitored for all introgression lines at several stages of development and compared to the performance of the *GFP* transgenes in Col-0 lines. Seven out of 40 introgression lines reproducibly showed significant effects on PTGS at two or more consecutive time points of the experiments. Five lines displayed more and two lines less silencing than the Col-0 reference transgenic lines. At least two independent introgression lines had been established for each of the selected alleles, number, position and size of introgressions had also been characterised for all lines, thus it was possible to delimit several genome regions carrying modulators of gene silencing. Diminished transgene silencing correlated with Gie-0 introgressions mapping to chromosome 1. A chromosome 5 region from accession Sq-8 and two overlapping introgressions on chromosome 2 from accessions Cvi-0 and Shahdara also modulated *GFP* silencing. Strikingly, the regions correlating with enhanced transgene silencing did not carry the introgressed alleles of the candidate genes.

ZUSAMMENFASSUNG

Hoch-exprimierte Transgene werden leicht post-transkriptionellem Gen-Silencing (PTGS) unterworfen. Durch die Charakterisierung von *Arabidopsis thaliana*-Mutanten wurden viele der in diesem Prozess involvierten Gene identifiziert, vergleichsweise wenig ist dagegen über die Rolle natürlicher Variation bei PTGS bekannt. Um erste Einsichten dazu zu gewinnen, wurde allele Diversität in 12 Genen analysiert, die eine Rolle bei PTGS spielen. Amplicon-Sequenzierung der Kandidatengene wurde für 25 *A. thaliana* Akzessionen durchgeführt, die Teilbereiche der genetischen Diversität in dieser Art abdecken. Die Akzession Col-0 diente als Referenz in allen Sequenzvergleichen. Unterschiede in Bezug auf SNP-Frequenzen wurden zwischen den 12 Kandidatengenen sowie unter den 25 Akzessionen für die einzelnen Kandidatengene verzeichnet. Bei den sechs konserviertesten Genen wiesen alle Akzessionen SNP-Frequenzen auf, die niedriger als 0.5% waren. Im Fall der divergentesten Gene, *AGO7*, *HEN1*, *NRPE1* und *WEX*, wurden SNP-Frequenzen zwischen 1.1% und 4.4% beobachtet, allerdings nur für bestimmte Akzessionen. Basierend auf den Resultaten der Diversitätstudien wurden 19 Allele, die sich durch hohe SNP-Frequenzen und/oder große Insertionen/Deletionen in Exonregionen auszeichneten, für weitere Studien ausgewählt. Transkriptuntersuchungen zeigten, dass alle 19 Allele ausgeprägt wurden. In der Mehrzahl der Fälle unterschieden sich die Transkriptionsniveaus in oberirdischen Keimlingsgeweben der Akzessionen nicht mehr als dreifach von den für die Referenz Col-0 etablierten Werten. Um herauszufinden, ob und in welchem Maße die ausgewählten Allele PTGS beeinflussen, wurden die Allele in transgene Col-0-Linien introgressiert, die sechs Kopien des *GFP*-Gens unter der Kontrolle des CaMV 35S Promotors trugen. *GFP*-Silencing wurde in Introgressionslinien zu verschiedenen Entwicklungszeitpunkten beobachtet und mit dem Verhalten der *GFP*-Transgene in Col-0-Linien verglichen. Sieben von 40 Introgressionslinien zeigten reproduzierbar signifikante Effekte auf PTGS an zwei oder mehr aufeinanderfolgenden Zeitpunkten der Experimente. Fünf Linien wiesen mehr und zwei Linien weniger Silencing auf als die transgenen Col-0-Referenzlinien. Für jedes ausgewählte Allel waren mindestens zwei unterschiedliche Introgressionlinien etabliert worden, außerdem waren Zahl, Position und Größe der Introgressionen in allen Linien bestimmt worden, daher konnten mehrere Genomregionen eingegrenzt werden, die Modulatoren des Gensilencing trugen. Vermindertes Gensilencing korrelierte mit Gie-0-Introgressionen, die

auf Chromosom 1 lokalisiert waren. Eine Region des Chromosoms 5 der Akzession Sq-8 und zwei überlappende Introgressionen auf Chromosom 2 der Akzessionen Cvi-0 und Shahdara modulierten ebenfalls *GFP*-Silencing. Bemerkenswerterweise wiesen die Regionen, die mit verstärktem Transgen-Silencing korrelierten, nicht die introgressierten Allele der Kandidatengene auf.

REFERENCES

1001 Genomes Consortium (2016). 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*, *166*(2), 481-491.

Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouché, N., Gasciolli, V., & Vaucheret, H. (2006). DRB4-dependent *TAS3* trans-acting siRNAs control leaf morphology through AGO7. *Current Biology*, *16*(9), 927-932.

Alonso, J. M., & Ecker, J. R. (2006). Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. *Nature Reviews Genetics*, *7*(7), 524-536.

Alonso-Blanco, C., Aarts, M. G., Bentsink, L., Keurentjes, J. J., Reymond, M., Vreugdenhil, D., & Koornneef, M. (2009). What has natural variation taught us about plant development, physiology, and adaptation? *The Plant Cell*, *21*(7), 1877-1896.

Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C. J., & Koornneef, M. (1999). Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences USA*, *96*(8), 4710-4717.

Alonso-Blanco, C., Peeters, A. J., Koornneef, M., Lister, C., Dean, C., van den Bosch, N., ... & Kuiper, M. T. (1998). Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. *The Plant Journal*, *14*(2), 259-271.

Al-Shehbaz, I. A., & O'Kane Jr, S. L. (2002). Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). In: The *Arabidopsis* Book *1*, e0001, *American Society of Plant Biologists*.

Arlt, M., & Schmidt, R. (2006). *Gene-Silencing* in transgenen Pflanzen. Tätigkeitsbericht 2006. Max-Planck-Institut für Molekulare Pflanzenphysiologie. Potsdam. Max-Planck-Gesellschaft. In: Jahrbuch. MPG, München.

Arlt, M. (2007). Studien zur Initiation, Ausbreitung und Übertragbarkeit verschiedener Varianten von posttranskriptionellem Transgensilencing anhand molekularer Charakteristika in *Arabidopsis thaliana*. PhD thesis, University of Potsdam.

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., ... & Jiang, R. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, *465*(7298), 627-631.

Bakker, E. G., Toomajian, C., Kreitman, M., & Bergelson, J. (2006). A genome-wide survey of R gene polymorphisms in *Arabidopsis*. *The Plant Cell*, *18*(8), 1803-1818.

Bateman, A. (2002). The SGS3 protein involved in PTGS finds a family. *BMC Bioinformatics*, *3*: 21.

Baulcombe, D. (2004). RNA silencing in plants. *Nature*, *431*(7006): 356-363.

Baumberger, N., & Baulcombe, D. C. (2005). *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences USA*, *102*(33), 11928-11933.

Baxter, I., Brazelton, J. N., Yu, D., Huang, Y. S., Lahner, B., Yakubova, E., ... & Vitek, O. (2010). A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS Genetics*, *6*(11), e1001193.

Benfey, P. N. and Mitchell-Olds T. (2008). From genotype to phenotype: systems biology meets natural variation. *Science*, *320*(5875): 495-497.

Bergelson, J., Buckler, E. S., Ecker, J. R., Nordborg, M., & Weigel, D. (2016). A proposal regarding best practices for validating the identity of genetic stocks and the effects of genetic variants. *The Plant Cell*, *28*(3), 606-609.

Blevins, T., Rajeswaran, R., Shivaprasad, P. V., Beknazariants, D., Si-Ammour, A., Park, H. S., ... & Pooggin, M. M. (2006). Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Research*, *34*(21), 6233-6246.

Bolha, L., Dusanic, D., Narat, M., & Oven, I. (2012). Comparison of methods for relative quantification of gene expression using real-time PCR. *Acta Agriculturae Slovenica*, *100*(2), 97.

Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R., & Zhu, J. K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell*, *123*(7), 1279-1291.

Boutet, S., Vazquez, F., Liu, J., Béclin, C., Fagard, M., Gratias, A., ... & Vaucheret, H. (2003). *Arabidopsis HEN1*: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Current Biology*, *13*(10), 843-848.

Brodersen, P., & Voinnet, O. (2006). The diversity of RNA silencing pathways in plants. *Trends in Genetics*, *22*(5), 268-280.

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y. Y., Sieburth, L., & Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, *320*(5880), 1185-1190.

Buhtz, A., Springer, F., Chappell, L., Baulcombe, D. C., & Kehr, J. (2008). Identification and characterization of small RNAs from the phloem of *Brassica napus*. *The Plant Journal*, *53*(5), 739-749.

Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., ... & Wang, X. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics*, *43*(10), 956-963.

Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Gilbert, K. B., Montgomery, T. A., Nguyen, T., ... & Carrington, J. C. (2012). Functional analysis of three *Arabidopsis* ARGONAUTES using slicer-defective mutants. *The Plant Cell*, *24*(9), 3613-3629.

Chellappan, P., Vanitharani, R., Ogbe, F., & Fauquet, C. M. (2005). Effect of temperature on geminivirus-induced RNA silencing in plants. *Plant Physiology*, *138*(4), 1828-1841.

Clark, R. M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., ... & Chen, H. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science*, *317*(5836), 338-342.

Clerkx, E. J., El-Lithy, M. E., Vierling, E., Ruys, G. J., Blankestijn-De Vries, H., Groot, S. P., ... & Koornneef, M. (2004). Analysis of natural allelic variation of *Arabidopsis* seed germination and seed longevity traits between the accessions Landsberg *erecta* and Shakdara, using a new recombinant inbred line population. *Plant Physiology*, *135*(1), 432-443.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., & Scheible, W. R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiology*, *139*(1), 5-17.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., & Baulcombe, D. C. (2000). An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell*, *101*(5), 543-553.

Dalmay, T., Horsefield, R., Braunstein, T. H., & Baulcombe, D. C. (2001). *SDE3* encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. *The EMBO Journal*, *20*(8), 2069-2077.

Daniell, H. (2002). Molecular strategies for gene containment in transgenic crops. *Nature Biotechnology*, *20*(6), 581-586.

De Bolle, M. F., Butaye, K. M., Coucke, W. J., Goderis, I. J., Wouters, P. F., van Boxel, N., ... & Cammue, B. P. (2003). Analysis of the influence of promoter elements and a matrix attachment region on the inter-individual variation of transgene expression in populations of *Arabidopsis thaliana*. *Plant Science*, *165*(1), 169-179.

Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K. D., Carrington, J. C., & Voinnet, O. (2006). Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science*, *313*(5783), 68-71.

Dellaporta, S. L., Wood, J., & Hicks, J. B. (1983). A plant DNA minipreparation: version II. *Plant Molecular Biology Reporter*, *1*(4), 19-21.

Devert, A., Fabre, N., Floris, M., Canard, B., Robaglia, C., & Crété, P. (2015). Primer-dependent and primer-independent initiation of double stranded RNA synthesis by purified *Arabidopsis* RNA-Dependent RNA Polymerases RDR2 and RDR6. *PloS One*, *10*(3), e0120100.

Dunoyer, P., Himber, C., & Voinnet, O. (2005). DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nature Genetics*, *37*(12), 1356-1360.

Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A., & Voinnet, O. (2007). Intra- and intercellular RNA interference in *Arabidopsis thaliana* requires components of the microRNA and heterochromatic silencing pathways. *Nature Genetics*, *39*(7), 848-856.

Dunoyer, P., Schott, G., Himber, C., Meyer, D., Takeda, A., Carrington, J. C., & Voinnet, O. (2010). Small RNA duplexes function as mobile silencing signals between plant cells. *Science*, *328*(5980), 912-916.

Eamens, A., Vaistij, F. E., & Jones, L. (2008). NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in *Arabidopsis*. *The Plant Journal*, *55*(4), 596-606.

Edwards, K., Johnstone, C., & Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, *19*(6), 1349.

El-Lithy, M. E., Bentsink, L., Hanhart, C. J., Ruys, G. J., Rovito, D., Broekhof, J. L., ... & Koornneef, M. (2006). New *Arabidopsis* recombinant inbred line populations genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. *Genetics*, *172*(3), 1867-1876.

Elvira-Matelot, E., Bardou, F., Ariel, F., Jauvion, V., Bouteiller, N., Le Masson, I., ... & Vaucheret, H. (2016). The nuclear ribonucleoprotein SmD1 interplays with splicing, RNA quality control, and posttranscriptional gene silencing in *Arabidopsis*. *The Plant Cell*, *28*(2), 426-438.

Fagard, M., & Vaucheret, H. (2000). (Trans) gene silencing in plants: how many mechanisms?. *Annual Review of Plant Biology*, *51*(1), 167-194.

Fagard, M., Boutet, S., Morel, J. B., Bellini, C., & Vaucheret, H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proceedings of the National Academy of Sciences USA*, *97*(21), 11650-11654.

Fahlgren, N., Montgomery, T. A., Howell, M. D., Allen, E., Dvorak, S. K., Alexander, A. L., & Carrington, J. C. (2006). Regulation of *AUXIN RESPONSE FACTOR3* by *TAS3* ta-siRNA affects developmental timing and patterning in *Arabidopsis*. *Current Biology*, *16*(9), 939-944.

Feldmann, K. A. (1991). T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. *The Plant Journal*, *1*(1), 71-82.

Forsbach, A., Schubert, D., Lechtenberg, B., Gils, M., & Schmidt, R. (2003). A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. *Plant Molecular Biology*, *52*(1), 161-176.

Francisco-Mangilet, A. G., Karlsson, P., Kim, M. H., Eo, H. J., Oh, S. A., Kim, J. H., ... & Manavella, P. A. (2015). THO2, a core member of the THO/TREX complex, is required for microRNA production in *Arabidopsis*. *The Plant Journal*, *82*(6), 1018-1029.

Fu, J., Keurentjes, J. J., Bouwmeester, H., America, T., Verstappen, F. W., Ward, J. L., ... & Johannes, F. (2009). System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nature Genetics*, *41*(2), 166-167.

Fukunaga, R., & Doudna, J. A. (2009). dsRNA with 5′ overhangs contributes to endogenous and antiviral RNA silencing pathways in plants. *The EMBO Journal*, *28*(5), 545-555.

Fusaro, A. F., Matthew, L., Smith, N. A., Curtin, S. J., Dedic-Hagan, J., Ellacott, G. A., ... & Waterhouse, P. M. (2006). RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. *EMBO Reports*, *7*(11), 1168-1175.

Gabel, H. W., & Ruvkun, G. (2008). The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi. *Nature Structural & Molecular Biology*, *15*(5), 531-533.

Gan, X., Stegle, O., Behr, J., Steffen, J. G., Drewe, P., Hildebrand, K. L., ... & Kahles, A. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature*, *477*(7365), 419-423.

Garcia, D., Garcia, S., Pontier, D., Marchais, A., Renou, J. P., Lagrange, T., & Voinnet, O. (2012). Ago hook and RNA helicase motifs underpin dual roles for SDE3 in antiviral defense and silencing of nonconserved intergenic regions. *Molecular Cell*, *48*(1), 109-120.

Gasciolli, V., Mallory, A. C., Bartel, D. P., & Vaucheret, H. (2005). Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Current Biology*, *15*(16), 1494-1500.

Gazzani, S., Lawrenson, T., Woodward, C., Headon, D., & Sablowski, R. (2004). A link between mRNA turnover and RNA interference in *Arabidopsis*. *Science*, *306*(5698), 1046-1048.

Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nature Reviews Genetics*, *10*(2), 94-108.

Glazov, E., Phillips, K., Budziszewski, G. J., Meins, F., & Levin, J. Z. (2003). A gene encoding an RNase D exonuclease-like protein is required for post-transcriptional silencing in *Arabidopsis*. *The Plant Journal*, *35*(3), 342-349.

Graham, N., & May, S. (2011). Bioinformatics Resources for *Arabidopsis thaliana*. In: Genetics and Genomics of the Brassicaceae. R. Schmidt and I. Bancroft (editors), *Springer, New York*, 585-596.

Gregory, B. D., O'Malley, R. C., Lister, R., Urich, M. A., Tonti-Filippini, J., Chen, H., ... & Ecker, J. R. (2008). A link between RNA metabolism and silencing affecting *Arabidopsis* development. *Developmental Cell*, *14*(6), 854-866.

Gy, I., Gasciolli, V., Lauressergues, D., Morel, J. B., Gombert, J., Proux, F., ... & Mallory, A. C. (2007). *Arabidopsis* FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors. *The Plant Cell*, *19*(11), 3451-3461.

Hamilton, A. J., & Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, *286*(5441), 950-952.

Han, Y., & Grierson, D. (2002). Relationship between small antisense RNAs and aberrant RNAs associated with sense transgene mediated gene silencing in tomato. *The Plant Journal*, *29*(4), 509-519.

Haubold, B., Kroymann, J., Ratzka, A., Mitchell-Olds, T., & Wiehe, T. (2002). Recombination and gene conversion in a 170-kb genomic region of *Arabidopsis thaliana*. *Genetics*, *161*(3), 1269-1278.

Havecker, E. R., Wallbridge, L. M., Hardcastle, T. J., Bush, M. S., Kelly, K. A., Dunn, R. M., ... & Baulcombe, D. C. (2010). The *Arabidopsis* RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *The Plant Cell*, *22*(2), 321-334.

Henderson, I. R., Zhang, X., Lu, C., Johnson, L., Meyers, B. C., Green, P. J., & Jacobsen, S. E. (2006). Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nature Genetics*, *38*(6), 721-725.

Hernandez-Pinzon, I., Yelina, N. E., Schwach, F., Studholme, D. J., Baulcombe, D., & Dalmay, T. (2007). SDE5, the putative homologue of a human mRNA export factor, is required for transgene silencing and accumulation of trans-acting endogenous siRNA. *The Plant Journal*, *50*(1), 140-148.

Herr, A. J., Jensen, M. B., Dalmay, T., & Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science*, *308*(5718), 118-120.

Herr, A. J., Molnàr, A., Jones, A., & Baulcombe, D. C. (2006). Defective RNA processing enhances RNA silencing and influences flowering of *Arabidopsis*. *Proceedings of the National Academy of Sciences USA*, *103*(41), 14994-15001.

Himber, C., Dunoyer, P., Moissiard, G., Ritzenthaler, C., & Voinnet, O. (2003). Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *The EMBO Journal*, *22*(17), 4523-4533.

Hiraguri, A., Itoh, R., Kondo, N., Nomura, Y., Aizawa, D., Murai, Y., ... & Fukuhara, T. (2005). Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*. *Plant Molecular Biology*, *57*(2), 173-188.

Hoffmann, M. H. (2002). Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). *Journal of Biogeography*, *29*(1), 125-134.

Holtorf, S., Apel, K., & Bohlmann, H. (1995). Comparison of different constitutive and inducible promoters for the overexpression of transgenes in *Arabidopsis thaliana*. *Plant Molecular Biology*, *29*(4), 637-646.

Horton, M. W., Hancock, A. M., Huang, Y. S., Toomajian, C., Atwell, S., Auton, A., ... & Nordborg, M. (2012). Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics*, *44*(2), 212-216.

Hou, X., Li, L., Peng, Z., Wei, B., Tang, S., Ding, M., ... & Qu, L. J. (2010). A platform of high-density INDEL/CAPS markers for map-based cloning in *Arabidopsis*. *The Plant Journal*, *63*(5), 880-888.

Huang, Y., Ji, L., Huang, Q., Vassylyev, D. G., Chen, X., & Ma, J. B. (2009). Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature*, *461*(7265), 823-827.

Hunter, C., Sun, H., & Poethig, R. S. (2003). The *Arabidopsis* heterochronic gene *ZIPPY* is an *ARGONAUTE* family member. *Current Biology*, *13*(19), 1734-1739.

Hunter, C., Willmann, M. R., Wu, G., Yoshikawa, M., de la Luz Gutiérrez-Nava, M., & Poethig, S. R. (2006). Trans-acting siRNA-mediated repression of *ETTIN* and *ARF4* regulates heteroblasty in *Arabidopsis*. *Development*, *133*(15), 2973-2981.

Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. *Nature Reviews Molecular Cell Biology*, *9*(1), 22-32.

Jauvion, V., Elmayan, T., & Vaucheret, H. (2010). The conserved RNA trafficking proteins HPR1 and TEX1 are involved in the production of endogenous and exogenous small interfering RNA in *Arabidopsis*. *The Plant Cell*, *22*(8), 2697-2709.

Jauvion, V., Rivard, M., Bouteiller, N., Elmayan, T., & Vaucheret, H. (2012). RDR2 partially antagonizes the production of RDR6-dependent siRNA in sense transgene-mediated PTGS. *PLoS One*, *7*(1), e29785.

Juenger, T., Pérez-Pérez, J. M., Bernal, S., & Micol, J. L. (2005). Quantitative trait loci mapping of floral and leaf morphology traits in *Arabidopsis thaliana*: evidence for modular genetic architecture. *Evolution & Development*, *7*(3), 259-271.

Kalantidis, K., Schumacher, H. T., Alexiadis, T., & Helm, J. M. (2008). RNA silencing movement in plants. *Biology of the Cell*, *100*(1), 13-26.

Kalantidis, K., Tsagris, M., & Tabler, M. (2006). Spontaneous short-range silencing of a *GFP* transgene in *Nicotiana benthamiana* is possibly mediated by small quantities of siRNA that do not trigger systemic silencing. *The Plant Journal*, *45*(6), 1006-1016.

Kanno, T., Huettel, B., Mette, M. F., Aufsatz, W., Jaligot, E., Daxinger, L., ... & Matzke, A. J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nature Genetics*, *37*(7), 761-765.

Kastenmayer, J. P., & Green, P. J. (2000). Novel features of the XRN-family in *Arabidopsis*: evidence that AtXRN4, one of several orthologs of nuclear Xrn2p/Rat1p, functions in the cytoplasm. *Proceedings of the National Academy of Sciences USA*, *97*(25), 13985-13990.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Zhu, J. K., ... & Jin, H. (2006). A pathogen-inducible endogenous siRNA in plant immunity. *Proceedings of the National Academy of Sciences USA*, *103*(47), 18002-18007.

Kawakatsu, T., Wakasa, Y., Yasuda, H., & Takaiwa, F. (2012). RNA silencing induced by an artificial sequence that prevents proper transcription termination in rice. *Plant Physiology*, *160(2)*, 601-12.

Kennedy, S., Wang, D., & Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature*, *427*(6975), 645-649.

Ketting, R. F., Haverkamp, T. H., van Luenen, H. G., & Plasterk, R. H. (1999). *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell*, *99*(2), 133-141.

Keurentjes, J. J., Fu, J., Terpstra, I. R., Garcia, J. M., van den Ackerveken, G., Snoek, L. B., ... & Jansen, R. C. (2007). Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proceedings of the National Academy of Sciences USA*, *104*(5), 1708-1713.

Kim, S., Plagnol, V., Hu, T. T., Toomajian, C., Clark, R. M., Ossowski, S., ... & Nordborg, M. (2007). Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, *39*(9), 1151-1155.

Knee, E. M., Rivero, L., Crist, D., Grotewold, E., & Scholl, R. (2011). Germplasm and molecular resources. In: Genetics and Genomics of the Brassicaceae. R. Schmidt and I. Bancroft (editors), *Springer, New York*, 437-467.

Kohli, A., González-Melendi, P., Abranches, R., Capell, T., Stoger, E., & Christou, P. (2006). The quest to understand the basis and mechanisms that control expression of introduced transgenes in crop plants. *Plant Signaling & Behavior*, *1*(4), 185-195.

Koornneef, M., & Meinke, D. (2010). The development of *Arabidopsis* as a model plant. *The Plant Journal*, *61*(6), 909-921.

Koornneef, M., Alonso-Blanco, C., & Vreugdenhil, D. (2004). Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology*, *55*, 141-172.

Koornneef, M., C. Alonso-Blanco and P. Stam (2006). Genetic Analysis. In: *Arabidopsis* Protocols. J. Salinas and J. Sanchez-Serrano (editors), *Humana Press*, *Totowa, Methods in Molecular Biology 323*, 65-77.

