Functional Characterization of Orphan Genes in *Arabidopsis thaliana*

Volume I of I

Sandesh Swamidatta, M.Sc.

A thesis submitted to National University of Ireland Galway

For the degree of Doctor of Philosophy

Under the supervision of Prof. Charles Spillane

College of Science, School of Natural Sciences

Discipline of Botany and Plant Science

September 2014

Table of Contents

List of Tables

List of Figures

Declaration

I certify that this thesis is my own work, and that I have not used this work in the course of another degree, either at National University of Ireland Galway, or elsewhere.

Signed: 

Sandesh Swamidatta

Acknowledgements

I would like to thank my supervisor Prof. Charles Spillane for giving me an opportunity to do my PhD in his research group and for all his support and encouragement throughout my PhD.

To college of Science, NUI Galway and Science Foundation Ireland for the scholarship during my PhD.

To my GRC committee members – Prof. Brian McStay and Dr. Heinz Peter Nasheuer.

I am thankful to Prof. Ken Feldmann, University of Arizona, USA for all his help, encouragement and support as collaborator. I would also like to thank Prof JM Mulet, IBMCP, Valencia for all the help with the yeast work.

A special thanks to Dr. Galina Brychkova for all the help and support and Dr. Peter Mckeown for all the good advice, and helping to improve the presentation of my thesis.

To Reetu Tuteja and Martin Braud for helping me with computational analysis.

To all my lab mates for valuable discussions, continuous support and encouragement.

Finally and importantly my heartfelt thanks to my parents and family back in India for their encouragement and support throughout my educational career so far.

Abbreviations

ABA : Abscisic acid

ABRC : Arabidopsis Biological Resource Centre, Ohio State University

AGI : Arabidopsis Genome Initiative

AGRIKOLA : Arabidopsis Genomic RNAi Knock-Out Line Analysis

ANOVA : Analysis of variance

ASC : Ascorbic acid

BHT : Butylated hydroxy toluene

BLAST : Basic Local Alignment Search Tool

cDNA : Complementary DNA

CDS : Coding sequence

Col-0 : Columbia (ecotype of A. thaliana)

CTAB : Cetyltriethylammoniumbromide

DAB : Diaminobenzidine

DNA : Deoxyribose nucleic acid

EDTA : Ethylenediaminetetraacetic acid

GA : Gibberellic acid

GPD : Glyceraldehyde-3-phosphate dehydrogenase

GSH : Glutathione

GSTs : Gene sequence tags

GUS : β-glucuronidase

H_2O_2 : Hydrogen peroxide

MA : Mugeneic acid

MS media : Murashige and Skoog media

NA : Nicotianamine

NASC : Nottingham Arabidopsis Stock Centre

NBT : Nitroblue tetrazolium

ORF : Open reading frame

PAC : Paclobutrazol

PAGE : Polyacrylamide gel electrophoresis

PCR : Polymerase chain reaction

PEG : Polyethylene glycol

RNA : ribonucleic acid

ROS : Reactive oxygen species

SDS : Sodium dodecyl sulphite

SiR : Sulphite reductase

TAIR : The Arabidopsis Information Resource

T-DNA : Transferred DNA

TILLING : Targeting Induced Local Lesions in Genomes

UTR : Untranslated region

WT : Wild type

YPD media : Yeast extract, Peptone and Dextrose media

Summary of contents

All organisms are known to contain a unique set of genes called ‘orphan genes’, which do not share any similarity with genes or protein domains in other evolutionary lineages. Such orphan genes can impart unique traits to the organism resulting in lineage specific evolutionary innovation. In this work Brassicaceae specific orphan genes in *Arabidopsis thaliana* were analysed for their functionality in abiotic stress responses. Abiotic stress responsive orphan genes were identified based on their differential expression under various stress conditions and were functionally screened using three different approaches; loss-of-function screen, gain-of-function screen and gain-of-function screen by heterologous expression. From the loss of function screen, one orphan gene (*ASR50)* was identified which confers tolerance to salt stress when knocked out. The mutant *asr50-1* also shows high photosynthetic efficiency under salt stress compared to wild type. The gain-of-function screen in *A. thaliana* resulted in two promising candidates, *ASR35* and *ASR63*, which conferred tolerance to salt stress and iron deficiency respectively. The third approach, a gain-of-function screen by heterologous expression using yeast stress sensitive mutants, identified a total of 27 orphan genes conferring varied degrees of stress tolerance or sensitivity to different abiotic stress conditions. Importantly, this approach also indicated that orphan gene functionality is transferable to other lineages.

The overexpression lines of *ASR63* (*35S::ASR63*) were further analysed to identify the possible mechanism of tolerance conferred under iron deficiency. The results suggested that the *35S::ASR63* lines accumulated more iron in shoots and in the mature seeds compared to wild type control. Furthermore, the transgenic lines were found to produce higher levels of the antioxidants glutathione and ascorbic acid in the shoots which possibly maintains favourable redox balance under iron deficiency for the normal uptake of iron. In addition, the seeds of *35S::ASR63* lines show germination sensitivity to ABA and the GA inhibitor PAC which needs further characterization. Our hypothesis is that *ASR63* is a *de novo* gene originated

due to the divergent transcription of a “bifunctional” promoter of an upstream gene. Overall, this study highlights the importance of orphan genes and adds to the increasing evidences for functionality of orphan genes giving rise to evolutionary innovations within lineages.

Chapter 1: General introduction

1.1 The origin of new genes

The study of origin of new genes is one of the exciting topics in the field of molecular evolution. Since evolution is a continuous process, genes are gained and lost during the course of evolution. This is evident from the fact that number of genes in a genome vary greatly among different species (Ranz and Parsch, 2012), ranging from as low as 110 genes in the endosymbiont *Candidatus* Tremblaya princeps (López-Madrigal et al., 2011) to over 40000 genes in cultivated rice (Goff et al., 2002; Yu et al., 2002). Even closely related species like chimpanzees and humans differ in 6.5% of their genome of about 22,000 genes (Demuth et al., 2006). Such divergence is observed despite the fact that both have separated from a common ancestor only about six million years ago, suggesting that both chimpanzees and humans have a proportion of genes unique to their genomes. Throughout the course of evolution new genes have been created within the lineages of organisms. This would also partly explain why there is an enormous diversity among organisms in terms of morphology and physiology. Clearly, the origin of new genes is one of the important processes of evolutionary innovation which contribute to adaptive evolution in all organisms.

Although the interests in studying the evolutionary innovation by origin of new genes with new functions attracted scientists in the past, more in depth studies have only recently become possible due to the advent of genomic technologies that are deciphering the complete genomes of many diverse species (Kaessmann, 2010). With the availability of genome sequences it became possible to answer questions such as, how do new genes arise? What are the evolutionary mechanisms of formation of new genes? and how they evolve and acquire new functions?

1.2 Mechanisms of gene evolution

1.2.1 Gene Duplication

Gene duplication is one of the major driving forces that provide raw material for the emergence for new genes. The importance of gene

duplication in evolution and its potential for establishing genetic novelty was described by Susumu Ohno in 1970 (Ohno, 1970) who stated that gene duplication is the single most important factor in evolution. This set the tone for subsequent studies into gene duplication events/mechanisms. From various studies it is now clear that, once a gene is duplicated, it can take one of the following fates; one copy of the gene get converted into a pseudogene or be lost from the genome (non-functionalization) (Prince and Pickett, 2002), it can transform into a complementary copy (subfunctionalization), preserved as a functionally redundant copy or evolve into novel functions (neofunctionalization)(Lynch and Conery, 2000; Prince and Pickett, 2002).

The duplication of genes can occur in several different ways. DNA based duplication mechanisms were most widely studied although intronless duplicated gene copies may also originate from RNA based mechanism. In other words, duplication can occur through small scale events such as duplication of chromosomal fragments containing whole genes or partial fragments (called segmental duplication) which occur mainly due to the misguided recombination events (also called replication slippage). Larger scale duplication events are caused by whole genome duplication or polyploidization mechanisms (Conant and Wolfe, 2008; Van de Peer et al., 2009). The RNA based mechanism is called retroduplication or retrotransposition, discussed in the next section.

1.2.2 Retrotransposition

This is a RNA mediated duplication mechanism whereby mRNA of the source gene (parental gene) is reverse transcribed and gets integrated into a new locus in the genome creating a new gene. Unlike duplication, retrotransposition only copies the coding sequence without any promoter elements. Hence, the duplicated copy has to recruit new regulatory element to be functional or else it will be lost through mutation into a pseudogene (Kaessmann et al., 2009; Long et al., 2003). Retrogenes have been systematically characterised in several genomes including plants (Wang et

al., 2006), mammals (Emerson et al., 2004; Vinckenbosch et al., 2006) and fruit flies (Zhou et al., 2008).

1.2.3 Exon shuffling

Exon shuffling is a process of formation of new genes by combinations of two or more exons from different genes or by the duplication of exon within the gene. The exons of different genes can be brought together by either illegitimate recombination or retroposed exon elements (Long et al., 2003). At the RNA level, exon shuffling can occur by the formation of chimeric RNA from different gene sources through the process of trans-splicing or transcription slippage. Exon shuffling is one of the important mechanisms of evolution of new genes and approximately about 19% of the eukaryotic genes have been estimated to be formed by exon shuffling (Long et al., 2003; Long et al., 1995).

1.2.4 *De novo* formation of genes

For a long time it has been thought that new genes evolve exclusively by duplication or recombination of existing genes (Neme and Tautz, 2014). The emergence of new genes from non-coding sequences was believed to be unlikely (Wu and Zhang, 2013). In such *de novo* origination, new genes are formed from previously non-coding DNA regions. Formation of new genes *de novo* occurs through the formation of cryptic functional sites (for example, transcription initiation regions, splice sites and polyadenylation sites) from randomly occurring sequence combinations which eventually come under regulatory control to produce distinct processed RNA transcripts. This RNA would then acquire a functional open reading frames (ORFs) to code for a new protein (Tautz and Domazet-Loso, 2011). The most stringent criterion for identifying the *de novo* formed gene is to show that the corresponding genomic region of the gene is present in out-group organisms, but as a non-coding stretch that is neither transcribed nor translated. One recent study in yeast show that *de novo* originated ORFs can be transcribed and translated which would provide adaptive potential and that over

evolutionary time some of these translated ORFs (called as 'proto-genes') would be retained and evolve into novel genes (Carvunis et al., 2012). In addition, Abrusan (2013) has showed that essential genes are already present in proto-genes suggesting proto-genes can rapidly acquire essential functions. Both non-coding RNA and protein coding genes can arise through *de novo* origination mechanisms. However *de novo* protein coding genes can also be formed through noncoding RNA intermediate. For example, analysis of the yeast de novo gene *BSC4* showed that the orthologous, but non protein-coding, loci of *BSC4* gene in the out-group species are expressed (Cai et al., 2008).

Although *de novo* evolution of genes has been considered unlikely for quite a long time, recent studies suggest that d*e novo* origin of transcripts and genes might have been an active process throughout evolution (Tautz and Domazet-Loso, 2011). The availability of sequenced genomes of closely related species made it possible for the systematic identification of *de novo* evolution of genes (Ding et al., 2012). As a result, d*e novo* origination of genes has received increased attention in recent years. At present, *de novo* origination is considered one of the most important mechanisms of origin of genes and has been studied in various species including *Drosophila* (Begun et al., 2007; Levine et al., 2006; Reinhardt et al., 2013; Zhou et al., 2008), mammals (Meunier et al., 2013), humans (Knowles and McLysaght, 2009; Li et al., 2010a; Wu et al., 2011), mouse (Heinen et al., 2009), yeast (Cai et al., 2008; Carvunis et al., 2012; Li et al., 2010b), viruses (Sabath et al., 2012) as well as in plants such as rice (Xiao et al., 2009) and *Arabidopsis thaliana* (Donoghue et al., 2011; Felippes et al., 2008; Silveira et al., 2013). Collectively, these studies have indicated that *de novo* gene formation played an important role during the evolution of new genes and several case studies have confirmed the importance of these genes. Since *de novo* origination is capable of producing entirely different protein from that of existing genome, it can be assumed to provide genetic material for functional innovation in all organisms.

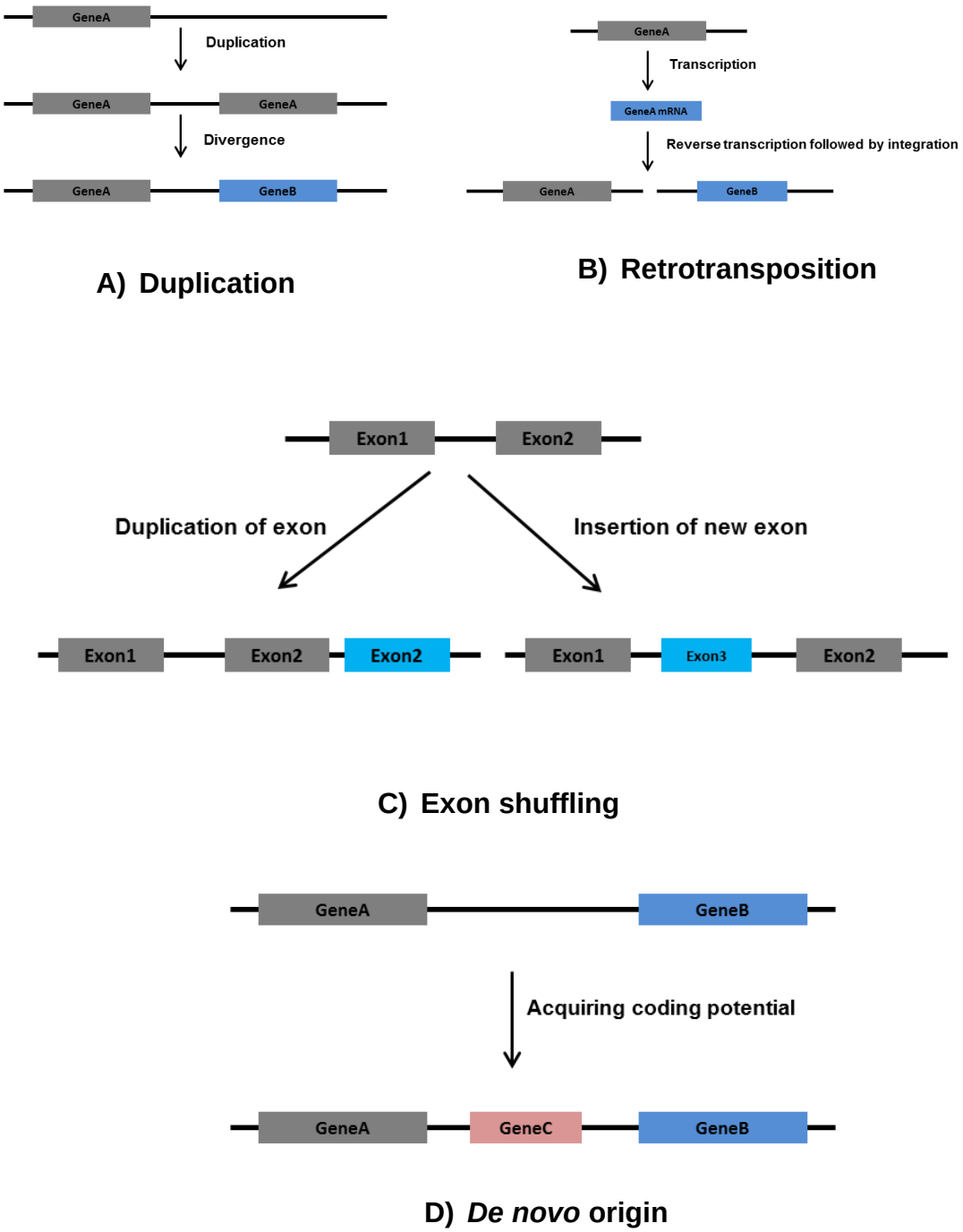


Figure 1: Schematic representation of mechanisms of origin of genes.

A) Duplication and divergence mechanism where GeneA duplicates and the duplicated copy undergo divergence and evolve new function to become GeneB. B) Retrotransposition wherein mRNA of a gene is reverse transcribed and get integrated into genome forming a new gene. C) Exon shuffling in which a new exon is formed by duplication of existing exon or by insertion of exon from a different gene through illegitimate recombination. D) *De novo* origin of gene in which a non-coding region acquire coding potential and regulatory regions to form new functional gene.

Although the above mentioned mechanisms are the major ways of formation of new genes, there are also other mechanisms of new gene formation namely, transposon exaptation, Gene fusion/fission, overprinting and horizontal gene transfer. Gene formation through transposon exaptation results from co-opting or exaptation of transposable element into ORFs. Several cases of this mechanism have been observed in mammals (Cordaux et al., 2006; Krull et al., 2005; Zemojtel et al., 2007). In the *Arabidopsis thaliana* genome 7.8% of all the expressed genes are estimated to contain exapted DNA from transposable elements (Lockton and Gaut, 2009). The importance of transposons in plant evolution has recently been extensively reviewed (Lisch, 2013).

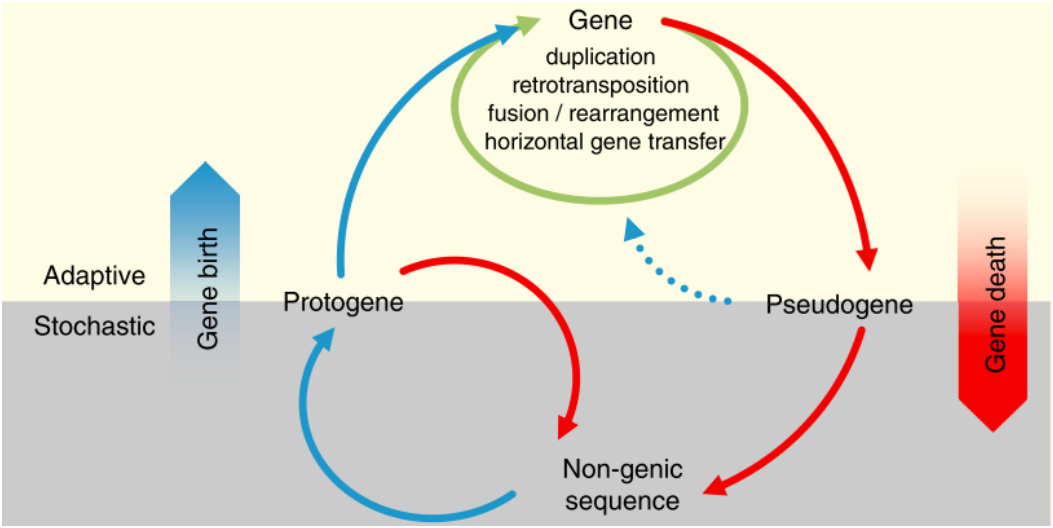


Figure 2: Schematic representation of life cycle of genes.
Adapted from Neme and Tautz, 2014. Blue arrows represent transitions which lead, either partially or completely, to a new gene, and are therefore dubbed processes of gene birth. Red arrows represent the loss of features which result in the degradation of the genic potential of a sequence. Green arrows represent the processes which increase the gene repertoire from existing genes. Raw material for genes is stochastically generated as protogenes, entities that have gene-like properties (i.e. stable expression or translation), but may still lack a proper function. Once a protogene is able to perform a function that has an adaptive advantage, it will become fixed in a population. Gene loss through pseudogenization can lead to the death of a gene in a lineage, when the selective pressure upon it is released. This will also be the case for protogenes which have not fully developed into genes.

Overprinting refers to the formation of a gene within an existing gene. Approximately 30% of microbial genes are formed through overprinting which is likely selected for to reduce genome size (Johnson and Chisholm, 2004).

Finally, horizontal gene transfer is the exchange of genetic material between closely or distantly related species and has played a major role in evolution of bacterial lineages (Ochman, 2001). From the growing number of studies on novel gene evolution and mechanisms, it is clear that new genes can originate either from re-arrangement of pre-existing genes or from non-genic sequences. Based on the studies so far, a general life cycle of a gene is given by Neme and Tautz, 2014 (Fig. 2).

1.3 Orphan genes and evolution of genetic novelty

Orphan genes are defined as protein coding genes with no significant sequence similarity to any other proteins/peptides in the databases (Donoghue et al., 2011; Fischer and Eisenberg, 1999; Wilson et al., 2005). Orphan genes can also include paralogous families of orphans within a species or they can be found within a specific taxonomic group or lineage and can also be called taxonomically restricted genes (TRGs) (Wilson et al., 2005). Although orphan genes were initially thought to be an artefact of the limited number of available sequence datasets, the total number of orphans has continued to increase as more sequence data became available. Orphan genes represent up to one-third of the genes in all genomes including archaea, bacteria and viruses (Domazet-Loso and Tautz, 2003). The term orphan gene was first used by Bernard Dujon in a review article on yeast genome sequencing (Dujon, 1996) to reflect a surprising aspect of the yeast genome which was the occurrence of large numbers of genes (about 26% of the genome) with no homologs in other species. He introduced the term ‘orphans’ for such genes (Neme and Tautz, 2014). A synonymous term ORFans (Orphan Open Reading Frames) is also used in the microbial literature (Fischer and Eisenberg, 1999). One study has suggested that orphan genes comprise about 10-20% of all the genes in a genome based on a study of 30 published genomes (Khalturin et al., 2009). Since orphan genes represent a substantial fraction of every genome, it is important to realise that total number of orphans across all lineages by far exceeds the known number of gene families (Tautz and Domazet-Loso, 2011).

Although orphan gene definition was extended to include ortholgous orphans or taxonomically restricted orphans (also called as TRGs), recent literature defines orphan genes as species specific genes and a subset of TRGs (Arendsee et al., 2014). Since orphan genes are defined based on sequence comparisons, the percentage estimates of orphan genes in various species typically ranges from 5-15% (Arendsee et al., 2014). These variations can be attributable to varying evolutionary distance between focal species and it's nearest sequenced species. The variations sometimes also arise from quality of sequence datasets and the methods used in identification of orphans. Hence depending on the methods and parameters used the number of orphans reported will differ. For example, the number of orphan genes in *A.thaliana* from three independent studies reported to be 958, 1430 and 1324 respectively (Donoghue et al., 2011; Guo, 2013; Lin et al., 2010). However it was also observed that with each new genome sequenced the number of orphans increase nearly linearly while the number of genes common to most species (non-orphans) quickly reaches saturation (Wilson et al., 2005). In this study when referring to orphan genes it will be in reference to taxonomically restricted or lineage specifc orphan genes as defned in our previous study (Donoghue et al., 2011).

While the concept of orphan genes by definition looks simple, the concept is complex operationally. The study of orphan genes is challenging for both bioinformatic and wet-lab based approaches to their functional characterization. The lack of homology to other genes means that homology-based functional classifications are not possible, which renders the majority of homology or evolutionary conservation based bioinformatics approaches redundant (Donoghue et al., 2011). However, the genomic features and context of such Lineage-Specific Genes (LSGs) can provide some preliminary clues regarding possible modes of evolution of LSGs. To date, some general genomic characteristics that have been identified for LSGs include; short length, fewer introns, atypical GC content, and increased evolutionary rates (Domazet-Loso and Tautz, 2003; Toll-Riera et al., 2009; Wilson et al., 2007). The detection method and the reference set of genomes considered for detection of orphan genes can also influence the results.

Another aspect of orphan genes studied to date is how orphan genes originate and acquire genetic functionality. One of the models proposed for the origin of orphan genes is the duplication-divergence model (Domazet-Loso and Tautz, 2003). This model proposes that after duplication, one copy of the gene can diverge by accumulating mutations and acquire a new function through an adaptive phase. The gene can diverge beyond a point at which it can still be recognized by BLAST tools and thus be classified as an orphan gene (Tautz and Domazet-Loso, 2011). Such duplications can occur through recombination mediated events or viruses/ transposon-mediated processes. Earlier studies on evolution of new genes in mammals (McCarrey and Thomas, 1987) and *Drosophila* (Long and Langley, 1993) revealed that newly originated genes specifically expressed in testis. This forms the basis of 'out of the testis hypothesis' which suggests that new genes arise predominantly in testis (particularly spermatocytes and spermatids) which is facilitated by the open chromatin states (mainly due to demethylation of CpG dinucleotides in promoter regions) in testis germ cells (Sassone-Corsi, 2002) allowing the transcription of new genes. Once transcribed, new genes can evolve more efficient promoters and acquire diverse expression pattern eventually acquiring specific functions in other (somatic) tissues (Kaessmann, 2010). This has been supported by many examples in mammals and fruit flies where new genes of all kinds (RNA genes, segmental duplicates, chimeric genes and *de novo* genes) show testis specific expression (for e.g., Heinen et al., 2009; Levine et al., 2006; She et al., 2004).

De novo origination is another important mechanism of formation of orphan genes whose study has gained importance in recent years. A *De novo* evolved gene can form functional protein through an intermediate step as RNA gene. An intermediate RNA gene could eventually acquire a functional ORF to code for a completely new protein. Several studies have now confirmed the *de novo* origin of genes by functional studies in yeast (Cai et al., 2008; Li et al., 2010b), mouse (Heinen et al., 2009), humans (Knowles and McLysaght, 2009; Li et al., 2010a), *Drosophila melanogaster* (Levine et

al., 2006; Zhao et al., 2014), as well as in rice (Xiao et al., 2009) and *Arabidopsis* (Silveira et al., 2013). New genes can also gain function by acquiring different expression patterns which allow them to have a different spatial and temporal expression pattern than their parental gene (Ranz and Parsch, 2012). Although gaining novel expression patterns seems to be particularly important for genes originating through RNA-based duplication mechanisms (Bai et al., 2008), genes originating from any other mechanisms (described in the previous section) can also acquire different expression patterns.

Since genes cannot function alone, any newly arisen gene has to be integrated into the respective functional pathway if it is to be functionally relevant. However, few cases have been reported on how novel genes acquire functions through pathway integration. Matsuno et al. (2009) identified two Brassicaceae orphan genes *CYP98A8* and *CYP98A9* which arose by retroposition and duplication followed by neofunctionalisation via selective and local amino acid replacement leading to a novel phenolic pathway in Brassicaceae. This was the first study showing how origination of new genes could lead to the invention of a new metabolic pathway. Another study in yeast by Li et al. (2010b) demonstrated how a newly originated *de novo* gene integrated into canonical mating pathway and facilitated the organism's adaptation. Two other case studies in *Drosophila melanogaster* show that recently originated genes can acquire novel functions and gene networks different from those of their ancestral copy (Chen et al., 2012; Ding et al., 2010).

1.3.1 Do some orphan genes have function?

Apart from identification and origin, determining the functionality of orphan genes is another challenge. The usual strategy of homology based inferences for orphan genes is not possible or very limited because of the inherent lack of homology of these proteins to any known domains. Another approach is to find a bias in expression patterns. Since functionally relevant

orphan genes are considered to function in lineage specific adaptations, it could be expected that their expression and function would be related to specific developmental stages. In the *Drosophila* lineage, a number of orphan genes found to be X-linked and exhibit testis biased expression (Begun et al., 2007; Levine et al., 2006). Orphans are also known to be differentially expressed under different stress conditions as shown in our previous study in *Arabidopsis thaliana* (Donoghue et al., 2011), and also in rice where orphan genes show increased expression in either hormone treated or injured tissues compared to normal controls (Guo et al., 2007). A recent study in zebrafish shows that orphan genes have highly tissue-preferred expression and are also highly temporally restricted (Yang et al., 2013). Moreover they also find that orphan genes in zebrafish are specifically enriched for early stage embryos and early larval stages. In *Drosophila* species, Chen et al. (2010), using RNA interference (RNAi), systematically studied newly evolved genes (evolved within 35 million years) and found that 30% of them had lethal effects when knocked down. Although most of the genes were not orphans, two out of the 16 *de novo* originated genes investigated showed lethal phenotypes and one showed a semi-lethal phenotype. This work shows how some newly evolved genes can quickly become essential.

Some orphan genes may acquire functions that facilitate lineage specific adaptation to the environment (Tautz and Domazet-Loso, 2011). In *Hydra* species, 35% of genes up–regulated in response to *Pseudomonas aeruginosa* are identified as ORFans. In particular, Periculin-1 has been identified as novel bactericidal *Hydra* peptide with a role in host defence (Bosch et al., 2009). A species specific trait related to feeding in *Hydra sp.* is also controlled by a family of orphan genes (Khalturin et al., 2008). A family of orphan genes encoding surface antigens are involved in host parasite interactions in the protozoa *Plasmodium* (Kuo and Kissinger, 2008). In plants, a *de novo* originated gene in rice negatively regulates resistance to bacterial blight (Xiao et al., 2009). Another study in rice, an O*ryza* specific gene *Ehd4* has been shown to be involved in photoperiodic control of flowering time (Gao et al., 2013). All of these studies support the prediction

that (where they occur) orphan gene functions are often related to lineage specific adaptation, and more supporting examples are being added by the fast growing number of functional studies.

Orphan genes constitute a part of the 'dispensable genome' in the 'pan-genome concept'. Pan-genome was initially described for bacterial species (Tettelin et al., 2005) and the concept has been extended to plants (Hirsch et al., 2014; Morgante et al., 2007) and human genomes (Li et al., 2010c). Pan-genome includes a core genome containing genes present in all strains and a dispensable genome containing strain-specific (orphan/*de novo* genes) and partially shared DNA sequences (Morgante et al., 2007). Although the dispensable portion of the genome, by definition, is non-essential, this definition needs to be redefined considering the growing evidence of orphan gene functionality in many species. Moreover a large part of the dispensable genome consists of structural variants or copy number variants which also contribute towards determining plant phenotypes and shaping genome evolution (Marroni et al., 2014). However, genome and transcriptome sequencing of additional strains of a population will further help to redefine orphan genes as well as the concept of pan-genome.

1.4 Defining Brassicaceae specific orphan genes in *Arabidopsis thaliana*

Despite several genomes having been sequenced, orphan gene studies in plant species have been limited until recently. Yang et al. (2009) computationally identified species specific genes in *Arabidopsis, Oryza* and *Populus* genomes wherein they found that orphan genes show tissue specific expression patterns and are also upregulated by stress. Lin et al. (2010) also identified lineage-specific genes in Brassicaceae and categorised them as conserved Brassicaceae specific genes (CBSGs) and Arabidopsis lineage specific genes (ALSGs). They found that CBSGs and ALSGs are highly methylated in floral tissues, and that both CBSGs and ALSGs are fast evolving at the protein level. Our lab's previous study describes the genome wide analysis of Brassicaceae specific orphan genes

(referred as Lineage specific genes/LSGs in the study) in *Arabidopsis thaliana* along with a comprehensive elucidation of the mechanisms of origin for the majority of the orphan genes in the *Arabidopsis thaliana* genome (Donoghue et al., 2011). Our study also employed position-specific methods that can detect weaker homologous relationships that would otherwise be missed by the standard BLAST algorithms. Moreover mitochondrial and chloroplast genomes were also screened by our lab for orphan genes which were not considered in the previous two studies. Our study identified 1789 genes (of which 28 are mitochondrial) which are Brassicaceae specific that also contain a subset of 958 genes which are *Arabidopsis thaliana* specific. These numbers slightly differed from the previous studies of *Arabidopsis thaliana* LSGs which can be attributable to the databases searched and the additional use (in our study) of position specific methods. All of the 1789 orphan genes were examined in detail for their evolutionary mechanisms of origin. Four different evolutionary scenarios were investigated to elucidate the mechanisms of origin of these orphan genes, namely:

1) Duplication: By identifying a significant hit to an evolutionarily conserved gene the possible origin of an orphan gene can be traced to that conserved "parental" gene. In total, 417 orphan genes (22.25%) were found to have originated from duplication mechanisms based on protein/CDS hits to a non-orphan gene (Fig.3a).