Koornneef, M., Reymond, M., & Alonso-Blanco, C. (2011). Natural variation in *Arabidopsis thaliana*. In: Genetics and Genomics of the Brassicaceae. R. Schmidt and I. Bancroft (editors), *Springer, New York*, 123-151.

Kotakis, C., Vrettos, N., Kotsis, D., Tsagris, M., Kotzabasis, K., & Kalantidis, K. (2010). Light intensity affects RNA silencing of a transgene in *Nicotiana benthamiana* plants. *BMC Plant Biology*, *10*(1), 220.

Kumakura, N., Takeda, A., Fujioka, Y., Motose, H., Takano, R., & Watanabe, Y. (2009). SGS3 and RDR6 interact and colocalize in cytoplasmic SGS3/RDR6-bodies. *FEBS Letters*, *583*(8), 1261-1266.

Laibach, F. (1943). *Arabidopsis thaliana* (L.) Heynh. als Objekt für genetische und entwicklungsphysiologische Untersuchungen. *Botanisches Archiv*, *44*, 439-455.

Lanfranco, L. (2002). Engineering crops, a deserving venture. *Rivista di Biologia*, *96*(1), 31-54.

Le Masson, I., Jauvion, V., Bouteiller, N., Rivard, M., Elmayan, T., & Vaucheret, H. (2012). Mutations in the *Arabidopsis* H3K4me2/3 demethylase JMJ14 suppress posttranscriptional gene silencing by decreasing transgene transcription. *The Plant Cell*, *24*(9), 3603-3612.

Lechtenberg, B., Schubert, D., Forsbach, A., Gils, M., & Schmidt, R. (2003). Neither inverted repeat T-DNA configurations nor arrangements of tandemly repeated transgenes are sufficient to trigger transgene silencing. *The Plant Journal*, *34*(4), 507-517.

Li, J., Yang, Z., Yu, B., Liu, J., & Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Current Biology*, *15*(16), 1501-1507.

Liang, D., White, R. G., & Waterhouse, P. M. (2012). Gene silencing in *Arabidopsis* spreads from the root to the shoot, through a gating barrier, by template-dependent, nonvascular, cell-to-cell movement. *Plant Physiology*, *159*(3), 984-1000.

Lisec, J., Meyer, R. C., Steinfath, M., Redestig, H., Becher, M., Witucka-Wall, H., ... & Willmitzer, L. (2008). Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations.*The Plant Journal*, *53*(6), 960-972.

Lister, C., & Dean, C. (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *The Plant Journal*, *4*(4), 745-750.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods*, *25*(4), 402-408.

Lloyd, A. (2003). Vector construction for gene overexpression as a tool to elucidate gene function. *Plant Functional Genomics*, 329-344.

Lloyd, A. M., Barnason, A. R., Rogers, S. G., Byrne, M. C., Fraley, R. T., & Horsch, R. B. (1986). Transformation of *Arabidopsis thaliana* with *Agrobacterium tumefaciens*. *Science*, *234*(4775), 464-466.

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., ... & Voronin, V. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics*, *45*(8), 884-890.

Loudet, O., Chaillou, S., Camilleri, C., Bouchez, D., & Daniel-Vedele, F. (2002). Bay-0 × Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *Theoretical and Applied Genetics*, *104*(6-7), 1173-1184.

Loudet, O., Gaudon, V., Trubuil, A., & Daniel-Vedele, F. (2005). Quantitative trait loci controlling root growth and architecture in *Arabidopsis thaliana* confirmed by heterogeneous inbred family. *Theoretical and Applied Genetics*, *110*(4), 742-753.

Luo, J., & Hall, B. D. (2007). A multistep process gave rise to RNA polymerase IV of land plants. *Journal of Molecular Evolution*, *64*(1), 101-112.

Luo, Z., & Chen, Z. (2007). Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in *Arabidopsis*. *The Plant Cell*, *19*(3), 943-958.

Mallory, A. C., & Vaucheret, H. (2009). ARGONAUTE 1 homeostasis invokes the coordinate action of the microRNA and siRNA pathways. *EMBO Reports*, *10*(5), 521-526.

Mallory, A. C., Hinze, A., Tucker, M. R., Bouché, N., Gasciolli, V., Elmayan, T., ... & Laux, T. (2009). Redundant and specific roles of the ARGONAUTE proteins AGO1 and ZLL in development and small RNA-directed gene silencing. *PLoS Genetics*, *5*(9), e1000646.

Mallory, A., & Vaucheret, H. (2010). Form, function, and regulation of ARGONAUTE proteins. *The Plant Cell*, *22*(12), 3879-3889.

Margis, R., Fusaro, A. F., Smith, N. A., Curtin, S. J., Watson, J. M., Finnegan, E. J., & Waterhouse, P. M. (2006). The evolution and diversification of Dicers in plants. *FEBS Letters*, *580*(10), 2442-2450.

Matzke, A. J., & Matzke, M. A. (1998). Position effects and epigenetic silencing of plant transgenes. *Current Opinion in Plant Biology*, *1*(2), 142-148.

Matzke, M. A., & Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics*, *15*(6), 394-408.

Matzke, M. A., Primig, M., Trnovsky, J., & Matzke, A. J. M. (1989). Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *The EMBO Journal*, *8*(3), 643.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., & Matzke, A. J. (2009). RNA-mediated chromatin-based silencing in plants. *Current Opinion in Cell Biology*, *21*(3), 367-376.

Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., & Koornneef, M. (1998). *Arabidopsis thaliana*: a model plant for genome analysis. *Science*, *282*(5389), 662-682.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nature Reviews Genetics*, *14*(7), 447-459.

Mermigka, G., Verret, F., & Kalantidis, K. (2015). RNA silencing movement in plants. *Journal of Integrative Plant Biology, 58*(4), 328-342.

Mette, M. F., Aufsatz, W., Van der Winden, J., Matzke, M. A., & Matzke, A. J. M. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *The EMBO Journal*, *19*(19), 5194-5201.

Meyer, P. (1995). Understanding and controlling transgene expression. *Trends in Biotechnology*, *13*(9), 332-337.

Meyer, P. (2013). Transgenes and their contributions to epigenetic research. *The International Journal of Developmental Biology*, *57*(6-8), 509-515.

Meyer, P., & Saedler, H. (1996). Homology-dependent gene silencing in plants. *Annual Review of Plant Biology*, *47*(1), 23-48.

Meyer, R. C., Hönig, G., Brandt, R., Arana-Ceballos, F., Neitsch, C., Reuter, G., ... & Kuhlmann, M. (2015). Overexpression of *Arabidopsis thaliana ERI*, the homolog of *C. elegans Enhancer of RNAinterference*, leads to enhanced growth. *Frontiers in Plant Science*, *6*, 531.

Meyer, R. C., Kusterer, B., Lisec, J., Steinfath, M., Becher, M., Scharr, H., ... & Altmann, T. (2010). QTL analysis of early stage heterosis for biomass in *Arabidopsis*. *Theoretical and Applied Genetics*, *120*(2), 227-237.

Meyerowitz, E. M. (1987). *Arabidopsis thaliana*. *Annual Review of Genetics*, *21*(1), 93-111.

Meyerowitz, E. M., & Pruitt, R. E. (1985). *Arabidopsis thaliana* and plant molecular genetics. *Science*, *229*(4719), 1214-1218.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., ... & Chen, S. (2008). Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5′ terminal nucleotide. *Cell*, *133*(1), 116-127.

Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. *Nature*, *457*(7228), 413-420.

Molnar, A., Melnyk, C. W., Bassett, A., Hardcastle, T. J., Dunn, R., & Baulcombe, D. C. (2010). Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. *Science*, *328*(5980), 872-875.

Morel, J. B., Godon, C., Mourrain, P., Béclin, C., Boutet, S., Feuerbach, F., ... & Vaucheret, H. (2002). Fertile hypomorphic *ARGONAUTE* (*ago1*) mutants impaired in post-transcriptional gene silencing and virus resistance. *The Plant Cell*, *14*(3), 629-639.

Moreno, A. B., de Alba, A. E. M., Bardou, F., Crespi, M. D., Vaucheret, H., Maizel, A., & Mallory, A. C. (2013). Cytoplasmic and nuclear quality control and turnover of single-stranded RNA modulate post-transcriptional gene silencing in plants. *Nucleic Acids Research*, *41*(8), 4699-4708.

Mourrain, P., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., ... & Rémoué, K. (2000). *Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*, *101*(5), 533-542.

Muangsan, N., Beclin, C., Vaucheret, H., & Robertson, D. (2004). Geminivirus VIGS of endogenous genes requires SGS2/SDE1 and SGS3 and defines a new branch in the genetic pathway for silencing in plants. *The Plant Journal*, *38*(6), 1004-1014.

Napoli, C., Lemieux, C., & Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *The Plant Cell*, *2*(4), 279-289.

Nguyen, A. H., Matsui, A., Tanaka, M., Mizunashi, K., Nakaminami, K., Hayashi, M., ... & Seki, M. (2015). Loss of *Arabidopsis* 5'-3'exoribonuclease AtXRN4 function enhances heat stress tolerance of plants subjected to severe heat stress. *Plant and Cell Physiology*, *56*(9), 1762-1772.

Nordborg, M., & Weigel, D. (2008). Next-generation genetics in plants. *Nature*, *456*(7223), 720-723.

Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., ... & Jakobsson, M. (2005). The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology, 3*(7), e196.

Olmedo, G., Guo, H., Gregory, B. D., Nourizadeh, S. D., Aguilar-Henonin, L., Li, H., ... & Ecker, J. R. (2006). *ETHYLENE-INSENSITIVE5* encodes a 5'→ 3' exoribonuclease required for regulation of the EIN3-targeting F-box proteins EBF1/2. *Proceedings of the National Academy of Sciences USA*, *103*(36), 13286-13293.

Onodera, Y., Haag, J. R., Ream, T., Nunes, P. C., Pontes, O., & Pikaard, C. S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, *120*(5), 613-622.

Ossowski, S., Schneeberger, K., Clark, R. M., Lanz, C., Warthmann, N., & Weigel, D. (2008). Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Research*, *18*(12), 2024-2033.

Păcurar, D. I., Păcurar, M. L., Street, N., Bussell, J. D., Pop, T. I., Gutierrez, L., & Bellini, C. (2012). A collection of INDEL markers for map-based cloning in seven *Arabidopsis* accessions. *Journal of Experimental Botany*, *63*(7), 2491-2501.

Palauqui, J. C., Elmayan, T., Pollien, J. M., & Vaucheret, H. (1997). Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *The EMBO Journal*, *16*(15), 4738-4745.

Parent, J. S., Jauvion, V., Bouché, N., Béclin, C., Hachet, M., Zytnicki, M., & Vaucheret, H. (2015). Post-transcriptional gene silencing triggered by sense transgenes involves uncapped antisense RNA and differs from silencing intentionally triggered by antisense transgenes. *Nucleic Acids Research*, *43*(17), 8464-8475.

Parent, J. S., Martínez de Alba, A. E., & Vaucheret, H. (2012). The origin and effect of small RNA signaling in plants. *Frontiers in Plant Science*, *3*, 179.

Park, W., Li, J., Song, R., Messing, J., & Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Current Biology*, *12*(17), 1484-1495.

Patil, B. L., & Fauquet, C. M. (2015). Light intensity and temperature affect systemic spread of silencing signal in transient agroinfiltration studies. *Molecular Plant Pathology*, *16*(5), 484-494.

Peláez, P., & Sanchez, F. (2013). Small RNAs in plant defense responses during viral and bacterial interactions: similarities and differences. *Frontiers in Plant Science*, *4*, 343.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L., & Poethig, R. S. (2004). *SGS3* and *SGS2*/*SDE1*/*RDR6* are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. *Genes & Development*, *18*(19), 2368-2379.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Research*, *29*(9), e45-e45.

Platt, A., Horton, M., Huang, Y. S., Li, Y., Anastasio, A. E., Mulyati, N. W., ... & Dunning, M. (2010). The scale of population structure in *Arabidopsis thaliana*. *PLoS Genetics*, *6*(2), e1000843.

Plchova, H., Hartung, F., & Puchta, H. (2003). Biochemical characterization of an exonuclease from *Arabidopsis thaliana* reveals similarities to the DNA exonuclease of the human Werner syndrome protein. *Journal of Biological Chemistry*, *278*(45), 44128-44138.

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M. A., ... & Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes & Development*, *19*(17), 2030-2040.

Pumplin, N., & Voinnet, O. (2013). RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nature Reviews Microbiology*, *11*(11), 745-760.

Qi, Y., Denli, A. M., & Hannon, G. J. (2005). Biochemical specialization within *Arabidopsis* RNA silencing pathways. *Molecular Cell*, *19*(3), 421-428.

Ramachandran, V., & Chen, X. (2008). Degradation of microRNAs by a family of exoribonucleases in *Arabidopsis*. *Science*, *321*(5895), 1490-1492.

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., & Bartel, D. P. (2002). MicroRNAs in plants. *Genes & Development*, *16*(13), 1616-1626.

Ríos, G., Lossow, A., Hertel, B., Breuer, F., Schaefer, S., Broich, M., ... & Farrás, R. (2002). Rapid identification of *Arabidopsis* insertion mutants by non-radioactive detection of T-DNA tagged genes. *The Plant Journal*, *32*(2), 243-253.

Rotmistrovsky, K., Jang, W., & Schuler, G. D. (2004). A web server for performing electronic PCR. *Nucleic Acids Research*, *32*(suppl 2), W108-W112.

Ruiz-Ferrer, V., & Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. *Annual Review of Plant Biology*, *60*, 485-510.

Salathia, N., Lee, H. N., Sangster, T. A., Morneau, K., Landry, C. R., Schellenberg, K., ... & Queitsch, C. (2007). Indel arrays: an affordable alternative for genotyping. *The Plant Journal*, *51*(4), 727-737.

Schauer, S. E., Jacobsen, S. E., Meinke, D. W., & Ray, A. (2002). DICER-LIKE1: blind men and elephants in *Arabidopsis* development. *Trends in Plant Science*, *7*(11), 487-491.

Scheid, O. M., Paszkowski, J., & Potrykus, I. (1991). Reversible inactivation of a transgene in *Arabidopsis thaliana*. *Molecular and General Genetics*, *228*(1-2), 104-112.

Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, *3*(6), 1101-1108.

Schneeberger, K., Ossowski, S., Ott, F., Klein, J. D., Wang, X., Lanz, C., ... & Henz, S. R. (2011). Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proceedings of the National Academy of Sciences USA*, *108*(25), 10249-10254.

Scholl, R. L., May, S. T., & Ware, D. H. (2000). Seed and molecular resources for *Arabidopsis*. *Plant Physiology*, *124*(4), 1477-1480.

Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S., & Schmidt, R. (2004). Silencing in *Arabidopsis* T-DNA transformants: the predominant role of a gene-specific RNA sensing mechanism versus position effects. *The Plant Cell*, *16*(10), 2561-2572.

Schuler, G. D. (1997). Sequence mapping by electronic PCR. *Genome Research*, *7*(5), 541-550.

Schwarte, S., Wegner, F., Havenstein, K., Groth, D., Steup, M., & Tiedemann, R. (2015). Sequence variation, differential expression, and divergent evolution in starch-related genes among accessions of *Arabidopsis thaliana*. *Plant Molecular Biology*, *87*(4-5), 489-519.

Shin, M. R., Natsuume, M., Matsumoto, T., Hanaoka, M., Imai, M., Iijima, K., ... & Kodama, H. (2014). Sense transgene induced post-transcriptional gene silencing in tobacco compromises the splicing of endogenous counterpart genes. *PloS One*, *9*(2), e87869.

Simon, M., Loudet, O., Durand, S., Bérard, A., Brunel, D., Sennesal, F. X., ... & Camilleri, C. (2008). Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. *Genetics*, *178*(4), 2253-2264.

Smith, C. J. S., Watson, C. F., Bird, C. R., Ray, J., Schuch, W., & Grierson, D. (1990). Expression of a truncated tomato polygalacturonase gene inhibits expression of the endogenous gene in transgenic plants. *Molecular and General Genetics*, *224*(3), 477-481.

Smith, L. M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F. K., Herr, A. J., ... & Baulcombe, D. C. (2007). An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in *Arabidopsis*. *The Plant Cell*, *19*(5), 1507-1521.

Somerville, C., & Koornneef, M. (2002). A fortunate choice: the history of *Arabidopsis* as a model plant. *Nature Reviews Genetics*, *3*(11), 883-889.

Souret, F. F., Kastenmayer, J. P., & Green, P. J. (2004). AtXRN4 degrades mRNA in *Arabidopsis* and its substrates include selected miRNA targets. *Molecular Cell*, *15*(2), 173-183.

Szittya, G., Silhavy, D., Molnár, A., Havelda, Z., Lovas, Á., Lakatos, L., ... & Burgyán, J. (2003). Low temperature inhibits RNA silencing-mediated defence by the control of siRNA generation. *The EMBO Journal*, *22*(3), 633-640.

Tanzer, M. M., Thompson, W. F., Law, M. D., Wernsman, E. A., & Uknes, S. (1997). Characterization of post-transcriptionally suppressed transgene expression that confers resistance to tobacco etch virus infection in tobacco. *The Plant Cell*, *9*(8), 1411-1423.

The *Arabidopsis* Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814): 796-815.

Tinland, B. (1996). The integration of T-DNA into plant genomes. *Trends in Plant Science*, *1*(6), 178-184.

Törjék, O., Meyer, R. C., Zehnsdorf, M., Teltow, M., Strompen, G., Witucka-Wall, H., ... & Altmann, T. (2008). Construction and analysis of 2 reciprocal *Arabidopsis* introgression line populations. *Journal of Heredity*, *99*(4), 396-406.

Törjék, O., Witucka-Wall, H., Meyer, R. C., von Korff, M., Kusterer, B., Rautengarten, C., & Altmann, T. (2006). Segregation distortion in *Arabidopsis* C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *Theoretical and Applied Genetics*, *113*(8), 1551-1561.

Van der Krol, A. R., Mur, L. A., Beld, M., Mol, J. N., & Stuitje, A. R. (1990). Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *The Plant Cell*, *2*(4), 291-299.

Vaucheret, H. (2008). Plant argonautes. *Trends in Plant Science*, *13*(7), 350-358.

Vaucheret, H., & Fagard, M. (2001). Transcriptional gene silencing in plants: targets, inducers and regulators. *Trends in Genetics*, *17*(1), 29-35.

Vaucheret, H., Béclin, C., & Fagard, M. (2001). Post-transcriptional gene silencing in plants. *Journal of Cell Science*, *114*(17), 3083-3091.

Vaucheret, H., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., ... & Vernhettes, S. (1998). Transgene-induced gene silencing in plants. *The Plant Journal*, *16*(6), 651-659.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A. C., ... & Crété, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Molecular Cell*, *16*(1), 69-79.

Voinnet, O. (2005). Non-cell autonomous RNA silencing. *FEBS Letters*, *579*(26), 5858-5871.

Voinnet, O., & Baulcombe, D. C. (1997). Systemic signalling in gene silencing. *Nature*, *389*, 553.

Voinnet, O., Vain, P., Angell, S., & Baulcombe, D. C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. *Cell*, *95*(2), 177-187.

Wassenegger, M., & Krczal, G. (2006). Nomenclature and functions of RNA-directed RNA polymerases. *Trends in Plant Science*, *11*(3), 142-151.

Wassenegger, M., Heimes, S., Riedel, L., & Sänger, H. L. (1994). RNA-directed *de novo* methylation of genomic sequences in plants. *Cell*, *76*(3), 567-576.

Waterhouse, P. M., Wang, M. B., & Lough, T. (2001). Gene silencing as an adaptive defence against viruses. *Nature*, *411*(6839), 834-842.

Weigel, D. (2012). Natural variation in *Arabidopsis*: from molecular genetics to ecological genomics. *Plant Physiology*, *158*(1), 2-22.

Weigel, D., & Mott, R. (2009). The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biology*, *10*(5), 107.

Weigel, D., & Nordborg, M. (2005). Natural variation in *Arabidopsis*. How do we find the causal genes? *Plant Physiology*, *138*(2), 567-568.

Werle, E., Schneider, C., Renner, M., Völker, M., & Fiehn, W. (1994). Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Research*, *22*(20), 4354.

Willmann, M. R., Endres, M. W., Cook, R. T., & Gregory, B. D. (2011). The functions of RNA-dependent RNA polymerases in *Arabidopsis*. In: *The Arabidopsis* Book *9*, *e0146*, *American Society of Plant Biologists*.

Xie, Z., Allen, E., Wilken, A., & Carrington, J. C. (2005). DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences USA*, *102*(36), 12984-12989.

Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., ... & Carrington, J. C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biology*, *2*(5), e104.

Yang, S. J., Carter, S. A., Cole, A. B., Cheng, N. H., & Nelson, R. S. (2004). A natural variant of a host RNA-dependent RNA polymerase is associated with increased susceptibility to viruses by *Nicotiana benthamiana*. *Proceedings of the National Academy of Sciences USA*, *101*(16), 6297-6302.

Yelina, N. E., Smith, L. M., Jones, A. M., Patel, K., Kelly, K. A., & Baulcombe, D. C. (2010). Putative *Arabidopsis* THO/TREX mRNA export complex is involved in transgene and endogenous siRNA biosynthesis. *Proceedings of the National Academy of Sciences USA*, *107*(31), 13948-13953.

Yoo, B. C., Kragler, F., Varkonyi-Gasic, E., Haywood, V., Archer-Evans, S., Lee, Y. M., ... & Lucas, W. J. (2004). A systemic small RNA signaling system in plants. *The Plant Cell*, *16*(8), 1979-2000.

Yoshikawa, M., Peragine, A., Park, M. Y., & Poethig, R. S. (2005). A pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*. *Genes & Development*, *19*(18), 2164-2175.

Yu, A., Saudemont, B., Bouteiller, N., Elvira-Matelot, E., Lepère, G., Parent, J. S., ... & Vaucheret, H. (2015). Second-site mutagenesis of a hypomorphic *argonaute1* allele identifies *SUPERKILLER3* as an endogenous suppressor of transgene posttranscriptional gene silencing. *Plant Physiology*, *169*(2), 1266-1274.

Yu, B., Bi, L., Zhai, J., Agarwal, M., Li, S., Wu, Q., ... & Chen, X. (2010). siRNAs compete with miRNAs for methylation by HEN1 in *Arabidopsis*. *Nucleic Acids Research*, *38*(17), 5844-5850.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., ... & Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science*, *307*(5711), 932-935.

Zeller, G., Clark, R. M., Schneeberger, K., Bohlen, A., Weigel, D., & Rätsch, G. (2008). Detecting polymorphic regions in *Arabidopsis thaliana* with resequencing microarrays. *Genome Research*, *18*(6), 918-929.

Zhang, D., & Trudeau, V. L. (2008). The XS domain of a plant specific SGS3 protein adopts a unique RNA recognition motif (RRM) fold. *Cell Cycle*, *7*(14), 2268-2270.

Zhang, H., & Zhu, J. K. (2011). RNA-directed DNA methylation. *Current Opinion in Plant Biology*, *14*(2), 142-147.

Zhang, X., Niu, D., Carbonell, A., Wang, A., Lee, A., Tun, V., ... & Jin, H. (2014). ARGONAUTE PIWI domain and microRNA duplex structure regulate small RNA sorting in *Arabidopsis*. *Nature Communications*, *5*, 5468.

Zhang, X., Zhang, X., Singh, J., Li, D., & Qu, F. (2012). Temperature-dependent survival of *Turnip crinkle virus*-infected *Arabidopsis* plants relies on an RNA silencing-based defense that requires DCL2, AGO2, and HEN1. *Journal of Virology*, *86*(12), 6847-6854.

Zheng, X., Zhu, J., Kapoor, A., & Zhu, J. K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *The EMBO Journal*, *26*(6), 1691-1701.

Zilberman, D., Cao, X., & Jacobsen, S. E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*, *299*(5607), 716-719.

Zuo, Y., & Deutscher, M. P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Research*, *29*(5), 1017-1026.

SUPPLEMENTARY DATA

Supplementary table 1. Amplicons used in allelic diversity studies in 26 *Arabidopsis thaliana* accessions.

Candidate gene	Amplicon	Forward primer (5'->3')	Reverse primer (5'->3')
AGO1	**AGO1-1**	TCGTGCTTCTGAGTCGTTTG	GAAGACGACTTCCAAGGTGAG
	AGO1-2	GTGTCAATCGTGCTGTGATG	GTACTTCCGGCGCATGTTTC
	AGO1-3	AGGCAAACCCTGTGATTCAG	CAAGCAAACATATGCATCCAAG
	AGO1-4	TGGTGATAAAGCCTCATTGTG	TATTCACCGTTCCACCATTG
	AGO1-5a	GTATGGGTGTGGCACTTGAC	GCTTGAGCGCAAACTAATCC
	AGO1-6	CACCCTGGAGAGGATTCAAG	CCTGAATACCAGCATGACTACAG
	AGO1-7	GCTCAGAACCACAATGATCG	GAGTGATGAAATATCCAAACACACG
AGO7	**AGO7-1**	CAAAGCTTTCCAAAGCCAAC	GGTTGGCGAGGAGATAAATG
	AGO7-2	GGAGAGAAAGCAGCAACACC	GATTTGCAGACACGGCAAG
	AGO7-3	TCGAGTTTATGGGTTAACAGAGG	CGGGTCTAAGCAGTCTAGGG
	AGO7-4	TGACACAATGCTGCTTATACCC	CGTCTTTGGAGGACCAACAC
DCL4	**DCL4-1**	GAGGGAGACGTGAGATCGTG	ACCTCATTCGCTGCAATCTC
	DCL4-2	TTGTTTACAGCAAGCCAAGG	CCGGCAGTATTCAATTCTAGG
	DCL4-3	TGGTTCTGAAATTCCCTTGG	CAGCAAGCATCTAACAGAAACC
	DCL4-4	TGGCAAGATTCACAATAGCTG	GCTTCATTTCTAGGTTCTCAGTCC
	DCL4-5	CCTTAGCTTTATTGAGTGGCTACG	AGACTGAGGCATTCGAGCAC
	DCL4-6	GGCCTTGATATTCAGACATGC	GATCCTATCAGCGGGATGAG
	DCL4-7	AAGGCTGTTCACGAGGTGAC	TGTGCAGGATGCTTGAGTTC 20
	DCL4-8	TTGGTTGAGACTCTAGATTTGAGG	GTATGGAGGCAAAGGGTCTG
	DCL4-9	TGCATTCCTCAAGTTTGCTG	CAGGTTTCAGTTTCGGGAAG
	DCL4-10	GCAATAATTTGGAGACCCTTG	AAGGCAACAACCATTCTTGG
	DCL4-11	CCAGTCTTACAAGTGGGATCG	GCTCGAGATCATCAGCAAAG
ERI	**ERI-1**	CCCAGTCCCTGTACTCGAAAG	AAGATCAGCAGCAGCCACTC
	ERI-2	TTCGATGGAATGTCTGAACG	CATGCCACACACTACAATTCG
	ERI-3	GTTGCACTAGTCAATCCTTAATGC	AGGCGTAACCACATTGTCTTC
HEN1	**HEN1-1a**	CGGTGGTGGGAAGCATAC	TCCGAGCCTTTCAGCAATAC
	HEN1-2	TGTGGATCAGTTCCCGTTTC	GCAGCAATTACAGCTTGTCG
	HEN1-3	GATGCATGACATTGGCTTTG	TTCACTGGTGAAGGCAGTTC
	HEN1-4	CTGATGCTGCTGAAGCTTTG	TGCAGAGAATCATGAGTGGTG
	HEN1-5	TCATCGGTGTTGACATTTCG	TGAACACGTCATCACATTGC
NRPD1	**NRPD1-1**	TGCACATTGGAATTGGTCAC	TTGGTTTCAAACAGCTCTCG
	NRPD1-2	TTACCGAAGACCAGCCTGAG	AAGGACGGATCACCAACAAC
	NRPD1-3	TCATGAAAGACGTGCTGCTC	TCGAGACTGCAGATCAGAGG
	NRPD1-4	GAAGGATCTGCTTGGCTCAG	AGATTCAGAAGCGGCGAAG
	NRPD1-5a	GACTGATGGTCCCGTTGAAG	TTTGATCCCTTGATCACCTG
	NRPD1-6	TTTGGAACTGGATGCTATTCG	TAAACCCGTGAACCTGATGG
	NRPD1-7	GAGCCCAAGTCAATGCTTTC	ATACCGCAGAAGAGCCAGTC
NRPE1	**NRPE1-1**	GTGAGCTCAAATGTCTTTGTTTCTC	CCTCACAACAGACACCCAAC
	NRPE1-2	CAAATGCTGAGCCTTCTGTG	CATGGTGCTGAAACCATCAG
	NRPE1-3	AAAGAAACTCACGGCAAAGG	TCAGCTGTCCAGTATGTGAGC
	NRPE1-4	CATGAAGACAACACCGTGAAG	CTGCAATTGAATGAGCCATC
	NRPE1-5	GGCAGGATTTCTTCATCTGG	TCCAATCTTGCAGCAATGTC
	NRPE1-6	TCTTTCTGGAAGCCTCTACCTG	TCACCACTTTGCTTGACAGC

Candidate gene	Amplicon	Forward primer (5'->3')	Reverse primer (5'->3')
	NRPE1-7	TGATTCACGGATTTGCTCAG	CTGCGTTTGTCGTGGATATG
	NRPE1-8	GCGTGGATGTTGGTACAGG	GTCCCAAGAATTCCATGCAG
	NRPE1-9	AATTCTGACGTTGGGTCAGG	GAAATTGGGTCTCCATCAGG
	NRPE1-10	CCGTCTGCGTATGCCTTTC	AAACCGCCATTCTCTCCAC
SDE3	**SDE3-1**	GCAGACAAGGGAGAGATTGG	GTCTCCATGGACAAGGGAAG
	SDE3-2	TCCCTGATGATTTGAACGAAG	TCATTGTCCTTGATGCGAAC
	SDE3-3	GAGAAATGCTCGGGTTCTTG	GACCTGTTTCTCTTGACCTTGG
	SDE3-4	AACAGCGAATGACTGTGTGC	ACTTCCACCCATCAGACCAC
SDE5	**SDE5-1**	TGCAGCTTCAAATGCATAAC	ATGAAGGCATATGGGAAAGC
	SDE5-2	GGTATTGGCGACCTTAGCTG	TCTTGCTCCAACACTGCATC
	SDE5-3a	TGTTGCTTTGTTTCCCTTTC	TTCCTGCCAATGGAGCATAC
	SDE5-4	TACGCGAGCATTTGAATCAG	TCAATCTCATCAACCCGAATC
SGS3	**SGS3-0**	TCGCTGATCGGAGTATTTGAC	TCACTTGCAAGGTCGTCATC
	SGS3-1a	GACGAGCGTTGAGCAGAAAG	CCATACAGTTGGCGAACACC
	SGS3-2	AAAGTTCTTTGGCAGCTTGG	GATGCATTAACACGGACCAAC
	SGS3-3	TCGCCAACTGTATGGCTTC	AGCTGTTCCAAAGCCTCATC
WEX	**WEX-1**	CGTCAAATTGGATCGACGAC	ACAAAGCTTGCAGCAATGAG
	WEX-2b	TGGCTTGGATATTGAGTGGAG	GGTTGGCTAAATCTGAAAGATCC
	WEX-3	CCATGGATGCTCAATTAACG	GACGGTTAAAGGGACCAGAAC
XRN4	**XRN4-1**	ACCTTCAAGCTCGAGACCAC	ATGCAAACTCACTCCCTTCG
	XRN4-2a	CCGAAAGTCAATAGAAGCAATG	AGACAATGCCGTGTATTTGG
	XRN4-3	TCGCATTCTTGCAGTTTCTC	TGCCTCAGTGGTATTGTAGAGC
	XRN4-4	ATGCCTACACTGGAAATTCG	TGCTCAGTTTCCTCTGGATTG
	XRN4-5	TATGTGCCTGTGAGCTGGAG	TTTGAGCTGCAACAACCAAG
	XRN4-6	ATTGGAGAGCCCTTTAAACC	TGGCTGCAACAAACACAAAC
	XRN4-7	GCCACCACATGAATGTCTTC	GCTCGAGTCATCGTCATAAGC
	XRN4-8	TATAGCAGGGCCTTCTCTGG	CGAATAACATCCAATTGCCAAG

Supplementary table 2. Amplicons used for amplification of specific regions of candidate genes in selected accessions.

Accession	Amplicon	Forward primer (5'->3')	Reverse primer (5'->3')
Bor-4, Gie-0	AGO7-1a	CCAACAAAGGTCTCTCTCTCAATC	GAAACTATTCCGTTCCGTCTCC
Lp2-2, Sq-8	HEN1-2_3	ATGTTTGGCAAAGCTTCCTG	TCATTGGCATCATCAACTCC
Lp2-2, Sq-8	HEN1-3_4	CGGCTAATTGTGAATCCTCAG	TTCTGATGGCCCTTTCACTC
Lp2-2, Sq-8	HEN1-4f/HEN1-3pr	CTGATGCTGCTGAAGCTTTG	GGTCTCCTCCTCTCCGTCAT
Lp2-2, Sq-8	HEN1-5f/HEN1-3pr	TCATCGGTGTTGACATTTCG	GGTCTCCTCCTCTCCGTCAT
Lp2-2, Sq-8	HEN1-3pr	CGCCATAACTACAGCGTCGA	GGTCTCCTCCTCTCCGTCAT
Bor-4, Gie-0	NRPD1-2_3	TTGAAGGATGAACGGACTCG	AACAGCAGATCGGGTTCAAG
Bor-4, Gie-0	NRPD1-4_5	AAGGGAATATCGGGAAGCTG	CAAGTGACCCGTATTCGAACC
Bor-4, Gie-0	NRPD1-5a_6	TGTGGATTAGAAACCACTGCTC	ACTCCAAATTCTGCGGTCAC
Cvi-0, Shahdara, Kas-1	NRPE1-4_5	TTCCATTCGGAACTTGATTG	TTGTTCCTGACTGTGCAAGC
Cvi-0, Shahdara, Kas-1	NRPE1-5_6	CTAATGATCGGCGCGTAATC	TTTGCTCCCACAAGGATCTC
Baa-1, Lz-0, Ra-0, Ws-0	SDE3-2_3	GCTTTGTACACCGCGTAGAG	GAAGCTTGACCAGCCTCATC
Baa-1, Lz-0, Ra-0, Ws-0	SDE3-3pr	GGCACAAAGGAGAAAGATGTCAGG	GGTCTGCTTCTTTCCGTCCG
Kin-0	WEX-2csf/Wex-3r	GATGAAAGTGTAGATGCCTCAG	GACGGTTAAAGGGACCAGAAC
Mt-0, Sq-8	WEX-2bf/Wex-3r	TGGCTTGGATATTGAGTGGAG	GACGGTTAAAGGGACCAGAAC
Ang-0, Baa-1, Kin-0, Lp2-2, Shahdara, Sq-8	WEX-5pr	CCACGTCTCTGTTGCCTTTTC	CGGAGCAGCAGAGGAAGAAG
Ang-0, Baa-1, Kin-0, Lp2-2, Shahdara, Sq-8	WEX-3pr	AGGCTTGGGAACTGGGAGT	TACAGCAGTCACTCCGAAGC

Supplementary table 3. Oligonucleotide pairs for semi-quantitative RT-PCR and/or qRT-PCR of reference and candidate genes.

Amplicon	Forward primer(5'->3')	Reverse primer (5'->3')	Amplicon length (bp)
At1g58050-qRT-1	CTCTCGATTGCTGCTTTCTTGAG	AAGAGCAAGTTTCACCCGGTC	90
At3g18780-qRT-2	TAAGGTCGTTGCACCACCTG	CCTCATCATACTCGGCCTTGG	113
At4g33380-qRT-2	GCATAAACATCAGGAGTGTTGGC	CGCTTTTGGGTGGTTGTTTCTG	100
At4g34270-qRT-1	CGTGAAAACTGTTGGAGAGAAGC	TGGAAGCCTCTGACTGATGG	108
AGO7_RT3	CACTCAATGACCCGTTTCC	GAACCATCTTGACCACCAAAG	112
HEN1_RT2	TGACATCGGCACTTGCTTAG	ACCGCTGGAGAATTGTGTTG	151
NRPD1-RT1	GATTCTCGGCTAGGATTACCG	AACGCAGCTACCTCCTTGAG	146
NRPE1_RT1a	AATGTGAATCTTGCGGTGCG	AGGCTCAGCATTTGTTTCAACTC	121
SDE3_RT1	GAAGGTCTTACAGCAAGGAAC	CAAGGCCAGGAACTTCGAG	154
SDE5-RTS1	AACTTTACTTCCTTCTCCTCTCGTC	TTCTCCACGAGTCTGTGTGC	1282
SDE5-4	TACGCGAGCATTTGAATCAG	TCAATCTCATCAACCCGAATC	935
SDE5-RTS2	AGAAACCGAGCGAGCACAC	TCATAATGCTTTAAACGTTGCAAC	862
WEX_RT	CCACGAGGAGGATCCAAATC	CCTACCACCAAACCTCATTGC	142

Supplementary table 4. Indel markers used for the analysis of introgression lines. The table shows the alias names of the Indel marker and their location on the sequence maps of the *A. thaliana* chromosomes. In the column "PCR" the annealing temperatures for the PCR amplification are listed.

Indel marker	Chr.	Position (Mbp)	PCR	Forward primer (5'->3')	Reverse primer (5'->3')
Nga59	1	0	60	GCATCTGTGTTCACTCGCC	TTAAAACAGTAGCCCAGACCCG
UPSC_1-1021	1	1.0	60	AAAGAATCAGGGACGGGTTC	TCAGTCCCTCTTCGACGTTT
F19P19	1	1.2	58	CCACGTAGGTCAAGAAGAAGAAG	TGTCTGCTGCGATAGAGAGAG
IndRIL-I-2a	1	1.2	60	GTCAAATCCATCTTCTTCCTTGATAC	GGCTCTTACAGCGTTCCATC
T1G11	1	1.2	58	GAAGACAAAGCTCTGCAGTAATG	AATTGCATAAGGCACTTGAAAG
ATEAT1	1	1.4	58	CCACTGCGTGAATGATATG	CGAACAGAAAACATTAATTCCC
F12K11-2-IND	1	2.0	60	GTAGTCTCAATAAGTCAACCGTTATCC	AGTGCAATCTTTGAAACTTGAGG
NGA63	1	3.2	55	AACCAAGGCACAGAAGCG	ACCCAAGTGATCACCACC
1-1259	1	3.8	49	GATATTTGTTTTGCTAACAC	TAATAAAGTTCCAGCTTTGA
Ind_I_5	1	4.5	60	AAGCCAAGTACCTCCAAGCA	GATCATCCCAAGGTCATGCT
1-2653	1	8.0	52.5	CACTGCAACAAAGTGGAAAT	ATCCGTTTCAATATCCACAA
MSAT1.3	1	8.3	55	GGAACTGTTGTCTGGGTAAG	CGATTGCACTAAAAGCTCTC
CIW12	1	9.6	52.5	AGGTTTTATTGCTTTTCACA	CTTTCAAAAGCACATCACA
NGA248	1	9.9	58	TCTGTATCTCGGTGAATTCTCC	TACCGAACCAAAACACAAAGC
Ind_I_12	1	12.2	60	AAGCGGAAAGGGACGTAGAT	TGGTAGTACGGGTTTTGGTC
1-4276	1	13.0	52.5	TATTATCTTGACTGGTGTAT	TTGACAATTTCCTTTATCT
T27K12	1	15.9	58	GGAGGCTATACGAATCTTGACA	GGACAACGTCTCAAACGGTT
1-5335	1	16.2	52.5	AAGCGGTCCAGTCCTTAAGC	CGAGAGATCACCCATCTGAA
CIW1	1	18.4	50	ACATTTTCTCAATCCTTACTC	GAGAGCTTCTTTATTTGTGAT
1-6613	1	20.1	47	AAGACAACTTGCCTTGTG	AAGCAAAACAAATTCCAGTA
NGA280	1	20.9	60	CTGATCTCACGGACAATAGTGC	GGCTCCATAAAAAGTGCACC
Ind_I_21	1	20.9	60	CCTTAATCTTAGAAATTGCAAATCG	AGAGTGGCAGCGAAGAAGAA
F11P17-4615	1	22.6	58	TTTCAGTTTGATGATTTATTCGC	CGCAATCGATTTTATTTAAATCC
1-7539	1	22.9	55	GAATTCTGTAACATCCCATTTCC	GGTCTAATTGCCGTTGTTGC
F5I14-IND	1	24.4	60	CTTTTCTGCCTGAAATTGTCG	GGGTTTCCAGCATTACTTGTTG
MSAT1.13	1	25.8	60	CCTGGTCAAACCAGTTCAATC	ACCACCAGGCTCTGTAATGG
1-8645	1	26.3	51.5	GGACCGACGGTTACGAGAGT	TAACGGGCCGTTGCAAGA
MSAT1.1	1	26.4	58	TCTCCTCCTGATGCAAATTC	CGTCTCAGAGATGATATTGCTACC
UPSC_1-26627	1	26.6	60	GCAATTCATCAGCAGGAGGT	ATCAGGGAGCAAAATGCAAG
Ind_I_27	1	26.8	60	ATTCGATACCTCCCATGTGC	AAGTTTGCGCAATTGGTAGG
ATHATPASE	1	28.5	58	GTTCACAGAGAGACTCATAAACCA	CTGGGAACGGTTCGATTCGAGC
UPSC_1-29617	1	29.6	60	CCCGATAATCTTCCCCAACT	GATGGCCGACGAGTACAAAT
MSAT2.5	2	0.2	60	TGAGAGGGACAGATAGGAAGG	ATCAAAAGGGATACTGACAAAGC
Ind_II_1	2	0.6	60	TGTTCCTGCTCTTCCTCACA	AGAGAACGTGGTACCGATGG
Nga1145	2	0.7	60	CCTTCACATCCAAAACCCAC	GCACATACCCACAACCAGAA
MSAT2.26	2	1.9	60	TCTCCGATTGAGCCCCAAAG	CGGGGAAAGATGGGTTTTGA
Ind RIL II29a	2	1.9	60	CAGTAACATCACCGGTATTCATG	CGTGACGACGCCAAAAG
MSAT2.38	2	2.5	50	TGTAACGCTAATTTAATTGG	CGCTCTTTCGCTCTG
Ind_II_4	2	4.2	60	GTCCTGGAGATGGTGGACAG	GGCAAAACCCTAATGTGGAA
MSAT2.28	2	6.4	55	AATAGAAATGGAGTTCGACG	TGAACTTGTTGTGAGCTTTG
2_3475	2	6.8	50	ATGTTGTTGGGGTTCTTG	ATGATTTCCGAAGTTTAG
MSAT2.11	2	8.2	50	GATTTAAAAGTCCGACCTA	CCAAAGAGTTGTGCAA
2-4269	2	8.4	50	ATGTATTTGTTGCAAAATAA	TGCACAGAAGAAAAAACTA
MSAT2.36	2	8.7	55	GATCTGCCTCTTGATCAGC	CCAAGAACTCAAAACCGTT

Indel marker	Chr.	Position (Mbp)	PCR	Forward primer (5'->3')	Reverse primer (5'->3')
Ind_II_9	2	8.9	60	TGCATTTCAACACCAACAAT	AAACGTTTCAATCCGCTGAC
UPSC_2_9168	2	9.2	60	CGGAAACGAAGACGAGTGAT	CGCCCCCTAATTTTTCTTTT
UPSC_2-9637	2	9.6	60	CACCAGCTGCCAAGTGTGTA	TGTGCCCTGCAAAACAATAG
PLS7	2	9.8	55	GATGAATCTTCTCGTCCAAAAT	GACAAACTAAACAACATCCTTCTT
MSAT2.17	2	10.7	52.5	GATTCCACCATATGTGGAT	CTTCGTCACTGCCAATAC
MSAT2.41	2	11.1	55	GACTGTTTCATCGGATCCAT	ACAAACCATTGTTGGTCGTG
2-5887	2	11.5	50	TCCGATTCGATTAAACTC	TTATTTCCTATTTCAAGACT
Nga1126	2	11.7	60	CAGAAAAGTCGCAGATAATAACAGAG	CAGTTCCTTTATCGCTCCTTCAC
CZSOD2	2	11.9	50	GGATCTCAATATGTGTCAAC	GCATTACTCCGGTGTCGTC
MSAT2.4	2	13.8	52.5	TGGGTTTTTGTGGGTC	GTATTATTGTGCTGCCTTTT
UPSC_2-14568	2	14.6	60	GAGTATATTCTTTTTTCGTCGGCA	TCGTAATAGCGGTTTCACCAC
2-8295	2	16.3	50	ATGAACGGAGTAGCTATC	CGCGTAGAACATAATCTGTA
MSAT2.9	2	18.1	60	TCCCTCGTAAAGACCAAACC	CTCGTTGTTGTTGTGGCATT
UPSC_2-18415	2	18.4	58	AATGGGACAAAATGGGTGAA	ATTCATTGCTGTTGCGGTTT
Ind_II_19	2	19.0	60	GAACACACTTGCGCGTCTAA	CCAACTTTATTGGCCTCACAA
MSAT2.22	2	19.6	60	CGATCCAATCGGTCTCTCTG	TGGTAACATCCCGAACTTCTG
3-0089	3	0.2	51.5	CAATCTAACAAGGCAAAAG	CTCGCAACCAACCTACAAG
3-0186	3	0.4	50	TCTGTTAATCCGGGTTATG	TTCTTGCCTCTCAGATTAAA
3-0363	3	0.8	52.5	CATCCGAATGCCATTGTTC	AGCTGCTTCCTTATAGCGTCC
Nga172	3	0.8	60	CATCCGAATGCCATTGTTC	AGCTGCTTCCTTATAGCGTCC
Ind_III_1	3	1.4	60	GACGTGGAGCTGAAATCGAC	TGACAAAACACAGAAATGAGAGG
RIL-III-50	3	3.5	60	TGTACCAGGCAATTCCAGTTC	AACGACTTCGTGTTCGTTCC
UPSC_3-3716	3	3.7	60	TAATGGTGGCCCAATCTCAT	AATTCCAAATGGAGCCACAA
Ind RIL III 52	3	5.6	60	TGGCTAATTAAAGAACGGTCTCAC	CGTACGTCCATGCCTTTACC
3-2402	3	5.6	50	ACCTGTTCAGTCTATGTTAC	GGGAATTATTAACATTATCA
Ind_III_6	3	6.0	60	TCCTCTCATGCAGAAGATGCT	GATTGAGGTGGGCACAGATT
MSAT3.19	3	8.8	60	CGATCCAATTGACATTGAAACC	GGCTTGGCACAAACTGAGAG
Ind_III_10	3	9.5	60	TGAGCAAACAGTCGGTCAAG	CCTAGGTCAACCCAATTTCG
IND-RIL-III-56	3	9.9	60	CCGTAACCCAAGAATAGTTGAATC	GATAACTTGTACATGATTCACCAAAAT
3-4332	3	10.1	50	ATGAGCTTTAGGAGTGTGTA	AATTTTGTCCCAAAAGAATA
MSAT3.32	3	11.2	60	GCACTTGCAGCTTAACTTTGTG	GTGACTGTCAAACCGTCTCTTTC
Ind_III_12	3	11.9	60	GCTAACAGGGATATCAAATGTGC	CGGACGAGCTGACACTTGTA
IND-RIL-III-63	3	16.2	60	CCAAAGTTTCTAAAATAATCGAATGG	GTTAGCTACGGTTCCGGTGTTAC
IND-RIL-III-62a	3	16.2	60	AGAAGATAGTTTTGTTCACTGTATATGATTC	GCTAACTTCTCCCTACCAGAAGC
Ind_III_17	3	17.1	60	TGTGTTCTCACCGCAGATTG	GAACAAATATGTCTCCAAGCACA
MSAT3.21_IND	3	17.3	60	AAAAGATGTTCACGGACTAGGTG	CTATTGTCAACGACGGTGTCTATC
INDRIL-III-64a	3	18.5	60	GAAGCAAGATAGGTTGTAGATGAACTAG	AAGAAATGCAAGGCTTCACAC
3-8728	3	20.5	50	GAGCAACATTAAGGATAGAA	ATCTCATACTCATAATATGTAG
Ind_III_21	3	21.2	60	TTTGGTGACAGGGAAAGGTC	ATAGGCCTGCTCCTGGATCT
3-9924	3	23.3	52.5	AAATGCCAGGGGAATAGA	CAAACCATGCAATGATGC
INDRIL-IV70b	4	0.1	60	CGTCGCCAGCATTTATTCTTC	CTAAAATTGAAGAAGCCGAAAAGAC
FRI-IND	4	0.3	60	GAGTGGTTCGAATGAGATTGC	GGAACTTGATGTTGGTCGATG
4-0175	4	0.4	49	GGTGTCTTGGTCTTGACTAT	AAAGCAATGCCAGGGAAC
4-1384	4	2.5	49	CTGAATCTCCCAGTTTATTT	AAGATCCAGGACATATAAGC
Ind_IV_3	4	2.7	60	AAACAACTAAATCGCCGTCA	CCTGTCGGTTTGCTATTGGT
UPSC_4-2821	4	2.8	57	GAACTAATTTCTCGTAAAACAAACAA	AGCACATTGCTTGCATAAACT
INDRIL-IV-74	4	5.2	60	AAGATTTCGGGTCAGGTTTG	GAGATAACCTCGGTAAAACTTCCTG