2) Overprinting: Overprinting describes a mechanism that can create new genes via mutations occurring within a coding sequence that lead to the expression of a novel protein in another reading frame that overlaps the existing parental gene. 21 orphan genes (1.18%) were found to have overlapping gene models with non-orphan genes suggesting that they arose via an overprinting mechanism (Fig.3f).

3) Transposon exaptation: This refers to a process whereby transposons get inserted into functional loci leading to formation of new genes. 175 orphan genes (constitute 9.75%) were found to have originated by transposable elements in *A. thaliana* (Fig.3c).

4) *De novo* origin: The identification of *de novo* genes involved identifying nucleotide sequence matches (either non-coding or in a different reading frame) to non-Brassicaceae plant genomes for Brassicaceae specific orphans and to the *A. lyrata* genome for *A. thaliana* specific orphan genes. Of all the orphan genes analysed, 15 orphan genes (0.84%) displayed a significant match to non-Brassicaceae out-of-frame CDS sequence (Fig.3g) and 25 (1.4%) orphans displayed a significant match to non-Brassicaceae intergenic DNA (Fig.3e). Of the 958 *A*. *thaliana* orphan genes, 111 (6.2%) have out-of-frame hits to *A. lyrata* (Fig.3d) and 423 (23.65%) have hits to non-coding sequences (Fig.3b). While it is not clear how many of these differences represent gene birth in *A. thaliana* and how many represent gene loss in *A. lyrata* lineage, for two of the loci we provided evidence of gene birth in *A. thaliana* through conversion of noncoding sequences (Donoghue et al., 2011).

1.4.1 Genomic features of *Arabidopsis thaliana* orphan genes

Based on several studies of orphan genes in different genomes, there are some general trends in the characteristic features of orphan genes. Young genes are known to have high divergence rates (Alba and Castresana, 2005; Cai et al., 2006). *Arabidopsis* orphan genes were also found to be fast evolving which was consistent with the previous study by Lin et al. (2010). Orphan genes also have low expression levels and show a high extent of tissue specific expression. Older genes tend to encode longer proteins compared to new genes (Lipman et al., 2002). Orphan genes in *Arabidopsis* encode for shorter peptides (Donoghue et al., 2011; Lin et al., 2010). Also fewer introns, low GC content, fewer paralogs and enrichment for secretary peptides (Lin et al., 2010) were found to be general features of orphan genes. In addition we found that *A. thaliana* orphan genes are enriched for defensin like genes and other cysteine-rich peptides (Donoghue et al., 2011).

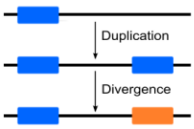
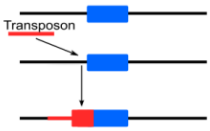
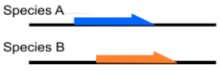
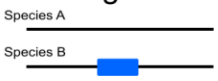
	Number of ORFans expression support loci (gene model)
a. Duplication 	417 (22.247%) 270 (180)
b. Intergenic hit in *A.lyrata* (*A.thaliana* specific) 	423 (23.645%) 86 (53)
c. Transposon Exaptation 	175 (9.782%) 115 (57)
d. Out-of-frame CDS hit in *A.lyrata* (*A.thaliana* specific) 	111 (6.204%) 63 (41)
e. Intergenic hit non-Brassicaceae 	25 (1.397%) 15 (12)
f. Overlapping gene models 	21 (1.179%) 19 (14)
g. Out-of-frame CDS hit non-Brassicaceae 	15 (0.838%) 10 (9)
h. Unknown origins Brassicaceae specific ?	478 (26.719%) 323 (273)
i. Unknown origins *A.thaliana* specific ?	178 (9.950%) 57 (30)

Figure 3: Summary of evidence for evolutionary origins of *Arabidopsis thaliana* lineage-specific genes. The number of orphan genes that fit each evolutionary scenario analysed, plus the number of LSGs without elucidated origins. Support for gene model expression provided by an EST or cDNA consistent with that of gene model (as listed by TAIR v8). Support of expression at the locus provided by EST, cDNA or microarray probeset (TAIR and ATH1 affymetrix microarray). Figure adapted from Donoghue *et al*. (2011).

The earlier predictions that orphan gene functions relate to lineage specific adaptations to the environment have also now been supported by a growing number of case studies (explained in section 1.3.1). In our previous study, enrichment for stress-responsive differentially expressed orphans was identified using the AtGenExpress expression data series for stress, pathogen infection, growth condition treatment, chemical treatments and hormone treatments (Donoghue et al., 2011). The stress responsiveness of orphan genes was also suggested in previous studies as explained earlier in section 1.3.1 (Khalturin et al., 2009; Xiao et al., 2009). In our previous study, out of 497 orphan genes (which had unique probe sets in ATH1 microarray) analysed, a total of 137 orphan genes were found to be responsive to various stress conditions (Donoghue et al., 2011) where in a total 130 orphan genes were up-regulated and 103 down-regulated with an intersection of 96 (across all treatments). The total number of differentially expressed orphan and non-orphan genes for each individual treatment, tissue and time point was found and hypergeometric tests performed (p-value < 0.05) to identify if orphan genes were enriched in either the significantly up-regulated or down-regulated genes. It was found that orphan genes are enriched for abiotic stress responsiveness wherein out of 137 stress responsive genes, 106 genes were responsive to one or more abiotic stress conditions (Donoghue et al., 2011). Nine different abiotic stress conditions were investigated namely, wounding, UV-B, salt, oxidative, osmotic, heat, genotoxic, drought and cold stress treatments. The summary of differential expression of the Arabidopsis orphan genes for abiotic stress treatments is shown in Figure 4 (Data from Donoghue et al. (2011)).

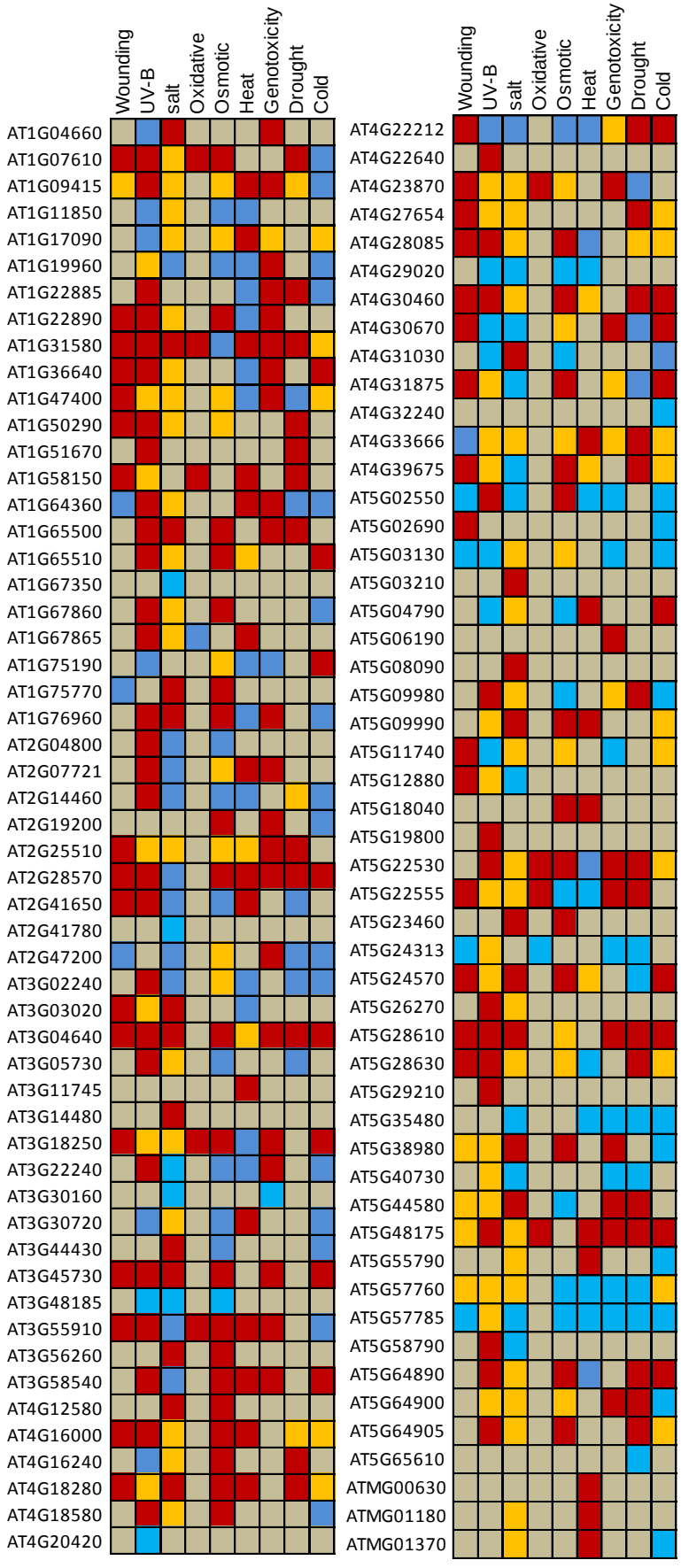


Figure 4: Summary of differential expression of the abiotic stress responsive orphan genes. Differential expression of abiotic stress responsive orphan genes for nine different abiotic stresses tested is shown. Significant up or down regulation are mentioned in colours red and blue respectively. Up and down regulated (in orange colour) category means the gene is up or down regulated at different time points. Data from Donoghue et al. (2011).

1.4.2 *Arabidopsis thaliana* as a model organism for functional study of plant orphan genes

Since the pioneering studies in the 1960s, *A. thaliana* has now emerged as the most popular plant model organism and is the most intensively studied plant on the planet. The enormous amount of resources for this plant (genomics/ transcriptomics/proteomics etc.) make it ideal for studying orphan genes. *A. thaliana* is a small mustard weed commonly known as thale cress belonging to the family Brassicaceae and it has been used in experimental research from more than half a century (The Arabidopsis InformationResource). In 1943, Friedrich Laibach for the first time proposed the suitability of *Arabidopsis* as a model organism for genetic and developmental studies in plants (Somerville and Koornneef, 2002). The major features for suitability in experimental research being that it was easy to cultivate in restricted space, had prolific seed production and developed rapidly, had abundant natural variation, produced fertile hybrids and had a relatively low chromosome number (Laibach, 1943). The availability of genome information and development of powerful molecular biology methods provided a great opportunity for discoveries and during the past quarter of a century *Arabidopsis* has fundamentally changed plant biology (Somerville, 2000). Now *Arabidopsis* is the most thoroughly studied flowering plant (Koornneef and Meinke, 2010; Resource) and is popularly known as the 'Botanical Drosophila' (Leonelli, 2007). *A. thaliana* has well annotated genome, extensive publicly available data and also a good number of well characterised genes which will help to characterise the functionality of orphan genes. Apart from that, *Arabidopsis* also has large populations of mutagenized lines, cDNA and ORF clones which are publicly available through stock centres. These resources can be a great aid in identifying and understanding the orphan gene functions.

Plant genomes are dynamic and genetic novelty can arise via processes such as gene duplication and polyploidization (Osborn et al., 2003). It is well known that polyploidization has played a significant role in evolution of plant genomes. *A. thaliana* shows evidence of three ancient polyploidy events (Fig 5). Polyploidization can introduce genome novelty

through rapid genetic and epigenetic changes and dosage-regulated gene expression changes (Soltis et al., 2009). Duplication events (whole genome duplication/polyploidy or local duplication events), which are known to be frequent in plant genomes, are considered as a major source for genome evolution leading to genomic novelty. Nearly 35% of the vascular plants are known to be recent polyploids (Wood et al., 2009) and in *Arabidopsis* about 25% of the genes are the product of whole genome duplications (Blanc et al., 2003) and nearly 16% are tandem duplicates (Rizzon et al., 2006). Susumu Ohno had proposed that gene duplication was the single most important factor to play a major role in evolution (Ohno, 1970). After duplication one copy of the duplicated gene can acquire beneficial mutations overtime and result in evolution of novel function. This became the basis of duplication and divergence model of new gene origination. Now it has been well established that gene duplication is an important source of raw material for evolution of novel genes.

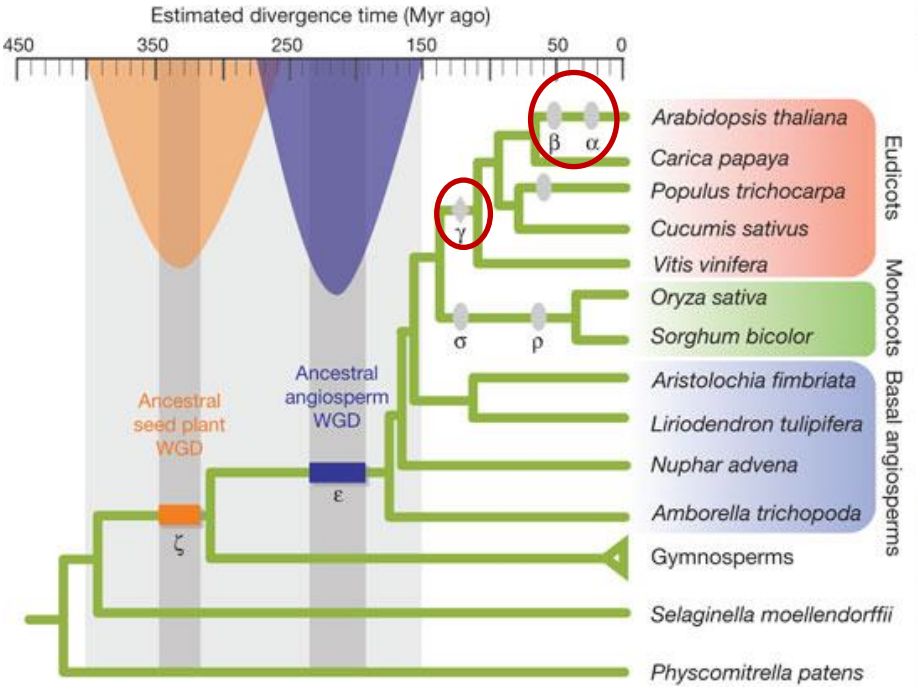


Figure 5: Polyploidy events in the evolution of land plants. Three polyploidy events in the evolution of eudicots are shown in red circles. Two ancestral whole genome duplication events in seed plants and angiosperms are also shown. Figure adapted from Jiao et al. (2011).

Apart from a few computational studies (Donoghue et al., 2011; Lin et al., 2010; Yang et al., 2009) in plants, there have been no systematic genome-wide studies on plant orphan genes to date. However, there have been a few functional studies on individual orphan genes. Li et al. (2009) identified the *A. thaliana* gene QQS (for Qua-Quine Starch) and showed that it is involved in regulation of starch biosynthesis in *Arabidopsis* leaves. Another study on the same gene has shown that it is under tight epigenetic control and differentially expressed across natural *Arabidopsis* populations (Silveira et al., 2013). Xiao et al. (2009) has shown that a *de novo* originated rice-tribe specific gene *OsDR10* negatively regulates pathogen induced defense response pathway. This study also suggests that novel genes can be quickly recruited to new biological functions. Another study in rice found a novel *Oryza* specific gene *Ehd4* is involved in photoperiodic regulation of flowering time in rice (Gao et al., 2013).

In our lab's previous study we showed that orphan genes in *A. thaliana* are enriched for abiotic stress responses (Donoghue et al., 2011). Based on this we hypothesised that the abiotic stress responsiveness of a subset of orphan genes could be indicative of such genes being functionally involved in abiotic stress tolerance in the Brassicaceae lineage. This became the basis for this PhD thesis work, to study the functionality of the Brassicaceae specific orphan genes in *A. thaliana* in relation to abiotic stress responses. To test our hypothesis, we used a range of genetic and molecular tools to investigate the following scientific objectives.

1) To functionally determine the abiotic stress relevance of the abiotic stress responsive orphan genes by reverse genetic approach (loss of function screen) in *A. thaliana*. This involved screening of T-DNA insertion lines for orphan genes for different abiotic stress conditions to identify which orphan genes had functional roles in abiotic stress tolerance.

2) To functionally determine the abiotic stress relevance of the abiotic stress responsive orphan genes in *A. thaliana* using a gain of function approach. This involved cloning and expression of orphan gene ORFs under a constitutive promoter (*CaMV 35S* promoter). The transgenic overexpression lines are then screened for abiotic stress phenotypes.

3) To functionally determine the abiotic stress relevance of the Arabidopsis abiotic stress responsive orphan gene through a gain of function screen in yeast (*Saccharomyces cerevisiae*) using abiotic stress-sensitive yeast strains. This involved screening of the orphan genes in yeast heterologous system to determine the possible functionality of *Arabidopsis* orphan genes in the context of abiotic stress. To fulfil this objective, *Arabidopsis* abiotic stress responsive orphan genes were cloned and expressed under a yeast constitutive promoter (GPD promoter) in stress sensitive yeast strains and subjected to different abiotic stress conditions to screen for abiotic stress functionality.

In this PhD thesis, I investigated the orphan genes which were shown to be responsive to abiotic stresses (Fig.4) in Donoghue et al. (2011). The orphan genes were re-analysed with recent datasets to make sure that they are still Brassicaceae specific. Since the orphan gene definition is based on sequence homology, sequencing of more and more related species and outgroup species makes the identification of orphan genes more robust and also increases the power to elucidate the evolutionary origins of orphan genes. Hence it is important to continually redefine the orphan gene set whenever a related species or a close out group species sequence data is available. This will help to re-confirm the lineage specificity of the orphan genes in question. Since the initiation of this PhD study, several *Arabidopsis* close relatives have been sequenced, including recently sequenced cleome species (*Tarenaya hassleriana*), the closest outgroup of Brassicaceae. From the reanalysis of the 109 genes under study in this PhD thesis, 12 genes

were found to be not Brassicaceae specific any more, based on BLAST (BLASTP and TBLASTN) sequence analysis (with e-value cut-off of 0.001). The complete list of orphan genes investigated in this work is given in Table 1 and non-orphans in Table 2, although all these genes were screened together. Each gene was given a code name starting with the prefix ASR (for Abiotic Stress Responsive) followed by number.

The following PhD chapters describe the scientific effort to investigate the functionality of the orphan genes in *Arabidopsis thaliana*. The more in depth characterization of one of the most promising candidate ASR orphan genes is explained in the final two chapters. Overall this PhD aims to advance our general understanding of the functions of plant orphan genes.

Table 1: List of abiotic stress responsive Brassicaceae specific genes used in the study

Gene name	Gene ID		Gene name	Gene ID		Gene name	Gene ID
ASR1	AT1G22890		*ASR38*	AT1G58150		*ASR75*	AT5G09980
ASR2	AT1G47400		*ASR39*	AT1G64360		*ASR76*	AT5G09990
ASR3	AT1G51670		*ASR40*	AT1G65510		*ASR78*	AT5G12880
ASR4	AT1G65500		*ASR41*	AT1G67860		*ASR79*	AT5G19800
ASR5	AT3G03020		*ASR42*	AT1G67865		*ASR80*	AT5G22555
ASR6	AT3G04640		*ASR43*	AT1G75190		*ASR81*	AT5G24313
ASR7	AT3G11745		*ASR44*	AT1G75770		*ASR82*	AT5G24570
ASR8	AT3G14480		*ASR45*	AT1G76960		*ASR83*	AT5G26270
ASR11	AT3G45730		*ASR46*	AT2G04800		*ASR84*	AT5G28610
ASR13	AT4G12580		*ASR47*	AT2G07721		*ASR85*	AT5G28630
ASR14	AT4G18280		*ASR48*	AT2G14460		*ASR86*	AT5G38980
ASR15	AT4G22640		*ASR49*	AT2G19200		*ASR87*	AT5G40730
ASR16	AT4G28085		*ASR50*	AT2G25510		*ASR88*	AT5G44580
ASR17	AT4G30460		*ASR52*	AT2G41650		*ASR89*	AT5G48175
ASR18	AT4G33666		*ASR53*	AT2G47200		*ASR90*	AT5G55790
ASR19	AT5G03210		*ASR54*	AT3G02240		*ASR92*	AT5G57785
ASR20	AT5G06190		*ASR55*	AT3G05730		*ASR94*	AT5G64900
ASR21	AT5G08090		*ASR57*	AT3G30720		*ASR95*	ATMG01180
ASR22	AT5G18040		*ASR58*	AT3G44430		*ASR96*	ATMG01370
ASR23	AT5G22530		*ASR59*	AT3G55910		*ASR97*	AT1G19500
ASR24	AT5G23460		*ASR60*	AT3G58540		*ASR98*	AT1G25097
ASR25	AT5G64890		*ASR61*	AT4G16000		*ASR100*	AT2G33010
ASR26	AT5G64905		*ASR62*	AT4G16240		*ASR101*	AT2g41280
ASR28	AT1G04660		*ASR63*	AT4G18580		*ASR102*	AT3G30160
ASR29	AT1G07610		*ASR64*	AT4G22212		*ASR103*	AT3G48185
ASR30	AT1G09415		*ASR65*	AT4G23870		*ASR104*	AT4G20420
ASR31	AT1G11850		*ASR66*	AT4G27654		*ASR105*	AT4G32240
ASR32	AT1G17090		*ASR67*	AT4G30670		*ASR106*	AT5G29210
ASR33	AT1G19960		*ASR68*	AT4G31030			
ASR34	AT1G22885		*ASR69*	AT4G31875			
ASR35	AT1G31580		*ASR71*	AT5G02550			
ASR36	AT1G36640		*ASR72*	AT5G02690			
ASR37	AT1G50290		*ASR73*	AT5G03130			

Table 2: List of abiotic stress responsive genes used in the study that are not Brassicaceae specific (as of 2014)

Gene name	Gene ID
ASR9	AT3G18250
ASR10	AT3G22231
ASR12	AT3G56260
ASR27	ATMG00630
ASR51	AT2G28570
ASR56	AT3G22240
ASR70	AT4G39675
ASR74	AT5G04790
ASR77	AT5G11740
ASR91	AT5G57760
ASR93	AT5G58790
ASR99	AT1G67350

Chapter 2: Loss of function screen of Brassicaceae specific orphan genes in *Arabidopsis thaliana* for abiotic stress responses

2.1 Introduction

2.1.1 Analysis of gene functions in *Arabidopsis thaliana*

The possibility of transforming *Arabidopsis* with *Agrobacterium tumefaciens* (Clough and Bent, 1998; Feldmann and Marks, 1987) made it more popular for genetic studies which enabled production of *Arabidopsis* mutants with defects in phenotypic growth (Leonelli, 2007). This led to the development of transfer DNA (T-DNA) insertional mutant lines which greatly helped for genome wide screening of genes for identifying gene functions. With the sequencing of the genome in the year 2000 by Arabidopsis Genome Initiative (AGI), *Arabidopsis* obtained a pre-eminent status in plant biology. Being a small genome of about 125 Mb, *Arabidopsis thaliana* contains 27,416 protein coding genes, 4827 pseudogenes or transposable element genes and 1359 ncRNAs (33,602 genes in all, 41,666 gene models) according to the latest version of 'The Arabidopsis Information Resource (TAIR)', TAIR10.

After the genome sequencing of *Arabidopsis thaliana*, the major focus became the identification of all the genes and to find the functions of the genes. Although about 10% of 27,000 were already characterized by that time, identifying the functions of remaining 90% of the genes was a big challenge (Alonso and Ecker, 2006). To achieve this goal, two major approaches were followed: (1) Forward genetic screens and (2) Reverse genetic screens (Page and Grossniklaus, 2002). Forward genetics starts with identification of the mutant phenotype and then finding the corresponding gene whereas reverse genetics starts with the gene of interest and then leading to the corresponding mutant (Fig.6). Although there are several methodologies in each approach (reviewed in Alonso and Ecker, 2006), most efforts were focussed on generating loss-of-function mutant lines for reverse genetic screens. However methodologies for gain-of-function screens by overexpressing the genes have also gained importance (Bolle et al., 2011). Since the focus of this chapter is on loss-of-function screen, only methods for

loss-of function screen are described here and the gain of function approach is described in detail in chapter 3.

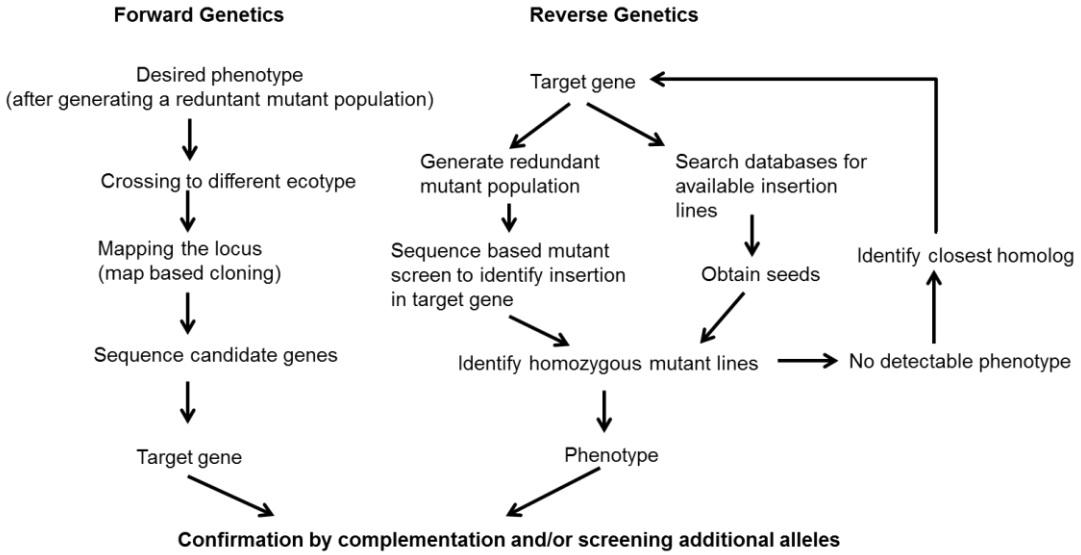


Figure 6: Overview of Forward and Reverse Genetics approaches.

Individual steps in forward genetics (left) and reverse genetics (right) approach are mentioned. Figure modified from (Ostergaard and Yanofsky, 2004).

2.1.3 Methods for loss-of-function (or knock out) approach

This is a straightforward approach comprising of set of methods designed to interfere with the normal activity of the gene thereby abolishing its function (Bouche and Bouchez, 2001). This is usually done by introducing point mutations or short insertion/deletions, typically by physical or chemical mutagen or by inserting larger DNA sequences like T-DNA or transposable elements through transformation methods (Bolle et al., 2011).

2.1.3.1 Loss-of function by point mutation and short insertion/deletion

This approach uses chemical (e.g. ethyl methane sulfonate) or physical mutagens (e.g. fast neutrons) to introduce point mutations (and/or insertion/deletion) in the genome. The commonly used chemical mutagen ethyl methane sulfonate (EMS) typically introduces point mutations in high frequencies in the genome. Although most mutations are expected to be

recessive, dominant mutations can be obtained through EMS mutagenesis at a low frequency (McConnell et al., 2001). It is possible to find a mutation in any given gene by screening fewer than 5000 plants in the M1 generation (Feldmann et al., 1994; Greene et al., 2003). However to obtain a total loss of function in a gene a large number of plants need to be screened. Since there is a high probability of unwanted background mutations with chemical or radiation based mutagenesis, several rounds of backcrossing to the wild type are required to recover a line which has only the mutations in the gene of interest.

The development of TILLING (Targeting Induced Local Lesions IN Genomes) technique helped the use of EMS mutagenesis in reverse genetics approach (McCallum et al., 2000). TILLING mainly helps in identifying the mutation in the EMS mutant populations with the use of a specific nuclease called CEL1 which cleaves heteroduplex formed by denaturing and renaturing of the amplified target gene. Since PCR amplification uses primers with fluorescent probes (both forward and reverse primer with different probes), the cleaved fragments are easily detected on the denaturing acrylamide gels. Although TILLING is efficient in identifying the point mutation in the target gene, it is still labour intensive and hence a secondary choice for reverse genetic screen when there is availability of sequence-indexed mutant libraries for the target species (Alonso and Ecker, 2006).

2.1.3.2 Loss-of-function by insertional mutagenesis

This is one of the widely used methodologies for functional characterization of genes for both forward genetics and reverse genetics approaches. Apart from *Arabidopsis* (Alonso et al., 2003), insertional mutagenesis has been successfully used in many species including bacteria (Maguin et al., 1996), *Drosophila* (Cooley et al., 1988), Zebra fish (Golling et al., 2002) and rice (Jeon et al., 2000). Insertional mutagenesis in *Arabidopsis* is usually performed using transposons (Long et al., 1993) or T-DNA (Alonso et al., 2003) as a mutagen which gets inserted in the genome disrupting the

gene in which it is inserted into (Parinov and Sundaresan, 2000). Unlike point mutations from EMS mutagenesis, T-DNA insertions in the coding region are more likely to completely abolish gene function. Another advantage with insertional mutagenesis is that, the insertion site can be easily determined as the inserted sequence is already known.

Apart from the above methods, in the past few years new methodologies and tools are emerging for targeted mutagenesis. Use of engineered nucleases like zinc finger nucleases (ZFNs) and Transcription Activator-like Effector Nucleases (TALENs) have been shown to be effective in *Arabidopsis* (Joung and Sander, 2013; Zhang et al., 2010). Recently another technique called CRISPR/cas system (Clustered Regulatory Interspaced Short Palindromic Repeats) was also demonstrated to be efficient for targeted genome editing in *Arabidopsis* and other plant species like Rice and wheat (Feng et al., 2013; Shan et al., 2013).

2.1.3.3 Loss-of-function by gene silencing (knockdown) approach

The usefulness of the loss of function approach by insertional mutagenesis is limited when the null mutation causes lethality or when there is genetic redundancy. Transgene mediated gene silencing approaches can help to overcome these problems by reducing the expression of gene/s. Silencing is usually achieved by post transcriptional down regulation of transcripts by small RNAs that act in a sequence-specific manner by binding to complementary RNA (Bolle et al., 2011). Although several approaches of gene silencing by small RNAs have been described (Ossowski et al., 2008), gene silencing by RNA interference approach is widely used. This uses a binary vector with hairpin RNA into which gene sequence tags (GSTs) can be cloned (Chuang and Meyerowitz, 2000). A collection of silenced *Arabidopsis* lines were generated by AGRIKOLA (Arabidopsis Genomic RNAi Knock-Out Line Analysis) consortium by systematically transforming RNAi vectors into *Arabidopsis* (Hilson et al., 2004). These lines are made available to researchers through Nottingham Arabidopsis Stock Centre (NASC). Another popular approach is through artificial microRNAs which

exploits endogenous miRNA precursors to generate sRNAs that direct gene silencing (Alvarez et al., 2006).