Indel marker	Chr.	Position (Mbp)	PCR	Forward primer (5'->3')	Reverse primer (5'->3')
Nga8	4	5.6	60	GAGGGCAAATCTTTATTTCGG	CATCCTTTAGTGAAACAACGGAAC
Nga1111	4	6.1	58	TGGTTCGGTTACAATCGTGT	AGTTACCAGATTGAGCTTTGAGC
UPSC_4-6222	4	6.2	60	CAGAACCAAGCTGCAATGAA	CCTTCGATGTCTTCGCTGAT
UPSC_4-6362	4	6.4	58	TGTCCACTGGATTGGAGATT	TGGCTCTTTCTAATCAAAATAGTGT
MSAT4.25	4	7.0	50	GAATGGTTGTTGATAGTTGA	AAATTTCAGGAGGTGATAGA
MSAT4.35	4	7.5	55	CCCATGTCTCCGATGA	GGCGTTTAATTTGCATTCT
Ind_IV_8	4	7.6	60	TTGATTAGGCCACGGTTAGG	TGTCGCTTCTTCAGATGTGG
MSAT4.15	4	9.3	55	TTTCTTGTCTTTCCCCTGAA	GACGAAGAAGGAGACGAAAA
4-5268	4	9.7	49	TTCGGAGAAAGAAACGACAT	ATGGAACTATTCAGGCATTA
UPSC_4-11022	4	11.0	60	CCCAACCAAACTGAACCAAC	CAACTTTCTTGCGCCATTTC
UPSC_4-11152	4	11.2	60	CTGAGTGGACGTTTTGGTTG	GCATCTATGAGTGCATGAGCA
CIW7	4	11.5	55	AATTTGGAGATTAGCTGGAAT	CCATGTTGATGATAAGCACAA
MSAT4.18	4	12.0	60	TGTAAATATCGGCTTCTAAG	CTGAAACAAATCGCATTA
UPSC_4-12273	4	12.3	60	TCAACCAATCGCCTTAGTCA	TTTCCATTTGCATGCTCGTA
4-7366	4	13.7	49.3	GACGTCGTTTCAATAACTAA	AGTTCCATTCCACGAATCTT
Ind_IV_15	4	14.6	60	GGTTATTCGACGCACTTGCT	TACTCCCGTGATCGGTCAAT
UPSC_4-14985	4	15.0	60	TGGCTGCAGCGAATAACTAA	TGGTGCTGGTGAAACCAATA
INDRIL-IV-86	4	17.8	60	TTGACAAGAAAATTGGCTCAATC	GAAACAACCCAAATAATTAGTCACCTAC
4-9963	4	18.5	49.3	ATCCGATCTCAAACAGAGTC	GTCGGGTTTCTGTATCTCC
CTR1.2	5	1.0	50	CCACTTGTTTCTCTCTCTAG	TATCAACAGAAACGCACCGA
Nga249	5	2.8	58	GGATCCCTAACTGTAAAATCCC	TACCGTCAATTTCATCGCC
Ind_V_5	5	5.4	60	CCTTTGAAAAACCGCCATTA	AGATCTCATACCGCCGGAGT
Nga106	5	5.4	60	GTTATGGAGTTTCTAGGGCACG	TGCCCCTTTTGTTCTTCTCC
5-2862	5	7.7	50	TTCATGAGAGCGGCATTC	GCAAAATGTTTGGACAATTA
Nga139	5	8.4	60	GGTTTCGTTTCACTATCCAGG	AGAGCTACCAGATCCGATGG
Ind_V_9	5	9.5	60	TGTGGCACAGGGTTTGTAAG	AAAGCCAGCCAATGTTTCAC
5-5037	5	13.5	50	CACAGGCCATTGGATGTA	TGTTAGAACCCACCATTTG
MSAT5.2	5	14.0	55/52.	TCTCAGACATGGAAATCTTGT	GGCATTTTTAACACTTTCAAA
PHYC.3	5	14.0	55	AAACTCGAGAGTTTTGTCTAGATC	CTCAGAGAATTCCCAGAAAAATCT
5-6437	5	17.3	50	AAGGATCTCGTCTTCAATAG	GTACTTAGCGTCGCACAC
5-7443	5	20.1	50	CCTGTTCCAATGAATATG	TGTAGCTGCTGAGTTGTC
Nga129	5	20.1	60	AAATCGTAAAACCTATAGAGAAACATCG	CAACACTGAAGATGGTCTTGAGG
Ind_V_22	5	22.4	60	CACATCTGAAGCTGTGTTGCTCGT	CGCTAACGCTCTTTGGCGATCTTT
INDRIL-V-112	5	26.9	60	GAGCGAGACATAAGCAATCG	TGACCATGCTGTCACTTTACTG
K8K14-IND	5	26.9	60	CTAAGTATGCCACATAACTGAATTTTTG	TGGTGGAAACTTCGTCTTCTG
Ind_V_27	5	27. 0	60	AAATGATATCCGAGCAACACG	TGGTCGGGTCAATTTCAACT

Supplementary table 5. Indel markers and allele-specific oligonucleotides for selected accessions and candidate genes. Length polymorphisms in the listed accessions are observed in comparison to Col-0. * or ** indicates the annealing temperature for the PCR amplification is 50°C or 64 °C, respectively.

A	Accession	Amplicon	Forward primer (5'->3')	Reverse primer (5'->3')
	Lp2-2, Sq-8	HEN1-1alp2	AATGTCTTCAAGAAGAAGAAGGATTC	TCCACAGTAAGATCATCATTCTGAG
	Cvi-0, Kas-1, Shahdara	NRPE1-10lp	GGTAATGGAGGTGACGACTTTC	GGAGACTGAGATGATGGAGACTG
	Baa-1, Ws-0	SDE3-4alp	GCGAATGAGAATGGTGAATGGTC	CTTTCTCCTTTGTGCCACCATTC
	Ang-0, Baa-1, Lp2-2, Shahdara, Sq-8	WEX-2b	TGGCTTGGATATTGAGTGGAG	GGTTGGCTAAATCTGAAAGATCC
B	**Accession**	**Amplicon**	**Forward primer (5'->3')**	**Reverse primer (5'->3')**
	Lp2-2, Sq-8	HEN1-2a	GATGATATCCAGGTGATTATATTGTCA	CAGACATAAAACTGAGCGAAATTG
	Bor-4, Gie-0	AGO7-1a	CCAACAAAGGTCTCTCTCTCAATC	GAAACTATTCCGTTCCGTCTCC
	Bor-4, Gie-0	AGO7-2a	GGTGTGCTTTCCCACACAG	AGCTTGATCATCCGTAAGCTTC
	Col-0	AGO7-1H*	GGTCTAAAGGTTTAGGATCC	TTTGTGTGGGAATACACAG
	Lz-0, Ra-0	SDE3-2b	CAATAAACCTCTTGTATTTTACCATGAAG	TGATGCGAACTCCCTCCAG
	Col-0	SDE3-2H1**	AACTCAATAAACCTCTTGTACCATGAAT	CGAACTCCCTCTAAACAAAGGAGT
	Bor-4, Gie-0	NRPD1-5bs1**	AATGTCTTTGTATTTGTCTGAAAC	CGTCTTTATTTTTCATCCGGC
	Col-0	NRPD1-5bsH1**	TTGTTTCATCTTTAAACGAGCAG	CGATAATAACAAAACAACGATCTG
	Kin-0	WEX-2cs	GATGAAAGTGTAGATGCCTCAG	CAATTCCAATACCTACCTATACATATAC
	Col-0	WEX-2cH1	GAATAGAGCTTTATGGTCGTTG	GCAAACAACAAGAACTTCGAC

Supplementary table 6. Primer sequences for the analysis of *GFP* T-DNA lines. Combination of oligonucleotides LB1c and LB2 were used to analyse the presence of a particular T-DNA locus. Combinations of oligonucleotides LB2 and RB2 were used to assess the presence of the empty donor site of a particular T-DNA locus.

Amplicon	Oligonucleotide	Sequences (5'->3')
F8 T-DNA	**mgfpf8 LB2**	GCTGAATTCCATGATCTCACTG
	Lb1c	TGGGTATCTGGGAATGGCGAAATC
F18 T-DNA	**mgfpf18 LB2**	GAGGCGTTGCTTCTTCTCTAAC
	Lb1c	TGGGTATCTGGGAATGGCGAAATC
F128 T-DNA	**mgfpf128 LB2**	TCTCATCTTTGCTGTGATGTAGAAC
	Lb1c	TGGGTATCTGGGAATGGCGAAATC
R127 T-DNA	**Lb1c**	TGGGTATCTGGGAATGGCGAAATC
	mgfpr127 LB2	TCGTTAACGTGGACGAAATC
F8 EDS	**mgfpf8 LB2**	GCTGAATTCCATGATCTCACTG
	mgfpf8 RB2	GCTGCTAAGGCTACTTGCGACG
F18 EDS	**mgfpf18 LB2**	GAGGCGTTGCTTCTTCTCTAAC
	mgfpf18 RB2	CTAGACATCAACACAAACATACAC
F128 EDS	**mgfpf128 LB2**	TCTCATCTTTGCTGTGATGTAGAAC
	mgfpf128 RB2	TTTGAAAGACTTGCCGTGTAAC
R127 EDS	**mgfpr127 RB2**	AAACCCTATCAGAGACCGATTG
	mgfpr127 LB2	TCGTTAACGTGGACGAAATC

Supplementary table 7. Regions of candidate genes and ORFs sequenced in all 26 accessions. The positions refer to the Col-0 gene and ORF sequences that were retrieved from TAIR. The ORF corresponding to splicing variant 1 was used if more than one splicing variant was available.

	Amplicon											
	0	**1 or 1a**	**2, 2(a) or 2(b)**	**3 or 3a**	**4**	**5 or 5a**	**6**	**7**	**8**	**9**	**10**	**11**
AGO1		528-1517	1449-2444	2200-3203	3001-4026	3858-4899	4685-5681	5456-6399				
AGO1-ORF		602-1431	1581-2367	2280-3093	3065-3948	3912-4819	4758-5608	5538-6290				
AGO7		92-1055	901-1948	1807-2802	2653-3548							
AGO7-ORF		203-970	985-1875	1869-2740	2716-3491							
DCL4		175-1179	1057-2012	1912-2912	2839-3843	3662-4648	4547-5565	5432-6437	6333-7293	7028-8044	7761-8758	8665-9707
DCL4-ORF		253-1088	1103-1961	1995-2846	2904-3639	3801-4549	4615-5483	5542-6332	6532-7230	7097-7981	7827-8654	8732-9637
ERI		5-714	504-1455	1265-2185								
ERI-ORF		50-683	590-1398	1320-2132								
HEN1		122-1121	768-1767	1673-2774	2586-3695	3522-4485						
HEN1-ORF		196-1067	844-1695	1741-2386	2940-3623	3594-4375						
NRPD1		1330-2344	2027-3024	2952-3937	3755-4752	4621-5726	5556-6543	6311-7281				
NRPD1-ORF		1390-2201	2104-2965	3017-3852	3821-4633	4682-5663	5615-6483	6384-7216				
NRPE1		500-1411	1216-2227	2113-3083	2912-3902	3809-4781	4679-5686	5546-6538	6316-7305	7171-8184	8026-9012	
NRPE1-ORF		553-1355	1275-2164	2174-3019	3002-3792	3922-4644	4744-5626	5626-6420	6400-7222	7312-8104	8113-8936	
SDE3		418-1354	1064-1999	1898-2819	2666-3669							
SDE3-ORF		498-1279	1138-1929	1964-2736	2743-3551							
SDE5		679-1577	1420-2322	2117-3121	2684-3618							
SDE5-ORF		715-1545	1532-2264	2171-3054	2748-3559							
SGS3	295-1048	829-1775	1112-1908	1761-2714								
SGS3-ORF		319-1028	891-1712	1175-1828	1833-2646							
WEX		48-1005	667-1464	1329-2151								
WEX-ORF		121-936	752-1348	1380-2080								
XRN4		55-1005	750-1877	1667-2693	2533-3531	3358-4353	4210-5199	5056-6055	5848-6685			
XRN4-ORF		134-943	886-1808	1752-2592	2595-3476	3460-4243	4267-5122	5118-5990	5909-6638			

Supplementary table 8. Compilation of SNPs and Indels detected in 26 accessions for 12 candidate genes. Col-0 gene and ORF sequences were retrieved from TAIR and used as reference for polymorphism detection. If more than one splicing variant was listed in TAIR, the ORF corresponding to splicing variant 1 was used for the analysis. All positions given in the table refer to the Col-0 reference sequences in the manually edited sequence alignments. Nucleotides shown in italics represent Indels. For Indels that consist of microsatellite sequences the number and identity of repeat units are given. For deletions spanning more than 10 bp only the first and last three nucleotides are shown but the lengths of all deletions (Δ) relative to the Col-0 sequences are also listed in the table. Asterisks indicate linked polymorphisms that affect the same codon in one or several accessions. For all SNPs and/or Indels present in ORF sequences the relevant codons and the corresponding amino acids are shown according to IUPAC codes. One Indel caused a frame shift in the ORF, this is indicated as "FS". Accessions showing a particular polymorphism are marked by "x".

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
AGO1-1	663//664		*T*													x												x				
AGO1-1	670	T	A													x												x				
AGO1-1	749	T	C	55	TCT-S	CCT-P								x																		
AGO1-1	1000-1011	*AGG...TTC*	*12 bp Δ*	306-317	GGA...TCT -GGGPS	GGT-G				x																						
AGO1-1	1083	C	T	389	TCT-S	TTT-F																			x	x						
AGO1-1	1207	T	G	513	CCT-P	CCG-P					x			x	x	x		x						x				x		x		x
AGO1-1	1316	T	C	622	TTG-L	CTG-L			x				x									x										
AGO1-1	1386	A	G													x												x				
AGO1-1	1402	G	T																						x	x						
AGO1-1	1408	T	A																						x	x						
AGO1-2	1659	*T*	*1 bp Δ*														x															
AGO1-2	2106	T	G					x	x				x				x					x										
AGO1-2	2258	A	C	1212	CGA-R	CGC-R			x				x									x										
AGO1-2/3	2301	C	A																x													
AGO1-3	2512	T	C								x				x	x		x	x					x				x		x		
AGO1-3	2660	A	G	1431	CAA-Q	CAG-Q					x													x								
AGO1-3	2922	A	T									x																				
AGO1-3	2963	G	T									x																				
AGO1-3	3013	C	A																						x	x						
AGO1-3	3043	G	T					x	x				x				x					x										
AGO1-3	3043	G	A																													x
AGO1-3/4	3076	A	G																						x	x						
AGO1-4	3113	T	A								x			x	x	x		x	x					x				x		x		
AGO1-4	3119	A	G								x				x	x		x	x					x				x		x		

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
AGO1-4	3299-3300	*TG*	*2 bp Δ*																													x
AGO1-4	3354	T	A						x				x									x										
AGO1-4	3394	T	C	1626	GCT-A	GCC-A											x															
AGO1-4	3494	C	G														x															
AGO1-4	3552	C	T					x	x				x				x					x										
AGO1-4	3722	C	T					x	x		x		x	x	x	x	x	x	x			x		x				x		x		x
AGO1-4	3801	T	C												x				x													
AGO1-4	3843	A	G					x																								
AGO1-4/5a	3924	G	T														x															
AGO1-4/5a	3941	T	G								x				x	x		x	x					x				x		x		
AGO1-5a	3964	T	A						x				x									x										
AGO1-5a	3979	A	G					x	x		x		x	x	x	x	x	x	x			x		x				x		x		x
AGO1-5a	4094	T	C	1878	CTT-L	CTC-L										x												x				
AGO1-5a	4142	*A*	*1 bp Δ*					x	x		x		x	x	x	x	x	x	x			x		x				x		x		x
AGO1-5a	4191	T	A						x				x									x										
AGO1-5a	4600	A	T	2208	ACA-T	ACT-T		x																								
AGO1-5a/6	4777	T	C					x	x		x		x	x	x	x	x	x	x			x		x				x		x		x
AGO1-5a/6	4780	T	C														x															
AGO1-6	4825	T	C					x																								
AGO1-6	4852	G	A	2343	CAG-Q	CAA-Q			x																							
AGO1-6	4996	A	G					x	x				x				x					x										
AGO1-6	5002	T	C					x																								
AGO1-6	5008	C	G																		x											
AGO1-6	5164	C	T											x																		
AGO1-6	5185	C	T					x									x															
AGO1-6	5203//5204		*T*						x				x									x										
AGO1-6	5279	T	C	2586	GAT-D	GAC-D																										x
AGO1-6	5318//5319		*A*																	x												
AGO1-6	5330	T	C					x	x				x				x					x										
AGO1-6	5349	C	G															x												x		
AGO1-6	5355	A	T					x	x		x		x	x	x	x	x	x	x			x		x				x		x		x
AGO1-6	5357	T	A					x	x		x		x	x	x	x	x	x	x			x		x				x		x		x
AGO1-6	5364	A	T																													x
AGO1-6	5455	T	C	2679	TTT-F	TTC-F																										x

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
AGO1-6	5523	T	C																													x
AGO1-6/7	5556	A	T																												x	
AGO1-7	5691-5700	*CAA...AAA*	*10 bp Δ*					x	x		x		x	x	x	x	x	x	x			x		x				x		x		
AGO1-7	5712	*G*	*1 bp Δ*																	x												
AGO1-7	5727	A	T						x				x									x										
AGO1-7	5934	C	T																													x
AGO1-7	5936	A	G																													x
AGO1-7	5946	T	C					x	x		x		x	x	x	x	x	x	x			x		x				x		x		x
AGO1-7	5989	A	T								x				x	x		x	x					x				x		x		
AGO1-7	6056	T	G																													x
AGO1-7	6072-6073	*TT*	*2 bp Δ*																													x
AGO1-7	6073	*T*	*1 bp Δ*						x		x	x	x	x	x	x	x	x	x	x		x		x	x	x		x		x		
AGO1-7	6092	T	G	2937	GTT-V	GTG-V																										x
AGO1-7	6095	C	G	2940	CCC-P	CCG-P																										x
AGO7-1_1a	203	G	A	53	AGT-S	AAT-N				x	x				x							x			x		x					
AGO7-1_1a	227	T	C	77	CTC-L	CCC-P				x	x				x							x			x		x					
AGO7-1_1a	258	C	T	108	ACC-T	ACT-T			x																				x			
AGO7-1_1a	313	G	A	163	GCC-A	ACC-T														x												
AGO7-1_1a	339	C	T	189	TAC-Y	TAT-Y					x				x																	
AGO7-1_1a	401-403	*CTC*	*3 bp Δ*	251-253	CCTCAT-PH	CAT-H					x				x																	
AGO7-1_1a	403//404		*CTC*	253//254	CAT-H	CCTCAT-PH								x																		
AGO7-1_1a	421	C	A	271	CCT-P	ACT-T																x			x							
AGO7-1_1a	477	C	G	327	CAC-H	CAG-Q					x				x																	
AGO7-1_1a	505	T	A											x																		
AGO7-1_1a	507-509	*CAT*	*3 bp Δ*								x				x																	
AGO7-1_1a	515	C	T								x				x																	
AGO7-1_1a	521//522		*T*								x				x																	
AGO7-1_1a	525	C	G								x				x																	
AGO7-1_1a	550//551		*TGTTTTTTCAT*								x				x																	
AGO7-1_1a	553	T	C								x				x																	
AGO7-1_1a	554	C	T								x				x																	
AGO7-1_1a	580	G	T								x				x																	
AGO7-1_1a	595	A	T								x				x																	

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
AGO7-1_1a	598//599		*T*																											x		
AGO7-1_1a	602	G	A								x				x																	
AGO7-1_1a	624-652	*TCT...TTC*	*29 bp Δ*									x	x				x															x
AGO7-1_1a	631	G	C								x				x																	
AGO7-1_1a	643	C	G					x			x			x	x	x			x	x	x		x	x		x		x				
AGO7-1_1a	648	T	C								x				x																	
AGO7-1_1a	649	A	T								x				x																	
AGO7-1_1a	660	A	G								x				x																	
AGO7-1_1a	671	A	T								x				x																	
AGO7-1_1a	678	*A*	*1 bp Δ*								x				x																	
AGO7-1_1a	694-696	*TCT*	*3 bp Δ*								x				x																	
AGO7-1_1a	745	T	A																	x	x			x		x						
AGO7-1_1a	753	G	T																								x					
AGO7-1_1a	892	A	C																											x		
AGO7-1_1a	907	A	G	360	AGA-R	AGG-R					x				x																	
AGO7-1_1a	917	C	A	370	CAC-H	AAC-N								x																		
AGO7-1_1a	923//924		*AACAGC*	376//377	CAG-Q	CAACAGCAG-QQQ					x				x																	
AGO7-1_1a	937	A	T	390	ATA-I	ATT-I					x				x																	
AGO7-1_1a	957	T	C	410	GTG-V	GCG-A											x															x
AGO7-1_1a	962	G	A	415	GGA-G	AGA-R																x			x							
AGO7-2	1009	G	T	462	CCG-P	CCT-P					x				x																	
AGO7-2	1036	C	T	489	GTC-V	GTT-V					x				x																	
AGO7-2	1042	T	C	495	TAT-Y	TAC-Y					x				x																	
AGO7-2	1078	G	T	531	TCG-S	TCT-S					x				x																	
AGO7-2	1102	T	C	555	AAT-N	AAC-N					x				x																	
AGO7-2	1111	A	C	564	ATA-I	ATC-I					x				x																	
AGO7-2	1117	A	T	570	CCA-P	CCT-P					x				x																	
AGO7-2	1139-1141*	AGA*	CGG*	592-594*	AGA-R*	CGG-R*					x				x																	
AGO7-2	1162	T	G	615	GTT-V	GTG-V					x				x																	
AGO7-2	1168	A	G	621	ACA-T	ACG-T					x				x																	
AGO7-2	1171	T	A	624	GAT-D	GAA-E					x				x																	
AGO7-2	1174	T	G	627	CGT-R	CGG-R					x				x																	
AGO7-2	1180	C	T	633	AGC-S	AGT-S					x				x																	

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
AGO7-2	1210	C	T	663	GGC-G	GGT-G					x				x																	
AGO7-2	1240	T	A	693	TTT-F	TTA-L					x				x																	
AGO7-2	1245	G	A	698	GGA-G	GAA-E					x				x																	
AGO7-2	1490	A	G	943	ACA-T	GCA-A																					x					
AGO7-2	1591	T	G	1044	CAT-H	CAG-Q					x				x																	
AGO7-2	1648	T	C	1101	AGT-S	AGC-S					x				x																	
AGO7-2	1678	G	T	1131	CGG-R	CGT-R					x				x																	
AGO7-2	1681	C	T	1134	CTC-L	CTT-L					x				x																	
AGO7-2	1690	T	C	1143	CTT-L	CTC-L					x				x																	
AGO7-2	1692	C	A	1145	ACG-T	AAG-K					x				x																	
AGO7-2	1727	T	C	1180	TTA-L	CTA-L					x				x																	
AGO7-2	1747	G	A	1200	GTG-V	GTA-V					x				x																	
AGO7-2	1813	T	C	1266	GTT-V	GTC-V					x				x																	
AGO7-2	1816	T	C	1269	TAT-Y	TAC-Y					x				x																	
AGO7-2	1831	A	G	1284	GAA-E	GAG-E					x				x																	
AGO7-2	1855	T	G	1308	CCT-P	CCG-P					x				x																	
AGO7-2	1864	A	T	1317	GAA-E	GAT-D					x				x																	
AGO7-2	1868	A	G	1321	AAA-K	GAA-E			x																							
AGO7-2/3	1870	A	G	1323	AAA-K	AAG-K					x				x																	
AGO7-2/3	1871	T	C	1324	TAT-Y	CAT-H					x				x																	
AGO7-3	2017	A	G	1470	GGA-G	GGG-G					x				x																	
AGO7-3	2023	A	T	1476	CTA-L	CTT-L					x				x																	
AGO7-3	2024-2026*	TCA*	ACG*	1477-1479*	TCA-S*	ACG-T*					x				x																	
AGO7-3	2044	G	A	1497	AAG-K	AAA-K																								x		
AGO7-3	2116	C	T	1569	GTC-V	GTT-V					x				x																	
AGO7-3	2137	A	T								x				x																	
AGO7-3	2175	A	C								x				x																	
AGO7-3	2178	G	A								x				x																	
AGO7-3	2182	T	G																											x		
AGO7-3	2186	T	A																								x					
AGO7-3	2198	G	T								x				x																	
AGO7-3	2261	G	A	1632	TTG-L	TTA-L					x				x																	
AGO7-3	2285	A	T	1656	CCA-P	CCT-P					x				x																	
AGO7-3	2336	T	A	1707	TTT-F	TTA-L					x				x																	

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
AGO7-3	2450	A	T	1821	GGA-G	GGT-G					x				x																	
AGO7-3	2454	T	G	1825	TTC-F	GTC-V					x				x																	
AGO7-3	2480	T	C	1851	AGT-S	AGC-S		x	x	x	x	x	x	x	x	x	x			x	x	x		x	x	x	x	x	x	x		x
AGO7-3	2482	C	T	1853	ACC-T	ATC-I															x			x								
AGO7-3	2492	A	G	1863	GAA-E	GAG-E					x				x																	
AGO7-3	2499	C	T	1870	CAC-H	TAC-Y					x				x																	
AGO7-3	2515	T	A	1886	ATT-I	AAT-N																								x		
AGO7-3	2525	C	T	1896	CTC-L	CTT-L					x				x																	
AGO7-3	2531	A	G	1902	TCA-S	TCG-S					x				x																	
AGO7-3	2561	G	T	1932	TCG-S	TCT-S																								x		
AGO7-3	2563	A	G	1934	AAC-N	AGC-S					x				x																	
AGO7-3	2570	C	T	1941	CTC-L	CTT-L					x				x																	
AGO7-3	2582	T	C	1953	ATT-I	ATC-I					x				x																	
AGO7-3	2612	C	T	1983	TAC-Y	TAT-Y					x				x																	
AGO7-3	2624	G	A	1995	AAG-K	AAA-K								x																		
AGO7-3	2631	T	G	2002	TCA-S	GCA-A					x				x																	
AGO7-3	2654	G	C	2025	GTG-V	GTC-V					x				x																	
AGO7-3	2686	A	T	2057	AAG-K	ATG-M				x																						
AGO7-3	2695	C	G	2066	TCT-S	TGT-C					x				x																	
AGO7-3	2708	G	A	2079	TCG-S	TCA-S					x				x																	
AGO7-4	2768	G	A	2139	TCG-S	TCA-S					x				x																	
AGO7-4	2786	T	A	2157	CCT-P	CCA-P					x				x																	
AGO7-4	2801	C	T	2172	CCC-P	CCT-P					x				x																	
AGO7-4	2885	A	C	2256	GTA-V	GTC-V					x				x																	
AGO7-4	2933	G	A	2304	AGG-R	AGA-R					x				x																	
AGO7-4	3002	T	C	2373	TTT-F	TTC-F					x				x																	
AGO7-4	3011	G	T	2382	GCG-A	GCT-A					x				x																	
AGO7-4	3026	G	A	2397	CCG-P	CCA-P																							x			
AGO7-4	3047	G	A	2418	AGG-R	AGA-R					x				x																	
AGO7-4	3056	T	G	2427	GTT-V	GTG-V					x				x																	
AGO7-4	3101	G	C	2472	TCG-S	TCC-S					x				x																	
AGO7-4	3110	T	A	2481	ACT-T	ACA-T					x				x																	
AGO7-4	3119	G	A	2490	TCG-S	TCA-S					x				x																	
AGO7-4	3128	A	C	2499	CAA-Q	CAC-H					x				x																	

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
AGO7-4	3179	A	G	2550	ACA-T	ACG-T					x				x																	
AGO7-4	3194	C	T	2565	TGC-C	TGT-C					x				x																	
AGO7-4	3200	A	T	2571	CCA-P	CCT-P					x				x																	
AGO7-4	3209	G	A	2580	GAG-E	GAA-E					x				x																	
AGO7-4	3215	A	C	2586	ATA-I	ATC-I					x				x																	
AGO7-4	3224	T	C	2595	GGT-G	GGC-G					x				x																	
AGO7-4	3246	A	G	2617	ACT-T	GCT-A					x				x																	
AGO7-4	3254	G	A	2625	CCG-P	CCA-P					x				x																	
AGO7-4	3326	T	C	2697	ATT-I	ATC-I					x				x																	
AGO7-4	3329	A	G	2700	CTA-L	CTG-L					x				x																	
AGO7-4	3347	C	T	2718	TTC-F	TTT-F					x				x																	
AGO7-4	3380	T	C	2751	AAT-N	AAC-N					x				x																	
AGO7-4	3416	A	T	2787	ATA-I	ATT-I					x				x																	
AGO7-4	3424	T	C	2795	GTG-V	GCG-A																	x									
AGO7-4	3470	A	T	2841	CTA-L	CTT-L					x				x																	
DCL4-1	322	G	A	124	GAC-D	AAC-N																						x				
DCL4-1	338	A	T	140	CAC-H	CTC-L													x													
DCL4-1	348-350	*TGC*	*3 bp Δ*	150-152	GCTGCC-AA	GCC-A		x					x				x															
DCL4-1	360	G	A	162	AAG-K	AAA-K			x	x																x						
DCL4-1	409	C	T	211	CTT-L	TTT-F								x																		
DCL4-1	498	G	C	300	TCG-S	TCC-S		x					x				x															
DCL4-1	535	T	C	337	TCA-S	CCA-P						x																				
DCL4-1	567	C	T	369	ATC-I	ATT-I								x																		
DCL4-1	668	G	A					x					x				x															
DCL4-1	670	A	G					x					x				x															
DCL4-1	702	C	T					x					x				x															
DCL4-1	762	G	A																								x					
DCL4-1	843	G	A	418	GTT-V	ATT-I		x					x	x			x															
DCL4-1	905	G	T	480	GAG-E	GAT-D		x					x				x															
DCL4-1	983	G	A					x					x				x															
DCL4-1	1019	A	T					x					x	x			x															
DCL4-2	1184	*T*	*1 bp Δ*																	x												
DCL4-2	1192	*T*	*1 bp Δ*																												x	
DCL4-2	1214	G	A											x																		

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
DCL4-2	1291	T	C											x																		
DCL4-2	1301-1304	*GTTT*	*4 bp Δ*					x					x	x			x															
DCL4-2	1314	G	T																								x					
DCL4-2	1460	A	G	795	GCA-A	GCG-A		x					x	x			x															
DCL4-2	1511	T	C					x			x		x	x		x	x		x		x		x	x			x	x	x			x
DCL4-2	1522	T	A																	x												
DCL4-2	1535	*T*	*1 bp Δ*																x													
DCL4-2	1582	G	C	827	AGT-S	ACT-T										x												x				
DCL4-2	1681	C	T									x																				
DCL4-2	1736	C	A					x					x				x															
DCL4-2	1786//1787		*T*																													
DCL4-2	1803	A	G									x																				
DCL4-2	1806	T	C											x																		
DCL4-2	1826	T	C					x					x	x			x															
DCL4-3	2205	T	G					x					x	x			x															
DCL4-3	2458	A	G					x					x	x			x															
DCL4-3	2569	T	A	1168	TCT-S	ACT-T		x			x		x	x		x	x		x		x		x	x			x	x	x			x
DCL4-3	2596	C	G	1195	CTG-L	GTG-V		x	x	x	x	x	x	x	x	x	x		x	x	x		x	x	x	x	x	x	x	x	x	x
DCL4-3	2683	T	C																				x									
DCL4-3	2705	G	T					x					x	x			x															
DCL4-3	2762	G	T					x					x	x			x															
DCL4-3	2790	A	C									x																				
DCL4-3	2795	T	G											x																		
DCL4-4	2925	A	G																										x			
DCL4-4	2932	G	A																		x		x	x								x
DCL4-4	2943//2944		*T*					x					x	x			x															
DCL4-4	3272	T	C													x												x				
DCL4-4	3275	*T*	*1 bp Δ*											x																		
DCL4-4	3275//3276		*T*								x										x		x	x								x
DCL4-4	3290//3291		*AA*									x																				
DCL4-4	3296	T	G																						x							
DCL4-4	3319//3320		*T*								x																					
DCL4-4	3349	T	A													x												x	x			
DCL4-4	3395	G	T																													x

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
DCL4-4	3487	G	A	1369	GAC-D	AAC-N									x																	
DCL4-4	3616	G	T					x					x	x			x															
DCL4-5	3958	T	A					x					x	x			x															
DCL4-5	3975	C	G					x					x				x															
DCL4-5	4097	G	C								x					x			x		x		x	x			x	x	x			x
DCL4-5	4368	T	G									x																				
DCL4-5	4382	G	A													x												x				
DCL4-5	4458	T	C								x					x			x		x		x	x			x	x	x			x
DCL4-5	4462	C	T					x					x				x															
DCL4-6	4686	G	A					x					x	x			x															
DCL4-6	4698	C	G											x																		
DCL4-6	4740	T	G																				x									
DCL4-6	5031	G	T					x			x		x	x		x	x		x		x		x	x			x	x	x			x
DCL4-6	5069	C	T												x																	
DCL4-6	5336	C	T	2277	AAC-N	AAT-N		x					x	x			x															
DCL4-6	5392	T	C					x					x	x			x															
DCL4-7	5615	A	G	2471	GAG-E	GGG-G								x																		
DCL4-7	5703	T	C	2559	GAT-D	GAC-D		x					x	x			x															
DCL4-7	5872	T	A	2659	TCT-S	ACT-T																					x					
DCL4-7	6002	C	G	2789	TCT-S	TGT-C										x												x				
DCL4-7	6180	C	T	2967	AGC-S	AGT-S																									x	
DCL4-8	6750	G	A	3192	TCG-S	TCA-S					x																					
DCL4-8	6786	G	A	3228	GAG-E	GAA-E					x										x		x	x								x
DCL4-8	6795	C	T	3237	CTC-L	CTT-L								x																		
DCL4-8	6814	C	A	3256	CAT-H	AAT-N																									x	
DCL4-8	6830	C	T	3272	TCG-S	TTG-L		x					x				x															
DCL4-8	6864	G	A	3306	AGG-R	AGA-R		x			x		x	x		x	x		x		x		x	x			x	x	x			x
DCL4-8	7057	A	T	3407	CAC-H	CTC-L									x																	
DCL4-9	7291	T	C					x					x	x			x															
DCL4-9	7347	G	A									x																				
DCL4-9	7365	A	G					x					x				x															
DCL4-9	7497	G	A	3602	AGT-S	AAT-N		x					x				x															
DCL4-9	7716	A	G	3821	CAA-Q	CGA-R								x																		
DCL4-9	7738	G	A	3843	AGG-R	AGA-R																									x	

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
DCL4-9	7753	C	T	3858	CTC-L	CTT-L																		x								
DCL4-9/10	7955	G	A																		x			x								x
DCL4-10	8019	C	T	4038	TTC-F	TTT-F								x																		
DCL4-10	8098	G	A	4117	GTT-V	ATT-I															x			x								x
DCL4-10	8244	C	T																	x												
DCL4-10	8313	A	C																		x			x								x
DCL4-10	8375-8384	*GAT...AGA*	*10 bp Δ*					x					x	x			x															
DCL4-10	8388	G	A					x					x	x			x															
DCL4-10	8498	C	A					x					x	x			x															
DCL4-10	8501	A	T											x																		
DCL4-10	8647	A	G	4395	AAA-K	AAG-K		x	x	x	x	x	x	x	x	x	x		x	x	x		x	x	x	x	x	x	x	x	x	x
DCL4-11	8896	A	G					x					x				x															
DCL4-11	8943	G	C											x																		
DCL4-11	9021	G	T	4633	GGC-G	TGC-C					x					x					x		x	x				x	x			x
DCL4-11	9103	A	C	4715	AAC-N	ACC-T		x					x				x										x					
DCL4-11	9125	T	C	4737	TTT-F	TTC-F										x												x				
DCL4-11	9274	A	G																x													
DCL4-11	9392	G	A	4921	GAA-E	AAA-K																x										
DCL4-11	9454	G	A						x	x															x	x				x		
DCL4-11	9461	C	A																				x									
DCL4-11	9620	C	T	5033	GCA-A	GTA-V					x					x					x		x	x				x	x			x
ERI-1	139	T	C	75	TAT-Y	TAC-Y				x										x						x			x			x
ERI-1	143	G	A	79	GCC-A	ACC-T			x																							
ERI-1	242-256	*TCT...TCT*	*15 bp Δ*	178-192	5x TCT-SSSSS			x				x											x									
ERI-1	245-256	*TCT...TCT*	*12 bp Δ*	181-192	4x TCT-SSSS									x																		
ERI-1	251-256	*TCTTCT*	*6 bp Δ*	187-192	TCTTCT-SS				x				x				x		x					x	x						x	
ERI-1	254-256	*TCT*	*3 bp Δ*	190-192	TCT-S											x						x										
ERI-1	256//257		*TCT*	192//193		TCT-S									x													x				
ERI-1	256//257		*3x TCT*	192//193		3x TCT-SSS															x											
ERI-1	256//257		*4x TCT*	192//193		4x TCT-SSSS				x										x						x			x			
ERI-1	256//257		*5x TCT*	192//193		5x TCT-																										x

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)- amino acid(s)	Polymorphic codon(s)- amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
						SSSSS																										
ERI-1	256//257		*6x TCT*	192//193		6x TCT-SSSSSS												x														
ERI-1	363	T	A					x		x		x								x			x	x		x			x			x
ERI-1	398//399		*A*					x	x	x		x	x	x		x	x		x	x		x	x	x	x	x			x		x	x
ERI-1	477	A	T					x	x	x		x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x		x	x
ERI-2	741	T	A	381	AAT-N	AAA-K										x																
ERI-2	887	A	C						x	x			x	x		x	x		x	x		x		x	x	x			x		x	x
ERI-2	1019	C	T					x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x
ERI-2	1084	A	C											x																		
ERI-2	1134	C	T	507	AGC-S	AGT-S								x																		
ERI-2	1206	C	G						x										x			x		x	x						x	
ERI-2/3	1329	T	G							x										x						x			x			x
ERI-3	1539	T	G	658	TGG-W	GGG-G		x	x	x		x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x		x	x
ERI-3	1550	C	A	669	AAC-N	AAA-K			x										x			x		x	x						x	
ERI-3	1710	A	T	732	GTA-V	GTT-V			x																							
ERI-3	1835-1838	*TATT*	*4 bp Δ*						x																							
ERI-3	1841	A	G					x				x											x									
HEN1-1a	232	T	C	116	CTA-L	CCA-P		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
HEN1-1a	245	T	A	129	ATT-I	ATA-I																							x			
HEN1-1a	254	A	T	138	AAA-K	AAT-N							x							x	x									x	x	
HEN1-1a	290	T	A	174	CCT-P	CCA-P																							x			
HEN1-1a	293	A	C	177	GAA-E	GAC-D																							x			
HEN1-1a	387	*T*	*1 bp Δ*																										x			
HEN1-1a	402-403	*AT*	*2 bp Δ*									x										x			x							
HEN1-1a	403//404		*AT*													x		x														
HEN1-1a	403//404		*2x AT*														x											x				
HEN1-1a	403//404		*3x AT*																				x									
HEN1-1a	403//404		*4x AT*						x																							
HEN1-1a	403//404		*5x AT*											x	x																	
HEN1-1a	403//404		*6x AT*							x										x												
HEN1-1a	403//404		*7x AT*																								x				x	
HEN1-1a	403//404		*10x AT*																							x						x
HEN1-1a	403//404		*14x AT*																		x											

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
HEN1-1a	403//404		*>13x AT*																x													
HEN1-1a	403//404		*>15 AT*																											x		
HEN1-1a	403//404		*>22x AT*																										x			
HEN1-1a	403//404		*>25x AT*										x																			
HEN1-1a	419	C	A																										x			
HEN1-1a	438	G	A																										x			
HEN1-1a	439	C	A																										x			
HEN1-1a	447	T	A					x																								
HEN1-1a	452	G	A																												x	
HEN1-1a	758	T	G	399	GAT-D	GAG-E			x							x								x				x				
HEN1-1a	782	C	T	423	CCC-P	CCT-P															x								x			
HEN1-1a/2	903	G	T	544	GCA-A	TCA-S															x								x			
HEN1-1a/2	942	A	G	583	AAA-K	GAA-E															x								x			
HEN1-1a/2	999	C	G	640	CTA-L	GTA-V															x								x			
HEN1-1a/2	1019	G	A	660	GCG-A	GCA-A													x				x			x						x
HEN1-1a/2	1036	G	A	677	TGT-C	TAT-Y						x																				
HEN1-1a/2	1039	T	C	680	ATC-I	ACC-T													x							x						x
HEN1-2	1168	T	A																		x								x			
HEN1-2	1184	C	G																		x								x			
HEN1-2	1330	C	G	887	ACT-T	AGT-S			x											x	x								x	x		
HEN1-2	1372	C	T	929	GCA-A	GTA-V			x		x		x							x										x	x	
HEN1-2	1410	G	A	967	GAT-D	AAT-N													x		x					x			x			x
HEN1-2	1475	C	T	1032	ACC-T	ACT-T		x																								
HEN1-2	1509//1510		*AATTTC*																		x								x			
HEN1-2	1554	A	G							x					x																	
HEN1-2	1568	A	G																		x								x			
HEN1-2	1601	C	G															x														
HEN1-2	1637	G	A							x					x																	
HEN1-2	1671	C	T																		x								x			
HEN1-2	1677	C	T													x												x				
HEN1-3	1790	A	G	1135	ACA-T	GCA-A		x																								
HEN1-3	1886	G	C	1231	GCA-A	CCA-P											x															
HEN1-3	1909	C	T	1254	GAC-D	GAT-D															x								x			
HEN1-3	1982	G	A	1327	GAA-E	AAA-K						x																				

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
HEN1-3	1985	G	A	1330	GAT-D	AAT-N						x																				
HEN1-3	2003	C	A	1348	CAT-H	AAT-N																					x					
HEN1-3	2116	A	G	1461	TTA-L	TTG-L															x								x			
HEN1-3	2119	G	A	1464	AAG-K	AAA-K															x								x			
HEN1-3	2287	G	A	1632	CTG-L	CTA-L															x								x			
HEN1-4	2947	G	A											x																		
HEN1-4	2952	T	C					x																								
HEN1-4	3008	T	C																										x			
HEN1-4	3116	A	T																		x								x			
HEN1-4	3122	G	A										x																			
HEN1-4	3226	G	A	2070	GCG-A	GCA-A															x								x			
HEN1-4	3256	A	G	2100	CAA-Q	CAG-Q															x								x			
HEN1-4	3315	C	T																		x								x			
HEN1-4	3344	A	G																		x								x			
HEN1-4	3384	C	T								x																					
HEN1-4	3416	G	A								x																					
HEN1-5	3634	C	G							x					x														x			
HEN1-5	3778	C	A	2284	CTA-L	ATA-I													x		x					x						x
HEN1-5	3925	C	T								x																					
HEN1-5	3930	A	T					x																								
HEN1-5	3931	T	A					x																								
HEN1-5	3932	C	T					x																								
HEN1-5	3957	G	A											x																		
HEN1-5	3969//3970		A													x								x				x				
HEN1-5	4241	A	C	2642	AAC-N	ACC-T													x		x					x						x
HEN1-5	4314	T	A	2715	ATT-I	ATA-I													x		x		x			x						x
NRPD1-1	1416	T	C					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-1	1453	T	C	15	TGT-C	TGC-C		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-1	1538	A	G					x							x		x				x					x						
NRPD1-1	1598	T	C					x		x		x	x		x	x	x				x	x				x		x				
NRPD1-1	1749	G	T					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-1	1831	C	T	213	TTC-F	TTT-F					x												x									
NRPD1-1	1952	T	G					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-1/2	2143	A	G					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
NRPD1-1/2	2153-2154*	TG*	CC*	368-369*	TTG-L*	TCC-S*		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-1/2	2160	C	T	375	ACC-T	ACT-T		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2221	G	A	436	GTT-V	ATT-I		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2256	C	A	471	CTC-L	CTA-L		x		x		x	x		x	x	x			x	x	x				x	x	x	x			x
NRPD1-2	2307	G	A	522	CCG-P	CCA-P		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2417	T	C					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2430	*A*	*1 bp Δ*											x																		
NRPD1-2	2430	A	T					x				x	x		x		x				x	x				x						
NRPD1-2	2442//2443		*T*					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2517	G	T	656	GGT-G	GTT-V								x																		
NRPD1-2	2590	G	A	729	CGG-R	CGA-R		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2632	T	G																					x								
NRPD1-2	2649	T	C															x													x	
NRPD1-2	2681//2682		*A*					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2789	T	C					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2790	*T*	*1 bp Δ*					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2794	G	A					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2798	C	T					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2827//2828		*ATT*					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2845	T	A					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-2	2853	T	G						x		x							x	x				x							x	x	
NRPD1-2	2890	T	C	861	AGT-S	AGC-S								x																		
NRPD1-2	2936	C	T	907	CTT-L	TTT-F			x										x											x		
NRPD1-3	3053	A	G	1024	AGT-S	GGT-G		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-3	3178	G	A	1149	TTG-L	TTA-L			x									x	x											x	x	
NRPD1-3	3309	T	C	1280	ATC-I	ACC-T									x																	
NRPD1-3	3436	G	T	1407	GAG-E	GAT-D		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-3	3467	A	C	1438	AGA-R	CGA-R		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-3	3783	A	G	1754	GAA-E	GGA-G								x																		
NRPD1-4	4121	G	A	2092	GAA-E	AAA-K		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-4	4283	T	A	2254	TCA-S	ACA-T							x																			
NRPD1-4	4288	A	T	2259	GCA-A	GCT-A																	x									
NRPD1-4	4371	G	A	2342	GGA-G	GAA-E						x																				
NRPD1-4	4434	C	T	2405	ACG-T	ATG-M														x												