2.1.4 Resources for loss-of-function screen by insertional mutagenesis in *Arabidopsis thaliana*

Because of the ease of transformation of *Arabidopsis,* it is easy to generate large number of insertional mutant lines. Because the insertion sequence (tag) is known, DNA flanking the insertion can be easily identified by PCR amplification (Liu and Whittier, 1995) and sequencing. Large numbers of insertional mutant lines were developed from various *Arabidopsis* research groups all over the world. These lines are available to researchers through stock centres. The Arabidopsis Biological Resource Centre (ABRC - https://abrc.osu.edu/) and the European Arabidopsis Stock Centre (uNASC- http://arabidopsis.info/) are the major stock centres from which all the resource materials for *Arabidopsis* research can be obtained. The resources available at these centres are generated by the *Arabidopsis* research community and made available to *Arabidopsis* researchers. Other centres that distribute mutant lines include the Flanking Sequence Tags (FST) Project at Versailles (http://publiclines.versailles.inra.fr/) and RIKEN bioresource centre, Japan (http://www.brc.riken.jp/lab/epd/Eng/) which maintains separate databases for maintaining and distributing *Arabidopsis* resource materials. The insertional mutant lines used in this study are obtained from the above sources.

Our previous study identified an enrichment of orphan genes for abiotic stress responses in *Arabidopsis thaliana* (Donoghue et al., 2011). We hypothesised that some orphan genes in *Arabidopsis* could have lineage specific roles in abiotic stress responses. This chapter describes the testing of this hypothesis using a loss-of-function approach in *Arabidopsis thaliana.* T-DNA and transposon insertion lines for abiotic stress responsive orphan genes were obtained from stock centres (mentioned in material and methods) and used for screening mutants under different abiotic stress conditions. The complete list of genes and the mutant lines used are given in

the Table-3**.** The main objective of this chapter was to identify and evaluate any abiotic stress responsive phenotypes for orphan gene mutants by a comprehensive loss of function screen using *Arabidopsis thaliana* insertional mutant lines.

2.3 Results

2.3.1 Selection of *Arabidopsis* orphan genes for abiotic stress screens by loss-of-function approach

In our previous study Brassicaceae specific orphan genes were found to be enriched for abiotic stress responses in *Arabidopsis* (Donoghue et al., 2011). For the analysis, 497 orphan genes which had unique probe sets on ATH1 affymetrix microarray were analysed for abiotic stress treatments that included cold, drought, genotoxic, heat, osmotic, oxidative, salt, UV-B and wounding. A total of 96 orphans were upregulated, 90 down-regulated with an intersection of 80 across all treatments. To investigate these genes for abiotic stress responses, we obtained available insertional mutant lines for these genes. We could obtain insertion lines for 79 of the abiotic stress responsive orphan genes of which 66 genes were down regulated and 13 genes were upregulated under abiotic stress conditions. The complete list of genes and the corresponding insertional mutant lines analysed are given in the Table 2. The genes were named with numbers with prefix ASR for 'Abiotic Stress Responsive'. Respective wild type controls were used for each insertion line, for example Col-0 for SALK and SAIL lines, Noessen for RIKEN lines and Wassilewskija for FLAG lines.

Table 3 : List of abiotic stress responsive orphan genes and their insertion lines used in this study.

Allele name	Gene ID	Insertion line ID	Accession
asr1-1	AT1G22890	SALK_062788	Col-0
asr1-2	AT1G22890	pst17863	Noessen
asr4-1	AT1G65500	psh13246	Noessen
asr5-1	AT3G03020	SALK_067389	Col-0
asr5-2	AT3G03020	pst04492	Noessen
asr6-1	AT3G04640	SALK_087773	Col-0
asr7-1	AT3G11745	SM_3_17126	Col-0
asr8-1	AT3G14480	FLAG_261C03	WS
asr9-1	AT3G18250	SALK_052341	Col-0
asr12-1	AT3G56260	SAIL_586_F09	Col-0
asr13-1	AT4G12580	SAIL_1264_F11	Col-0
asr14-1	AT4G18280	SALK_112442	Col-0
asr15-1	AT4G22640	FLAG_199B04	WS
asr16-1	AT4G28085	SALK_065775	Col-0
asr17-1	AT4G30460	SALK_140721	Col-0
asr17-2	AT4G30460	SALK_101746	Col-0
asr18-1	AT4G33666	pst15749	Noessen
asr19-1	AT5G03210	pst02006	Noessen
asr19-2	AT5G03210	pst12265	Noessen
asr20-1	AT5G06190	SAIL_139_B12	Col-0
asr21-1	AT5G08090	pst02070	Noessen
asr22-1	AT5G18040	pst02304	Noessen
asr24-1	AT5G23460	SALK_054529	Col-0
asr25-1	AT5G64890	FLAG_020B08	WS
asr25-2	AT5G64890	SALK_136904	Col-0
asr26-1	AT5G64905	FLAG_338B09	WS
asr26-2	AT5G64905	SALK_017813	Col-0
asr28-1	AT1G04660	SALK_011240	Col-0
asr28-2	AT1G04660	psh11293	Noessen
asr28-3	AT1G04660	psh11357	Noessen
asr31-1	AT1G11850	FLAG_235C12	WS
asr31-2	AT1G11850	pst15639	Noessen
asr34-1	AT1G22885	SALK_004703	Col-0
asr35-1	AT1G31580	SALK_088330	Col-0
asr35-2	AT1G31580	pst11525	Noessen
asr37-1	AT1G50290	SALK_041815	Col-0
asr37-2	AT1G50290	SALK_041799	Col-0
asr39-1	AT1G64360	SALK_063861	Col-0
asr39-2	AT1G64360	SALK_052233	Col-0
asr41-1	AT1G67860	SALK_041709	Col-0

Allele name	Gene ID	Insertion line ID	Accession
asr41-2	AT1G67860	SALK_093353	Col-0
asr42-1	AT1G67865	SALK_037677	Col-0
asr44-1	AT1G75770	SALK_072169	Col-0
asr44-2	AT1G75770	SAIL_869_E08	Col-0
asr45-1	AT1G76960	SALK_034899	Col-0
asr45-2	AT1G76960	SALK_098782	Col-0
asr46-1	AT2G04800	SALK_043103	Col-0
asr47-1	AT2G07721	SAIL_412_B02	Col-0
asr48-1	AT2G14460	FLAG_322B08	WS
asr48-2	AT2G14460	SALK_010561	Col-0
asr49-1	AT2G19200	SAIL_599_A12	Col-0
asr50-1	AT2G25510	SALK_132861	Col-0
asr50-2	AT2G25510	pst20121	Noessen
asr52-1	AT2G41650	GT_5_85109	Ler
asr52-2	AT2G41650	GT_5_85308	Ler
asr53-1	AT2G47200	SALK_029461	Col-0
asr53-2	AT2G47200	SM_3_38195	Col-0
asr54-1	AT3G02240	FLAG_035H03	WS
asr55-1	AT3G05730	SALK_031670	Col-0
asr55-2	AT3G05730	SAIL_742_B09	Col-0
asr56-1	AT3G22240	SALK_012158	Col-0
asr58-1	AT3G44430	FLAG_206D09	WS
asr58-2	AT3G44430	pst13878	Noessen
asr59-1	AT3G55910	SALK_006010	Col-0
asr60-1	AT3G58540	SALK_139216	Col-0
asr60-2	AT3G58540	SALK_024511	Col-0
asr61-1	AT4G16000	pst12951	Noessen
asr63-1	AT4G18580	SALK_000006	Col-0
asr64-1	AT4G22212	SALK_023543	Col-0
asr66-1	AT4G27654	SM_3_33171	Col-0
asr67-1	AT4G30670	pst20446	Noessen
asr68-1	AT4G31030	SAIL_52_A08	Col-0
asr69-1	AT4G31875	SM_3_28996	Col-0
asr70-1	AT4G39675	SALK_142090	Col-0
asr70-2	AT4G39675	SALK_111156	Col-0
asr71-1	AT5G02550	SALK_114601	Col-0
asr71-2	AT5G02550	SALK_018287	Col-0
asr72-1	AT5G02690	SALK_076269	Col-0
asr72-2	AT5G02690	SALK_075533	Col-0
asr73-1	AT5G03130	SALK_117524	Col-0
asr74-1	AT5G04790	SAIL_311_C06	Col-0
asr76-1	AT5G09990	SALK_021970	Col-0
asr77-1	AT5G11740	psh19150	Noessen

Allele name	Gene ID	Insertion line ID	Accession
asr78-1	AT5G12880	pst01485	Noessen
asr78-2	AT5G12880	pst00591	Noessen
asr79-1	AT5G19800	pst01636	Noessen
asr79-2	AT5G19800	pst19884	Noessen
asr80-1	AT5G22555	pst01699	Noessen
asr82-1	AT5G24570	FLAG_532C10	WS
asr83-1	AT5G26270	SAIL_563_E03	Col-0
asr84-1	AT5G28610	SAIL_894_C08	Col-0
asr84-2	AT5G28610	SAIL_857_C08	Col-0
asr85-1	AT5G28630	pst10912	Noessen
asr85-2	AT5G28630	pst12379	Noessen
asr88-1	AT5G44580	SALK_080439	Col-0
asr88-2	AT5G44580	pst14679	Noessen
asr90-1	AT5G55790	SM_3_34379	Col-0
asr90-2	AT5G55790	pst02687	Noessen
asr91-1	AT5G57760	SAIL_18_H07	Col-0
asr93-1	AT5G58790	SALK_020797	Col-0
asr93-2	AT5G58790	SALK_000171	Col-0
asr94-1	AT5G64900	FLAG_488B02	WS
asr94-2	AT5G64900	FLAG_102A09	WS
asr99-1	AT1G67350	pst19506	Noessen
asr99-2	AT1G67350	pst14303	Noessen
asr102-1	AT3G30160	SAIL_739_G07	Col-0
asr102-2	AT3G30160	pst17121	Noessen
asr103-1	AT3G48185	SM_3_26757	Col-0
asr103-2	AT3G48185	pst04523	Noessen
asr104-1	AT4G20420	SALK_031864	Col-0
asr104-2	AT4G20420	SALK_006217	Col-0
asr105-1	AT4G32240	SALK_022095	Col-0
asr106-1	AT5G29210	pst05570	Noessen
asr107-1	AT5G35480	FLAG_545E11	WS
asr107-2	AT5G35480	SAIL_18_A08	Col-0
asr109-1	AT5G65610	psh17835	Noessen

2.3.2 Abiotic stress screening of 79 T-DNA lines of orphan genes identified 45 promising candidates (Pre-screen)

T-DNA knockout lines for each of the 79 orphan genes were subjected to a range of abiotic stresses and the resistance/sensitivity was scored for each lines. The stress treatments included salt, osmotic (mannitol), drought (ABA, PEG treatments), heat (36 $^{\circ}$C) and genotoxic stresses. Lines showing differential phenotypic responses were considered stress responsive (Fig. 7-10). Since the lines were segregating, a 3:1 ratio of resistance to sensitivity (or vice versa) was expected. T-DNA lines were selected based on this scoring for each stress analysed.

Among the 79 genes analysed, a total of 45 orphan genes were selected as promising genes for one or more stress treatments. The Number of genes considered promising for each stress treatments were; 9 genes for salt, 8 for ABA treatment, 18 for osmotic (mannitol); 10 for drought (PEG); 10 for genotoxic and 14 genes for heat stress (Table 4). The representative images from the individual screens are shown in Figures 2, 3, and 4. Out of 45 genes, 10 genes showed multiple stress responses, with response to more than one stress treatment. This step helped to narrow down the number of genes to 45 for further screening and confirmation. Since the insertion lines were segregating, it was necessary to genotype them and rescreen to confirm whether the phenotype was real. Out of 45, we could obtain homozygous lines for 23 lines and they were reanalysed for the same abiotic stress conditions for which they showed phenotype in the previous screen for confirmation.

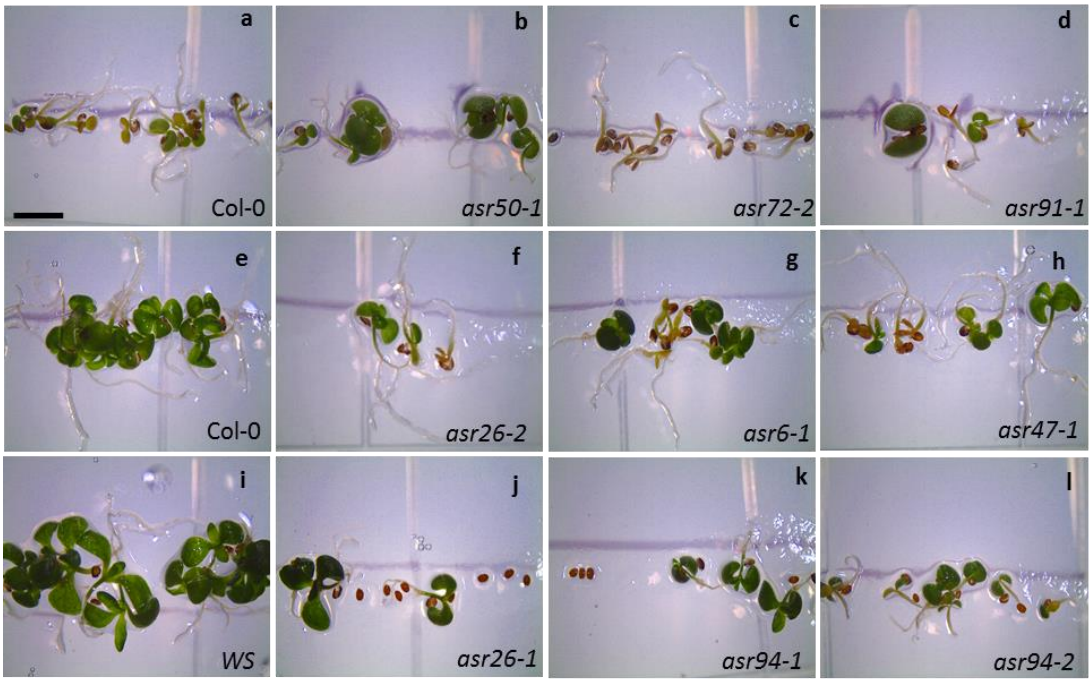


Figure 7: Abiotic stress pre-screen of ASR orphan genes - Salt and osmotic stress.

a to d- Salt screen - wild type (a) and mutants (b-d) grown on media containing 150mM NaCl. *asr50-1* and *asr91-1* segregating for resistance while *asr72-2* segregating for sensitivity.
e to l – Osmotic stress screen (Mannitol) - mutants (**f-h** in Col-0 background and **j-l** in WS background) segregating for sensitivity to mannitol at 375 mM. Scale bar =2.5 mm.

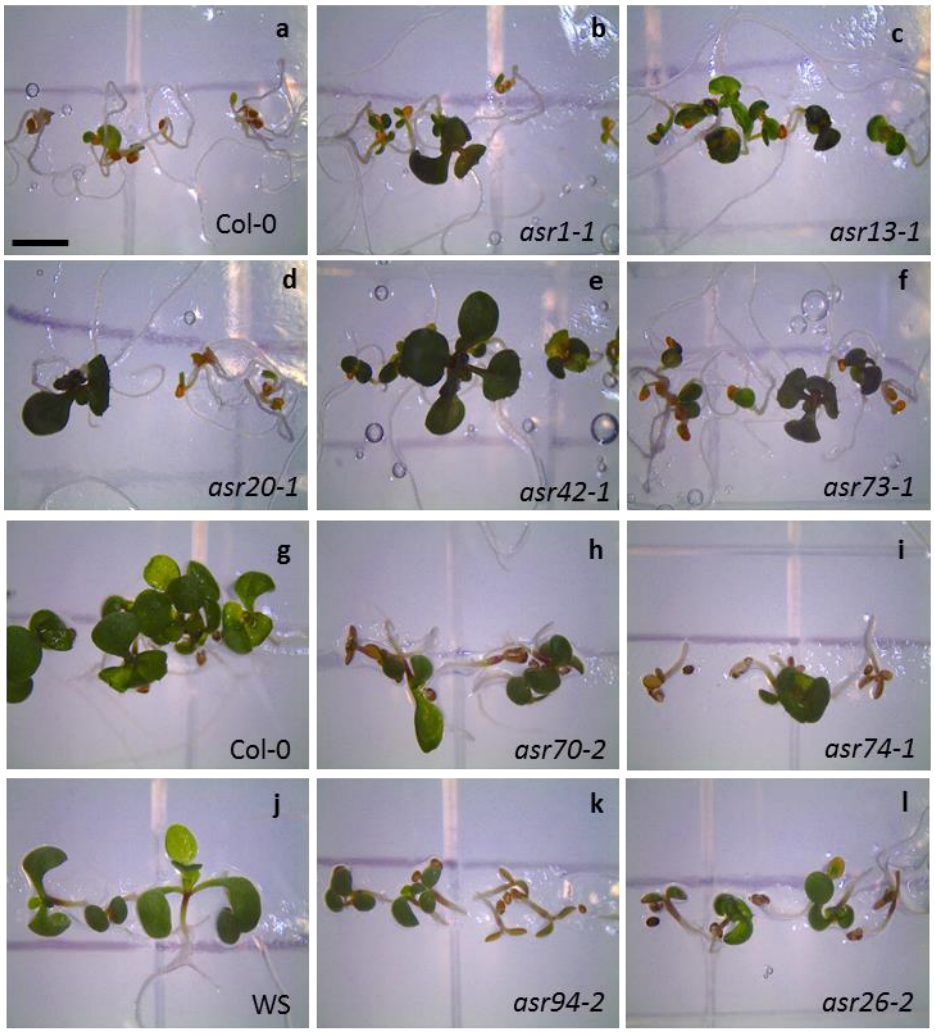


Figure 8: Abiotic stress pre-screen of ASR orphan genes - surrogate drought stress.

a to f - Drought screen (PEG)- mutant lines (b-f) segregating for tolerance to PEG (20%) compared to wild type Col-0 (a).
g to l - Drought screen (ABA)- Mutant lines (h and i) segregating for sensitive phenotype compared to wild type Col-0 (a). Similarly k and l which are in WS background segregating for sensitive phenotype compared to wild type WS (j) at ABA concentration of 1.5µM.
Scale bar =2.5 mm.

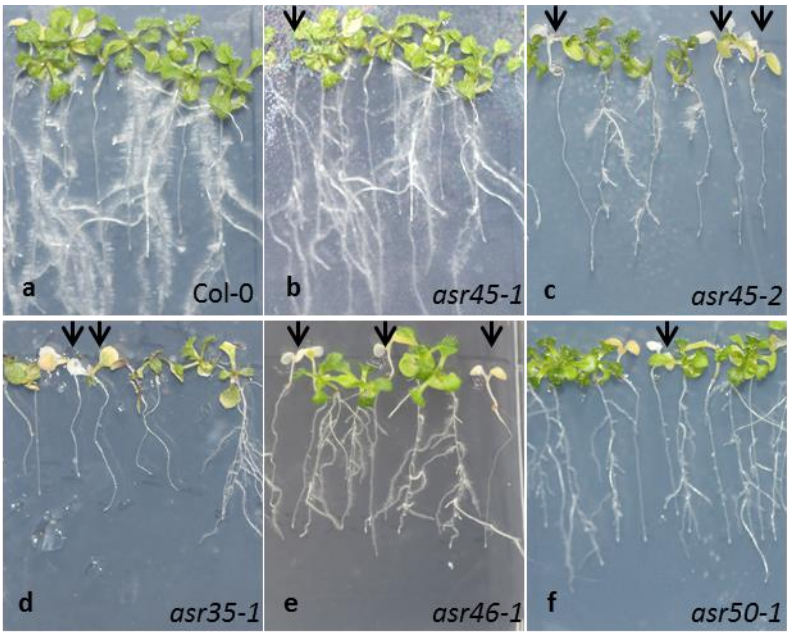


Figure 9: Abiotic stress pre-screen of ASR orphan genes - genotoxic stress. (a) Wild type Columbia-0, (b) to (f) T-DNA lines segregating for sensitive phenotype (dead seedlings indicated by black arrows). 7 days old seedlings were transplanted to half strength MS media supplemented with 0 ppm or 150 ppm Methyl methane sulphonate. Pictures were taken 7 days after transplant.

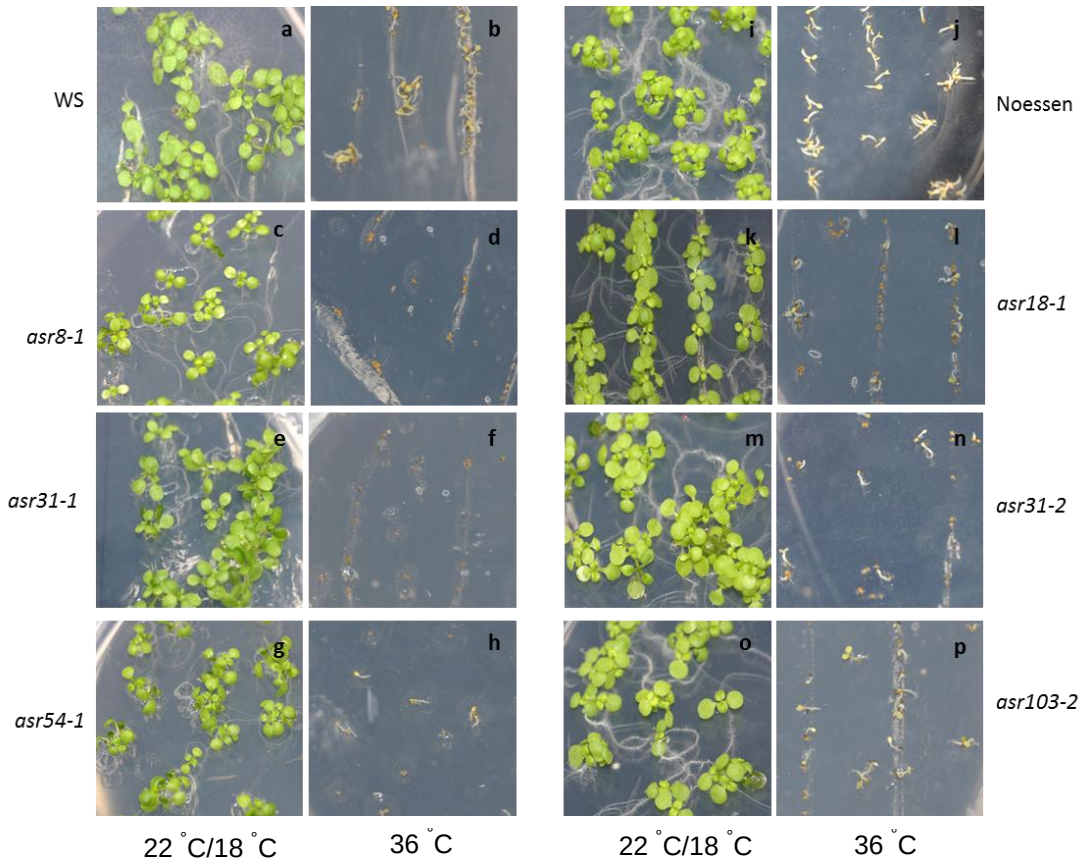


Figure 10: Abiotic stress pre-screen of ASR orphan genes - Heat stress. Wild types WS (a), Noessen (i) and mutant lines(c,e,b,k,m,o) germinated at normal growth conditions shows 100% germination. But under 36 ^{0}C, mutant lines (d,f,h,I,n,p) showed germination defect where as wild types (b and j) shows 100% germination, although failed to grow eventually. The T-DNA insertion lines *asr31-1* and *asr31-2* are for the same gene in WS and Noessen backgrounds respectively.

Table 4: Summary of abiotic stress pre-screen

Gene ID	Allele name	Insertion line ID	Ecotype	H	S	A	M	P	G
AT1G64360	*asr39-1*	SALK_063861	Col-0					+	
	asr39-2	SALK_052233	Col-0					+	
AT1G67865	*asr42-1*	SALK_037677	Col-0					+	
AT1G22890	*asr1-1*	SALK_062788	Col-0					+	
AT1G31580	*asr35-1*	SALK_088330	Col-0					+	+
AT1G75770	*asr44-1*	SALK_072169	Col-0		+			+	
	asr44-2	SAIL_869_E08	Col-0			+		+	
AT1G11850	*asr31-1*	FLAG_235C12	WS	+			+		
	asr31-2	pst15639	Noessen	+					
AT1G76960	*asr45-1*	SALK_034899	Col-0						+
	asr45-2	SALK_098782	Col-0			+			+
AT2G47200	*asr53-1*	SALK_029461	Col-0						
	asr53-2	SM_3_38195	Col-0		+		+		
AT2G14460	*asr48-1*	FLAG_322B08	WS	+					
AT2G25510	*asr50-1*	SALK_132861	Col-0		+				+
AT2G04800	*asr46-1*	SALK_043103	Col-0		+				+
AT2G07721	*asr47-1*	SAIL_412_B02	Col-0				+	+	
AT3G04640	*asr6-1*	SALK_087773	Col-0				+		
AT3G05730	*asr55-1*	SALK_031670	Col-0						+
AT3G11745	*asr7-1*	SM_3_17126	Col-0		+			+	
AT3G14480	*asr8-1*	FLAG_261C03	WS	+					
AT3G03020	*asr5-2*	pst04492	Noessen	+					
AT3G58540	*asr60-2*	SALK_024511	Col-0						+
AT3G48185	*asr103-2*	pst04523	Noessen	+					
AT3G56260	*asr12-1*	SAIL_586_F09	Col-0				+		
AT3G30160	*asr102-1*	SAIL_739_G07	Col-0	+		+	+		
AT3G02240	*asr54-1*	FLAG_035H03	WS	+					
AT4G30460	*asr17-2*	SALK_101746	Col-0						+
AT4G39675	*asr70-1*	SALK_142090	Col-0						
	asr70-2	SALK_111156	Col-0			+			
AT4G31875	*asr69-1*	SM_3_28996	Col-0		+				
AT4G12580	*asr13-1*	SAIL_1264_F11	Col-0				+	+	
AT4G33666	*asr18-1*	pst15749	Noessen	+			+		
AT4G31030	*asr68-1*	SAIL_52_A08	Col-0				+		
AT5G64900	*asr94-1*	FLAG_488B02	WS				+		
	asr94-2	FLAG_102A09	WS			+	+		
AT5G64905	*asr26-1*	FLAG_338B09	WS	+			+		
	asr26-2	SALK_017813	Col-0			+	+		

AT5G35480	*asr107-1*	FLAG_545E11	WS	+					
	asr107-2	SAIL_18_A08	Col-0				+		
AT5G44580	*asr88-2*	pst14679	Noessen				+		
AT5G64890	*asr25-1*	FLAG_020B08	WS	+					
AT5G03130	*asr73-1*	SALK_117524	Col-0				+		
AT5G09990	*asr76-1*	SALK_021970	Col-0						+
AT5G23460	*asr24-1*	SALK_054529	Col-0			+			
AT5G02690	*asr72-2*	SALK_075533	Col-0		+				
AT5G55790	*asr90-1*	SM_3_34379	Col-0		+				
	asr90-2	pst02687	Noessen	+					
AT5G06190	*asr20-1*	SAIL_139_B12	Col-0				+	+	
AT5G57760	*asr91-1*	SAIL_18_H07	Col-0		+			+	
AT5G04790	*asr74-1*	SAIL_311_C06	Col-0			+	+		
AT5G26270	*asr83-1*	SAIL_563_E03	Col-0				+		
AT5G29210	*asr106-1*	pst05570	Noessen				+		
AT5G12880	*asr78-1*	pst01485	Noessen						+
AT5G65610	*asr109-1*	psh17835	Noessen	+					

Table 4: Summary of abiotic stress pre-screen for orphan genes. List of selected genes from pre-screen along with their mutant lines. '**+**' indicates lines which showed response to the stress treatment. Blank cells mean no response (or similar to wild type controls). H-Heat;S-Salt;A-Abscisic acid;M-mannitol;P-Poly ethylene glycol;G-genotoxic.

2.3.3 Rescreening with homozygous T-DNA lines

Among the 45 genes selected from the preliminary screen, we isolated homozygous lines for 23 genes that were then rescreened. The lines were analysed for the stress conditions for which they showed responses in the pre-screen. From rescreening of the homozygous lines we could identify phenotypes for two genes namely *ASR50* (AT2G25510) for salt stress and *ASR70* (AT4G39675) for ABA sensitivity (Table 5).

Table 5: List of homozygous T-DNA lines screened for abiotic stresses

Sl. NO.	Gene ID	Allele name	Insertion ID	Stresses conditions analysed	phenotype confirmed
1	AT1G22890	*asr1-1*	SALK_062788	PEG	
2	AT1G31580	*asr35-1*	SALK_088330	Genotoxic	
3	AT1G76960	*asr45-2*	SALK_098782	ABA, Genotoxic	
4	AT2G47200	*asr53-1*	SALK_029461	PEG, Salt	
		asr53-2	SM_3_38195	Salt, Mannitol	
5	AT2G04800	*asr46-1*	SALK_043103	Salt, Genotoxic	
6	**AT2G25510**	***asr50-1***	**SALK_132861**	**Salt, Genotoxic**	**Salt tolerance**
7	AT3G05730	*asr55-1*	SALK_031670	Genotoxic	
8	AT3G58540	*asr60-2*	SALK_024511	Genotoxic	
9	AT3G56260	*asr12-1*	SAIL_586_F09	Mannitol	
10	AT3G30160	*asr102-1*	SAIL_739_G07	Heat, ABA, Mannitol	
11	AT4G30460	*asr17-2*	SALK_101746	Genotoxic	
12	**AT4G39675**	***asr70-1***	**SALK_111156**	**ABA, Genotoxic**	**ABA sensitive**
		asr70-2	**SALK_142090**	**ABA, Genotoxic**	**ABA sensitive**
13	AT4G31875	*asr69-1*	SM_3_28996	Salt	
14	AT5G03130	*asr73-1*	SALK_117524	Mannitol	
15	AT5G06190	*asr20-1*	SAIL_139_B12	Mannitol, PEG	
16	AT5G04790	*asr74-1*	SAIL_311_C06	ABA, Mannitol	
17	AT5G26270	*asr83-1*	SAIL_563_E03	Mannitol	
18	AT1G11850	*asr31-1*	FLAG_235C12	Mannitol	
19	AT5G57760	*asr91-1*	SAIL_18_H07	Salt, PEG	
20	AT5G23460	*asr24-1*	SALK_054529	ABA	
21	AT1G75770	*asr44-1*	SALK_072169	Salt, PEG	
22	AT2G07721	*asr47-1*	SAIL_412_B02	Mannitol, PEG	
23	AT4G12580	*asr13-1*	SAIL_1264_F11	Mannitol, PEG	

Table 5: List of homozygous lines screened for abiotic stresses.
Confirmed phenotypes for the T-DNA lines are mentioned in the last column. No obvious phenotypes were observed for the remaining lines for the respective stresses analysed. ABA - Abscisic acid; PEG - Poly Ethylene Glycol.

2.3.4 The Brassicaceae specific orphan gene *ASR50* (AT2G25510.1) show tolerance to salt in loss of function screen in *Arabidopsis thaliana*

The orphan gene *ASR50* showed segregation for salt tolerance in the pre-screen. It was identified as salt stress responsive in our previous study as it was differentially expressed under salt stress in the microarray experiment (Donoghue et al., 2011; Kilian et al., 2007). *ASR50* is annotated as an unknown protein with an ORF of 312 bases with one intron in the 5' UTR encoding a protein of 103 amino acids (TAIR). The *asr50-1* T-DNA line showed salt tolerance phenotype in the pre-screen (Fig-7b). To further confirm the salt stress phenotype, a homozygous line of *asr50*-1 was identified and the T-DNA insertion site was confirmed by left border sequencing. The T-DNA was found to be inserted in the intron of the 5' UTR (Fig-11A). The knockout status of the mutant line was confirmed by semi quantitative RT PCR using RNA extracted from seedlings from wild type and *asr50-1* (Fig-11B).

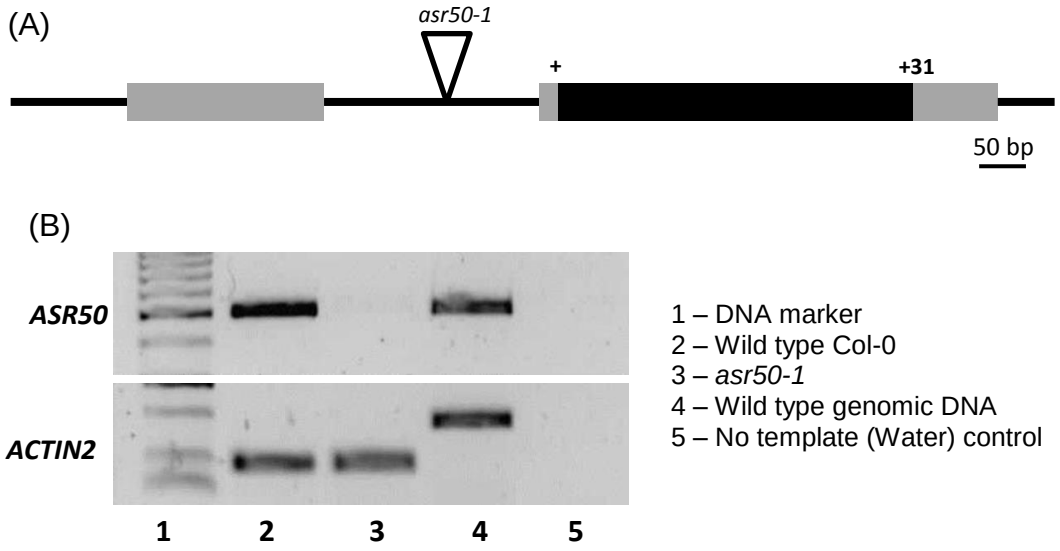


Figure 11 : Analysis of T-DNA insertion at *ASR50* (AT2G25510.1) locus. A) Characterization of T-DNA insertion sites in *asr50-1*. Triangle represents the T-DNA insertion site determined by T-DNA left border sequencing. Grey boxes indicates UTR region and black box indicates exon. The line between two grey boxes is the only intron in which T-DNA resides in *asr50-1.*B) RT-PCR analysis with *ASR50* specific primers. RNA extracted from 12 day old seedlings. No expression of *ASR50* in *asr50-1* line (Lane 3, upper panel). *ACTIN2* was used as reference gene (lower panel).

To investigate the salt tolerance of the homozygous mutant line, we germinated the homozygous mutant line and wild-type control on half strength MS medium supplemented with 0 and 150 mM NaCl. The salt tolerance of the mutant line was scored in terms of percentage green seedlings 16 days after stratification. The wild type showed about 34% of green seedlings *asr50-1* showed higher percentage of green seedlings compared to wild type control 16 days after stratification (Fig.12).

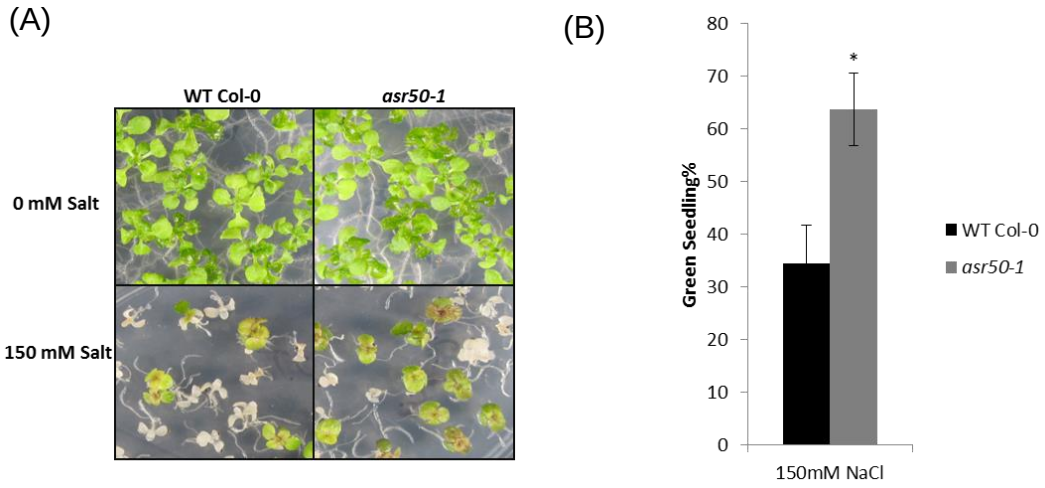


Figure 12: Salt tolerance of *asr50-1*. (A) & (B) Survival percentage of wild type and mutant lines grown on 0.5X MS media supplemented with 150 mM NaCl. Data represents mean±SEM (n is approx. 100/experiment) from two independent experiments with triplicates. (Student's t test,*P<0.05).

2.3.5 *asr50-1* maintain higher photosynthetic efficiency under salt stress

Chlorophyll fluorescence is an effective parameter for revealing early signs of stress and is a suitable way to screen for stress tolerance in plants (Chaerle et al., 2007). Measurement of total chlorophyll content indicated that *asr50-1* has higher chlorophyll content compared to wild-type control under salt stress (Fig.13B).

To evaluate whether the tolerance to photo inhibition is altered in the mutants under salt stress, we measured the PSII maximum operating efficiency (Fv/Fm). Without salt stress, the Fv/Fm ratio was similar in both wild-type and mutant lines. However, under salt stress, *asr50-1* showed

significantly higher Fv/Fm ratio compared to wild type control suggesting the tolerance to salt stress (Figure 13A).

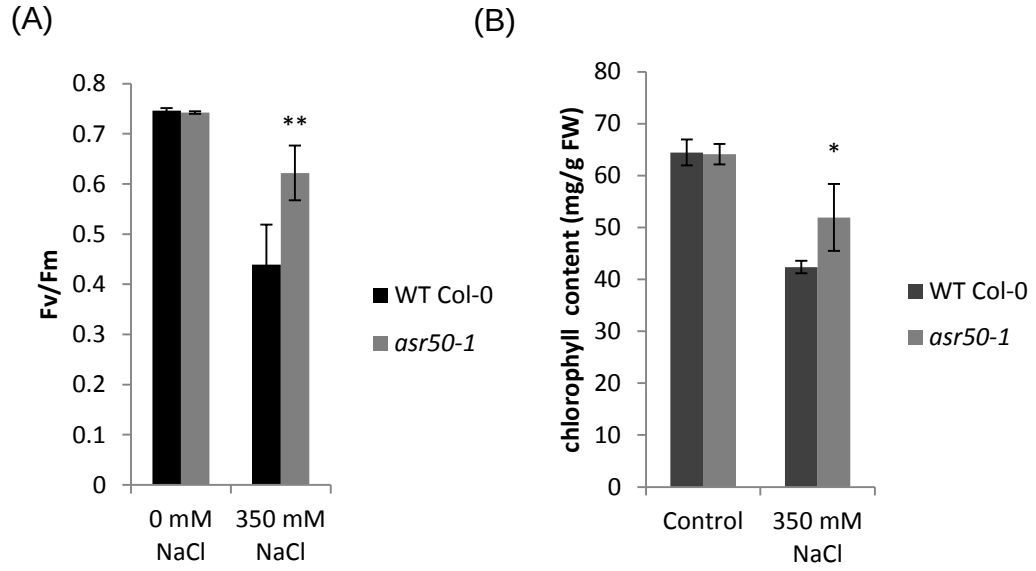


Figure 13: Comparison of photosynthetic potential of mutant and control under high salinity stress. Fv/Fm (A) and total chlorophyll content (B) were measured. For Fv/Fm leaves from fifteen plants per line were measured (Values are mean± SEM, **P< 0.01, *P<0.05).

2.3.6 Expression of stress responsive genes in *asr50-1*

To understand the molecular mechanism of *asr50-1* in salt stress, expression of stress responsive marker genes, *CBF1, CBF3, RD29A*, *KIN1* and *COR15A* were analysed under salt stress in wild type and mutant lines by real-time PCR. The expression level of *CBF1, RD29A*, *KIN1* and *COR15A* was similar in *asr50-1* mutant compared to wild type both under normal and salt stress conditions (only *CBF1* is shown in Fig. 14). The *CBF3* expression was similar in both wild type and *asr50-1* under normal conditions but under salt stress *CBF3* expression in *asr50-1* found to be up regulated compared to wild type control (Fig. 14B).

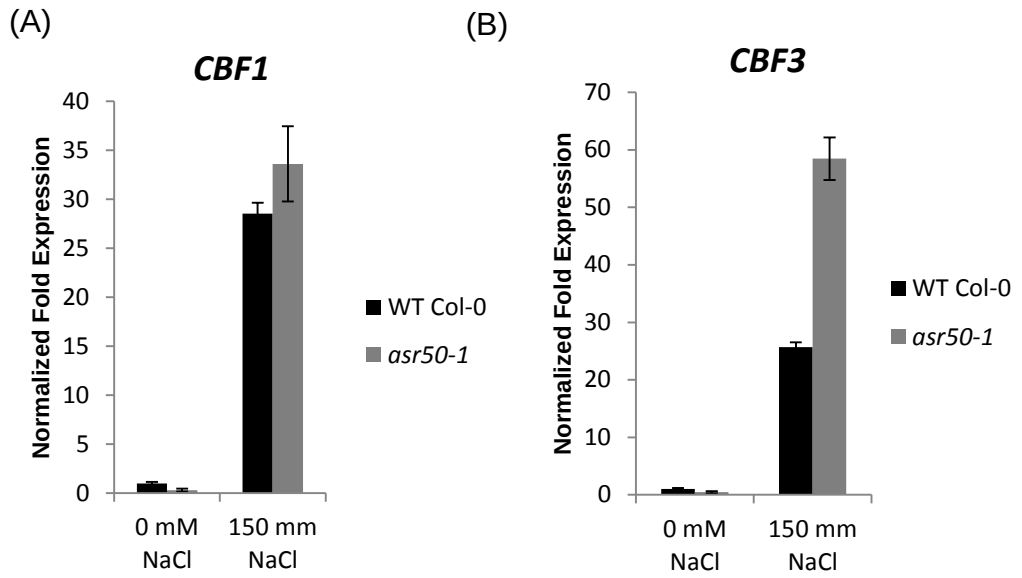


Figure 14: Expression of stress responsive genes in wild type and *asr50-1* seedlings induced by 150 mM NaCl. RNA extracted from 12 day old seedlings treated with 0 mM or 150 mM NaCl for 3 hours. *ACTIN2* was used as internal control.

2.3.7 *ASR70* is a Brassicales specific gene

In our previous study (Donoghue et al., 2011) we described *ASR70* (AT4G39675) as a Brassicaceae specific gene and originated from duplication mechanism as it has a BLASTN hit to a non-orphan gene in *Arabidopsis*. As there are several genomes sequenced since our previous study we rechecked to confirm its Brassicaceae specific orphan status. We found that *ASR70* ORF has a BLASTP hit to an uncharacterized protein (TCM_001203) from *Theobroma cacao* which was sequenced in 2013 (Motamayor et al., 2013). The threshold e-value cutoff followed for BLASTP was $10e^{-3}$ as in the previous study. Hence *ASR70* cannot be called as a Brassicaceae specific gene but can be considered as Brassicales specific gene following the threshold e-value cutoff of $10e^{-3}$.

ASR70 is annotated as an unknown protein, having a single exon with 213 base pairs long. In both alleles of *ASR70* analysed, *asr70-1* and *asr70-2,* T-DNA insert was found in the exonic region (Fig. 15A). RT PCR analysis of the mutant lines confirmed that *ASR70* was knocked out in both the mutant lines (Fig. 15B).

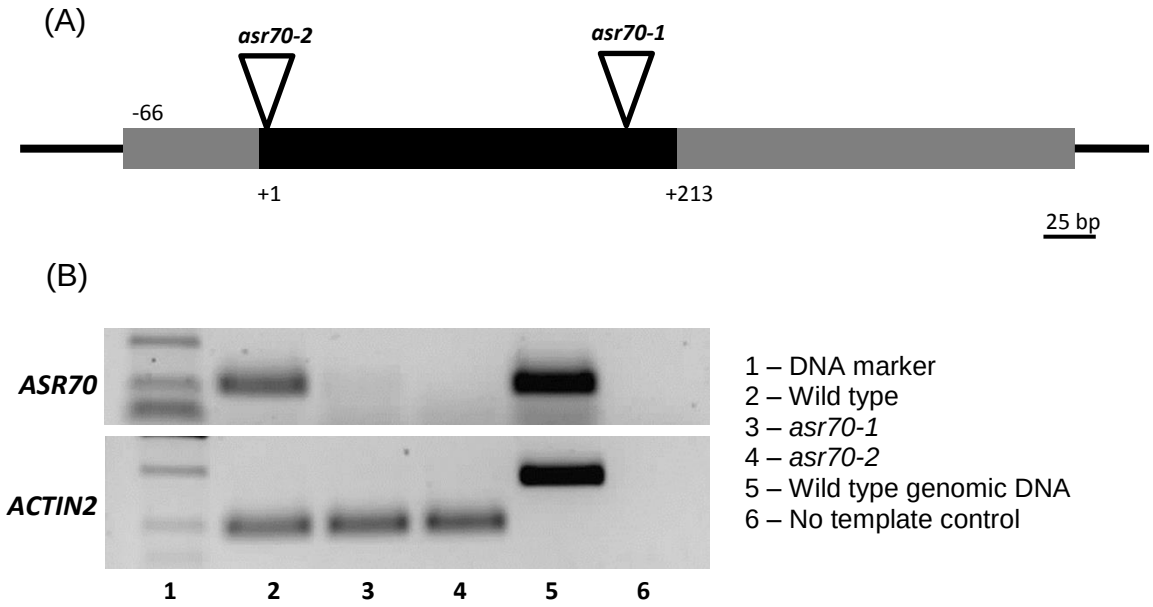


Figure 15: Analysis of T-DNA insertion at *ASR70* locus. **A**)Characterization of T-DNA insertion sites in mutants *asr70-1* and *asr70-2*. Triangle represents the T-DNA insertion site determined by sequencing. Grey boxes indicates UTR region and black box indicates exon. **B**) RT-PCR analysis with *ASR70* specific primers. RNA extracted from 12 day old seedlings. No expression of *ASR70* in *asr70-1* and *asr70-2* lines (Lane 3 and 4, upper panel). *ACTIN2* was used as reference gene (lower panel).

2.3.8 *The asr70 mutant lines* are responsive to abscisic acid

The plant growth regulator ABA affects many aspects of plant growth and development including water relations and tolerance to variety of environmental stresses (Borsani et al., 2002; Finkelstein et al., 2002; Himmelbach et al., 2003; Zhu, 2002). The lines which are sensitive to ABA are the candidates for drought tolerance. To determine the ABA response *asr70-1* and *asr70-2*, the seeds of wild type and mutant lines were germinated on half strength MS (Murashige and Skoog) medium supplemented with 1% sucrose containing 0 µM, 1.5 µM, 2 µM and 3 µM ABA, and compared for the differences in germination rate. Germination rates were monitored in terms of radicle emergence every day up to 10 days.

We found that the two independent insertion lines (alleles) for the gene *ASR70* (AT4G39675), *asr70-1* and *asr70-2,* showed ABA sensitive phenotype was measured as slower germination rate under 1.5 uM ABA treatment (Fig. 16A). Subsequently, cotyledon greening percentage was

scored 12 days after germination. Both the lines showed significantly lower percentages of green cotyledons in all three treatments analysed (Fig.16C and D). Inhibition of seedling root growth is a typical action of ABA (Ryu et al., 2010). To determine the effect of ABA on post germination growth, root growth of mutant lines on ABA medium was measured. The mutant (*asr70-2*) and the wild-type seeds germinated on half strength MS medium, grown for 5 days and then transferred to half strength MS medium with 0, 5, 10, 20 or 40 µM ABA. The root length measurements were taken 7 days after the transplant. However, *asr70-2* did not show any difference in reduction of root growth compared to wild type control (Fig.16B).

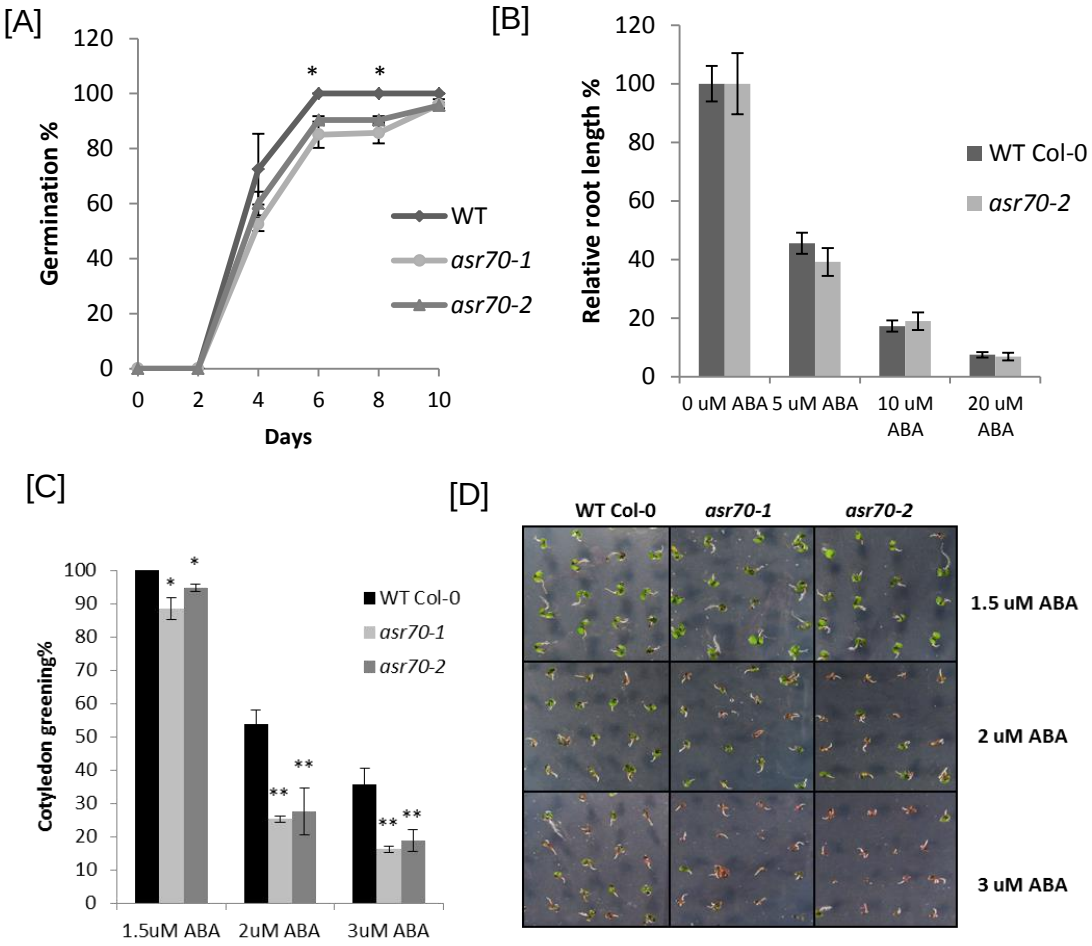


Figure 16: Green cotyledon percentage of *asr70-1* and *asr70-2* in response to ABA. [A] Germination rates of *asr70-1, asr70-2* and wild type Col-0 at 1.5 µM ABA. [B] Relative root growth of *asr70-2* compared to wild type control at different ABA concentrations which shows no response. [C] Green cotyledon percentage at different ABA concentrations. Observations were taken 12 days after stratification. [D] Phenotypes of wild-type, *asr70-1* and *asr70-2* lines under 1.5 µM, 2 µM and 3 µM ABA, 12 days after stratification. Values are mean ±SEM (Student's t-test, *P<0.05. **P<0.01).

2.4 Discussion

Abiotic stresses (e.g. cold, drought, heat and salinity) adversely affect plant growth and crop production worldwide. Understanding the mechanisms by which plants transmit the signals to cellular machinery to elicit adaptive responses is of fundamental importance to develop more stress tolerant crops to improve production efficiency (Huang et al., 2011). In the face of climate change, understanding plant responses to changing environmental conditions is even more important (Reddy et al., 2011). In our previous study we found that orphan genes are enriched for abiotic stress responses using the AtGenExpress expression data series for stress (Donoghue et al., 2011). In this study we identified two genes *ASR50* and *ASR70* to be involved in abiotic stress responses in *Arabidopsis thaliana* based on strong abiotic stress phenotypes detected via loss of function alleles. While *ASR50* is a Brassicaceae specific gene, *ASR70* is specific to the Brassicales lineage.

2.4.1 Brassicaceae specific orphan gene *ASR50* involved in salt stress responses in Ara*bidopsis thaliana*

Salt stress resulting from saline soils or irrigation water is a major factor limiting agricultural productivity worldwide (Munns and Tester, 2008). In our screen for salt stress, a loss of function allele (*asr50-1)* of *ASR50* showed a salt tolerant phenotype compared to wild-type control in terms of ability to maintain higher green seedling percentage which correlated with higher chlorophyll content and higher operating efficiency of PSII under salt stress. Chlorophyll fluorescence is a rapid and non-intrusive tool to screen for salinity tolerance (Maxwell and Johnson, 2000). Higher chlorophyll fluorescence under salt stress in the *asr50-1* mutant line compared to wild type control suggests the ability of the mutant line to maintain the higher operating efficiency of PSII, which enhances the tolerance of mutant line for increased salt content in the soil.

To further elucidate a possible mechanism of the *ASR50* gene in salt stress response, expression of stress responsive transcription factors *CBF1*

and *CBF3* were analysed in mutant lines and wild type under salt stress condition using real-time PCR. Overexpression of *CBF1* and *CBF3* in transgenic *Arabidopsis* have been reported previously which induce the expression of CRT/DRE-containing genes leading to freezing, dehydration and salt tolerance (Gilmour et al., 2000; Jaglo-Ottosen et al., 1998; Kasuga et al., 1999; Liu et al., 1998). The elevated expression of *CBF3* in *asr50-1* may help explain the salt tolerant phenotype of the mutant line and also could indicate that *ASR50* could be acting in the same pathway

The next step experiments that need to be done to understand the functionality of *ASR50* gene are (a) elucidation of evolutionary origin (b) identification and screening of additional loss of function alleles; (c) complementation of the *asr50-1* phenotype; (d) determining whether overexpression of *ASR50* increases the salt sensitivity (i.e. a dosage response test); and (e) determining whether *CBF3* overexpression contributes to the *asr50*-1 phenotype through generation and analysis of an *asr50 cbf3* double mutant line.

2.4.2 Loss of function of *ASR70* delays germination in *Arabidopsis thaliana*

The *Arabidopsis thaliana* gene *ASR70* was initially screened as Brassicaceae specific orphan gene. But we later confirmed that *ASR70* is not Brassicaceae specific but can be considered as Brassicales specific gene as it has a BLASTP hit to an uncharacterised protein of *Theobroma cacao* which is specific to Brassicales lineage. From the loss of function screening of *ASR70* knockout alleles, *asr70-1* and *asr70-2,* we showed that the knockout lines are sensitive to ABA in terms of slower germination rate and reduced green cotyledon percentage (Fig.16A, C and D). But *asr70-2* did not show any difference in terms of root growth compared to wild-type in response to ABA (Fig.16B). This is not surprising considering the inhibitory effect of ABA on root growth involves crosstalk between several hormonal pathways, such as ethylene, auxin, brassinosteroid, jasmonic acid, and also sugars (Beaudoin et al., 2000; Ghassemian et al., 2000), and the relationship between ABA and osmotic sensitivity is complex (Rosado et al., 2006). From the results presented here, it appears that *ASR70* has a potential role in

seed germination in *Arabidopsis thaliana*. However, this is a preliminary result and the *asr70* phenotype will have to be further confirmed by complementation experiment and/or by the gain of function approach by overexpressing *ASR70* under a constitutive promoter. Since this gene was found to be not a Brassicaceae specific orphan we did not pursue this further as it deviated from the objective of this thesis work.

The next step experiments that need to be done with the *ASR70* gene to further understand its role in abiotic stress response are (a) complementation of the *asr70* mutant lines with intact *ASR70;* (b) Analysis of *ASR70* overexpression lines to determine any dosage response to ABA; (c) to investigate the *asr70* mutant lines and *ASR70* overexpression lines under drought stress in transpiring conditions i.e. on soil.

2.5 Conclusions

The majority of the work on orphan genes so far has been focussed on studying the evolutionary origins and how these new genes could acquire functionality. Only recently has the focus shifted towards functionally characterizing orphan genes and determining whether these genes have essential functions or have acquired species specific traits. The main aim of our study in this chapter was to identify the functionality of Brassicaceae specific orphan genes in *Arabidopsis thaliana* in relation to abiotic stress responses by loss of function approach. Although this is not the first orphan gene study in *Arabidopsis*, we believe it is the first time any study has focussed on orphan genes in *Arabidopsis* related to specific traits, abiotic stresses in this case.

Although we found only two genes from a total of 23 genes screened with potential abiotic stress function that does not mean the other genes are unlikely to have any function in relation to the abiotic stress responses analysed. In conclusion, we identified two novel genes *ASR50* and *ASR70* having a function in salt tolerance and seed germination respectively in

Arabidopsis. The loss of function approach in *Arabidopsis* using T-DNA insertional mutant lines is a useful tool for characterising orphan gene functions. Further studies on how these orphan genes originated in the specific lineage and acquired respective functions will help to understand the importance of orphan gene evolution and the functional relevance of orphan genes.

Chapter 3: Gain of function screen of Brassicaceae specific orphan genes in *Arabidopsis thaliana* for abiotic stress responses

3.1 Introduction

3.1.1 Gain of function as a genetic tool

Many genes belong to gene families where there is genetic redundancy between gene family members. In such situations, it is not easy to identify the gene function by disrupting one gene. Because of this genetic redundancy loss of function screens are not always successful in dissecting gene function. Gain-of-function screening is a complementary method to loss of function screens and can be used to identify gene functions. Gain of function effects for a gene can be analysed by overexpressing the gene under the control of a strong promoter which result in gain-of-function phenotypes, especially when the expression is also ectopic (Bolle et al., 2011). Gain of function overexpression screening can also be done by random insertion of transcriptional enhancers in the genome (Ichikawa et al., 2006; Weigel et al., 2000). Gain-of-function approaches are advantageous for characterization of gene functions as they enable us to analyse individual gene family members. Unlike loss of function phenotypes, gain of function phenotypes segregate as dominant traits which also saves the additional genotyping step to obtain homozygous lines. Loss of function and gain of function phenotypes can often be complementary to each other (Bolle et al., 2011). Gain of function analysis has become an important tool for dissecting gene functions and has been successfully used for characterising genes responsible for various traits. Different approaches for gain of function analysis with examples are described in the following sections:

3.1.2 Gain of function through activation tagging

Activation tagging was the first gain of function system to be used in plants (Kakimoto, 1996; Weigel et al., 2000). In this method, a strong promoter or enhancer element containing T-DNA is randomly inserted into the genome. The promoter/enhancer element then activates the gene near the T-DNA insertion site. The promoter or enhancer elements from Cauliflower mosaic virus *35S* (*CaMV 35S*) promoter has been utilized in this

approach. The phenotypes observed in the transgenic lines can be attributable to the insertion event by sequencing the T-DNA flanking regions and mapping it onto the reference genome.

The activation tagging approach has been successfully utilised for identifying the roles of genes involved in stress tolerance, development and plant metabolism. The identification of *YUCCA* genes involved in auxin biosynthesis is one of the best examples of how activation tagging can be utilised for identifying genes related to plant development (Kim et al., 2007; Woodward et al., 2005; Zhao et al., 2001). The YUCCA family consists of 11 members in *Arabidopsis* and single mutants did not show any phenotype although double, triple and quadruple mutants analysed show deleterious effects on development. But the activation tagged *YUCCA* mutants show auxin overproduction phenotypes (Cheng et al., 2006). Activation tagging was also successfully used in identifying stress resistant genes, for example *HDG11* transcription factor. A mutant *edt1* showed enhanced drought tolerance (Aharoni et al., 2004) which was found be associated with increased expression of *HDG11* (Yu et al., 2008). Other examples include identification of transcriptional regulator of secondary metabolites in tomato (Mathews et al., 2003), isolation of key regulators of the terpenoid indole alkaloid pathway in *Catharanthus roseus* suspension cultures (van der Fits and Memelink, 2000) and identification of *Arabidopsis PAP1*, encoding a member of R2, R3 MYB transcription factor, which can enhance anthocyanin accumulation by overexpression (Borevitz et al., 2000).

A large numbers of activation tagged lines have been developed in *Arabidopsis* to date. The RIKEN activation tagged lines composed of more than 50,000 independently transformed *Arabidopsis* lines and more than 1000 putative mutants showing morphological phenotypes were isolated (Ichikawa et al., 2003; Nakazawa et al., 2003) and are available through RIKENBASE, Japan. Weigel et al. (2000) generated more than 30,000 transgenic lines using activation tagging vectors. Activation tagged lines were also developed in Rice. Jeong et al. (2006) generated 48,000 activation-tagged japonica rice plants and obtained 28,000 FSTs (Flanking

Sequence Tags) and made available through POSTECH Rice T-DNA Insertion Sequence Database (RISD).

3.1.3 Gain of function by overexpression

As mentioned in the earlier section, overexpression can also be achieved by cloning and expressing the gene of interest under a strong constitutive promoter. The importance of gene dosage for normal functioning of genes was known for a long time from studies from aneuploids in human genetic syndromes and mutant phenotypes in Drosophila and plants. But the overexpression strategy as a screening tool was started in yeast only when yeast transformation techniques developed (Hinnen et al., 1978). The easy transformation technique in *Arabidopsis* also enabled overexpression technique to be analysed in plant system. Although *CaMV 35S* promoter is the most widely used, several other plant promoters such as ubiquitin promoter are also in use for constitutive expression of genes. There are two ways of overexpressing a gene; one way is by overexpression of full length cDNAs and the other by overexpression of an open reading frame.