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
NRPD1-5a	4899	G	A	2775	ATG-M	ATA-I																							x			
NRPD1-5a	4918-4919*	TA*	AC*	2794-2795*	TAC-Y*	ACC-T*		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-5a	5016	G	T	2892	ATG-M	ATT-I			x										x											x		
NRPD1-5a	5123	C	T	2916	AAC-N	AAT-N			x										x											x		
NRPD1-5a	5198//5199		*A*								x																					
NRPD1-5a	5199	C	T																								x		x			
NRPD1-5a	5212	A	C																						x							
NRPD1-5a	5214-5215	*AT*	*2 bp Δ*					x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x
NRPD1-5a	5268	G	C	2978	CGG-R	CCG-P															x					x						
NRPD1-5a	5299	A	T	3009	TCA-S	TCT-S		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-5a	5314	G	A	3024	CAG-Q	CAA-Q		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-5a	5351	G	A	3061	GTT-V	ATT-I																		x								
NRPD1-5a	5356	C	T	3066	GAC-D	GAT-D		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-5a	5373	C	T	3083	ACA-T	ATA-I																			x							
NRPD1-5a	5548	C	T	3163	CAC-H	TAC-Y																					x		x			
NRPD1-5a	5558	T	G	3173	TTG-L	TGG-W												x													x	
NRPD1-6	5623	A	G					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-6	5647	A	G					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-6	5655	C	T					x		x		x	x		x	x	x			x	x	x				x	x	x	x			x
NRPD1-6	5675	A	G					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-6	5920	G	A	3443	CGA-R	CAA-Q																										x
NRPD1-6	5992//5993		*AAATTTCTT*						x										x											x		
NRPD1-6	5992//5993		*CAATTTCTT*					x		x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x		x	x
NRPD1-6	6268	T	G					x		x		x	x		x	x	x			x	x	x				x	x	x	x			x
NRPD1-6	6383	T	A	3741	GCT-A	GCA-A		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-6/7	6453	A	T												x																	
NRPD1-7	6608	G	A	3885	GCG-A	GCA-A					x																					
NRPD1-7	6620	C	A	3897	GAC-D	GAA-E																		x								
NRPD1-7	6641	T	C	3918	TTT-F	TTC-F		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NRPD1-7	6647	G	A	3924	TTG-L	TTA-L		x		x		x	x		x	x	x			x	x	x					x	x	x			x
NRPD1-7	6678	G	A	3955	GTA-V	ATA-I		x				x	x		x		x				x	x										
NRPD1-7	6795	C	G													x												x				
NRPD1-7	6979	T	C						x										x											x		
NRPE1-1	569	G	C													x												x				

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
NRPE1-1	572	A	G											x		x												x				
NRPE1-1	581	C	T											x		x												x				
NRPE1-1	628	C	A	33	GAC-D	GAA-E		x	x	x	x	x	x	x	x	x		x		x		x	x	x	x		x	x			x	
NRPE1-1	703	G	A								x																					
NRPE1-1	705	A	G					x	x	x	x	x	x	x	x	x		x		x		x	x	x	x		x	x			x	
NRPE1-1	794	T	A											x		x												x				
NRPE1-1	798	T	A											x																		
NRPE1-1	894	C	G													x																
NRPE1-1	914	C	T											x		x												x				
NRPE1-1	1052	A	G	201	AAA-K	AAG-K								x		x												x				
NRPE1-1	1076	G	T											x		x												x				
NRPE1-1	1101	A	G											x		x												x				
NRPE1-1	1129	G	A					x	x	x	x	x	x	x	x	x		x		x		x	x	x	x		x	x			x	
NRPE1-1	1252	A	G	310	ATA-I	GTA-V																						x				
NRPE1-1/2	1289	T	G										x																			
NRPE1-1/2	1294	T	A													x																
NRPE1-1/2	1295	C	A													x																
NRPE1-1/2	1296	C	T													x																
NRPE1-1/2	1323	A	G																									x				
NRPE1-1/2	1336	T	C											x		x												x				
NRPE1-2	1392	G	A	354	TTG-L	TTA-L								x		x												x				
NRPE1-2	1422	*C*	*1 bp Δ*											x		x												x				
NRPE1-2	1487	T	G											x		x												x				
NRPE1-2	1546	A	C											x																		
NRPE1-2	1625	T	C											x		x												x				
NRPE1-2	1693	G	A											x		x												x				
NRPE1-2	1719	T	C																									x				
NRPE1-2	1884	C	T	513	TAC-Y	TAT-Y																						x				
NRPE1-2	1930	T	C											x		x												x				
NRPE1-2	1998	T	G											x		x												x				
NRPE1-2	1999	T	A											x		x												x				
NRPE1-2	2000	T	G											x		x												x				
NRPE1-2	2007	T	A											x		x												x				
NRPE1-3	2185	C	T	648	AAC-N	AAT-N								x		x												x				

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
NRPE1-3	2222	A	C	685	ACC-T	CCC-P																						x				
NRPE1-3	2240	T	G											x		x												x				
NRPE1-3	2345	G	A	724	GTT-V	ATT-I										x																
NRPE1-3	2404	G	C	783	TCG-S	TCC-S								x		x												x				
NRPE1-3	2421	G	T	800	AGT-S	ATT-I								x																		
NRPE1-3	2647	A	C	1026	CCA-P	CCC-P								x																		
NRPE1-3	2649	T	G	1028	ATA-I	AGA-R								x		x												x				
NRPE1-3	2704	T	C	1083	GGT-G	GGC-G								x		x												x				
NRPE1-3	2797	T	G	1176	ACT-T	ACG-T																						x				
NRPE1-3	2866	C	G	1245	CCC-P	CCG-P																						x				
NRPE1-4	3226	T	C	1605	TCT-S	TCC-S								x																		
NRPE1-4	3511	C	T	1890	GAC-D	GAT-D										x																
NRPE1-4	3577	A	T	1956	TCA-S	TCT-S		x	x	x		x	x		x					x		x	x		x		x				x	
NRPE1-5	4082	C	T	2461	CTG-L	TTG-L																						x				
NRPE1-5	4093	T	G	2472	GCT-A	GCG-A								x		x												x				
NRPE1-5	4255	G	A											x		x												x				
NRPE1-5	4281	*T*	*1 bp Δ*											x		x												x				
NRPE1-5	4295	G	A	2586	CTG-L	CTA-L								x		x												x				
NRPE1-5	4352	A	T	2643	TTA-L	TTT-F								x		x												x				
NRPE1-5	4354	A	G	2645	AAC-N	AGC-S													x													
NRPE1-5	4355	C	T	2646	AAC-N	AAT-N												x						x								
NRPE1-5	4445	A	T	2736	GCA-A	GCT-A								x		x												x				
NRPE1-5	4466-4472	*AGAATTG*	*7 bp Δ*											x		x												x				
NRPE1-5	4523	*T*	*1 bp Δ*					x		x	x			x	x	x		x						x				x				
NRPE1-5	4543	T	C											x		x												x				
NRPE1-5	4562	T	G											x		x												x				
NRPE1-5	4572	A	C											x		x												x				
NRPE1-5	4578	T	C	2757	TAT-Y	TAC-Y								x		x												x				
NRPE1-5	4589	C	A	2768	CCA-P	CAA-Q								x		x												x				
NRPE1-5	4612-4614	*ATT*	*3 bp Δ*	2791-2793	ATT-I									x		x												x				
NRPE1-6	4984	C	A													x																
NRPE1-6	5034	T	C											x		x												x				
NRPE1-6	5047	T	C					x	x	x		x	x		x					x		x	x		x		x				x	

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
NRPE1-6	5146	C	T	3020	ACG-T	ATG-M								x		x												x				
NRPE1-6	5165	C	T	3039	ACC-T	ACT-T								x																		
NRPE1-6	5216	A	G	3090	CCA-P	CCG-P								x		x												x				
NRPE1-6	5231	A	G	3105	ATA-I	ATG-M																			x							
NRPE1-6	5255	A	G																									x				
NRPE1-6	5316	A	T											x		x												x				
NRPE1-6	5326	C	A											x		x												x				
NRPE1-6	5348	A	T																									x				
NRPE1-6	5368	A	G											x		x												x				
NRPE1-6	5373	A	G											x		x												x				
NRPE1-6	5455	T	A											x		x												x				
NRPE1-6	5474	*A*	*1 bp Δ*											x		x												x				
NRPE1-6	5498	T	G																									x				
NRPE1-6	5507	C	A											x		x												x				
NRPE1-6	5530	A	G													x																
NRPE1-6	5614	G	A	3185	CGT-R	CAT-H										x		x						x								
NRPE1-6	5625	C	A	3196	CGC-R	AGC-S																	x									
NRPE1-7	5629	G	A	3200	CGA-R	CAA-Q								x		x												x				
NRPE1-7	5718	T	C	3289	TCA-S	CCA-P		x	x	x	x	x	x	x	x	x		x		x		x	x	x	x		x	x			x	
NRPE1-7	5723	G	A	3294	GTG-V	GTA-V										x																
NRPE1-7	5753	C	T	3324	ATC-I	ATT-I																			x							
NRPE1-7	5822	A	G											x		x												x				
NRPE1-7	5848	T	G					x		x	x	x	x	x	x	x		x		x		x	x	x	x		x	x			x	
NRPE1-7	5848	T	A						x																							
NRPE1-7	5861-5862	*AC*	*2 bp Δ*																									x				
NRPE1-7	5879	C	T											x		x												x				
NRPE1-7	5917	T	C											x		x												x				
NRPE1-7	6020	C	T	3450	CTC-L	CTT-L								x		x												x				
NRPE1-7	6129	C	T	3559	CTG-L	TTG-L												x						x								
NRPE1-7	6183	T	A											x		x												x				
NRPE1-7	6222	C	T											x		x												x				
NRPE1-8	6433	A	T																									x				
NRPE1-8	6666	A	G	3906	TTA-L	TTG-L								x		x												x				
NRPE1-8	6675	A	G	3915	GAA-E	GAG-E																							x			

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
NRPE1-8	6707	G	A	3947	AGC-S	AAC-N			x																							
NRPE1-8	6809	A	C	4049	AAT-N	ACT-T								x		x												x				
NRPE1-8	6823	G	A	4063	GAT-D	AAT-N																						x				
NRPE1-8	6945	A	T	4185	ACA-T	ACT-T								x		x												x				
NRPE1-8	6951	A	G	4191	CCA-P	CCG-P								x		x												x				
NRPE1-8	6992	C	G	4232	ACT-T	AGT-S								x		x												x				
NRPE1-8	7032	G	T	4272	TCG-S	TCT-S								x		x												x				
NRPE1-8	7078	G	T	4318	GGT-G	TGT-C								x		x												x				
NRPE1-9	7317	A	G	4557	ATA-I	ATG-M		x	x	x	x	x	x	x	x	x				x		x	x		x		x	x			x	
NRPE1-9	7356	A	C	4596	GGA-G	GGC-G								x		x												x				
NRPE1-9	7394-7402	*GGGGTGCTT*	*9 bp Δ*	4634-4642	TGG...TGG -WGAW	TGG-W								x		x												x				
NRPE1-9	7407	C	T	4647	GAC-D	GAT-D								x		x												x				
NRPE1-9	7424	C	G	4664	ACT-T	AGT-S										x												x				
NRPE1-9	7487	C	T	4727	CCT-P	CTT-L			x																							
NRPE1-9	7527	G	T	4767	ACG-T	ACT-T								x		x												x				
NRPE1-9	7565	C	G	4805	GCT-A	GGT-G								x		x												x				
NRPE1-9	7569	T	G	4809	GCT-A	GCG-A								x		x												x				
NRPE1-9	7590	C	T	4830	AAC-N	AAT-N								x		x												x				
NRPE1-9	7725	G	A	4965	GAG-E	GAA-E								x		x												x				
NRPE1-9	7847	G	A	5087	GGT-G	GAT-D										x																
NRPE1-9	7872	T	G	5112	AAT-N	AAG-K								x		x												x				
NRPE1-9	8062	T	C											x		x												x				
NRPE1-9	8067	C	A															x						x								
NRPE1-9	8101	T	C											x		x												x				
NRPE1-10	8137	T	G															x						x								
NRPE1-10	8146	C	G											x		x												x				
NRPE1-10	8154	G	A											x		x												x				
NRPE1-10	8215	G	A	5328	AAG-K	AAA-K								x																		
NRPE1-10	8251	T	C	5364	CTT-L	CTC-L		x	x	x	x	x	x	x	x	x		x		x		x	x	x	x		x	x			x	
NRPE1-10	8260	C	T	5373	GGC-G	GGT-G								x																		
NRPE1-10	8283	A	G											x		x												x				
NRPE1-10	8289	*C*	*1 bp Δ*											x		x												x				
NRPE1-10	8364	G	A											x														x				

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
NRPE1-10	8371	G	T	5389	GTT-V	TTT-F																						x				
NRPE1-10	8483	A	G	5501	AAA-K	AGA-R										x																
NRPE1-10	8495	G	A	5513	CGT-R	CAT-H			x													x			x							
NRPE1-10	8561	A	G	5579	AAC-N	AGC-S										x																
NRPE1-10	8643	A	T	5661	ACA-T	ACT-T										x																
NRPE1-10	8649	T	A	5667	ACT-T	ACA-T								x														x				
NRPE1-10	8653	T	A	5671	TCT-S	ACT-T								x														x				
NRPE1-10	8663-8668	*CATCTC*	*6 bp Δ*	5681-5686	CCATCTCAG -PSQ	CAG-Q								x														x				
NRPE1-10	8666	C	T	5684	TCT-S	TTT-F										x																
NRPE1-10	8694//8695		*ACTCAAGCTCAG*	5712//5713		ACTCAAGCTCA G-TQAQ								x														x				
NRPE1-10	8696	C	T	5714	GCT-A	GTT-V													x													
NRPE1-10	8695-8808	*GCT...CAG*	*114 bp Δ*	5713-5826	GCT...CAG -A...Q											x																
NRPE1-10	8727	T	G	5745	TCT-S	TCG-S		x	x	x	x	x	x	x	x					x		x	x		x		x	x			x	
NRPE1-10	8733	T	G	5751	TCT-S	TCG-S								x																		
NRPE1-10	8739	T	G	5757	TCT-S	TCG-S								x																		
NRPE1-10	8745	T	G	5763	TCT-S	TCG-S								x																		
NRPE1-10	8758	C	T	5776	CAG-Q	TAG-Stop								x																		
NRPE1-10	8752-8823	*CAG...TCT*	*72 bp Δ*	5770-5841	12x CAGTCT-12x QS																							x				
NRPE1-10	8788-8823	*CAG...TCT*	*36 bp Δ*	5806-5841	6x CAGTCT-6x QS			x		x					x																	
NRPE1-10	8794-8823	*CAG...TCT*	*30 bp Δ*	5812-5841	5x CAGTCT-5x QS						x																					
NRPE1-10	8800-8823	*CAG...TCT*	*24 bp Δ*	5818-5841	4x CAGTCT-4x QS				x									x	x													
NRPE1-10	8806-8823	*CAG...TCT*	*18 bp Δ*	5824-5841	3x CAGTCT-3x QS																			x	x					x		
NRPE1-10	8812-8823	*CAG...TCT*	*12 bp Δ*	5830-5841	2x CAGTCT-2x QS									x							x								x			
NRPE1-10	8818-8823	*CAGTCT*	*6 bp Δ*	5836-5841	CAGTCT-QS												x															x
NRPE1-10	8823//8824		*CAGTCT*	5841//5842		CAGTCT-QS														x			x				x				x	
NRPE1-10	8823//8824		*CAGTCTCAGTCT*	5841//5842		CAGTCTCAGTC T-QSQS						x	x									x										
NRPE1-10	8836	C	G	5854	CAG-Q	GAG-E								x														x				

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
NRPE1-10	8869-8874	*GCTCAG*	*6 bp Δ*	5887-5892	GCTCAG-AQ									x														x				
NRPE1-10	8874//8875		*GCTCAG*	5892//5893		GCTCAG-AQ		x	x	x	x	x	x		x	x		x		x		x	x	x	x		x				x	
NRPE1-10	8898	C	T	5916	TCC-S	TCT-S						x																				
NRPE1-10	8905	C	T	5923	CAG-Q	TAG-Stop								x														x				
NRPE1-10	8927	A	G											x														x				
SDE3-1	543	T	G	174	TCT-S	TCG-S			x	x						x							x			x	x	x				x
SDE3-1	549	T	C	180	ACT-T	ACC-T			x	x						x							x			x	x	x				x
SDE3-1	573	C	T	204	TCC-S	TCT-S			x	x						x							x			x	x	x				x
SDE3-1	782	A	T	413	AAG-K	ATG-M							x											x								
SDE3-1	791	A	G	422	GAC-D	GGC-G										x																
SDE3-1/2	1171	A	G											x								x			x							
SDE3-1/2	1202//1203		*TTTTA*											x								x			x							
SDE3-1/2	1203	C	G									x	x											x								
SDE3-1/2	1210	T	G											x								x			x							
SDE3-2	1366	T	A	909	GTT-V	GTA-V								x								x			x							
SDE3-2	1427	C	T											x								x			x							
SDE3-2	1464	A	T											x								x			x							
SDE3-2	1543	C	T	971	GCG-A	GTG-V													x										x			
SDE3-2	1557	A	T	985	ATG-M	TTG-L			x	x		x	x	x	x	x	x			x		x	x	x	x	x	x	x			x	x
SDE3-2	1569	T	C	997	TCT-S	CCT-P								x								x			x							
SDE3-2	1586	C	G	1014	CGC-R	CGG-R							x											x								
SDE3-2	1687	A	G	1115	AAC-N	AGC-S								x								x			x							
SDE3-2	1787	T	A	1215	ATT-I	ATA-I							x											x								
SDE3-2	1805	C	T	1233	TGC-C	TGT-C								x								x			x							
SDE3-2	1871	G	A	1299	GAG-E	GAA-E								x								x			x							
SDE3-3	1971-1973*	TTA*	CTG*	1399-1401*	TTA-L*	CTG-L*								x								x			x							
SDE3-3	2003	T	A	1431	ATT-I	ATA-I									x		x														x	
SDE3-3	2162	C	T	1590	AAC-N	AAT-N								x								x			x							
SDE3-3	2177	C	G	1605	AAC-N	AAG-K						x																				
SDE3-3	2199	C	T	1627	CTC-L	TTC-F								x								x			x							
SDE3-3	2246	T	C	1674	GCT-A	GCC-A								x								x			x							
SDE3-3	2510	C	G	1938	GAC-D	GAG-E						x																				
SDE3-3	2582	A	G	2010	CAA-Q	CAG-Q								x								x			x							
SDE3-4	2891	G	A	2319	CTG-L	CTA-L			x	x		x	x		x	x	x			x			x	x		x	x	x			x	x

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
SDE3-4	3001	T	C						x	x		x	x		x	x	x			x			x	x		x	x	x			x	x
SDE3-4	3132	C	G	2470	CCA-P	GCA-A																						x				
SDE3-4	3225-3226*	GG*	AA*	2563-2564*	GGA-G*	AAA-K*																					x	x				
SDE3-4	3323-3373	*ATG...TGG*	*51 bp Δ*	2661-2711	GGA...GGG -G...G	GGG-G				x																						x
SDE3-4	3347	A	T	2685	ACA-T	ACT-T														x												
SDE3-4	3373	*G*	*1 bp Δ*	2711	GGG-G	GGT-G&FS																						x				
SDE3-4	3449	C	T	2787	GCC-A	GCT-A						x	x											x								
SDE3-4	3485	C	T	2823	GGC-G	GGT-G											x														x	
SDE5-1	870	G	A	64	GAT-D	AAT-N																					x				x	
SDE5-1	891	T	C	85	TCT-S	CCT-P		x																		x						
SDE5-1	1242-1243*	GA*	TT*	436-437*	GAA-E*	TTA-L*			x	x		x	x		x	x	x			x		x	x		x				x			x
SDE5-1	1461	A	G																x													
SDE5-2	1552	C	T					x																		x						
SDE5-2	1616	T	C													x																
SDE5-2	1740	*T*	*1 bp Δ*													x																
SDE5-2	1842	*C*	*1 bp Δ*						x	x		x	x		x	x	x			x		x	x		x				x			x
SDE5-3a	2273	T	A														x															
SDE5-3a	2369	A	G	782	GAA-E	GGA-G																								x		
SDE5-3a	2406	A	T	819	GAA-E	GAT-D											x															
SDE5-3a	2418	C	A	831	GAC-D	GAA-E		x																		x						
SDE5-3a	2635	C	T																											x		
SDE5-4	3082-3084	*CTC*	*3 bp Δ*																x													
SDE5-4	3130	T	A					x	x	x		x	x	x	x	x	x	x	x	x		x	x		x	x	x	x	x	x	x	x
SDE5-4	3223	A	G	1239	AAA-K	AAG-K			x													x										
SDE5-4	3296	T	C					x	x	x		x	x	x	x	x	x		x	x		x	x		x	x			x	x		x
SDE5-4	3433	C	G					x	x	x		x	x	x	x	x	x		x	x		x	x		x	x			x			x
SGS3-0	355	A	C					x	x	x	x						x		x			x		x	x				x	x		
SGS3-0	550	A	C	53	TAT-Y	TCT-S		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x
SGS3-0	626	G	A	129	GAG-E	GAA-E															x											
SGS3-0	801	G	A	304	GGC-G	AGC-S			x																							
SGS3-0/1a	929	T	C	432	GGT-G	GGC-G								x				x								x	x					
SGS3-0/1a	1025	T	A	528	GAT-D	GAA-E												x								x	x					
SGS3-1a	1145	C	T	648	ATC-I	ATT-I				x																						

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
SGS3-1a/2	1514	C	T									x								x	x										x	
SGS3-1a/2	1539	C	T	969	CTC-L	CTT-L			x								x		x													
SGS3-1a/2	1642	A	T	1072	ATG-M	TTG-L			x								x		x													
SGS3-2	1791	G	A	1221	AAG-K	AAA-K					x													x								
SGS3-3	1986	A	G	1305	GTA-V	GTG-V						x		x				x		x	x					x	x				x	
SGS3-3	2235	T	C									x								x											x	
SGS3-3	2324	A	T	1541	CAG-Q	CTG-L						x								x											x	
SGS3-3	2397	T	C	1614	GTT-V	GTC-V									x																	
WEX1	128-129*	CG*	ATT*	88-89*	CGT-R*	ATT-I*			x	x													x		x				x			
WEX1	129	G	T	89	CGT-R	CTT-L		x			x	x	x		x	x	x	x	x	x	x	x		x		x	x	x		x	x	x
WEX1	145-147	*TTC*	*3 bp Δ*	105-107	TCTTCC-SS	TCC-S			x	x																						
WEX1	158	C	T	118	CCG-P	TCG-S			x	x																						
WEX1	162	C	G	122	ACC-T	AGC-S										x								x				x				
WEX1	184	C	A	144	GTC-V	GTA-V			x	x													x		x				x			x
WEX1	217	C	G	177	CCC-P	CCG-P																					x					
WEX1	238	G	A	198	TTG-L	TTA-L										x				x	x	x		x	x			x		x		
WEX1	244	T	C	204	CGT-R	CGC-R			x	x																						
WEX1	244	T	A	204	CGT-R	CGA-R																	x						x			x
WEX1	257	T	C	217	TCT-S	CCT-P			x	x						x		x		x	x	x		x	x			x		x		
WEX1	267	A	C	227	TAT-Y	TCT-S			x	x						x		x		x	x	x		x	x			x		x		
WEX1	278	C	T	238	CCT-P	TCT-S										x					x			x				x				
WEX1	286	C	T	246	TCC-S	TCT-S			x	x																						
WEX1	300	T	A						x	x																						
WEX1	307	T	A						x	x																						
WEX1	317	C	G																				x						x			
WEX1	320	G	A								x					x		x		x	x	x		x	x			x		x		x
WEX1	325//326		*G*						x	x																						
WEX1	329	C	A						x	x																						
WEX1	331	G	C						x	x	x					x	x	x		x	x	x	x	x	x			x	x	x		x
WEX1	338	T	A						x	x																						
WEX1	373	C	T						x	x																						
WEX1	378	C	G						x	x																						
WEX1	420	C	T																		x											
WEX1	454	G	A					x	x	x	x	x	x		x	x	x	x	x	x	x	x		x	x	x	x	x		x	x	x