3.1.3.1 cDNA overexpression

This strategy involves expression of the full length cDNA under a constitutive promoter. In *Arabidopsis*, more than 240,000 full length cDNAs (fl-cDNA) were generated (Sakurai et al., 2005; Seki et al., 2002). Since, fl-cDNAs contain necessary information for RNA and protein production, they are important resource for functional genomic studies. Apart from *Arabidopsis*, fl-cDNA resources are also available in other plant species such as Rice (Rice Full-Length c et al., 2003), maize, poplar (Nanjo et al., 2007), wheat (Ogihara et al., 2004) and barley (Sato et al., 2009). In *Arabidopsis*, the full length cDNA clones can be obtained through stock centres.

The cDNA overexpression has resulted in identification of functions of several genes. Banno et al. (2001) generated cDNA library from different *Arabidopsis* tissues, fused with *CaMV 35S* promoter and transformed into

Arabidopsis roots. Root cell cultures were then grown without cytokinin. This led to the identification of novel cDNA *Enhancer of Shoot Regeneration 1 (ESR1)*, a putative transcription factor with an AP2/EREBP domain with transactivation activity which found to have a critical role in shoot differentiation (Banno et al., 2001). LeClere and Bartel (2001) generated more than 30,000 transgenic lines from cDNA library driven by *CaMV 35S* promoter. They identified a co-suppression phenotype caused by overexpression of a truncated cDNA encoding chloroplast ferredoxin-$NADP^+$ reductase (FNR). Another group (Kuhn et al., 2006) also identified ABA sensitive mutants like, protein phosphatase type 2C (PP2C) proteins from the same population.

Another slightly modified cDNA overexpression system is the FOX (Full length cDNA Over-eXpression) hunting system. This was developed by Ichikawa et al. (2006) which involves cloning an equimolar ratio of a mixture of fl-cDNAs into a cloning vector driven by *CaMV 35S* promoter followed by *in planta* transformation of *Arabidopsis* with this cDNA expression library. This technique eliminates cosuppression phenotypes caused by truncated cDNAs. Moreover, in this case the phenotype can be easily associated with the gene as the fl-cDNA can be easily identified and isolated using vector specific primers. Ichikawa and colleagues generated more than 15,000 *Arabidopsis* FOX lines expressing fl-cDNAs. Not only in *Arabidopsis*, FOX lines were also generated in rice, more than 12,000 rice lines have been generated in which 13,980 independent fl-cDNAs were overexpressed under the control of the ubiquitin promoter (Nakamura et al., 2007).

3.1.3.2 ORF overexpression

In this strategy only an open reading frame (ORF) is overexpressed under a constitutive promoter. An ORF is the minimal unit of a gene encoding protein and lacks 5’ and 3’ untranslated regions (UTR’s). This approach is especially important for computationally predicted genes. For example, a set of small secreted peptides were predicted to be involved in plant development in *Arabidopsis* and overexpression of these 153 ORFs led to the identification of the *Epidermal Patterning Factor 1 (EPF1)* gene (Hara

et al., 2007) which is involved in stomatal patterning. For *Arabidopsis*, a large number of ORF clones were generated by various research groups. The SALK institute has generated more than 10,000 ORF clones in collaboration with the SSP (Salk, Stanford, PGEC) consortium and Invitrogen. Gong et al. (2004) generated a collection of ORF clones for *Arabidopsis* transcription factors. The *Arabidopsis thaliana* OrfeoME (ATOME) project has also produced 5,500 ORF clones. All of these ORF clones can be obtained from the Arabidopsis Biological Resource Center (ABRC), USA.

In this chapter, we have employed the ORF overexpression strategy for gain of function screening of *Arabidopsis thaliana* orphan genes to elucidate their functionality under abiotic stress. We have found previously that *Arabidopsis* orphan genes have short ORFs with less or no introns which is considered as a general feature of orphan genes(Donoghue *et al*., 2011). Although orphan genes are annotated, they have not been well characterized and hence ORF overexpression strategy would be more appropriate for gain of function screens. This chapter describes the results from abiotic stress screening of orphan gene overexpression lines in *Arabidopsis thaliana*.

3.2 Results

3.2.1 Gain-of-function screen of abiotic stress responsive orphan genes in *Arabidopsis thaliana*

An ORF overexpression approach was employed for the gain of function analysis of stress responsive orphan genes. The ORF of each of the orphan genes was amplified by PCR using primers attached with *attB* site at the 5' end and cloned under *CaMV 35S* promoter through Gateway cloning technology. The list of overexpression constructs made in this study is mentioned in Table 6 along with the primers used. All the constructs were transformed into the *Arabidopsis thaliana* Columbia-0 accession. The genes selected for overexpression were the most promising candidates from the loss-of-function screen and also from the gain-of-function screen by a

heterologous expression approach in stress sensitive yeast strains of *Saccharomyces cerevisiae* (Chapter 4) where they were selected based on resistant or sensitive phenotypic response to one or more of the abiotic stress conditions analysed (salt, drought, oxidative and temperature stress).

Table 6: **List of orphan gene overexpression constructs made in this study**

Gene ID	Gene name	Primer name	Gateway primer sequence (5'-3')
AT1G22890	*ASR-1*	1_AT1G22890_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTGCAAAAA CATCAAATTTAG
		1_AT1G22890_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATGGGCCTTGTC CTTTG
AT1G47400	*ASR-2*	2_AT1G47400_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGATGTCTTTTG TCGCAAAC
		2_AT1G47400_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACGCAGCAGGA GCATAATC
AT1G65500	*ASR-4*	4_AT1G65500_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGTATGAAGT CTCCAAATATTG
		4_AT1G65500_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAGGAGATTGGT CAGAAG
AT3G11745	*ASR-7*	7_AT3G11745_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCGCAGCCGA ATCTTCTAAG
		7_AT3G11745_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTTTAATTCTTC CAG
AT3G22231	*ASR-10*	10_AT3G22231_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAATCAATCCG CGCAAAATTAC
		10_AT3G22231_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACTCTGATGTAC AGAG
AT3G45730	*ASR-11*	11_AT3G45730_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGCTTCCAT CTCCTTAC
		11_AT3G45730_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAGCCTCCTCTCT TACC
AT3G56260	*ASR-12*	12_AT3G56260_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCAGAAAGAC TAGAAG
		12_AT3G56260_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAGCTTTATTCA TGTC
AT4G22640	*ASR-15*	15_AT4G22640_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTTACACAA ACAAAG
		15_AT4G22640_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAAAGATACCTT TACATG
AT1G31580	*ASR-35*	35_AT1G31580_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCATCTTCTA TAGTC
		35_AT1G31580_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGTGACTTGGTG AGTTTTTTG
AT1G64360	*ASR-39*	39_AT1G64360_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCGTTTAGAA AAGTAG
		39_AT1G64360_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGTAATTAGGGA AGTG
AT1G65510	*ASR-40*	40_AT1G65510_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGTTCGAAGA CATCAAC
		40_AT1G65510_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAGGAGAGTTGT AAGAAG
AT1G67860	*ASR-41*	41_AT1G67860_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCTTGATACGC TTATTG
		41_AT1G67860_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTCCCTGTCTT TCTG
AT1G67865	*ASR-42*	42_AT1G67865_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCTTGATAAGC TTATTATTGG
		42_AT1G67865_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACTCCCTGTCTT TCTG
AT2G04800	*ASR-46*	46_AT2G04800_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGATCTAAAC CTAGC
		46_AT2G04800_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAACTATAAAACC GTGTC
AT2G25510	*ASR-50*	50_AT2G25510_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGATGCACGCTT TACTATTC
		50_AT2G25510_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTTATTATAGC GGCC

AT2G28570	*ASR-51*	51_AT2G28570_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGTCTGTGAAGACCAAATCCTC
		51_AT2G28570_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAGTGGATTCGTTGTTC
AT2G41650	*ASR-52*	52_AT2G41650_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCGGCAGCGAGCTTG
		52_AT2G41650_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGGGCTGCTCCATACG
AT2G47200	*ASR-53*	53_AT2G47200_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCGTCGATAGTGATG
		53_AT2G47200_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACATCAAGTGATTCCTG
AT3G58540	*ASR-60*	60_AT3G58540_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCGGGTGCAGTAGCAAG
		60_AT3G58540_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACTCACTCTCCTCAATG
AT4G18580	*ASR-63*	63_AT4G18580_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCAACATTGGCACCAG
		63_AT4G18580_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACACAACAAACCACTC
AT4G27654	*ASR-66*	66_AT4G27654_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGACGTCGTTAGAATTGG
		66_AT4G27654_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGCTGTCACTCTTC
AT4G31875	*ASR-69*	69_AT4G31875_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGGAGACGAGCGAG
		69_AT4G31875_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAACGTTGGCGTCGAGAAG
AT4G39675	*ASR-70*	70_AT4G39675_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCGTCCATTGGAGCAAG
		70_AT4G39675_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACTTTTGTGATTTGAAC
AT5G44580	*ASR-88*	88_AT5G44580_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAAAGAAAAAAGTTTTC
		88_AT5G44580_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAGGGGGCTGGAGTGCG

The transformed plants were selected on hygromycin (25 mg/L) antibiotic in the T1 generation and confirmed for the presence of the construct by PCR (Fig.17). To select the single copy transgenic lines in the T1 generation we followed a qRT-PCR based method described by Chen and Lin (2010) as an alternative to southern blotting. The technique utilizes an endogenous calibrator (any gene with known copy number) and threshold cycle of the target sequence and the calibrator to calculate the copy number. The selected plants were further investigated for overexpression of the orphan gene by semi-quantitative RT-PCR and advanced to T3 generation for getting the homozygous transgenic lines. The single copy insertion lines were further confirmed in T2 by segregation analysis for the antibiotic hygromycin. The values for single copy insertion by qRT-PCR varied from 0.5 to 1.6 (with few exceptions of up to 1.9) that correlated with 3:1 segregation ratio (Table 7). At least two lines per gene construct were selected in order to use for the stress screening. The abiotic stress screening was performed for the transgenic lines of 15 genes out of the lines generated

for 24 genes. The complete list of transgenic lines selected for screening are mentioned in Table 7.

Figure 17: Schematic representation of the expression cassette used for ORF expression. General representation of primer binding sites for PCR confirmation of transgenic plants is mentioned (P1 and P2). P1-vector specific forward primer; P2-gene specific reverse primer.

3.2.2 Overexpression of the non-orphan *ASR51* gene causes severe morphological defects in *Arabidopsis thaliana*

One of the advantages of gain-of-function mutation is that the phenotype caused by the mutation segregate as dominant traits (Bouche and Bouchez, 2001; Weigel et al., 2000). Hence any visible phenotype can be easily identified in the first transgenic generation (T1 generation). All the overexpression lines were closely monitored for visible phenotypes in the T1 generation. Overexpression of the gene *ASR51* showed a very striking phenotype of dwarfism (Fig. 18A). More than 50% of the hygromycin selected PCR confirmed T1 plants showed a dwarf phenotype although the severity of phenotype varied among the independent transgenic lines. Variation in the phenotype is common in independent transgenic lines due to varied levels of expression of the transgene. The semi-quantitative RT-PCR from RNA extracted from rosette leaves of the independent transgenic lines showed that the severity of phenotype correlates with the expression level of *ASR51* (Fig. 18B) suggesting that the overexpression of *ASR51* is the direct cause of the visible phenotype.

Table 7: Data on transgenic lines selected for screening

Gene ID	Gene name	Selected T1 plant No.	Copy number (calculated based on qRT-PCR)	T2 segregation ratio (HygR:HygS)	T2 Homozygous line No
AT1G22890	*ASR1*	1.1	NC	2.9:1	*ASR1.1.7*
		1.5	NC	2.3:1	*ASR1.5.2*
AT1G47400	*ASR2*	2.3	1.0	2.5:1	*ASR2.3.5*
AT3G11745	*ASR7*	7.2	0.98	2.7:1	*ASR7.2.1*
		7.3	0.84	2.5:1	*ASR7.3.1; ASR7.3.2*
AT3G56260	*ASR12*	12.7	1.53	3.2:1	*ASR12.7.2 ; ASR12.7.5*
		12.8	0.60	2.9:1	*ASR12.8.4*
AT1G31580	*ASR35*	35.1	1.50	3.1:1	*ASR35.1.1*
		35.11	0.94	2.1:1	*ASR35.11.3; ASR35.11.5*
AT1G65510	*ASR40*	40.1	0.66	3.3:1	*ASR40.1.3;ASR40.1.5*
		40.2	0.58	3.3:1	*ASR40.2.1*
AT1G67860	*ASR41*	41.6	1.14	2.5:1	*ASR41.6.2*
		41.7	1.20	2.9:1	*ASR41.7.2*
AT1G67865	*ASR42*	42.9	1.27	3.6:1	*ASR42.9.2;ASR42.9.3*
		42.11	1.96	3:1	*ASR42.11.2; ASR42.11.6*
AT2G25510	*ASR50*	50.2	NC	2.9:1	*ASR50.2.4*
AT2G28570	*ASR51*	51.7	0.77	3.4:1	NS
		51.21	NC	NC	NS
		51.22	NC	NC	NS
AT2G41650	*ASR52*	52.1	0.78	3.2:1	*ASR52.1.5*
		52.4	1.86	3.6:1	*ASR 52.4.3;ASR52.4.4*
AT2G47200	*ASR53*	53.3	1.9	3.6:1	*ASR53.3.3; ASR53.3.5*
		53.6	2.7	2.9:1	*ASR53.6.5*
AT3G58540	*ASR60*	60.1	0.59	4.4:1	*ASR60.1.2*
		60.6	0.58	4.6:1	*ASR60.6.7 ; ASR 60.6.8*
AT4G18580	*ASR63*	63.3	0.51	3.5:1	*ASR63.3.2*
		63.5	0.53	2.9:1	*ASR63.5.1*
AT4G31875	*ASR69*	69.1	-	2.6:1	*ASR69.1.2;ASR69.1.5*
		69.2	-	3:1	*ASR69.2.2;ASR69.2.7*
		69.4	-	3.1:1	*ASR69.4.3*
AT5G44580	*ASR88*	88.5	0.65	3.1:1	*ASR88.5.2;ASR88.5.4*

The copy number and segregation data for the transgenic lines screened for abiotic stress are mentioned. The copy number was calculated as described in material and methods section. NC- not calculated; NS-not selected. T2 homozygous lines selected based on segregation in T3.

A)

B)

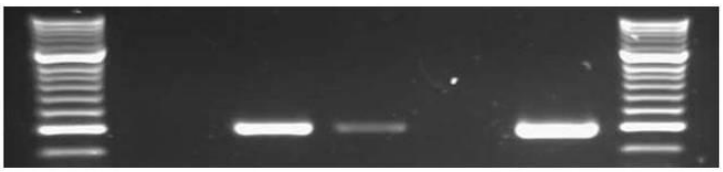


Figure 18: Effect of *ASR51* overexpression *Arabidopsis thaliana*. A) Dwarf phenotypes of *ASR51* overexpression lines (*ASR51.21* and *ASR51.22*) compared to wild type Col-0. B) Semi-quantitative RT-PCR to check the overexpression of *ASR51*. Severity of phenotype correlates with the expression level of the gene. WT- cDNA from wild type plant; *ASR51.21* and *ASR51.22* – cDNA from two independent transgenic lines; gDNA- genomic DNA control.

3.2.3 Overexpression of the non-orphan *ASR12* gene generates a salt tolerant phenotype

ASR12 (AT3G56260.1) was described as Brassicaceae specific gene in our previous study which was based on TAIR version 8, but the updated gene model (AT3G56260.2) in the TAIR version 10 leads to the detection of homologous protein sequences in species outside of Brassicaceae. Hence this gene is considered a non-orphan gene (Table 2). The expression analysis of this gene in our previous study revealed that this gene was upregulated under salt and osmotic stress conditions (Fig 19). The overexpression line was analysed for a range of abiotic stress conditions (as mentioned in methods section 3.5.2) to elucidate the stress responses of this gene. The overexpression lines of *ASR12* showed a salt tolerant phenotype with better root growth compared to wild type under salinity stress.

Figure 19: Expression of *ASR12* under salt and mannitol stress. Screenshot from Arabidopsis eFP browser. *ASR12* is induced after 3 hours of salt and mannitol treatment. The expression intensities are depicted in colour scale bar (right).

To investigate salt stress tolerance, 7 day old seedlings of *ASR12* overexpression lines and wild type were exposed to salt stress conditions for another 7 days (150 mM salt on 1xMS media), and the root length was scored as stress responsive phenotype. The overexpression lines showed higher relative root length percentage (15-16.5%) compared to wild type

(9%) under 150 mM salt (Fig-20). However, there was no difference in phenotypic responses to mannitol treatment.

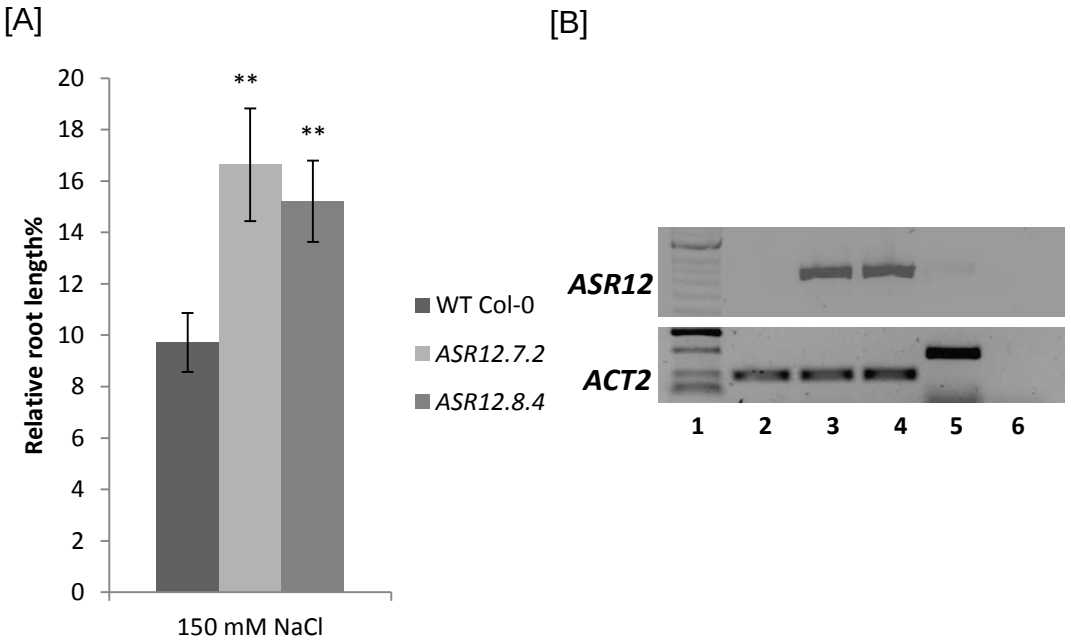


Figure 20: Abiotic stress screening of *ASR12* overexpression lines. [A] Salt stress response of *ASR12* overexpression lines. Two independent lines tested (*ASR12.7.2* and *ASR12.8.4*) show higher relative root growth compared to wild type control. Asterisks indicate significant differences between samples as determined by student's t test (** P < 0.01). [B] Semi quantitative RT-PCR to confirm the overexpression of *ASR12* in the transgenic lines. *Actin2* was used as internal control. 1 – DNA marker; 2 – Wild type; 3 – ASR12.7.2; 4 – ASR12.8.4; 5 – Wild type genomic DNA; 6 – No template control.

3.2.4 Overexpression of the Brassicaceae specific gene *ASR35* leads to salt stress tolerance

ASR35 is annotated as *ECS1*, encoding a cell wall protein (TAIR). This gene was found to be linked to a locus influencing resistance to *Xanthomonas campestris pv.campestris* (Xcc750)(Aufsatz et al., 1998). In that study, *ECS1* displayed an ecotype specific expression pattern. The expression analysis in our previous study showed that *ASR35* is upregulated under abiotic stress (UV-B, salt, oxidative and heat stress) and downregulated under osmotic stress (Fig-4) (Donoghue et al., 2011). We analysed an *ASR35* overexpression line for salt and osmotic stress responses. Although we did not see any phenotypic responses to osmotic

stress, the overexpression line (*ASR35.11.5*) showed tolerance to salinity in terms of better root growth compared to wild type. At 100 mM salt concentration *ASR35.11.5* showed no reduction in root length whereas wild type showed 25% reduction in root growth. At 150 mM salt, *ASR35.11.5* root growth was significantly higher (34%) than the wild type (19.5%) suggesting a salt tolerant phenotype (Fig- 21A).

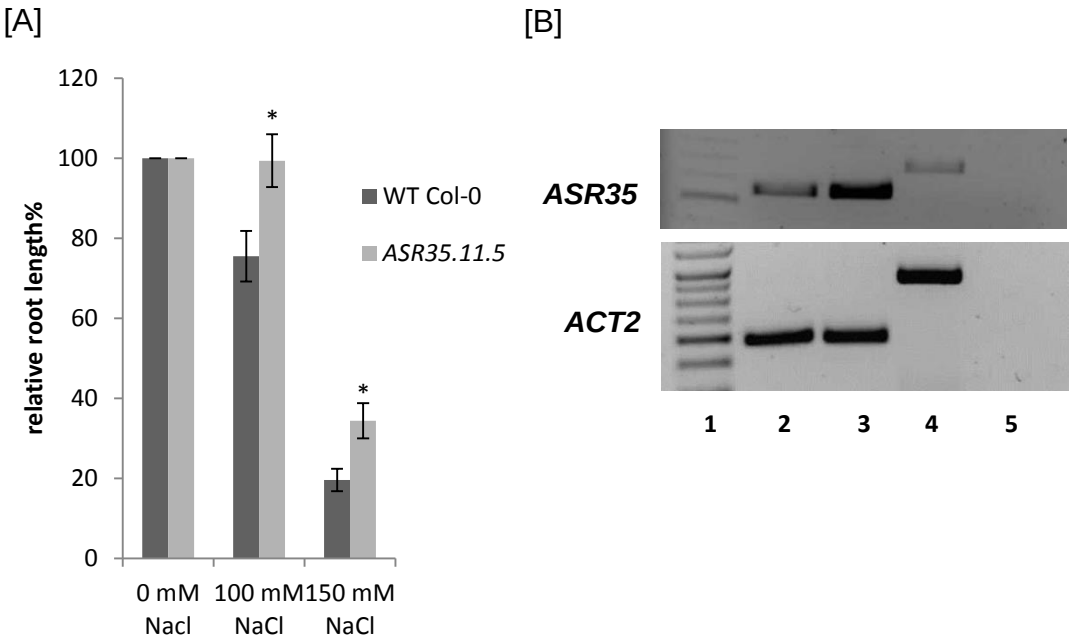


Figure 21: Abiotic stress screening of *ASR35* overexpression line. [A] Salt tolerant phenotype of *ASR35* overexpression line scored in terms of relative root growth. Asterisks indicate significant differences between samples as determined by t test (* $P < 0.05$). [B] Semi quantitative RT-PCR to confirm the overexpression of *ASR35* in the transgenic lines. *Actin2* was used as internal control. 1 – DNA marker; 2 – Wild type; 3 – ASR35.11.5; 4 – Wild type genomic DNA; 5 – No template control.

3.2.5 Overexpression of *ASR63* confers tolerance to iron deficiency conditions

Abiotic stress screening was extended to nutrient stress to determine whether any of the overexpression lines show responses to nutrient deficiency conditions. Co-expression analysis of the orphan genes using Genevestigator showed that expression of some of the orphan genes was correlated with expression of glutathione-s-transferase and thioredoxin family of proteins which are known to be involved in nutrient deficiency responses

apart from their role in abiotic stress. Orphan overexpression lines were screened under nitrogen, phosphorous, potassium and iron (Fe) deficiency conditions. In the screen for Fe deficiency, *ASR63* overexpression lines displayed improved growth phenotypes compared to wild type control.

To screen for iron deficiency responses, 7 day old seedlings of both wild type and transgenic lines were transferred to iron deficient (10 µM Fe) and iron sufficient conditions (100 µM Fe) at pH 5.7, pH 5.0 and pH 4.2. The seedlings were observed for growth phenotypes from 7 days after transfer. After 14 days in the iron deficient media at pH4.2, wild type seedlings started to show chlorosis symptoms while *ASR63* lines showed normal growth. Wild type plants also displayed slow growth under Fe deficient conditions at pH 4.2 (Fig-22A). To quantify the growth phenotype, total chlorophyll content was measured. The chlorophyll content of *ASR63* overexpression lines in the iron deficient conditions were similar to that of iron sufficient conditions in all three pH ranges analysed. However, wild type seedlings had significantly lower chlorophyll content compared to the transgenic lines under iron deficient conditions at pH 5.0 and pH 4.2 (Fig-23).

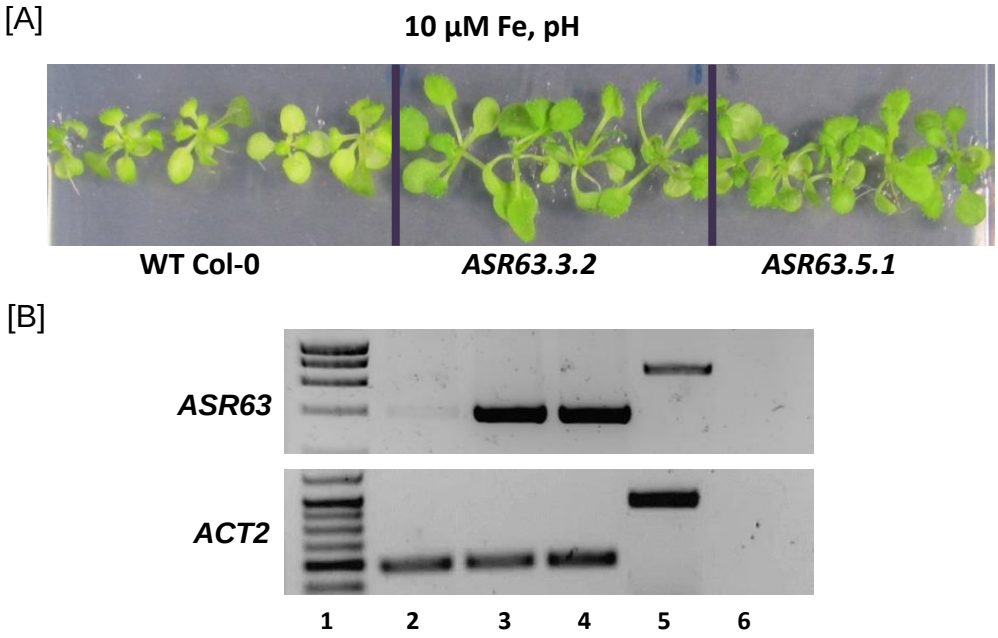


Figure 22: Iron deficiency response of *ASR63* overexpression lines. [A] Iron deficiency phenotype of *ASR63* overexpression lines under iron deficient condition at pH 4.2. [B] Semi quantitative RT-PCR to confirm the overexpression of *ASR63* in the transgenic lines. *Actin2* was used as internal control. 1 – DNA marker; 2 – Wild type; 3 – ASR63.3.2; 4 – ASR63.5.1; 5 – Wild type genomic DNA; 6 – No template control.

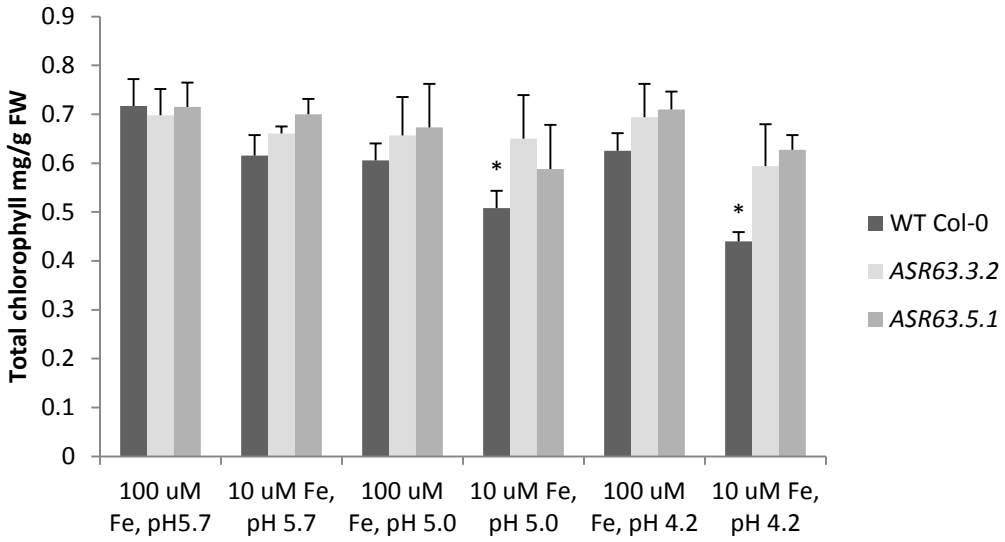


Figure 23: Total chlorophyll content of *ASR63* overexpression lines under iron deficiency. Seven day old seedlings were transferred to indicated media compositions and grown for 14 more days before collecting shoot tissue for chlorophyll estimation. Error bars indicate standard error of mean. Asterisks indicate significant differences from the control (* $P < 0.05$) as determined by the ANOVA, comparisons with control using Dunnett's method. Chlorophyll content of the wild type seedlings grown on 100 µM Fe at pH 5.7 was considered as control for comparison.

3.3 Discussion

3.3.1 Gain-of-function screen of novel Brassicaceae specific orphan genes in *Arabidopsis thaliana*

Gain of function overexpression screening allows analysis of genes belonging to gene families which is not possible through loss of function screen as single gene mutations may not show clear phenotypes due to genetic redundancy (Kondou et al., 2010; Nakazawa et al., 2003). Here we employed an open reading frame (ORF) overexpression approach for orphan genes to elucidate the effects on abiotic stress tolerance or sensitivity.

Overexpression of *ASR51* caused a severe dwarf phenotype (Fig. 17A). The phenotype was characterized by malformed aerial parts of the plants, with the rosette showing bunchy appearance with small crinkled leaves. Plants with severe phenotype shows shortened inflorescence stems.

Even though flowers appeared as normal, the siliques were shorter than those of wild type. Many dwarf mutants in *Arabidopsis* have previously been described, the majority of which are phytohormone related mutants. Quite similar to *ASR51* overexpression lines are the brassinosteroid insensitive mutants (Clouse et al., 1996; Li et al., 2001a; Mussig and Altmann, 2001). Hormone insensitive mutants usually have defects in proteins involved in hormone signalling and hence cannot be complemented by exogenous application of hormone. On the other hand hormone deficient mutants can be complemented by external hormone application as they have defects in proteins involved in hormone biosynthesis. For example gibberellin deficient mutants also show dwarf phenotypes and the phenotype can be complemented by spraying with gibberellic acid (Finkelstein and Zeevaart, 1994). The next step experiments that need to be done with the *ASR51* gene to elucidate its functionality are; (a) exogenous application of hormones such as gibberellic acid or brassinosteroid to *ASR51* overexpression lines to identify any potential link with these hormone biosynthetic pathways. (b) To analyse microarray datasets to find potential leads towards a developmental pathway (c) to generate and analyse knock out or knock down lines for *ASR51* to see if it shows any morphological defect (d) To generate transgenic *ASR51* with its own promoter to analyse any dosage effects of *ASR51* on plant morphology.

Although overexpression of *ASR51* had an interesting phenotype, these lines were not considered for further study as it was not Brassicaceae specific. However, our results demonstrate that *ASR51* is a novel gene potentially involved in plant development.

3.3.2 Overexpression of the genes *ASR12* and *ASR35* leads to enhanced salt stress tolerance in *Arabidopsis thaliana*

We identified two genes (*ASR12* and *ASR35*) in this research which potentially involved in salt tolerance in *Arabidopsis*. *ASR12* is not a Brassicaceae specific gene and from protein blast analysis we found that *ASR12* shows significant homology with proteins of non-brassicaceae

species including monocots like brachypodium, maize and wheat which are all annotated as hypothetical or uncharacterised proteins. Microarray data on *ASR12* response to abiotic stress indicated that it is up regulated under salt and mannitol treatment (Donoghue et al., 2011). However the overexpression lines for *ASR12* showed a tolerant phenotype to salt treatment but not to mannitol. Although the immediate responses (within few hours of treatment) to salt treatment are similar to that caused by non-ionic solutes (e.g. mannitol), the longer-term responses (from days to weeks) are purely ionic stress or salt specific (Munns and Tester, 2008). The response of *ASR12* lines to salt but not to mannitol suggest that the phenotype observed under salt stress is likely to be a salt-specific response but not an osmotic response.

The salt specific responses need to be confirmed by analysing the overexpression lines under transpiring conditions. It has been found that ionic stress responses, such as Na^+ toxicity in shoots, may not be manifested by the plants under non-transpiring conditions (i.e. on MS plates). Because low transpiring conditions do not allow build-up of high levels of Na^+ in the shoots, the response of shoot growth under salt accumulation may not be observed in this experimental set up (Verslues et al., 2006).

The follow on experiments that need to be done for *ASR12* are as follows: (a) elucidation of its evolutionary origins and distribution; (b) Comparative analysis under salt stress of seedlings of knockouts (-/-), wild type (+/+) and *ASR12* overexpression lines on MS plates; (c) Comparative analysis under salt stress of mature plants of knockouts, wild type and *ASR12* o/e lines on soil; (d) Use of yeast heterologous system to identify possible pathways the *ASR12* gene may be interacting with.

ASR35 (*ECS1*) is a brassicaceae specific orphan gene conferring tolerance to salt stress in *A. thaliana*. Similar to *ASR12* lines, *ASR35* overexpression also displayed a salt tolerant phenotype. Aufsatz and Grimm (1994) initially identified *ASR35* as a pathogen inducible gene against *Xanthomonas campestris pv. Campestris.* However subsequent analysis

found that *ASR35* is not the Xcc750 resistant gene itself but linked to a locus influencing resistance to *Xanthomonas campestris pv. campestris* (Xcc750) (Aufsatz et al., 1998). Microarray expression analysis for abiotic stress conditions revealed that the *ASR35* gene is up regulated for stress caused by wounding, UV-B, salt, heat, oxidative or drought. The *ASR35* gene also exhibits expression down regulation for osmotic stress and is differentially expressed (both up and down regulation at different time points) under cold stress (Donoghue et al., 2011). In this study, the overexpression line for *ASR35* displayed a salt tolerant phenotype. However, no differential phenotypes were observed under osmotic, heat and cold treatments.

This gene is described as a cell wall protein (TAIR) based on sub-cellular localisation experiments. Proteins associated with cell walls would be of special interest for stress related traits as cell wall are a primary barrier and first to come in contact with many environmental stimuli. Cell wall associated proteins responsive to stress could be potential candidates as cell wall 'sensors'. For instance, several wall associated kinases (WAKs) have been widely studied in *Arabidopsis* (Decreux and Messiaen, 2005) and the association of WAKs with pathogen associated genes is well established (Meier et al., 2010).

The follow on experiments that need to be done for *ASR35* are as follows: (a) Elucidation of its evolutionary origins and distribution; (b) Comparative analysis under salt stress of seedlings of knockout lines, wild type and *ASR35* overexpression lines on MS plates; (c) Comparative analysis under salt stress of mature plants of knockouts, wild type and *ASR35* overexpression lines on soil; (d) Use of yeast heterologous system to identify possible pathways the *ASR35* gene may be interacting with.

3.3.3 *ASR63* confers tolerance to iron deficient conditions in *A. thaliana*

Plant growth and performance depends on availability of sufficient amount of nutrients. Plants have developed sophisticated uptake pathways and distribution mechanisms for mineral nutrients. Nutrient deprivation leads to activation of signalling cascades of nutrient deficiency responses and also causes production of reactive oxygen species. The redox signals, the

indicators of cellular redox status and the antioxidant buffering systems are important in these responses (Kandlbinder et al., 2004). Gene expression correlation analysis using the Genevestigator tool revealed that *ASR2* and *ASR63* expression levels are correlated with expression levels of ferritin, glutathione and glutaredoxin family genes. Hence, orphan overexpression lines were also analysed for nutrient deficiency responses. From the nutrient deficiency screen *ASR63* overexpression led to enhanced tolerance to iron (Fe) starvation. *ASR63* lines were able to maintain higher chlorophyll content under Fe deficiency conditions. Because Fe is an important component of chlorophyll biogenesis, higher chlorophyll content under Fe deficient conditions suggests that *ASR63* lines are accumulating levels of Fe similar to that of Fe sufficient conditions. Under low pH Fe is more soluble and the better growth of *ASR63* lines under low Fe and low pH suggests efficient Fe uptake in *ASR63* lines compared to wild type control. The response of *ASR63* lines to Fe deficiency conditions was further investigated in detail in chapter 5.

3.4 Conclusions

The overexpression strategy for dissecting the gene function is a powerful tool which can identify the gene function for redundant genes which is not possible by loss of function screens. The current overexpression screen of 16 genes (14 orphans, 2 non orphans) has identified potential novel functions for two *Arabidopsis* orphan genes, one for salt tolerance (*ASR35*) and the other for tolerance to iron deficiency (*ASR63*). Another two non-orphan genes, *ASR51* and *ASR12* were found to be potentially involved in plant development and salt stress tolerance respectively. From the functional evolutionary genomics perspective it is necessary to know how all of these genes acquire novel functions and get integrated with the conserved pathways. Some orphan genes may acquire functions conferring improved adaptation to environment. Analysis of *Arabidopsis* orphan genes for stress related traits in other species (plants and others) could help to identify novel functionalities of these genes and possibly they can be used in translational research for crop improvement and biotechnology.

Chapter 4: Gain of function screening of Brassicaceae specific *Arabidopsis thaliana* orphan genes in Yeast (*Saccharomyces cerevisiae*) for abiotic stress tolerance

4.1 Introduction

4.1.1 Yeast, a eukaryotic model system

Yeasts belonging to the kingdom fungi and have been cultivated by humans for millenia for making bread, beer and wine. The budding yeast, *Saccharomyces cerevisiae*, is the most widely studied yeast although other species like the fission yeast, *Saccharomyces pombe* is also well studied and used as model systems. Since *S. cerevisiae* is used for fermentation, its study took more importance in the early days for metabolic research and also genetic studies. Like other model organisms yeast also has several advantages, such as non-pathogenicity, ease of propagation and manipulation in the laboratory and, from the point of genetic studies, mutant screening and segregation analysis is easier in yeast (Mell and Burgess, 2001).

Another reason for *S. cerevisiae* to become one of the most popular eukaryotic model organisms is that it has several 'firsts' in the eukaryotic world. It was the first eukaryote to be transformed by plasmids, it was also the first eukaryote for which precise gene knockouts were constructed, (Forsburg, 2001) and most importantly the first to have its genome sequenced (Goffeau et al., 1996). Yeast carries a compact genome of 12 megabase pairs with circa. 6000 genes spread across 16 linear chromosomes. Yeast genes are relatively smaller and with less introns and the genome has shorter intergenic regions compared to higher eukaryotes. In addition, all yeast share a similar cellular architecture and life cycle with multicellular eukaryotes such as plants and animals. Also many of the genes in yeast are related evolutionarily and functionally to genes in higher eukaryotes (Mell and Burgess, 2001). Overall, this makes yeast an attractive system to study eukaryotic biology.

The life cycle of yeast alternates between haploid and diploid states. The haploid cells divide by budding to produce identical daughter cells. The two haploid cells can also mate and fuse to produce a diploid. The diploid

cell can divide by budding or undergo meiosis to produce four haploid cells. The haploid yeast cells have two mating types called ‘a’ and ‘α’, each can only mate with the opposite mating type. Each mating type produces a pheromone detected by the opposite mating type and the mating types are governed by the locus *MAT,* accordingly mating types are named as either *MATa* or *MATα* (Bardwell, 2005). The mating requires a gene called HO endonuclease and mutating this gene prevents mating of the haploid cells. Most laboratory strains have mutations in HO endonuclease so that haploid cell can be stably maintained for genetic experiments.

The haploid life cycle of yeast is particularly important as this facilitates the recovery of recessive mutations, usually loss-of-function alleles which help to dissect gene function. In addition, the diploid cells can undergo meiosis like other sexual eukaryotes. Hence it is possible to study cell cycle control and phenomena like homologous recombination which are known to be highly conserved across eukaryotic taxa. This makes yeast a good model for studying fundamental eukaryotic biology.

Another important aspect of yeast which makes it a popular eukaryotic model system is the presence of extra chromosomal DNA, in particular a plasmid called the 2-micron circle. This plasmid contains protein coding genes known to code for proteins required for its high copy maintenance and segregation but has little apparent function for yeast (Mell and Burgess, 2001). This plasmid has been exploited by fusing with *Escherichia coli* plasmids to form shuttle vectors which allow it to be propagated and maintained in both bacteria and yeast. This, along with development of tools for gene isolation and cloning, facilitated analysis of genes and proteins in the yeast system. Genetic techniques like yeast-two hybrid assay (which makes it possible to find the interacting partner of a protein by utilizing multiple plasmids with genes for fusion proteins) has been used to test proteins of interest from any organism (Bartel and Fields, 1997).

With all these advantages, it is possible to use a wide range of genetic tools in yeast from classical genetics to mutant isolation to plasmid based

screens for functional analysis to understand basic cellular processes, genetic networks, metabolic processes and even human diseases. The utilization of yeast in scientific research and its contribution to science is immense and is reviewed in detail elsewhere (Botstein and Fink, 2011; Forsburg, 2001; Mell and Burgess, 2001). From the point of objective of this chapter, the gene overexpression strategy using yeast as a heterologous (meaning 'derived from different organism') system for gene function characterization is given the main focus here.

4.1.2 Heterologous overexpression as a tool for placing genes within genetic pathways

As mentioned in chapter 3, gain of function screens by overexpression is a popular tool for dissecting gene functions. The utilization of gene expression as a screening tool started only after the development of the yeast transformation techniques and construction of genomic libraries (Prelich, 2012). Gene expression in yeast became relatively simple due to the stable maintenance of the 2µ vector-based plasmids. Many types of plasmid-based vectors have been developed, from low copy to high copy, from unstable to stable and for accommodating a variety of insert sizes. This enables cloning of genes from any organisms into these vectors so they can be tested in yeast as a heterologous system.

A major advantage of using a yeast heterologous system is that it enables the study of gene functions across species barriers. One of the earliest examples of exploiting a heterologous system that crossed prokaryotic-eukaryotic boundaries was the isolation of *S. cerevisiae LEU2* and *HIS3* genes from a random genomic library by their ability to complement recessive mutations in *Escherichia coli leuB* and *hisB*, respectively (Ratzkin and Carbon, 1977; Struhl and Davis, 1977). Since then yeast has become a favourite heterologous host for understanding gene functions, finding deleterious or toxic effects of genes and also for the production of high value proteins (Yesilirmak and Sayers, 2009).

4.1.3 Mechanisms causing overexpression phenotypes

Similarly to gene knockouts or knockdowns, gene over-expression can also result in various types of mutant phenotypes which can fall into categories such as hypermorphic, hypomorphic, antimorphic or neomorphic effects according to Muller's classic criteria (Muller, 1932). From a number of overexpression studies over the years, there is now a more concrete understanding of mechanisms by which gene overexpression leads to different phenotypes. In a recent review Prelich (2012) described typical overexpression phenotypes separating them into two broad categories. They are (a) inhibition mechanisms and (b) activation mechanisms that target a protein, a protein complex or a whole pathway. These mechanisms are summarised here from the review.

One way that overexpression inhibits another protein is simply by reducing its steady state level. This can be caused by affecting the gene expression at the point of transcription or translation, or also by increasing the rate of its degradation. For example, overexpression of mammalian E3 ligase MKRN1 will lead to degradation of the hTERT telomerase subunit through proteolysis (Fig.24 panel I) (Kim et al., 2005). Another mechanism of inhibition is by competition with other macromolecules which can disrupt multiprotein complexes (Fig.24 panel II). A classic example for this is the overexpression of histone proteins H2A-H2B or H3-H4 pairs which lead to aberrant chromosome segregation and gene expression defects and can be rectified by co-overexpressing all four histone gene pairs (Clark-Adams et al., 1988; Meeks-Wagner and Hartwell, 1986). Sometimes overexpressed protein may compete for a shared subunit of two protein complexes resulting in inhibition of protein complex (Fig.24 panel III). The last type of inhibitory mechanism is a functional inactivation where the specific activity of a protein is reduced and has nothing to do with the protein level itself (Fig.24 panel IV).

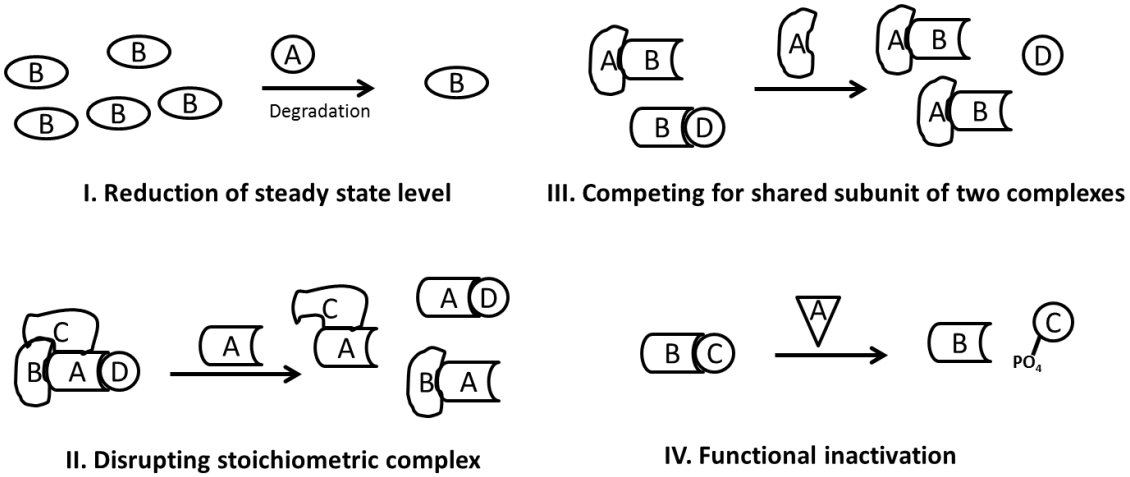


Figure 24: Mechanisms of inhibition by overexpression. Figure modified from Prelich (2012).

Activation of a pathway is another category by which overexpression can cause phenotypic effects. A simple example is the triggering of a pathway by expression of a key regulator (Fig.25 panel I). For example, overexpression of MyoD in mouse results in differentiation of fibroblasts into muscle (Davis et al., 1987) and overexpression of the *eyeless* gene causes development of ectopic eyes in *Drosophila* (Halder et al., 1995). Another way activation works is when overexpression increases the total activity of the protein beyond a threshold resulting in mutant phenotype (Fig.25 panel II). This has been exploited in identifying drug targets by adding the drug at a suboptimal dose, such that its target's activity becomes limiting, which is then overcome by overexpressing its target protein (Rine et al., 1983). Sometimes a pathway may be in an inactive state due to an inhibitor. Such pathways can be activated by overcoming or counteracting the inhibitor either by blocking its expression or causing its degradation through overexpression (Fig.25 panel III). The best example for this is the transcriptional activator Gal4 which is maintained in an inactive state by the Gal80 repressor. This can be overcome by overexpression of Gal4 itself (Nogi et al., 1984) or by overexpression of Gal3 which binds directly to Gal80 and releases the repression (Suzuki-Fujimoto et al., 1996). Another mechanism for overexpression effects is by increasing the total activity of the protein through post translational modification leading to an increase in the specific activity of the protein (Fig.25 panel IV). This has been found in stimulation of

transcriptional activity of estrogen receptor by overexpression of Ras-MAPK cascade (Kato et al., 1995). Finally, overexpression can occasionally lead to neomorphic phenotypes leading to a new function (by activating an unrelated pathway).

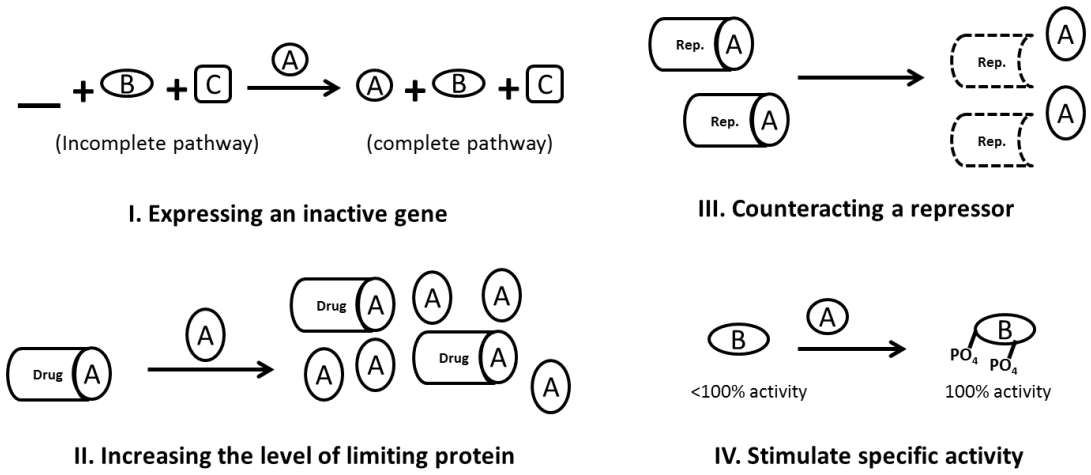


Figure 25: Mechanisms of activation by overexpression.
Figure modified from Prelich (2012).

4.1.3 Heterologous expression of plant genes in yeast

Expression of genes in a host other than the original source of the gene is referred to as heterologous expression. The development of yeast (*S. cerevisiae*) as a heterologous host dates back to yeast complementation experiments when development of shuttle vectors allowed for complementation of the yeast *leu2* mutation with *E. coli* DNA (Beggs, 1978). The first heterologous complementation in yeast using eukaryotic DNA was of *Drosophila melanogaster* cDNA library in yeast *ade-8* mutant (Henikoff et al., 1981). Since then, heterologous expression in yeast has been successfully used for expression of genes from various organisms for identifying gene functions and pathways. Yeast host systems used for heterologous expression are *S. cerevisiae*, *S. pombe* and also the methylotropic yeast *Pichia pastoris*. The *P. pastoris* system is preferred for large scale production of high value proteins and also for generating large amount of proteins for structural studies. On the other hand *S. cerevisiae* is a preferred host for understanding gene and protein function studies.

The molecular, genetic, and biochemical characteristics of yeast are similar to higher eukaryotes (Yesilirmak and Sayers, 2009). Although heterologous expression of plant genes in yeast was attempted in 1982, success only came a decade later when two plant cDNAs were functionally expressed to confirm their assumed function (Alderson et al., 1991; Hamilton et al., 1991). Since then yeast has become a preferred host for heterologous expression of plant genes. The largest group of plant proteins expressed and studied in yeast (in both *S. cerevisiae* and *S. pombe*) to date are membrane transporter proteins. *S. cerevisiae* mutants provide a convenient system for functional and kinetic studies of transporters (Gaber et al., 1999). Yeast heterologous system has also played a major role in identifying functionality of plant stress related proteins. It is already known that, at the cellular level, the mechanisms of stress tolerance are similar in yeast and plants (Mulet et al., 2004).

Many of the proteins involved in transport of Na^+, K^+, Ca^{2+} and Cl^- have been determined by functional complementation of transport deficient yeast mutants (Gaber et al., 1999). Functional expression of potassium transporter genes like *AtKT1* and *AtKT2* from *Arabidopsis* (Quintero and Blatt, 1997) and *HKT1* from wheat (Schachtman and Schroeder, 1994) also utilised *S*. *cerevisiae* mutants. Osmotic stress sensitive yeast mutants also helped to identify the determinants of plant stress tolerance. A putative mitogen-activated protein kinase (MAPK) was identified from *Pisum sativum* with 47% similarity to yeast Hog1p. The Hog1p is a MAP kinase in the yeast osmoregulatory pathway which controls glycerol accumulation (Popping et al., 1996). In addition to transporter proteins, the use of yeast has been pivotal in identifying proteins in plant signalling pathways such as calcium signalling and salt stress responses. For example, *Arabidopsis SAL1* was identified by complementation of the yeast ENA ATPase activity mutant (Quintero et al., 1996). Most of these studies took advantage of yeast episomal vectors as they can give high copy expression of the genes.

Gain of function screening in yeast by gene overexpression using episomal vectors has been shown to be a successful approach (Serrano,

1996) for functional characterization of genes, including genes related to abiotic stress responses. Abiotic stress screening in yeast is faster than in plants and, since abiotic stress sensitive mutant strains are available in yeast, plant stress responsive genes can be characterized by overexpression in yeast followed by functional selection approaches. This approach is of great advantage for genes from non-model organisms where genetic transformation studies are not yet possible. Abiotic stress tolerance genes from crop plants like pigeon pea (Priyanka et al., 2010) and Jatropha (Eswaran et al., 2010) have been identified using this approach.

In this chapter, yeast functional screens were used for *A*. *thaliana* orphan genes conferring abiotic stress tolerance using stress sensitive yeast strains. This work was done in collaboration with Prof. JM Mulet in IBMCP, Valencia, Spain.

4.2 Results

4.2.1 Gain of function screen of *Arabidopsis* orphan genes in *Saccharomyces cerevisiae*

To identify possible effects of Brassicaceae specific *A. thaliana* genes on abiotic stress responses in yeast, we performed a systematic gain-of-function analysis in stress sensitive mutant strains of *Saccharomyces cerevisiae*. In total, 96 plant orphan genes (*ASR1* to *ASR96,* Tables 1 and 2) were chosen for cloning for expression in yeast. All of the 96 genes selected for this gain of function yeast study are upregulated by one or more abiotic stress conditions (Donoghue et al., 2011). In total, 83 orphan genes were cloned (Table 8) into the yeast expression vector pAG426-GPD which contains the glyceraldehyde-3-phosphate dehydrogenase (GPD) promoter to allow high level constitutive gene expression (the remaining 13 gene ORFs couldn't be amplified due to technical issues). A step wise reduction of geneset in the screening process was mention in Fig.26. The cloned genes were analysed under different abiotic stress conditions using the stress sensitive yeast strains JM26 and JM162 are in the W303-1A background. W303-1A was used as a wild type control for the salt and drought screens.

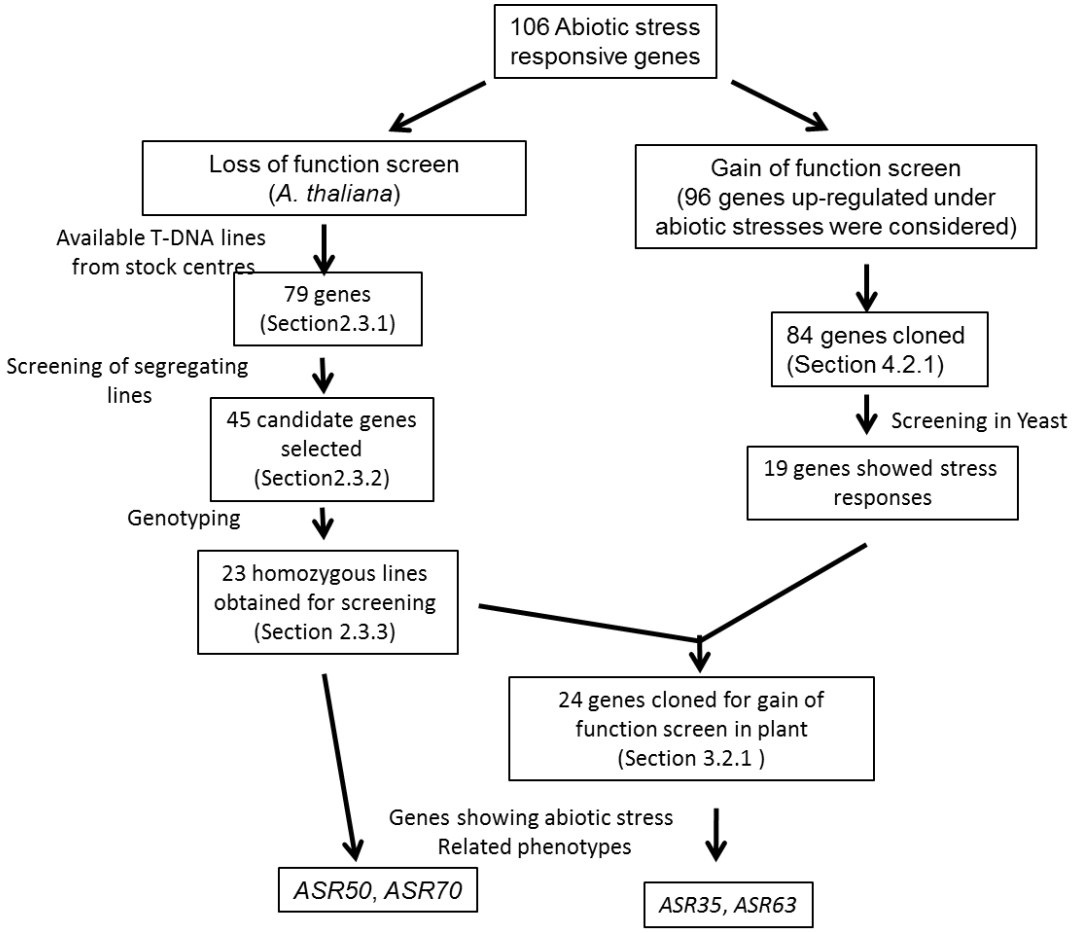


Figure 26: Stepwise reduction of geneset in screening. Reduction of total gene set in the screening process is mentioned. Details are mentioned in the respective section of the chapters noted in the box.

Table 8: List of orphan genes cloned for gain of function screen in yeast

A)

Gene name	Gene ID	Gene name	Gene ID	Gene name	Gene ID
ASR1	AT1G22890	*ASR29*	AT1G07610	*ASR58*	AT3G44430
ASR2	AT1G47400	*ASR30*	AT1G09415	*ASR59*	AT3G55910
ASR3	AT1G51670	*ASR31*	AT1G11850	*ASR60*	AT3G58540
ASR4	AT1G65500	*ASR32*	AT1G17090	*ASR61*	AT4G16000
ASR5	AT3G03020	*ASR33*	AT1G19960	*ASR63*	AT4G18580
ASR6	AT3G04640	*ASR35*	AT1G31580	*ASR64*	AT4G22212
ASR7	AT3G11745	*ASR36*	AT1G36640	*ASR65*	AT4G23870
ASR8	AT3G14480	*ASR37*	AT1G50290	*ASR66*	AT4G27654
ASR11	AT3G45730	*ASR38*	AT1G58150	*ASR67*	AT4G30670
ASR13	AT4G12580	*ASR39*	AT1G64360	*ASR68*	AT4G31030
ASR14	AT4G18280	*ASR40*	AT1G65510	*ASR69*	AT4G31875
ASR15	AT4G22640	*ASR41*	AT1G67860	*ASR71*	AT5G02550
ASR16	AT4G28085	*ASR42*	AT1G67865	*ASR72*	AT5G02690
ASR17	AT4G30460	*ASR43*	AT1G75190	*ASR73*	AT5G03130
ASR18	AT4G33666	*ASR44*	AT1G75770	*ASR75*	AT5G09980
ASR19	AT5G03210	*ASR45*	AT1G76960	*ASR76*	AT5G09990
ASR20	AT5G06190	*ASR46*	AT2G04800	*ASR78*	AT5G12880
ASR21	AT5G08090	*ASR48*	AT2G14460	*ASR81*	AT5G24313
ASR22	AT5G18040	*ASR50*	AT2G25510	*ASR82*	AT5G24570
ASR23	AT5G22530	*ASR52*	AT2G41650	*ASR84*	AT5G28610
ASR24	AT5G23460	*ASR53*	AT2G47200	*ASR85*	AT5G28630
ASR25	AT5G64890	*ASR54*	AT3G02240	*ASR86*	AT5G38980
ASR26	AT5G64905	*ASR55*	AT3G05730	*ASR88*	AT5G44580
ASR28	AT1G04660	*ASR57*	AT3G30720	*ASR96*	ATMG01370

B)

Gene name	Gene ID	Gene name	Gene ID	Gene name	Gene ID
ASR9	AT3G18250	*ASR51*	AT2G28570	*ASR77*	AT5G11740
ASR10	AT3G22231	*ASR56*	AT3G22240	*ASR91*	AT5G57760
ASR12	AT3G56260	*ASR70*	AT4G39675	*ASR93*	AT5G58790
ASR27	ATMG00630	*ASR74*	AT5G04790		

Table 8: List of genes cloned for gain of function screen in yeast. A) list of orphan genes B) List of non-orphan genes.

4.2.2 *Arabidopsis thaliana* orphan genes can enhance drought stress tolerance of stress sensitive yeast strains

To determine whether plant orphan genes confer drought tolerance in yeast, orphan gene constructs in a yeast expression vector were transferred to yeast strains W303-1A (as control) and JM162 (drought sensitive strain). The empty vector pAG426-GPD was also transferred into both strains to serve as a negative control. The transformants were spotted onto YPD media (control) and YPD supplemented with 1.7 M sorbitol using a replica plater (Sigma, R2383) with three dilutions (1:10, 1:100 and 1:1000). The plates were then incubated at 28°C for 2-3 days before analysing the effect of stress treatment. The stress tolerance or sensitivity was scored qualitatively based on the ability of the colonies to grow better (tolerance) or worse (sensitive) than the control.