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
WEX1	473	*A*	*1 bp Δ*						x	x																						
WEX1	478	G	T								x			x		x	x	x		x	x	x		x	x			x		x		x
WEX1	483	G	A						x	x	x			x		x	x	x		x	x	x		x	x			x		x		x
WEX1	500	G	A								x			x		x	x	x		x	x	x		x	x			x		x		x
WEX1	503	G	A						x	x																						
WEX1	522	A	T							x																						
WEX1	524	T	A								x			x		x	x	x		x	x	x		x	x			x		x		x
WEX1	620	A	C	343	ATT-I	CTT-L			x	x							x															
WEX1	625	A	G	348	AAA-K	AAG-K											x															
WEX1	657	T	G	380	ATA-I	AGA-R			x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX1	673	G	T	396	TTG-L	TTT-F			x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX1	707	G	A					x				x	x		x				x							x	x				x	
WEX1	712	T	C					x				x	x		x				x							x	x				x	
WEX1	718-723	*CCCTTC*	*6 bp Δ*					x				x	x		x				x							x	x				x	
WEX1	718-861	*CCC...TAT*	*144 bp Δ*						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX1	730	C	T														x															
WEX1	731	C	A														x															
WEX1	739-740	*GT*	*2 bp Δ*																												x	
WEX1	744	A	T														x															
WEX1/2b	827//828		*G*														x															
WEX1/2b	838	C	T									x																				
WEX1/2b	860	A	T														x															
WEX1/2b	863	T	C						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX1/2b	865	G	A						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX1/2b	872	G	T						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX1/2b	898	G	A						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX1/2b	899	A	G						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX1/2b	903	G	A						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX1/2b	907	T	G						x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX1/2b	910	A	T						x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX1/2b	911	T	A						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX1/2b	915	C	T						x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX1/2b	919	G	T														x															

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
WEX1/2b	927	*T*	*1 bp Δ*														x															
WEX2b	938	G	C													x				x	x	x		x				x				
WEX2b	941	G	T														x															
WEX2b	945	*A*	*1 bp Δ*						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX2b	957//958		*C*						x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX2b	959	C	G														x															
WEX2b	972	G	T														x															
WEX2b	981	A	G						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX2b	991	T	C						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX2b	1003//1004		*TCAGATGCTACTTT*														x															
WEX2b	1003//1004		*TCATATGCTACATT*											x											x							
WEX2b	1003//1004		*TCATATGCTACTTT*						x	x	x					x		x		x	x	x	x	x				x	x	x		x
WEX2b	1012//1013		*CTAATA*														x															
WEX2b	1020	C	A														x															
WEX2b	1028	C	T						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX2b	1030	A	T						x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX2b	1041	A	C														x															
WEX2b	1041-1042	*AT*	*2 bp Δ*																				x						x			
WEX2b	1054	A	G						x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX2b	1057	G	T						x	x																						
WEX2b	1081	C	G																												x	
WEX2b	1081	C	T						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX2b	1085	C	T																				x						x			
WEX2b	1096	G	T																								x					
WEX2b	1097	T	C						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX2b	1106	T	C						x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX2b	1110	A	C						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX2b	1127	T	A														x															
WEX2b	1131	T	G											x																		
WEX2b	1136	C	T						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX2b	1141	T	A														x															
WEX2b	1146	T	A						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX2b	1148	T	G						x	x	x			x		x		x		x	x	x	x	x	x			x	x	x		x
WEX2b	1160	C	G	435	CTC-L	CTG-L																	x						x			

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
WEX2b	1175	G	A	450	GCG-A	GCA-A			x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX2b	1187	A	C	462	ATA-I	ATC-I											x															
WEX2b	1198	G	A	473	AGT-S	AAT-N			x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX2b	1226	T	C	501	TTT-F	TTC-F			x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX2b	1250	C	T	525	CTC-L	CTT-L																								x		
WEX2b	1265	A	G	540	GAA-E	GAG-E			x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX2b	1297	T	G								x			x		x		x		x	x	x		x	x			x		x		x
WEX2b	1297-1298	*TT*	*2 bp Δ*						x	x							x						x						x			
WEX2b	1309	G	A						x	x							x						x						x			
WEX2b	1312	G	A						x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX2b	1329	C	T														x															
WEX2b	1340	C	A						x	x							x						x						x			
WEX2b	1348	G	A														x						x						x			
WEX3	1394	C	T	582	GAC-D	GAT-D			x	x																						
WEX3	1425-1427*	AGT*	TCC*	613-615*	AGT-S*	TCC-S*																	x						x			
WEX3	1439	T	A	627	GTT-V	GTA-V																	x						x			
WEX3	1442	G	T	630	GAG-E	GAT-D																	x						x			
WEX3	1466	A	T	654	CAA-Q	CAT-H											x															
WEX3	1474	G	A	662	GGT-G	GAT-D			x	x	x			x		x		x		x	x	x		x	x			x		x		x
WEX3	1475	T	C	663	GGT-G	GGC-G											x															
WEX3	1480//1481*	AT*	GTTCG*	668//669*	GAT-D*	GGTTCG-GS*											x															
WEX3	1481	T	G	669	GAT-D	GAG-E			x	x	x			x		x		x		x	x	x		x	x			x		x		x
WEX3	1483	A	G	671	AAA-K	AGA-R											x															
WEX3	1500	T	G	688	TCA-S	GCA-A											x															
WEX3	1533	T	C														x															
WEX3	1536	G	A														x															
WEX3	1537	T	A														x															
WEX3	1539	A	T						x	x	x			x		x		x		x	x	x		x	x			x		x		x
WEX3	1541	T	G														x															
WEX3	1542	*C*	*1 bp Δ*														x															
WEX3	1543	A	G														x															
WEX3	1547	T	C														x															
WEX3	1577	G	A														x															
WEX3	1584	T	G														x															

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
WEX3	1589	G	A														x															
WEX3	1618	A	C														x															
WEX3	1619//1620		*C*														x															
WEX3	1634	G	T														x															
WEX3	1636	A	G														x															
WEX3	1640	C	T													x												x				
WEX3	1642	A	T														x															
WEX3	1644	T	A						x	x	x			x		x	x	x		x	x	x	x	x	x			x	x	x		x
WEX3	1647	C	A														x															
WEX3	1655-1687	*CTG...ATT*	*33 bp Δ*														x															
WEX3	1701	C	T														x															
WEX3	1824	T	C	823	TAC-Y	CAC-H							x		x																	
WEX3	1827	A	C	826	AAG-K	CAG-Q							x		x		x															
WEX3	1833	A	G														x															
WEX3	1836	A	C														x															
WEX3	1838	A	T										x		x		x															
WEX3	1840	C	A										x		x		x															
WEX3	1842	A	G														x															
WEX3	1850	T	G										x		x																	
WEX3	1852	T	C						x	x	x			x		x		x	x	x	x	x	x	x	x			x	x	x		x
WEX3	1856	C	T										x		x																	
WEX3	1858	C	T										x		x																	
WEX3	1862	C	T										x		x																	
WEX3	1866	G	T										x		x																	
WEX3	1868	C	T										x		x																	
WEX3	1869	C	T														x															
WEX3	1875	G	A										x		x																	
WEX3	1881	A	G										x		x		x															
WEX3	1884	T	A										x		x																	
WEX3	1888	A	G										x		x																	
WEX3	1895	T	A										x		x																	
WEX3	1903	A	C										x		x																	
WEX3	1907//1908		*T*										x		x																	
WEX3	1908	G	A													x				x	x	x		x				x				

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
WEX3	1910	T	A										x		x		x															
WEX3	1909-1910	*TA*	*2 bp Δ*															x														
WEX3	1911	A	T																						x							
WEX3	1922	A	T														x															
WEX3	1930	G	C														x															
WEX3	1932	T	C														x															
WEX3	1938	C	T						x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x			x	x	x		x
WEX3	1939	G	A														x															
WEX3	1950	A	T														x															
WEX3	1951	G	A														x															
WEX3	1957	G	A														x															
WEX3	1962	A	G														x															
WEX3	1963	C	A														x															
WEX3	1966	T	G														x															
WEX3	1967	*A*	*1 bp Δ*														x															
WEX3	2011	T	C	849	AGT-S	AGC-S			x	x	x		x	x	x	x		x	x	x	x	x	x	x	x			x	x	x		x
WEX3	2013	G	A	851	GGC-G	GAC-D			x	x	x		x	x	x	x		x	x	x	x	x	x	x	x			x	x	x		x
WEX3	2017	A	T	855	TCA-S	TCT-S											x															
WEX3	2028	G	A														x															
WEX3	2033	A	G													x				x	x	x		x				x				
WEX3	2055	C	T															x														
WEX3	2064-2066	*TAG*	*3 bp Δ*														x															
WEX3	2067	C	A														x															
XRN4-1	498	T	G											x																		
XRN4-1	628	A	T									x												x					x			
XRN4-1/2a	912	T	A															x					x									
XRN4-2a	1404	T	C																								x					
XRN4-2a	1689	T	C					x	x	x	x	x	x		x		x	x		x	x		x			x			x		x	
XRN4-3	2068	C	T											x																		
XRN4-3	2383	*T*	*1 bp Δ*											x																		
XRN4-4	2627	G	T					x	x	x	x	x	x	x	x	x	x	x		x	x		x	x		x	x	x	x		x	x
XRN4-4	2694	C	A											x																		
XRN4-4	2938//2939		*TGAATTGCTT*											x																		

Amplicon(s)	Position in Col-0 gene	Col-0 nucleotide(s)	Polymorphic nucleotide(s)	Position in Col-0 ORF	Col-0 codon(s)-amino acid(s)	Polymorphic codon(s)-amino acid(s)	Col-0	Amel-1	Ang-0	Baa-1	Bor-4	C24	Cit-0	Cvi-0	Gie-0	Kas-1	Kin-0	Kl-5	Kno-18	LL-0	Lp2-2	Lz-0	Mt-0	Pu2-23	Ra-0	RRS-7	Sapporo-0	Shahdara	Sq-8	Tscha-1	Tsu-0	Ws-0
XRN4-4	2996	T	C					x	x	x	x	x	x	x	x	x	x	x		x	x		x	x		x	x	x	x	x	x	x
XRN4-4	3214	*T*	*1 bp Δ*																											x		
XRN4-5	3582-3586	*CTTCT*	*5 bp Δ*											x																		
XRN4-5	3718	C	T					x	x	x	x	x	x		x		x	x		x	x		x			x			x		x	
XRN4-5	3760//3761		*T*											x																		
XRN4-5	4044	A	T																					x								
XRN4-5	4112	T	A											x																		
XRN4-8	6257	T	G	2641	TCA-S	GCA-A		x																								
XRN4-8	6367	T	A	2751	GAT-D	GAA-E								x																		
XRN4-8	6573	G	T																					x								
XRN4-8	6624//6625		*TA*					x	x	x	x	x	x		x			x		x	x		x			x	x		x		x	
XRN4-8	6624//6625		*TATA*														x															

Supplementary table 9. cDNA information of twelve candidate genes

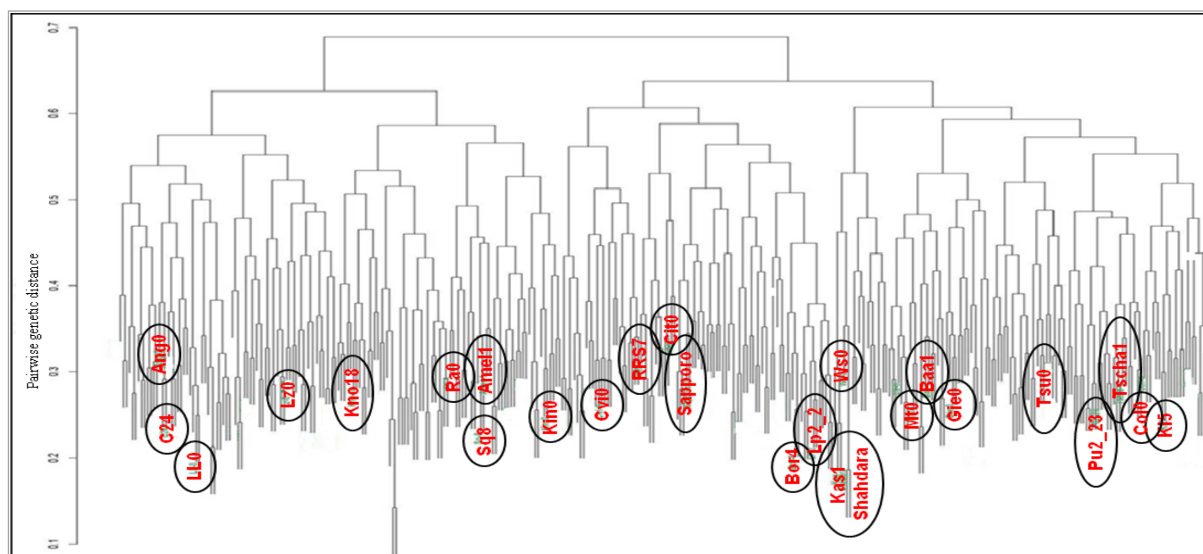
Gene of interest	No. of full-length cDNA(s)	Accession number	No. of partial cDNA(s)	Accession number
AGO1	1	U91995	6	AK227868
				BX815116
				AY080690
				BX818680
				AY600524
				BT000941
AGO7	1	AY394564	0	
DCL4	1	DQ118423	0	
ERI	1	AF419612	2	BX822510
				AY079112
HEN1	2	AF327068	0	
		AF531179		
NRPD1	1	DQ020657	1	AY826515
NRPE1	1	DQ020656	2	AY826516
				AY927744
SDE3	1	AK117698	0	
SDE5	0		1	BX825851
SGS3	1	BT002944	1	BT004380
XRN4	4	AF286718	2	BX814283
		AY064012		AY091411
		BT026022		
		BX815475		
WEX	2	AF531179	2	BT010908
		AJ404476		BX826662

Supplementary table 10. Screening for polymorphic Indel markers. The letters “a”, “b” and “c” indicate that the size of amplification products is very similar, longer or shorter when compared to Col-0, respectively. Multiple amplification products are indicated as “db”. The letter “n” indicates that amplification products were repeatedly not obtained. Asterisks indicate PCR amplification products need to be run on a NuSieve 3:1 agarose gel.

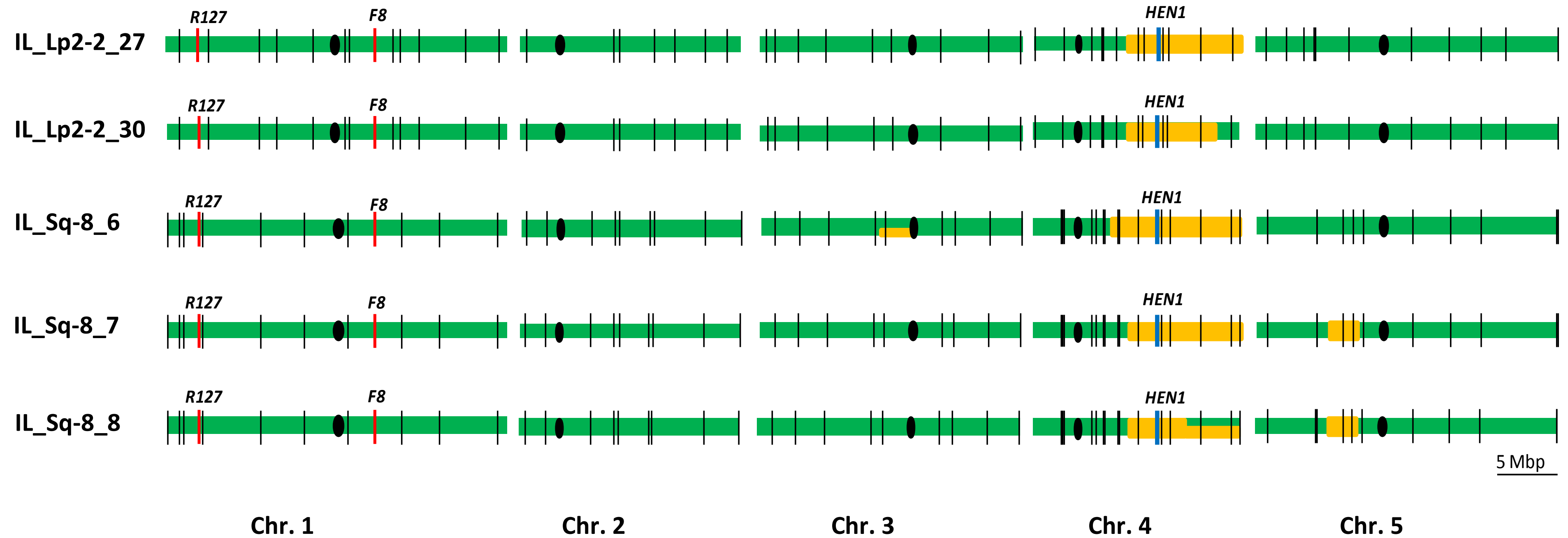
Indel marker	Col-0	Ang-0	Baa-1	Bor-4	Cvi-0	Gie-0	Kas-1	Kin-0	Lp2_2	Lz-0	Ra-0	Shahdara	Sq-8	Ws-0
Nga59	a	b	b	b	b	a*	b	b	a*	b	b	b	b	a*
1-0232	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F21B7	a	n	a	c	a*	c*	a*	n	a	n	n	a*	b*	n
UPSC_1-1021	a	c*	c*	c*	c*	c*	c	c	c*	c	c	c db	c db	c*
F19P19	a	c	c	c	c	c	c	c	c	c	c	c	c	c*
IndRIL-I-2a	a	c	a	a	a	a	c	a	a	a	a	c	a	a
T1G11	a	c*	c*	c	c*	c*	c*	c	c	a	a	c*	c*	c
ATEAT1	a	c*	c	c	c	c	c	c	c	c	c	c	c	c
F12K11-2-IND	a	a	a	c	c	c	c	c	a	c	c	c	a	c
F7G19	a	c*	a*	a*	c*	c*	a*	c*	a*	c*	c*	c*	a*	a*
NGA63	a	c	c	c	c	c	c	c	c	c	c	c	c	c
1-1259	a	c	c	a	a	a	a	a	c	c	c	c	a	a
Ind_I_5	a	a	n	c	a	n	n	c	n	n	c	a	n	a
1-2653	a	a	a	a	a	c	c*	a	a	a	a	c*	c	a
MSAT1.3	a	b*	c	c	c	c	c	a	c	c	c	c	c	c
CIW12	a	c	c*	b	c	c	c	b*	c	c	c	c	c	c
NGA248	a	c*	c	c	b	c	c*	c	c	c*	a*	c*	c*	c*
Ind_I_12	a	a	a	c	a	a	c	n	a	a	a	c	c	a
1-4276	a	n	c*	b*	a	n	c*	c*	c*	b	b*	a	b*	b
T27K12	a	b	c*	b	b*	b*	b	b*	b	b	b*	b	a*	b
1-5335	a	c	c	c	c*	a*	c	c	c	c	a	c	c	c
1-5380	a	a	a	a	a	a	a	c	a	a	a	a	a	a
INDRIL-I-15	a	a	a	a	a	a	a	c	a	a	a	a	a	a
Ind_I_17	a	a	a	a	b	a	a	a	a	a	a	a	a	a
CIW1	a	n	b	a	b	c	c	c	a	b*	b*	c	c*	c
1-6613	a	c	c	a	c	c	c*	c	b*	c	c	c*	c	c*
NGA280	a	c	c	c	c	c	c	c	c	c	c	c	c	c
Ind_I_21	a	c	c	a	a	a	c	a	a	c	c	c	c	c
F11P17-4615	a	c	c*	a	b*	c	c	c	c	c	c	c	c	c
1-7539	a	c	c	c	c*	c*	c*	c	c*	c*	c*	a	c*	c*
Ind_I_24	a	a	n	a	c	a	a	c	c	a	a	c	n	c
F5I14-IND	a	b	a	b	b	b	b	b	b	b	b	b	b	b
MSAT1.13	a	b*	a*	c	c	b*	c	c	b*	a*	c	c	c	b*
1-8645	a	c	c	a	a	c	a	c	c	c	c	a	c	c
MSAT1.1	a	c	c	c	c*	c	c	c*	c	a*	a	c	c*	a
UPSC_1-26627	a	b	b	b	b	b	b	b	b	b	b	b	b	b
Ind_I_27	a	a	a	a	a	a	a	a	b	a	a	a	a	b
ATHATPASE	a	c	c	c	c	c	c	c	c	c	c	c	c	c
UPSC_1-29617	a	a	a	b	b	b	b	a	b	a	a	b	b	a
1-9959	a	c	a	a	a	a	a	c	c	c	c	c	c	c
MSAT2.5	a	c	a	c	c	c	c	c	c	c	c	c	c	c
Ind_II_1	a	c	c	c	a	c	c	c	c	c	c	c	c	a
nga1145	a	c	c*	c	c	c	c	c	c	c	c	c	b*	c
MSAT2.26	a	n	c	a*	c*	c	c*	c	c*	c	c	c*	c	c
Ind RIL II29a	a	c	a	a	c	c	c	c	a	a	a	a	c	c
MSAT2.38	a	c	c	b	c*	c*	b*	c*	b*	c*	c	c*	c	c
Ind_II_4	a	a	a	a	a	c	c	c	a	a	a	c	a	a
MSAT2.28	a	c*	c*	c	n	c	c*	c*	n	c*	c*	c*	c	b*
2_3475	a	a	a	c	a	a	a	a	a	a	a	c	a	c
2-3728	a	a	a	a	a	a	a	a	a	a	a	a	a	a
MSAT2.11	a	c	a	c*	c	c	c	c	b	c	c	c	c	a
2-4269	a	c	c	c	c	c	c	c	c	c	c	c	c	c
MSAT2.36	a	b	b	a*	c	b	c*	b*	c	b*	b*	c*	b	b*