The drought screen in yeast identified nine genes conferring drought tolerant phenotypes in JM162 as indicated by better growth on stress media compared to JM162 transformed with empty vector. Although the wild type W303-1A strain transformed with the orphan gene constructs did not show any drastic differences in the growth in all constructs, for *ASR23*, *ASR34*, *ASR40* and *ASR72,* improved growth was visible compared to W303-1A with empty vector (Fig 27A). These genes also gave improved drought tolerant phenotypes in JM162 background (Fig 27B).

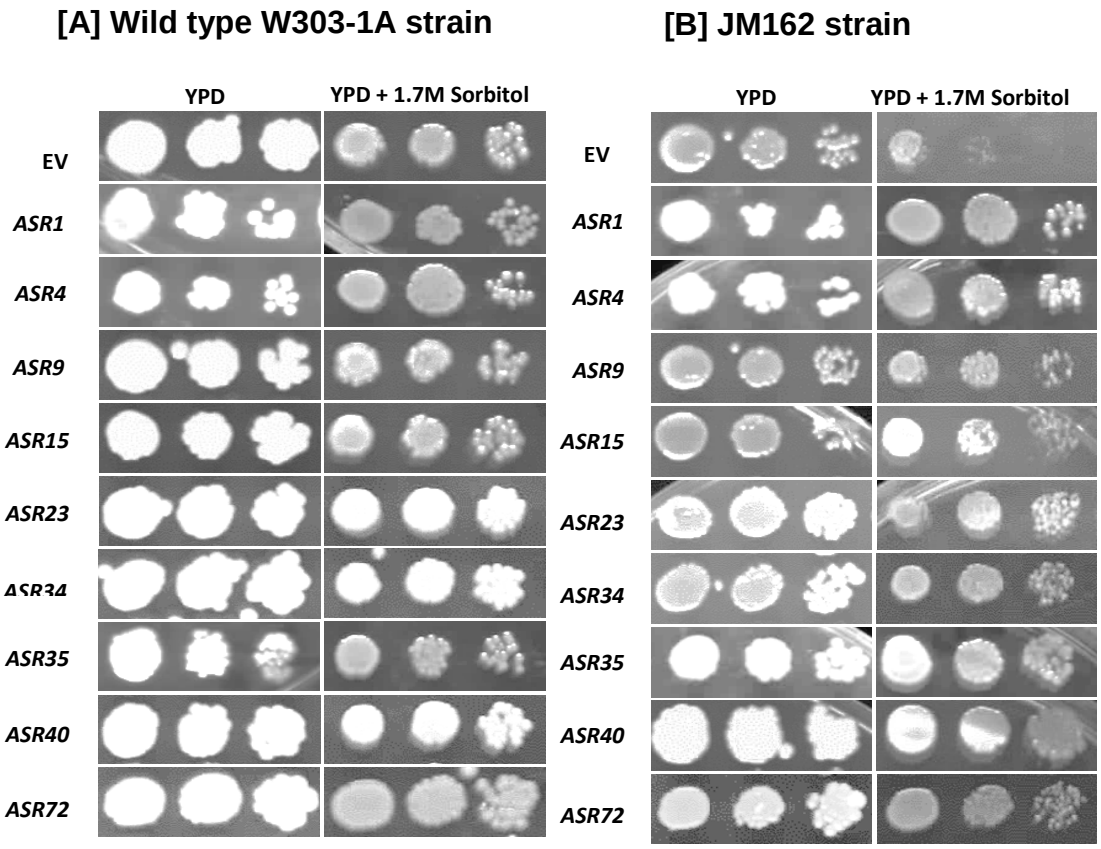


Figure 27: Screening of orphan genes for drought stress (Drought tolerant phenotypes). Empty vector (EV) and orphan gene constructs were transformed into yeast strains W303-1A (A) and JM162 (B) and plated on YPD media and YPD supplemented with 1.7 M Sorbitol in three dilutions (left to right - 1:10,1:100,1:1000). Stress tolerance or sensitivity was scored based on the growth of colonies.

From the yeast drought screen, a set of plant orphan genes conferring drought sensitive phenotypes in yeast was also isolated. In total four plant orphan genes conferred drought sensitive phenotypes as indicated by inhibition of growth compared to empty vector control (Fig.-28A and B). The growth inhibition phenotype was also visible in W303-1A background for *ASR2*, *ASR11* and *ASR18* (Fig.-28A) under drought stress. In JM162 background the colonies did not grow at all for yeast lines containing any of these four constructs.

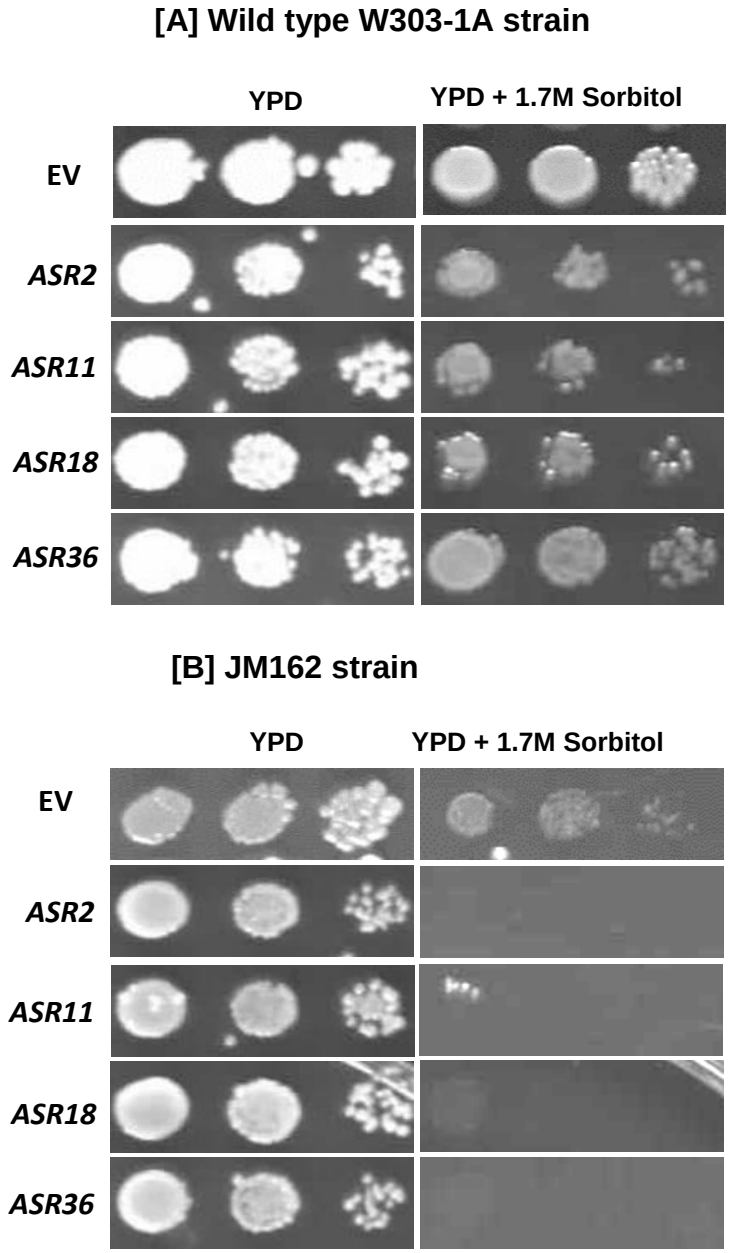


Figure 28: Screening of orphan genes for drought stress (Drought sensitive phenotypes). Empty vector (EV) and orphan gene constructs were transformed into yeast strains W303-1A (A) and JM162 (B) and plated on YPD media and YPD supplemented with 1.7 M Sorbitol in three dilutions (left to right - 1:10,1:100,1:1000).

4.2.3 Expression of five *Arabidopsis thaliana* orphan genes lead to enhanced sensitivity to oxidative stress in yeast

Oxidative stress can be caused by disruption of cellular redox homeostasis by formation of Reactive Oxygen Species (ROS). In plants, they are the unavoidable by-products of oxygenic photosynthesis and light-driven ROS production is potentially harmful as it can cause irreversible damage to photosynthetic components (Foyer and Shigeoka, 2011). In order to investigate whether orphan genes confer any improved resistance to oxidative stress in yeast, strain W303-1A was transformed with orphan gene constructs and plated in the same way as the drought screen described above. This screen used YPD media (control) and YPD supplemented with 5 mM hydrogen peroxide (H_2O_2) as an oxidative stress agent. Observations were taken after 2-3 days of growth at 28 $^{\circ}$C. In total five plant orphan genes showed oxidative stress phenotypes in yeast, all of which were causing increased sensitivity to oxidative stress (Fig.-29).

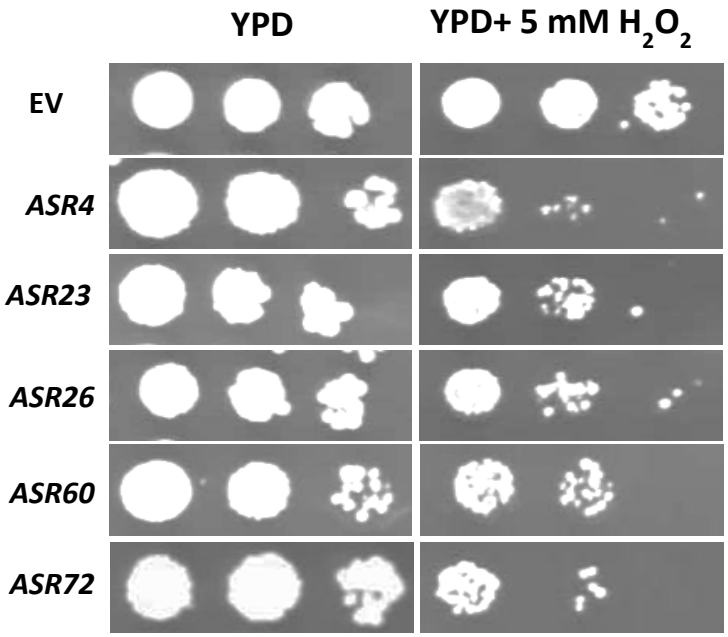


Figure 29: Screening of orphan genes for oxidative stress. Empty vector (pAG426-GPD) and orphan gene constructs were transferred to yeast strain W303-1A (WT) and plated on YPD media and YPD supplemented with 5 mM H_2O_2 in three dilutions (left to right - 1:10,1:100,1:1000).

4.2.4 Expression of *Arabidopsis* orphan genes generates altered salt and heat stress responses in yeast

Salt and heat stress were the two other abiotic stress conditions analysed for *Arabidopsis* orphan genes in yeast. The salt and heat screens were performed in the collaborating lab of Prof. J.M. Mulet in IBMCP, Valencia, Spain. Salt stress was analysed using the salt sensitive yeast strain JM26 which is a double mutant for ENA1 (Na^+-ATPase) and NHA1 (Na^+/H^+ antiporter) and hence defective in Na^+ extrusion which imparts salt sensitivity to the strain. The experimental set up was similar to the previous screens and the stress treatments given were 300 mM sodium chloride (NaCl) for salt stress and growth at 37 $^{\circ}$C for heat. The stress response was scored based on growth or inhibition of growth on stress media compared to control (strain transformed with empty vector). In total, four plant orphan genes displayed altered salt tolerant phenotypes at 300 mM salt in the JM26 background (Fig 30B). The wild type also seemed to show a tolerant phenotype with the four genes although the phenotype was only slightly better than the control (Fig 30A).

In the heat screen, the W303-1A strain was transformed with plant orphan gene constructs and analysed by growing them at 37 $^{\circ}$C and compared to their growth at 28 $^{\circ}$C. Seven plant orphan genes conferring heat tolerant phenotypes were isolated in this screen (Fig. 31), while no sensitivity phenotypes were observed. *ASR35*, *ASR60* and *ASR72* showed the strongest tolerance phenotypes while *ASR26*, *ASR34*, *ASR40* and *ASR59* showed weaker ones.

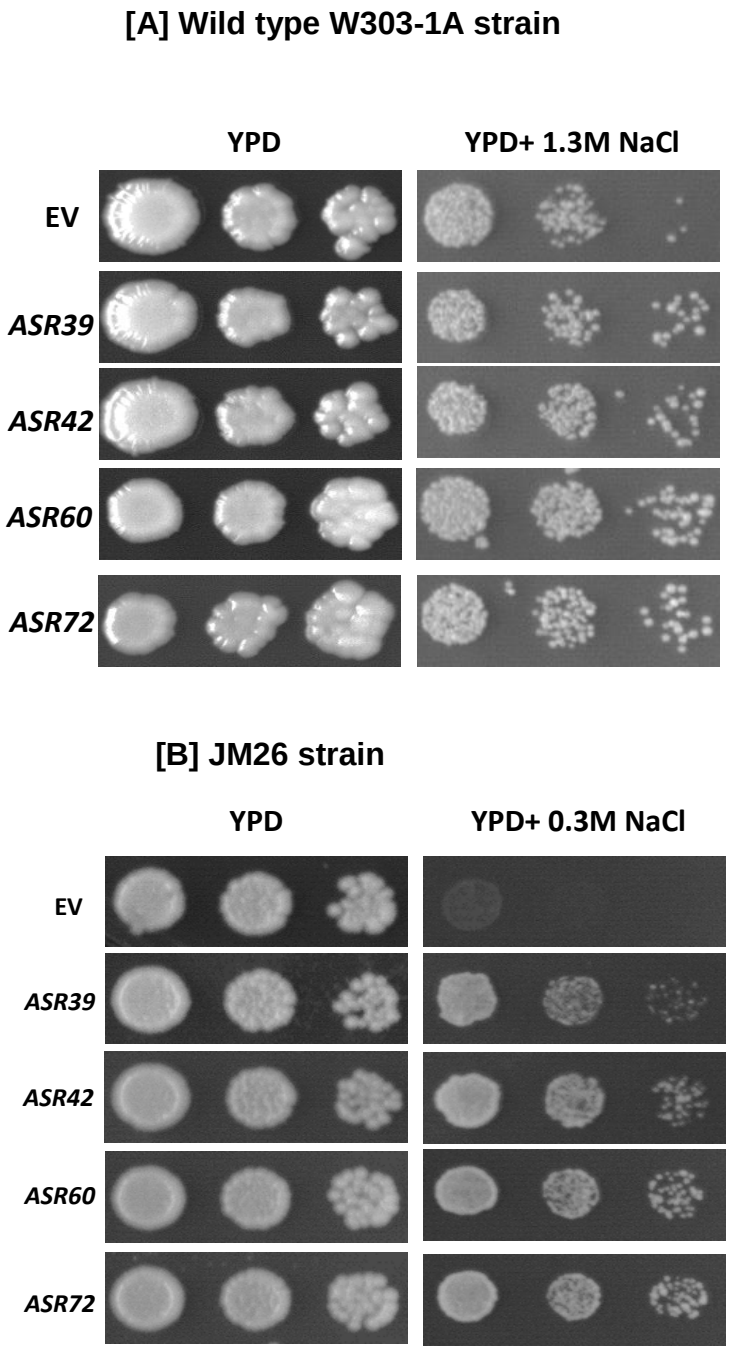


Figure 30: Screening of orphan genes for salt stress. Empty vector and orphan gene constructs were transformed to yeast strains W303-1A (A) and JM26 (B) and plated on YPD media and YPD supplemented with 1.3M NaCl for wild type (A) and 300 mM NaCl for JM26 (B) in three dilutions (left to right - 1:10,1:100,1:1000). (Data collected by JM Mulet)

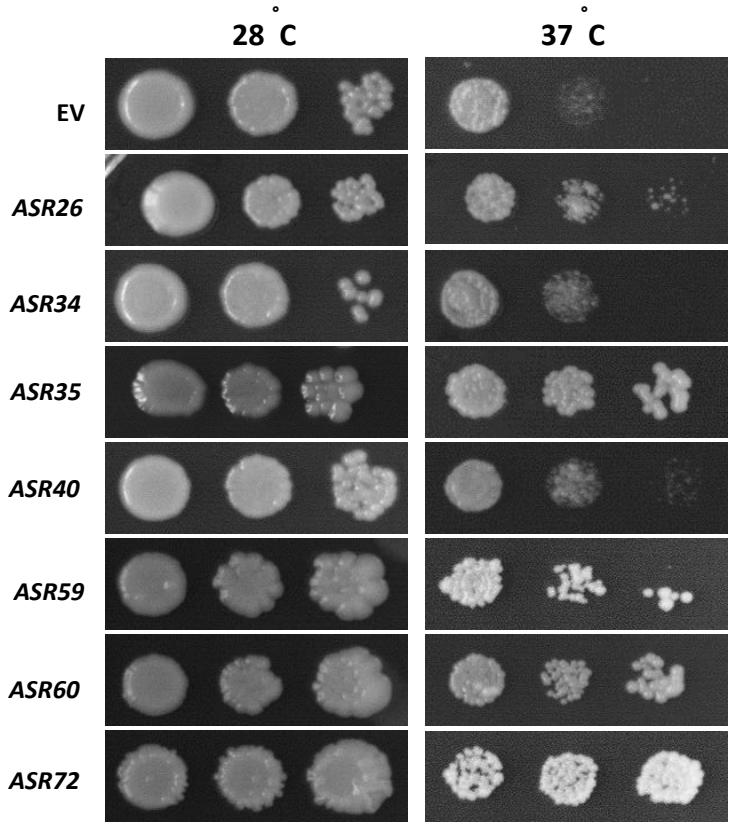


Figure 31: Screening of orphan genes for Heat stress. Empty vector and orphan gene constructs were transformed to yeast strains W303-1A (WT) and plated on YPD media in three dilutions (left to right - 1:10,1:100,1:1000) and incubated at 37 °C for growth. (Data collected by JM Mulet)

Interestingly, few genes in the yeast screens conferred improved responses to multiple stress conditions. However, the gene *ASR72* showed response in all the four stress conditions analysed, tolerance to drought, sensitive to oxidative stress, tolerance to salt stress and tolerance to heat stress. Similarly, *ASR60* showed tolerance to heat and salt stress and sensitivity to oxidative stress. Likewise *ASR34*, *ASR35* and *ASR40* showed tolerance to both drought and heat stress. The genes *ASR4*, *ASR10*, *ASR12* and *ASR23* showed tolerance to drought and sensitive to oxidative stress. The complete list of genes that showed responses to different stress conditions are given in Table 9.

Table 9: List of orphan genes conferring altered phenotypes to abiotic stress treatments in Yeast.

Gene Name	Gene ID	1.7M sorbitol	Oxidative stress	37 ºC	NaCl
ASR1	AT1G22890	+			
ASR2	AT1G47400	-			
ASR4	AT1G65500	+	-		
ASR9	AT3G18250	+			
ASR11	AT3G45730	-			
ASR15	AT4G22640	+			
ASR18	AT4G33666	-			
ASR23	AT5G22530	+	-		
ASR26	AT5G64905		-	+	
ASR34	AT1G22885	+		+	
ASR35	AT1G31580	+		+	
ASR36	AT1G36640	-			
ASR39	AT1G64360				+
ASR40	AT1G65510	+		+	
ASR42	AT1G67865				+
ASR52	AT2G41650	+			
ASR59	AT3G55910			+	
ASR60	AT3G58540		-	+	+
ASR66	AT4G27654				
ASR72	AT5G02690	+	-	+	+

Table 9: List of orphan genes conferring altered phenotypes to abiotic stress treatments in Yeast.

'**+**' indicates tolerant phenotype; '**-**' indicates sensitive phenotype while blank box indicates no response i.e. similar to control.

4.2.5 Measurement of intracellular sodium content to assess salt tolerance mechanism

The salt sensitive yeast strain JM26 used for salt stress screen was a double mutant for NHA1 and ENA1 transporters that are involved in sodium extrusion in *S. cerevisiae*. JM26 cannot survive on salt medium due to accumulation of sodium inside the cell. Measuring internal sodium content in JM26 strains overexpressing an orphan gene would give some hints about the mechanisms by which the orphan genes confer tolerance to increased sodium content of the media. The internal sodium content was measured for the four genes that showed a salt tolerant phenotype, *ASR39*, *ASR42*, *ASR60* and *ASR72* in both W303-1A and JM26 background. The sodium content in the wild type strain transformed with the orphan genes was found to be similar to the control (wild type with vector), except *ASR42* which showed slightly higher accumulation of sodium. However, in JM26 background, orphan gene overexpression lines showed different accumulation patterns compared to control (JM26 transformed with vector). *ASR42* in JM26 showed slightly higher accumulation of sodium (0.18 M) compared to control (0.16 M). However JM26 transformed with *ASR39*, *ASR60* or *ASR72* showed significantly lower sodium accumulation compared to control (Fig 32). This suggests that each of the plant orphan genes confers salt tolerance through different mechanisms.

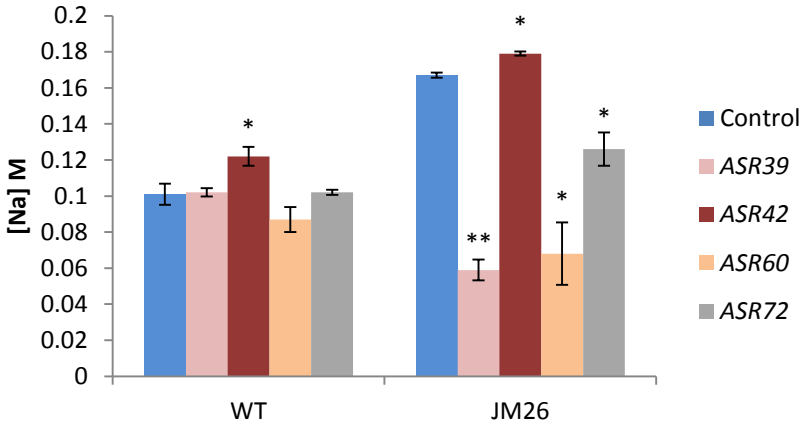


Figure 32: Measurement of intracellular sodium concentration. Internal sodium content of wild type (W303-1A) and JM26 stains overexpressing orphan genes measured and expressed in molar concentration. Control was the strain transformed with empty vector. Values are mean ± SEM (Student's t-test, *P<0.05.**P<0.01). (Data collected by JM Mulet)

4.3 Discussion

4.3.1 Gain-of-function screen of novel Brassicaceae specific orphan genes in Yeast

The yeast expression screening approach is an alternative approach to *in planta* approaches to identify plant gene functions and has been successfully used (Eswaran et al., 2010; Harrison et al., 2006; Kanhonou et al., 2001; Yamanaka et al., 2009). The underlying assumption is that at the cellular level, the stress tolerance mechanisms in yeast and plants are similar and therefore by overexpression in yeast and functional selection approaches, plant stress genes can be identified (Mulet et al., 2004). This assumption is based on a comparative evolutionary perspective which when interfaced with contemporary synthetic biology frameworks can suggest that the underlying “chassis” or biosynthetic framework is similar between plants and yeast. Where such similarities exist, there is potential that orphan genes from plants represent lineage-specific evolutionary innovations that can be moved across kingdoms to yeast and can function in the yeast genome chassis.

From this perspective, the Brassicaceae specific *Arabidopsis thaliana* orphan genes were investigated in yeast for effects on abiotic stress response. The evolutionary divergence (last common ancestor) of plants and yeast occurred about 1600 million years ago (Hedges, 2002). Hence, our study investigates whether some lineage-trapped orphan genes can display trans-Kingdom functionality in an organism that last shared a common ancestor over 1 billion years ago. The stress screens in yeast were carried out using stress sensitive yeast strains JM162 for drought and oxidative stress and JM26 for salt and heat stress. The plant orphan genes indeed conferred altered resistances to drought, salt, heat and oxidative stress treatments in yeast scored as tolerance or sensitivity as indicated in Table 9.

4.3.2 *Arabidopsis* orphan genes show drought and oxidative stress related phenotypes in Yeast

Drought stress screening of plant orphan genes was carried out using the yeast strain JM162 which is a mutant for glyceraldehyde-3-phosphate dehydrogenase (GPD) gene that results in disruption of the high osmolarity glycerol (HOG) response pathway. This gene disruption makes the strain sensitive for osmotic stress. In response to high osmotic external environment, yeast cells accumulate glycerol to adjust to the differences in internal and external water potential. The osmotic stress is perceived by membrane osmosensors and transferred via a HOG MAP kinase cascade which activates the glycerol biosynthetic pathway (Maeda et al., 1994). The genes complementing the JM162 strain under osmotic stress may be involved in the same pathway or act by activating/suppressing an alternative pathway to overcome the stress. Interestingly eleven orphan genes showed an enhanced tolerance phenotype to drought condition in yeast suggesting a possible role in drought stress response. Further, we also identified orphan genes which impart drought sensitive phenotypes to yeast.

These phenotypes were difficult to differentiate in the JM162 as the strain was already sensitive to the level of stress treatment applied. The phenotype was however apparent in the wild type background (W303-1A) for the genes *ASR2*, *ASR11* and *ASR18* where the growth of the colonies was reduced. The stress sensitive phenotype needs further analysis in the JM162 background under low sorbitol concentrations which could help to identify any epistatic effects of orphan gene overexpression in JM162. Epistasis experiments can be used to decipher the functional relationship between genes and genetic ordering of regulatory pathways (Phillips, 2008). Epistasis analysis is particularly useful when the genes analysed control a common process. For the orphan genes showing drought responsive phenotypes (tolerance or sensitivity), epistasis experiment using mutants for the HOG pathway genes will help to identify the functional order of action of the orphan genes in the pathway. In addition suppressor analysis by subjecting orphan gene overexpression strains to mutagenesis will identify the genetic suppressor of the phenotype conferred by orphan gene. Suppressor analysis

uses selection rather than screening (Forsburg, 2001) and hence it is particularly useful for suppressor analysis of orphan genes that are conferring sensitive phenotypes in yeast. Because the suppressor mutation causing tolerant phenotypes can be easily selected.

Oxidative stress is mainly caused by reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2), hydroxyl radicals ($^{\cdot}OH$) and superoxide ($O_2^{\cdot}$). ROS production is very common in all cells as a consequence of metabolic processes. Cells have antioxidant buffering systems to counteract the ROS compounds. Oxidative stress occurs when the ROS production exceeds the cells antioxidant buffering capacity. In plants, this often occurs under biotic and abiotic stress conditions which can cause damage to proteins, lipids and also DNA leading to oxidative stress. Hence oxidative stress plays a major role in plant stress.

In this study, five orphan plant genes conferred sensitivity to oxidative stress in yeast W303-1A upon overexpression. The oxidant defence response involves two types of systems, a non-enzymatic system involving ROS scavengers such as glutathione and ascorbic acid and an enzymatic defence system involving enzymes capable of removing oxygen radicals such as catalase, superoxide dismutase, pentose phosphate pathway enzymes and glutathione reductase. Yeast mutants deficient in forming ROS scavengers or the enzymes involved in detoxification are all show hypersensitive phenotypes to oxidative stress (Jamieson, 1998). Similar defence systems exist in plants. In plants photosynthesis also contributes to the ROS pool. ROS are also known to acts as secondary messenger in various cellular processes including stress tolerance (Neill et al., 2002; Yan et al., 2007). Hence oxidative stress responses are very broad and complex, and are also interlinked with other stress responses and signalling responses.

Although the orphan plant genes analysed in this study conferred stress responsive phenotypes they do need further analysis to identify the

possible mechanisms. In yeast glutathione (GSH) is an important antioxidant which acts as a bridge between the two defence mechanisms. Yeast mutants deficient in GSH have been shown to be sensitive to H_2O_2 (Izawa et al., 1995; Stephen and Jamieson, 1996). In addition glutaredoxins and thioredoxins are also required to give tolerance to H_2O_2 in yeast (Herrero et al., 2008; Luikenhuis et al., 1998). To start with, overexpression of the promising orphan genes in glutathione and glutaredoxin deficient mutants could reveal possible pathways in which they are acting.

4.3.3 Salt and heat stress responsiveness of *Arabidopsis* orphan genes in yeast

For the salt screen, the Na^+ sensitive yeast strain JM26 (Rausell et al., 2003) was used. Overexpression of four orphan plant genes in JM26 elicited tolerant phenotypes under 0.3 M salt (NaCl) treatment. This suggests that orphan plant genes could be activating some other pathways to tolerate NaCl accumulation inside the cell. It is known that in yeast three mechanisms act cooperatively to restrict the accumulation of sodium inside the cell: restriction of influx, sodium efflux and sequestration in vacuoles (Sychrova, 2004).

In JM26 sodium efflux is compromised due to mutations in *ENA1* and *NHA1*. To understand the mechanism, intracellular sodium concentrations were measured in mutants and wild type cells under salt stress (Fig 32). From the results at least two different mechanisms could be inferred. *ASR39, ASR60* and *ASR72* were found to be accumulating less sodium compared to that of empty vector control suggesting an efficient restriction of sodium influx. On the other hand *ASR42* accumulates slightly higher sodium levels compared to that of empty vector control suggesting a possible compartmentation in vacuole.

One approach for dissecting genetic pathways is to perform epistasis experiments which help in ordering of genes in a functional pathway. The orphan genes conferring salt stress phenotypes can be investigated in mutants like *trk1*, *trk2* and *hog1* to check whether the mutant phenotype will

be ablated upon overexpression of an orphan gene. Hog1 is a mitogen activated protein kinase that induces *ENA1* under sodium stress (Marquez and Serrano, 1996). Trk1 and Trk2 are the yeast high affinity potassium transporters. It has been shown that an influx of potassium will decrease membrane potential and thereby increase salt tolerance (Mulet et al., 1999). Hence using *trk1* and *trk2* mutants can help to decipher whether the phenotypes shown by overexpression of orphan genes under salt stress is due to sodium extrusion or regulation of membrane potential by Trk1 and Trk2 transporters.

Finally, a heat screen was performed using the wild type yeast strain W303-1A. The overexpression of orphan plant genes in yeast under high temperature stress identified seven promising candidate genes conferring tolerance to elevated temperature. In yeast heat shock response is primarily governed by two transcription factors Hsf1 and Msn2/4 (Verghese et al., 2012). Further, it is known that heat shock also activates the HOG pathway which is involved in osmotic stress response (Winkler et al., 2002). Epistasis experiments should be conducted for the heat tolerant seven orphan genes using mutants in the Hsf1 and Msn2/4 pathway which will help to decipher the regulatory action of the orphan gene in the pathway.

Overall, *Arabidopsis* orphan genes showing functional phenotypes in yeast is remarkable because yeast and plants diverged approximately 1600 million years ago (Hedges, 2002). Although it has been shown that newly evolved genes can quickly acquire functions within an organism (Chen et al., 2010), orphan genes showing functional phenotypes in such a distant species is quite a striking result.

However, it is important to note that orphan genes are defined based on sequence homology and not on protein structure or folding. Several metagenomic studies suggest that proteins arising in new environments show highly similar structures despite having divergent amino acid sequences with no recognisable sequence homologs (Godzik, 2011). For example, the crystal structure of a protein named ECX21941, which was

identified in a global oceanic sampling expedition show structural similarity to the Sm/LSm proteins, which are important RNA-binding proteins, despite no detectable sequence similarity (Das et al., 2009). Hence, structural studies on orphan genes could further add to the understanding of the orphan gene functionality at the protein level.