Indel marker	Col-0	Ang-0	Baa-1	Bor-4	Cvi-0	Gie-0	Kas-1	Kin-0	Lp2_2	Lz-0	Ra-0	Shahdara	Sq-8	Ws-0
Ind_II_9	a	c	c	c	c	c	n	c	c	c	c	c	c	c
UPSC_2_9168	a	c	c	c	a	c	c	c	a	a	a	c	c	a
UPSC_2-9637	a	c	c	c	c	c	c	c	c	c	c	c	c	c
PLS7	a	c	b	c	c	b	c*	a	b	c	c	c	b	b
MSAT2.17	a	b*	c	c	c	c*	a*	b*	c	b*	b*	c	b	c*
MSAT2.41	a	b	b	c	c	c	c	n	c	c*	c	c	c*	c
2-5887	a	a	c	a	a	c	c	c	a	a	a	c	c	a
nga1126	a	a*	a*	b*	a*	a*	b*	a*	b*	b*	b*	b*	a*	a
CZSOD2	a	c	c	c*	c	c	a	c	c	c	c	a	c*	c
Ind_II_13	a	a	a	a	a	a	a	a	a	a	a	a	a	a
MSAT2.4	a	b	c*	a*a	c	c	c	b	c*	b	b	c	n	n
UPSC_2-14568	a	a	a	a	b	b	b	b	b	a	a	b	b	b
Ind_II_16	a	a	c*	c*	a	a	c*	a	c*	c*	a	a	a	c*
2-8295	a	c*	c	c	c	c	c	c	c	c	c	c	c	c
MSAT2.9	a	c	c	c	c	c	c	c	c	c	c	c	c	c*
UPSC_2-18415	a	b	b	b	n	b	b	b	b	b	b	b	b	b
Ind_II_19	a	a	a	a	b*	b*	a	a	a	a	a	a	a	a
MSAT2.22	a	c*	a*	c*	c*	c	c*	c*	a*	c*	c*	c	c	c
3-0089	a	c	a	a	a	a	a	a	a	a	a	a	a	a
3-0186	a	a	c	a	a	a	a	a	a	a	a	a	a	c
3-0363	a	a	a	c*	b*	b	a*	c	b	a	a*	a	c*	b
nga172	a	a	a	c*	b	b	a	c	b	a	a	a	c*	b
Ind_III_1	a	c	a	c	c	c	c	c	c	a	a	c	c	c
RIL-III-50	a	b	a	a	b	a	a	b	b	a	a	a	b	a
UPSC_3-3716	a	c	c	c	c	c	c	c	c	c	c	c	c	c
Ind RIL III 52	a	b*	a	a	a*a	a	b*	b*	a	a	a	b*	a	a
3-2402	a	a	a	a	a	a	a	a	c	a	a	a	a	c
Ind_III_6	a	c	c	c	a	c	c	c	c	c	c	c	c	c
MSAT3.19	a	c	a*	a*	a*	b	a*	a*	c	b	b*	b	a	a*
MSAT3.19-IND	a	c*	a	n	n	c*	a*	a	a	a	c*	a*a	n	a
Ind_III_10	a	a	a	a	a*	a	c	a	a	a	a	c	a	a
IND-RIL-III-56	a	a	a	b	b	a	a	a	a	a	a	a	b	a
3-4332	a	c	c	c	c	c	c	c	c	c	c	c	c	c
MSAT3.32	a	b	b	b	b	c*	c*	b	b	b	b	c*	b	b
Ind_III_12	a	c	c	c	a	a	n	c	c	c	a	n	a	a
IND-RIL-III-63	a	b	b	b	b	n	b	b	b	b	b	n	b	b
IND-RIL-III-62a	a	b	b	b	b	b	b	b	b	b	b	b	b	b
Ind_III_17	a	c	a	a	c	a	a	c	a	a	a	c	c	a
MSAT3.21_IND	a	a	c	c	c	c	c	c	c	c	c	c	c	a
INDRIL-III-64a	a	b*	b*	a	b*	b*	b*	b*	b*	b*	b*	b*	a	b*
3-8728	a	n	b*	a*	a*	c	a*	a*	b*	c	c*	c*	c*	b*
Ind_III_21	a	a	a	a	a	a	a	a	a	a	a	c	a	c
3-9924	a	c	c	c	a	c	c	c	c	c	c	c	c	c
INDRIL-IV70b	a	c	a	c	c	a	c	a	c	a	a	a	a	c
FRI-IND	a	b	b	b	b	a	b	b	b	b	b	b	a	b
4-0175	a	b	b	b	a	b	b	a	b	a	b	b	a	b
Ind_IV_1	a	a	a	n	a	c	a	a	a	a	c	a	a	a
4-1384	a	b	a	b	b	b	b	b	b	b	b	b	b	b
Ind_IV_3	a	a	a	a	a	a	a	a	a	c	a	a	c	a
UPSC_4-2821	a	b	b	b	b	b	b	b	b	b	b	b	b	b
INDRIL-IV-74	a	b	b	b	b	b	b	b	b	b	a	b	b	b
nga8	a	b	a	b	c	b	b	b*	b	c*	b	c*	b	b
nga1111	a	b*	c	b	a*	a	b	b	b	b*	a*	b	a	b
UPSC_4-6222	a	c	c	c	a	c	c	c	c	c	c	c	c	c
UPSC_4-6362	a	c	c	c	c*	c	c*	c	c	c*	c*	c	c	c*
MSAT4.25	a	a	c*	c*	c	c*	c	c*	a*	c*	c	c*	a	c*
MSAT4.35	a	c	c	c	c	c	c	c*	c	c	c	c	c*	c
Ind_IV_8	a	c	n	a	n	a	c	c	n	n	a	c	c	c
MSAT4.15	a	c	c	c	c	c	c	c	c	c	c	c	c	a
4-5268	a	c	c	c	c	c	c	c	c	a	a	c	a	a

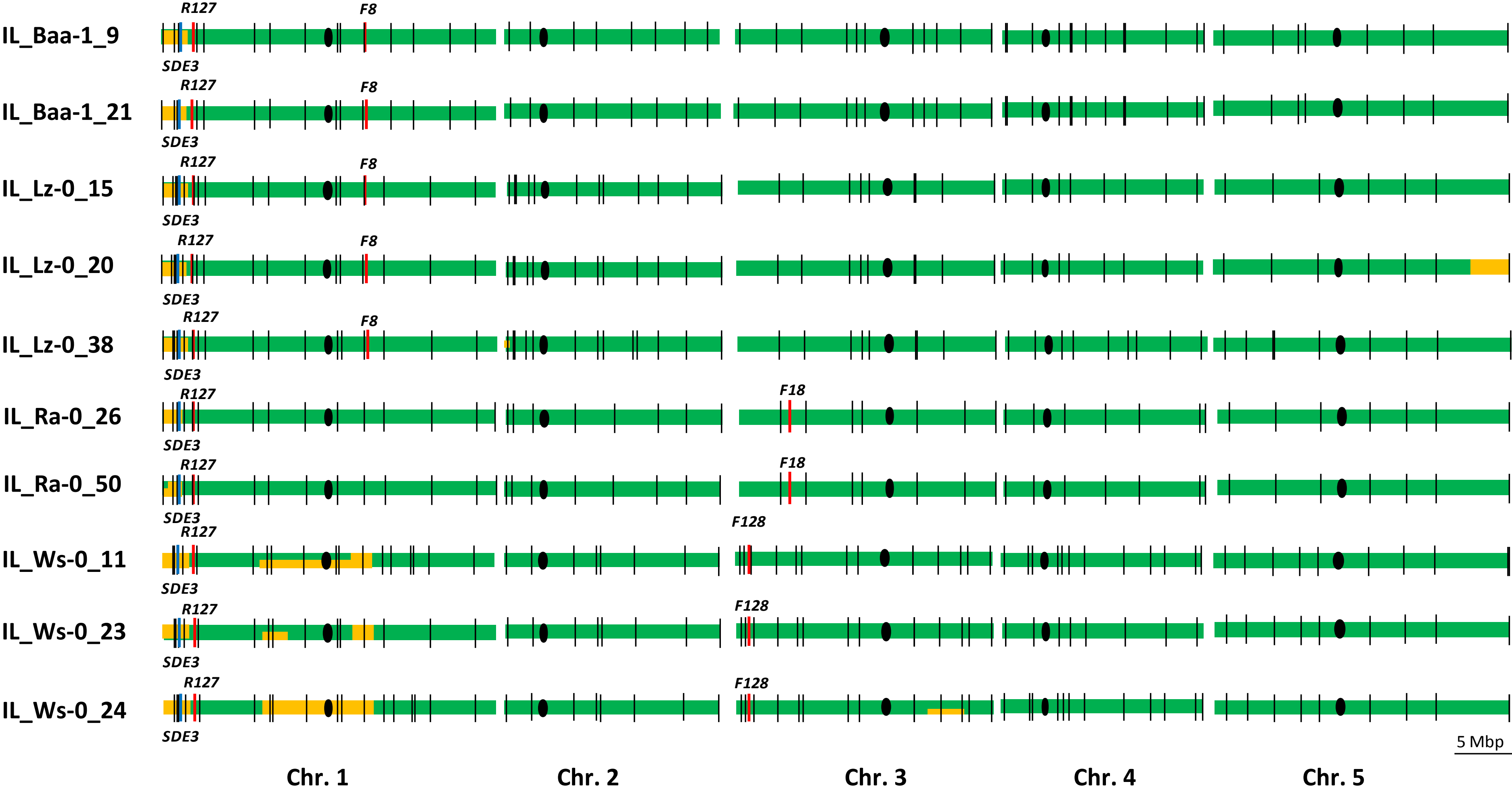
Indel marker	Col-0	Ang-0	Baa-1	Bor-4	Cvi-0	Gie-0	Kas-1	Kin-0	Lp2_2	Lz-0	Ra-0	Shahdara	Sq-8	Ws-0
INDRIL-IV79B	a	a	a	a	a	a	a	a	a	a	a	a	a	a
UPSC_4-11022	a	a	c	a	n	a	a	a	a	a	a	a	n	a
UPSC_4-11152	a	a	b	b	b*	b	b	b	a	b	b	b	a	b
CIW7	a	c	a	c*	c	a*	c	c	c*	c*	c	c	c	c*
MSAT4.18	a	c	c	c	c	c	c	c	c	c	c	c	a	c
UPSC_4-12254	a	a*	a*	a*	c*	a*	a*	a*	a*	a*	a*	a*	a*	a*
UPSC_4-12273	a	c*	c	a	c	c	c	n	c	c	c	c	c	c
4-7366	a	b	b	b	b	a	a	b	a	a	a	b	a	b
Ind_IV_15	a	a	a	a	a	a	c	a	a	a	a	c	a	a
UPSC_4-14985	a	b	b	b	b	b	b	b	b	b	a	b	b	b
INDRIL-IV-86	a	c	c	c	c	c	c	c	c	c	c	c	c	c
4-9963	a	c	c	a	c	c	c	c	a	a	c	c	c	c
CTR1.2	a	c	c	c	c	c	c	c	c	c	c	c	c	c
Ind_V_2	a	n	c	n	a	c	c	c	n	a	a	c	c	c
Nga249	a	c	a*	c	c	c	c	c	b*	c	c	c	c	c
5-1629	a	b	b	b	b	b	b	a	b	a	a	b	a	b
Ind_V_5	a	c	c	c	a	a	a	c	c	c	c	a	c	c
nga106	a	c	c	c	c	c	c	c	c	c	c	c	c	a
5-2862	a	c	c	c	c	c	c	c	c	c	c	c	c	c
nga139	a	c	c*	c	c*	c	c	c	c	c	c	a*	c*	c
Ind_V_9	a	c	a	a	c	a	c	c	a	c	c	c	c	c
5-3683	a	a	a	a	a	c	a	a	a	a	a	c	a	a
5-5037	a	c	a	c	a	c	a	a	a	a	a	a	a	a db
MSAT5.2	a	c	n	c*	n	c*	c	n	c	n	c	c	c	c
PHYC.3	a	b	b	b	b	a	b	b	a/b	b	b	b	b	a
5-6437	a	c*	c*	c*	c*	c*	c*	c*	b*	c*	c	c	c	b
Ind_V_18	a	a	a	a	a	a	a	a	a	a	a	a	a	a
5-7443	a	c	c	c	a	c	c	a	c	c	a	c	c	a
nga129	a	c	a*	c	c*	c	c	c	c	c	c	c	c	c
Ind_V_22	a	a	a	a	a	a	c	a	c	a	a	c	a	a
INDRIL-V-112	a	b	a	b	b	a	b	a	b	b	b	b	b	b
K8K14-IND	a	a	c	c	c	c	c	c	c	a	a	c	c	c
Ind_V_27	a	a	a	b	b	a	b	a	b	a	a	b	b	b



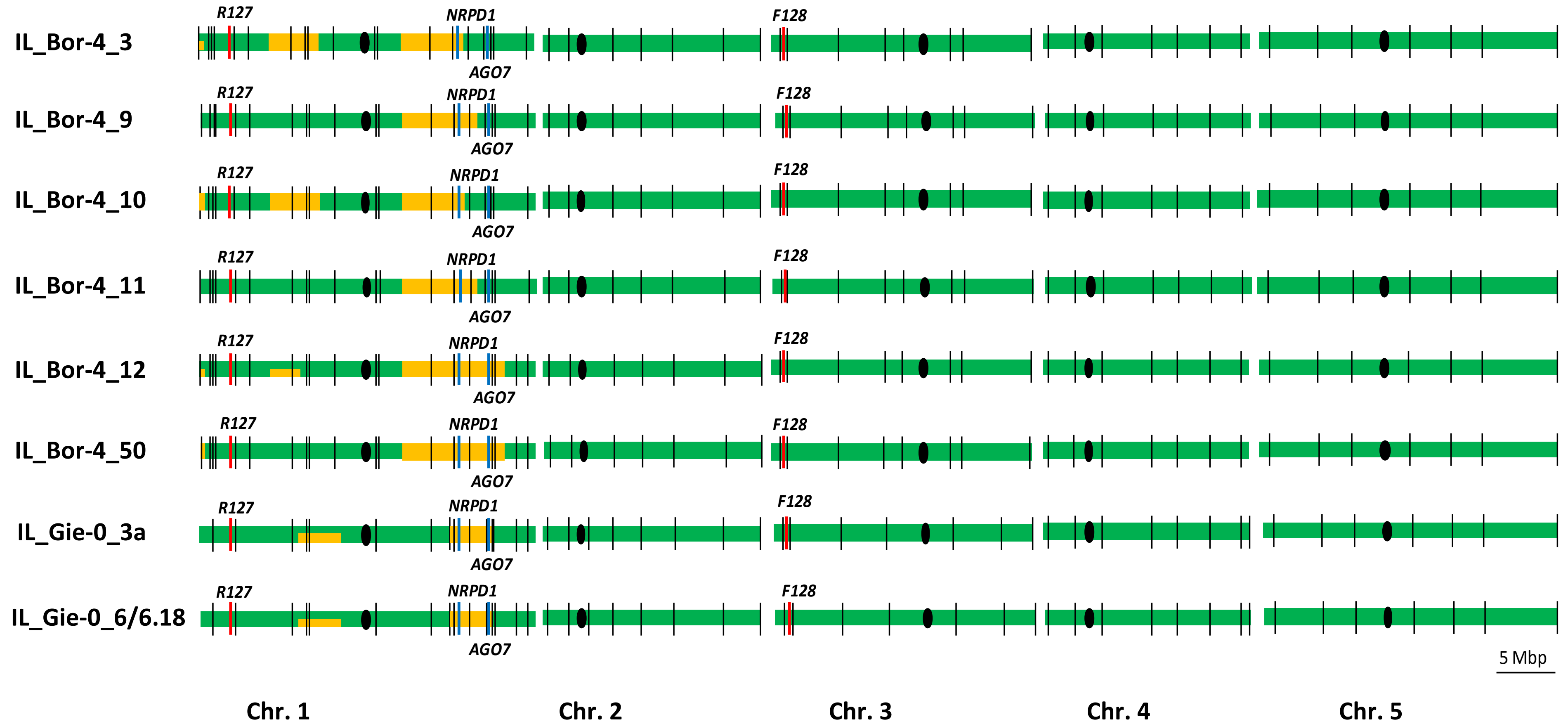
Supplementary figure 1. Pairwise genetic distances of 360 *A. thaliana* accessions using 149 SNPs. The accessions used in this study were coloured in red and circled in black (modified from http://borevitzlab.uchicago.edu).



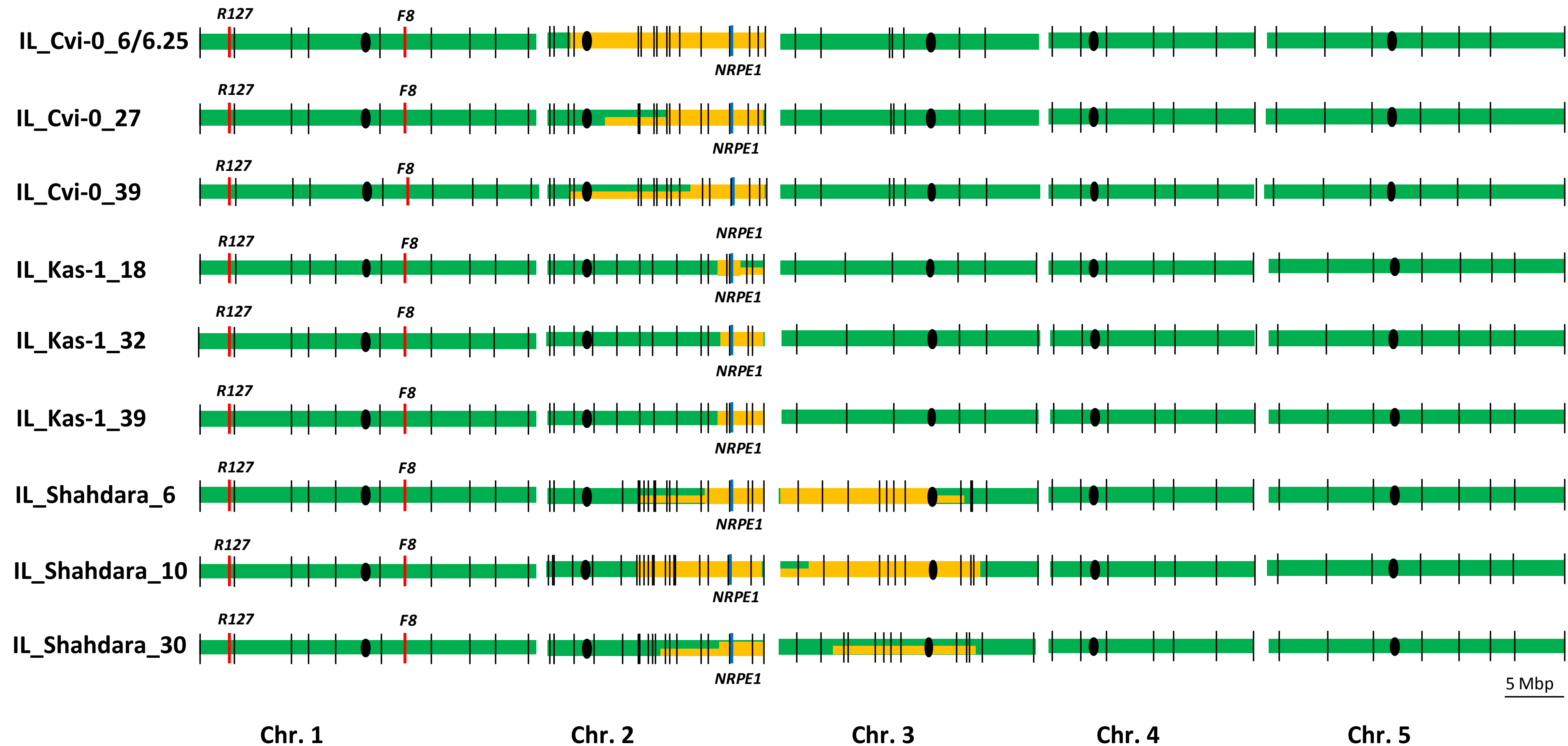
Supplementary figure 2. Characterisation of introgression lines that carry allelic variants of the *HEN1* gene with Indel markers. Centromeres are shown as black circles. The blue, red and black bars mark the map positions of candidate genes, *GFP* loci and Indel markers, respectively. Genome regions derived from Col-0 are shown in green colour and yellow colour represents segments of the other accessions.



Supplementary figure 3. Chromosome maps of introgression lines containing allelic variants of the *SDE3* gene. Map positions of the candidate gene, the *GFP* loci and Indel markers are indicated with blue, red and black bars, respectively. Black circles mark the centromeres. Genome regions of Col-0 are shown in green colour and yellow colour represents segments of the other accessions.



Supplementary figure 4. Graphical genotypes of introgression lines carrying allelic variants of the *AGO7* and/or *NRPD1* genes. Col-0 chromosome segments are shown in green colour and yellow colour represents genome regions of the other accessions. Black circles mark the map positions of the centromeres. *GFP* loci and candidate genes are indicated by red and blue bars, respectively. Map positions of Indel markers are shown as black bars.



Supplementary figure 5. Chromosomal location and sizes of introgressed segments for introgression lines containing allelic variants of the *NRPE1* gene. The segments from selected accessions are shown in yellow colour, green colour indicates genome regions derived from Col-0. Black circles show the positions of the centromeres. Loci showing the *GFP* and candidate genes are identified by the red and blue bars, respectively.

CURRICULUM VITAE

Personal information

Name:	Le Phuong Dung
Gender:	Female
Date of birth:	July 16th, 1985 in Thai Nguyen, Vietnam
Nationality:	Vietnamese
Marital status:	Married
Residential address in home country:	No. 949, Duong Tu Minh Street, Hoang Van Thu Ward, Thai Nguyen City, Thai Nguyen, Viet Nam
Residential address in Germany:	Selkeweg 4A, OT Gatersleben, 06466 Seeland
Email:	dung@ipk-gatersleben.de; lephuongdungkstn@gmail.com

Education and Employment:

10/2011-present:	PhD student in Research group Genome plasticity (Supervisor: Dr. Renate Schmidt), Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), OT Gatersleben, Seeland Germany. **PhD thesis**: Analysis of sense transgene-induced gene silencing in introgression lines reveals the presence of silencing modulators in *Arabidopsis thaliana* accession genomes.
10/2009-09/2011:	Lecturer at the Faculty of Biology, Thai Nguyen University of Education, Thai Nguyen, Vietnam.
09/2007-09/2009:	Master student at the Faculty of Biology, Thai Nguyen University of Education, Thai Nguyen, Vietnam. **Master thesis**: Cloning of the promoter of the gene encoding cinnamyl alcohol dehydrogenase (CAD) expressed in xylem of *Eucalyptus urophylla* S.T. Blake.
09/2003-05/2007:	Bachelor student at the Faculty of Biology, Thai Nguyen University of Education, Thai Nguyen, Vietnam.
09/2000-05/2003:	Thai Nguyen Specialised High school, Thai Nguyen, Vietnam.

Poster and Oral presentations:

- Loan, T.L., **Dung, P.L**. & R. Schmidt presented the poster "Natural variation of *Arabidopsis thaliana* gene involved in post-transcriptional transgene silencing at the Institute's Day, 24th-25th September 2012, IPK Gatersleben, Germany.
- Loan, T.L., **Dung, P.L**. & R. Schmidt presented the poster "*Arabidopsis thaliana* gene involved in post-transcriptional transgene silencing – Assessing natural variation and its impact" at the Institute's Day, 25th-27th September 2013, IPK Gatersleben, Germany.
- **Dung, P.L**., Loan, T.L. & R. Schmidt presented the poster "Natural variation of *Arabidopsis thaliana* genes involved in post-transcriptional transgene silencing" at The Plant Science Student Conference, 2nd-5th June 2014, IPK Gatersleben, Germany.
- Loan, T.L., **Dung, P.L**. & R. Schmidt presented the poster "Natural variants of *Arabidopsis thaliana* genes affect post-transcriptional transgene silencing" at the Institute's Day, 8th-10th October 2014, IPK Gatersleben, Germany.
- **Dung, P.L**., Loan, T.L. & R. Schmidt gave the talk "Characterisation of *Arabidopsis thaliana* introgression lines with an impact on post-transcriptional gene silencing" at The Plant Science Student Conference, 2nd-5th June 2015, IPB Halle, Germany.
- **Dung, P.L**., Loan, T.L. & R. Schmidt presented the poster "Characterisation of *Arabidopsis thaliana* introgression lines with an impact on post-transcriptional gene silencing" at the Institute's Day, 14th-16th October 2015, IPK Gatersleben, Germany.
- **Dung, P.L**., Loan, T.L. & R. Schmidt presented the poster "Characterisation of *Arabidopsis thaliana* introgression lines with an impact on post-transcriptional gene silencing" at The Plant Science Student Conference, 4nd-7th July 2016, IPK Gatersleben, Germany.

Erklärung

Hiermit erkläre ich, dass diese Arbeit bisher von mir weder an der Naturwissenschaftlichen Fakultät I der Martin-Luther-Universität Halle-Wittenberg noch einer anderen wissenschaftlichen Einrichtung zum Zwecke der Promotion eingereicht wurde.

Ferner erkläre ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die darin angegebenen Hilfsmittel und Hilfen benutzt und keine Textabschnitte eines Dritten ohne Kennzeichnung übernommen habe.

Gatersleben, June 2016

Le Phuong Dung