4.4 Conclusions

Considering that orphan genes arise in a taxon-specific manner and can be responsible for lineage specific innovations, analysing the effects of *Arabidopsis* orphan genes in yeast has revealed some exciting findings. The study resulted in identifying orphan plant genes which can confer stress related phenotypes in yeast. The work demonstrates that the yeast system can also be used for functional characterization of orphan genes even though they are highly diverse in terms of the evolutionary time scale. Future experiments focussing on how orphan genes acquire new functions and become integrated into the conserved pathways will be necessary and will add to the greater understanding of the evolution of genomic novelty in organisms. Most importantly, this work also indicated that orphan genes have some lineage specific innovations (e.g., stress tolerance in this case) which can be transferred to other lineages to confer novel traits. This suggests search for orphan genes in other plant species, such as those adapted to harsh environments, could result in finding novel stress tolerance genes. These genes could then be utilized for improvement of crops growing in stressful environments through transgenic or genome editing approaches.

Chapter 5: Brassicaceae specific orphan gene *ASR63* (AT4G18580) is involved in iron homeostasis in *Arabidopsis thaliana*

5.1 Introduction

Iron is the most common element by mass in the Earth's core and the fourth most common element in the crust (Darbani et al., 2013). The ability of iron to change its oxidation states in both hydrated states (ferrous and ferric ions in water) and in complexes with organic molecules makes it an excellent catalyst for biological functions. This also makes it an essential element for plants as iron is required for the biosynthesis of chlorophyll and also it is part of the photosynthetic reactions. Iron is also required for respiration, sulphate assimilation, hormone synthesis and nitrogen fixation as well as DNA synthesis and repair (Darbani et al., 2013; Miethke and Marahiel, 2007). In spite of being abundant in the environment, iron deficiency is a major medical as well as agricultural problem. According to the World Health Organization (WHO), about 30% of the world's population is anaemic, many due to iron deficiency. Similarly, in plants iron deficiency is a major constraint for growth of plants and is a common nutritional disorder worldwide (Schmidt and Buckhout, 2011).

As sessile organisms plants have developed different mechanisms for nutrient acquisition including uptake of iron. Based on their strategy of iron acquisition, they are categorized into two groups (Fig. 33). Strategy I plants (non-grass plants like *Arabidopsis*) use reduction of Fe whereas strategy II plants (graminaceous plants such as rice) use a chelation strategy for iron acquisition. The two main processes in the strategy I plants are the reduction of iron chelates at the root surface and the uptake of the reduced ferrous ions. The release of protons through H^+-ATPases of the (*AHA* gene family) lead to decreased soil pH and solubilises the ferric iron (Fe^{3+}) followed by the reduction of Fe^{3+} to Fe^{2+} form by the activity of ferric-chelate reductase (*AtFRO2*) (Robinson et al., 1999). The reduced iron is then transported into the root by the high affinity plasma membrane Iron Regulated Transporter, IRT1 (Eide et al., 1996) which is the major transporter of iron in plants (Vert et al., 2002). The strategy II plants (i.e. grasses) however, release phytosiderophores (PSs) such as mugineic acids (MAs) which binds to Fe^{3+} to acquire iron from soil. Only grasses can synthesise phytosiderophores

from nicotianamine (NA) (Hindt and Guerinot, 2012). The Fe^{3+}-MA complex will then be transported into the root by the yellow stripe 1 and yellow stripe 1-like transporters (Curie et al., 2001; Inoue et al., 2009).

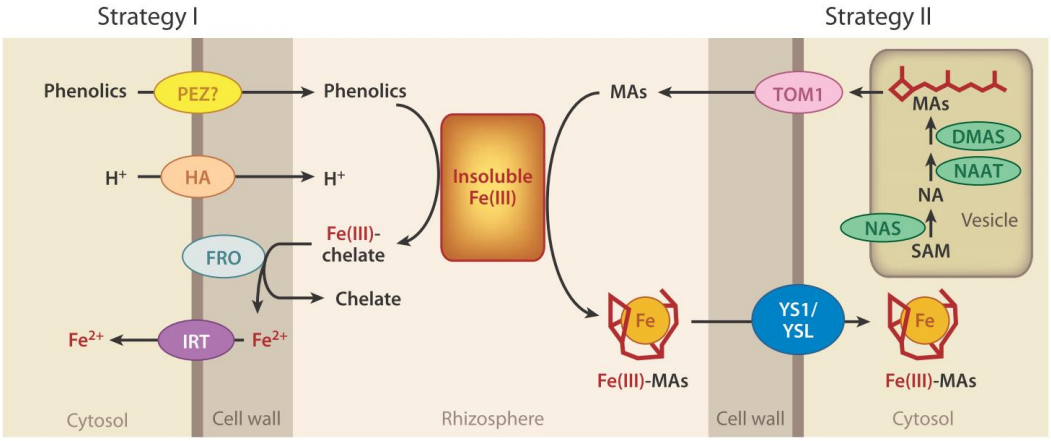


Figure 33: Iron acquisition strategies in higher plants. The mechanisms in strategy I (left) and Strategy II plants (right) are depicted. Ovals represent the transporters and enzymes that play central roles in these strategies, all of which are induced in response to iron deficiency. Abbreviations: DMAS, deoxymugineic acid synthase; FRO, ferric-chelate reductase oxidase; HA, H^+-ATPase; IRT, iron-regulated transporter; MAs, mugineic acid family phytosiderophores; NA, nicotianamine; NAAT, nicotianamine aminotransferase; NAS, nicotianamine synthase; PEZ, PHENOLICS EFFLUX ZERO; SAM, S-adenosyl-L-methionine; TOM1, transporter of mugineic acid family phytosiderophores 1; YS1/YSL, YELLOW STRIPE 1/YELLOW STRIPE 1–like. Figure is adapted from Kobayashi and Nishizawa (2012).

Excess of iron in plants is toxic due to production of cytotoxic hydroxyl radicals and also iron is not readily bioavailable as it forms insoluble complexes under aerobic conditions at neutral and alkaline pH (Grotz and Guerinot, 2006). On the other hand iron deficiency leads to drastic reduction in chlorophyll biosynthesis directly affecting the photosynthesis, leading to chlorosis, and can be lethal under prolonged deficiency conditions. Therefore iron homeostasis is crucial for plants. The majority of the genes involved in iron homeostasis and iron deficiency responses in plants have been elucidated during the past decade. In *Arabidopsis*, the basic helix-loop-helix (bHLH) transcription factor FIT (FER-like iron deficiency induced transcription factor), a functional ortholog of tomato FER, is required for regulating the iron deficiency response (Colangelo and Guerinot, 2004). FIT cannot function

alone and requires another bHLH transcription factor, AtbHLH38 or AtbHLH39, for function presumably forming heterodimers (Yuan et al., 2008). *AtbHLH38* and *AtbHLH39* belong to subgroup 1b of the *bHLH* genes, along with *AtbHLH100* and *AtbHLH101* and all the four genes are highly induced under iron deficiency conditions in both roots and leaves. The latter two genes have been found to function via an FIT-independent pathway (Sivitz et al., 2012). FIT, which is solely expressed in roots, forms a complex with AtbHLH38 or AtbHLH39 to positively regulate the downstream targets *IRT1* and *FRO2* under iron limited conditions, thereby increasing the iron uptake (Colangelo and Guerinot, 2004). This FRO2/IRT1 system is the major gateway for iron entry into the root epidermal cells. It has been found that FIT also interacts with ethylene signalling transcription factors ETHYLENE INSENSITIVE3 (EIN3) and ETHYLENE INSENSITIVE3-LIKE1 (EIN3/EIL1) linking the iron deficiency response with the ethylene signalling pathway (Lingam et al., 2011).

There are also other components identified for iron deficiency responses in *Arabidopsis*. Another bHLH transcription factor *POPEYE* (*PYE*), expressed in pericycle cells, is found to be important for root growth under iron deficiency and is proposed to be a positive regulator of growth and development under iron deficiency (Long et al., 2010). A *PYE* homolog, *ILR3*, encodes another bHLH protein which has been shown to be involved in metal ion mediated auxin sensing within *Arabidopsis* roots and is proposed to be involved in regulating expression of genes implicated in intracellular iron transport (Rampey et al., 2006). The proteins involved in intracellular iron transport are also important for iron homeostasis for example VIT1, a homolog of yeast Ccc1p, both involved in vacuolar iron influx (Kim et al., 2006; Li et al., 2001b). Similarly, other vacuolar iron transporters like NRAMP3/4, whose disruption causes sensitivity to iron deficiency (Lanquar et al., 2005) play important roles in iron homeostasis. Moreover, IRT1 is a nonspecific transporter which can transport Mn, Co, Cd, Zn and Ni under iron deficiency conditions (Vert et al., 2002). Hence the genes involved in detoxification and/or vacuolar sequestration of Zn like ZIF1 (Haydon and

Cobbett, 2007) or Co, Ni and excess Fe like FPN2 (Morrissey et al., 2009; Schaaf et al., 2006) are also the components of iron deficiency response.

The intercellular transportation of iron is also very important because of the low solubility and high reactivity of iron and hence iron should to be associated with suitable chelating molecules. Principle chelators inside the plants include citrate, nicotianamine (NA) and MAs as indicated by physiological and molecular studies. Once the iron is loaded into xylem, FRD3 (Ferric reductase defective3) facilitates the iron chelation to citrate and subsequent transport from root to shoot (Durrett et al., 2007). Furthermore, in young leaf tissue and developing seeds, iron is transported from xylem to phloem as Fe-nicotianamine complexes via the Yellow stripe-like1 (YSL1) transporter. The YSL family members are also present in nongraminaceoous plants and are involved with translocation of metal ions chelated with NA. Disruption of *FRD3*, *YSL* genes or genes encoding nicotianamine synthases (*NAS1*, *NAS2*, *NAS3* and *NAS4*) leads to chlorosis and increased sensitivity to iron deficiency (Klatte et al., 2009; Rogers and Guerinot, 2002; Waters et al., 2006). Another protein OPT3 (Oligopeptide transporter3) is known to be involved in iron accumulation in seeds and is a phloem specific transporter that mediates iron loading in to the phloem. It is also a component of iron signalling between roots and shoots (Mendoza-Cozatl et al., 2014; Zhai et al., 2014). The *opt3-1* null mutant causes embryo lethality in *Arabidopsis* (Stacey et al., 2002). These findings suggest the importance of proper iron distribution for normal growth and development of plants. Hence plants need tight regulation of iron uptake, transport in various organs and storage to ensure an optimal development by preventing both iron deficiency and toxicity.

In this study, we identified a *de novo* evolved gene, *ASR63* within the Brassicaceae lineage whose constitutive expression confers tolerance to iron deficiency in *Arabidopsis*. Our data indicates that the constitutive expression of the *ASR63* gene in *Arabidopsis thaliana* leads to increased production of antioxidants, glutathione and ascorbic acid, thereby maintaining a favourable redox status for optimum uptake of iron under iron-depleted conditions. We

further propose an evolutionary scenario for the *de novo* evolution of *ASR63* in which the activation of bidirectional transcription from the promoter of an adjacent gene lead to the origination of *ASR63* within the Brassicaceae lineage. Overall, we identified a Brassicaceae specific novel *de novo* gene which is potentially involved in iron homeostasis in *Arabidopsis thaliana.*

5.2 Results

5.2.1 *ASR63* is potentially a *de novo* evolved gene within Brassicaceae lineage

Previously we identified *ASR63* (AT4G18580) as a Brassicaceae specific gene in the *Arabidopsis thaliana* genome (Donoghue et al., 2011). This gene is annotated as producing an unknown protein with a coding sequence of 348 base pairs and is predicted to encode a protein of 115 amino acids (The Arabidopsis Information Resource, TAIR10). A NCBI blast search (blastp and tblastn) for *ASR63* like sequences found significant matches (E-value < 0.001) only in the Brassicaceae family, specifically in *Arabidopsis lyrata* (tblastn 78% identity), *Capsella rubella* (tblastn 54% identity), and *Eutrema salsugineum* (tblastn 64% and 51% identity respectively to two genes).

Syntenic relationships in gene order between species can provide important information on gene relationships (Cai et al., 2008). Hence we investigated the synteny of the genomic region containing the *ASR63* gene with genomic regions in other sequenced species of Brassicaceae and Cleomaceae (which is considered the closest outgroup family to Brassicaceae). This involved synteny analysis for the region comprising three genes upstream (approximately 4.5 Kb) and one gene downstream (approximately 6.5 Kb) of the *ASR63* locus in *Arabidopsis thaliana* compared with genomic regions in *Arabidopsis lyrata*, *Capsella rubella, Sisimbrium irio, Brassica rapa, Eutrema salsugineum, Aethionema arabicum* and the closest outgroup species to Brassicaceae *Tarenaya hassleriana* (belonging to Cleomaceae).

The synteny analysis indicated that the *ASR63* locus has syntenic matches in all the Brassicaceae species considered, but has no match in the out group species *Tarenaya hassleriana*. However, the sequences homologous to the flanking sequences of *ASR63* locus displayed synteny with this genome (Fig.34). The absence of any sequence similarity for *ASR63* in the outgroup species (*T*. *hassleriana)* could suggest that the *ASR63* locus has originated only in the Brassicaceae family (i.e. a *de novo* gene origin). Supporting this, we determined that the flanking genes at the *ASR63* locus display deep evolutionary conservation across eudicots, as well as the grasses (Fig. 35). In contrast, the *ASR63* gene is only found in *A. lyrata* (based on Phytozome; recently sequenced other Brassicaceae species like *B. rapa*, *C. rubella, E. salsugenium* etc., were not included). This supports the hypothesis that *ASR63* is a novel gene that is specific to the Brassicaceae family, and that *ASR63* may have arisen as a *de novo* gene within the Brassicaceae lineage.

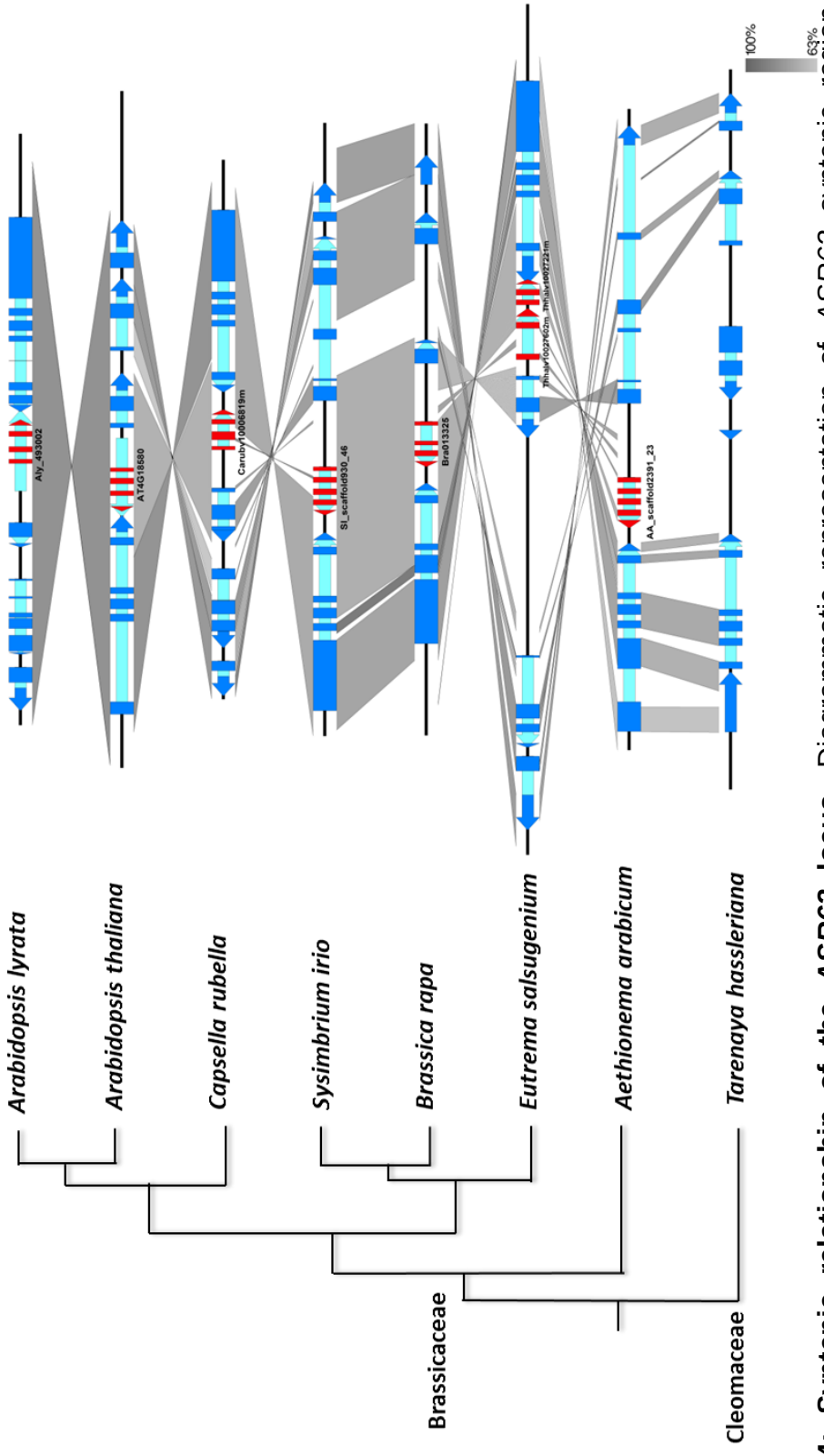


Figure 34: Syntenic relationship of the *ASR63* locus. Diagrammatic representation of *ASR63* syntenic region. *ASR63* homologs are depicted in red (exons) and labelled with annotated gene IDs, the flanking genes are depicted in blue. Light green colour depict introns, black lines depict intergenic regions. Grey shades depict the BLAST (blastn) matches (intensity range, 60-100% match) between the syntenic regions. Phylogenetic relationships of the species are represented by tree (not for scale).

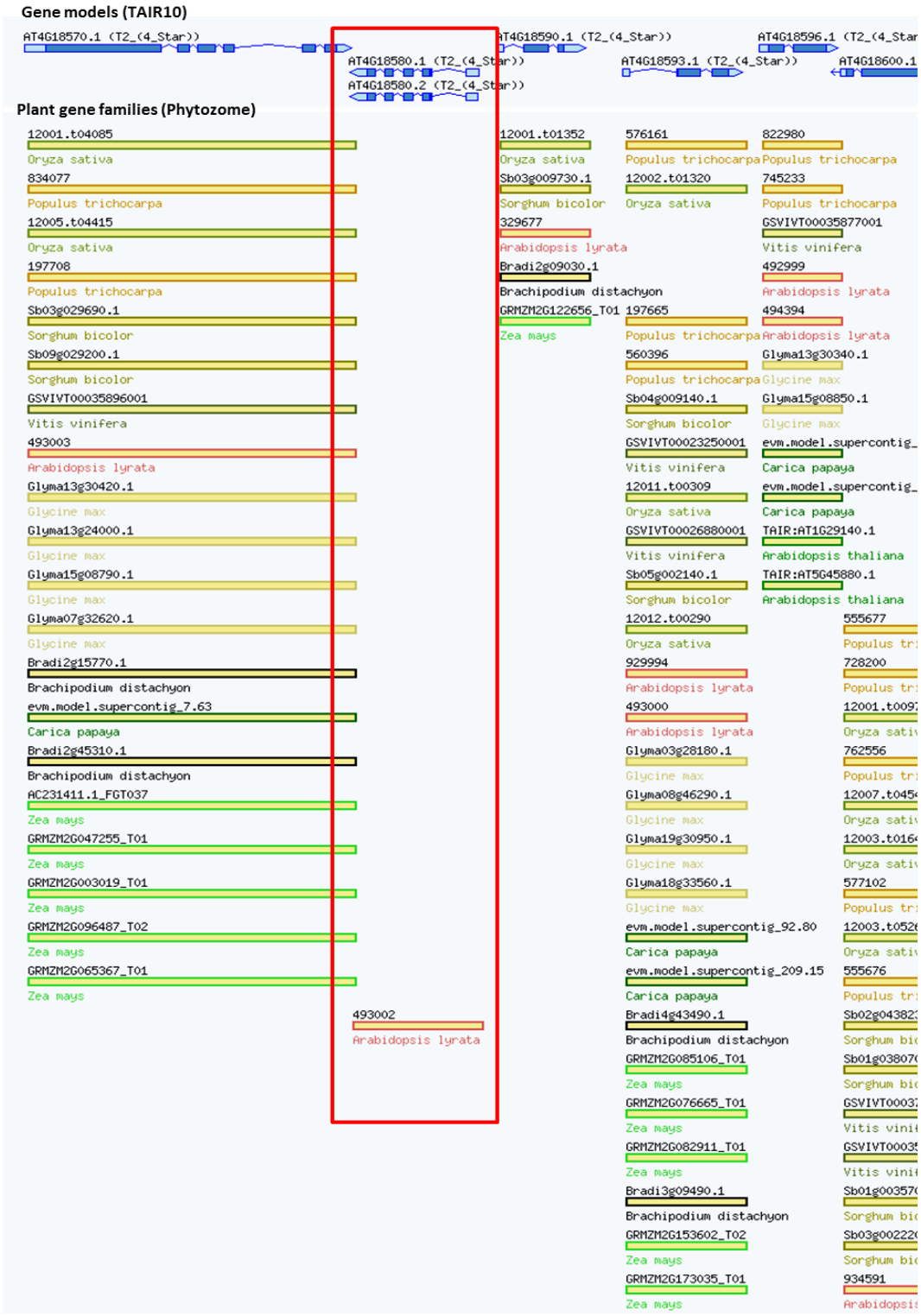


Figure 35: Conserved gene families at the *ASR63* locus. Conserved gene families at the *ASR63* locus (red box) are shown. The flanking genes show conservation in plant species outside of Brassicaceae while *ASR63* has match only within Brassicaceae. Upper panel shows the gene models (in blue, TAIR10). Screenshot from TAIR gbrowse.

Further, the intergenic region between the *ASR63* and the upstream gene AT4G18590 (Fig. 35) is just 179 base pairs suggesting that both the genes share the promoter region and the promoter is likely to be bidirectional. The intergenic regions between two transcriptional start sites (TSSs) lying in the head-to-head orientation which are not more than 1000 base pairs apart can be considered as putative bidirectional promoter (Adachi and Lieber, 2002). The gene AT4G18590 has conserved gene families across eudicots as well as in monocots. On the other hand we could not find any homologous sequences matching *ASR63* outside of Brassicaceae (Fig.34). Taken together our results indicate that *ASR63* has originated after the split of Brassicaceae from cleomaceae. Further we propose that *ASR63* is originated due to the activation of the promoter of AT4G18590 to function as bidirectional promoter. It is likely that the transcriptional activity at the *ASR63* locus due to the bidirectional promoter led to production of spurious transcript which overtime acquired the gene structure to become the stable protein coding gene in Brassicaceae and *A. thaliana* in particular.

5.2.2 *ASR63* shows expression in both shoot and roots specifically in vascular tissues

The transcriptome data shows that *ASR63* is expressed more or less ubiquitously at low level, although highest expression was shown to be in developing seeds and pollen (Arabidopsis eFP-browser). To analyse the *ASR63* gene expression, *ASR63* promoter (980 bp upstream of ATG) along with the gene excluding stop codon was amplified and fused with *GUS* gene in a plant binary vector pMDC163 (Curtis and Grossniklaus, 2003). *GUS* expression was analysed in seedling, leaf, root and inflorescences. The *GUS* expression was prominently seen in the leaves of young seedlings with 2-4 true leaves (approx. 2 weeks old), specifically in leaf veins. The expression was also seen in root vasculature and root tip although to a lesser extent. In the inflorescence the expression was observed in the flower petals especially in the vasculature (Fig. 36).

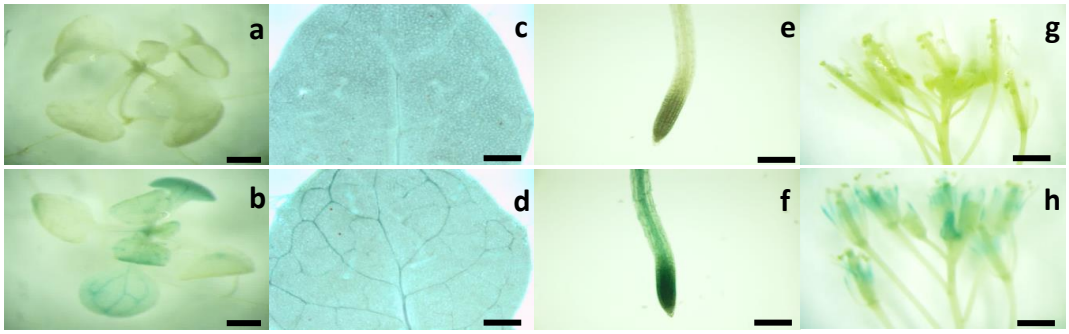


Figure 36: Histochemical analysis of *ASR63* promoter-GUS activity. GUS expression analysis of *ASR63* promoter-GUS fusion line (b,d,f and h), GUS expression was seen throughout the leaves (b) prominently in the vasculature (d), root (f) and flower petals (h). Wild type Col-0 plant was used as control (a,c,e and g).

5.2.3 Overexpression of *ASR63* leads to increased tolerance to iron depleted condition in *Arabidopsis thaliana*

Since we found that *ASR63* is responsive to low iron and low pH conditions, *ASR63* overexpression lines (*35S::ASR63*) were subjected to iron starvation screening as mentioned in the chapter 3. *35S::ASR63* lines indeed showed tolerance under low pH (pH 4.2) and iron starved (10 µM Fe) condition compared to wild type control (Chapter 3, Fig. 22). To further characterise the iron responsive phenotype of *ASR63* we subjected the *35S::ASR63* lines to 0 µM and 2 µM Fe (Fe refers to total iron in the media). The *35S::ASR63* lines exhibited increased tolerance to low Fe condition (2 µM Fe) both at pH 5.7 and pH 4.2 compared to wild type control (Fig. 37B). But they failed to grow normally under 0 µM Fe irrespective of the pH. The phenotypes were scored 14 days after transplanting to low pH and low Fe media. Under Fe deficiency condition the overexpression lines maintained higher chlorophyll content compared to wild type which was almost equivalent to normal conditions, suggesting that *ASR63* might have a role in Fe accumulation/uptake.

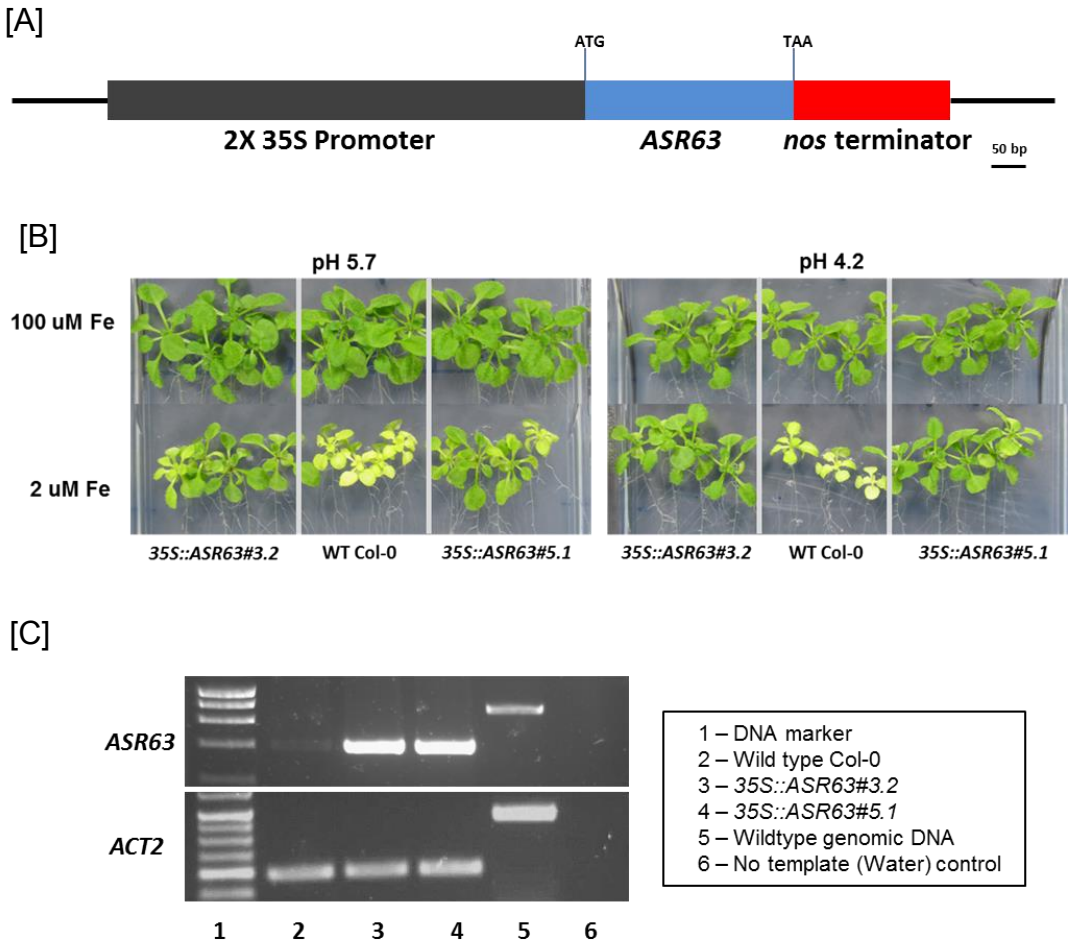


Figure 37: Screening of *35S::ASR63* lines for iron deficiency. [A] Schematic representation of *ASR63* overexpression construct. *ASR63* ORF was cloned to the binary vector pMDC32 under *CaMV35S* promoter.[B] Phenotypes of iron deficiency responses of WT Col-0 and two independent 35S::*ASR63* lines (*35S::ASR63#3.2* and *35S::ASR63#5.1*) under iron sufficient (100 µM Fe) and iron deficient (2 µM Fe) conditions at pH 5.7 and pH 4.2.[C] Semi quantitative RT-PCR showing over expression of *ASR63* in the transgenic 35S::*ASR63* lines (top panel - 25 PCR cycles). *Actin2* was used as internal control.

5.2.4 *ASR63* overexpression leads to increased accumulation of iron in vegetative tissues and seeds

Since *35S::ASR63* lines maintained almost normal levels of chlorophyll in the transgenic plants (Fig.23, chapter 3), we hypothesised that *ASR63* might be involved in uptake of iron. Hence we analysed iron accumulation in the vegetative tissues and seeds using a ferrozine assay. For the vegetative tissue, shoots from seedlings treated for 14 days under

iron depleted and iron replete conditions were used. Under iron replete conditions (100 µM Fe), leaf iron content in the *35S::ASR63* lines was found to be about 1.3 fold higher than the wild type plants in both pH 5.7 and pH 4.2. However, upon iron depletion, a drastic reduction in iron content in wild-type plants was observed, while *35S::ASR63* lines had the similar amount of iron as in normal condition irrespective of iron depletion and/or pH (Fig.38).

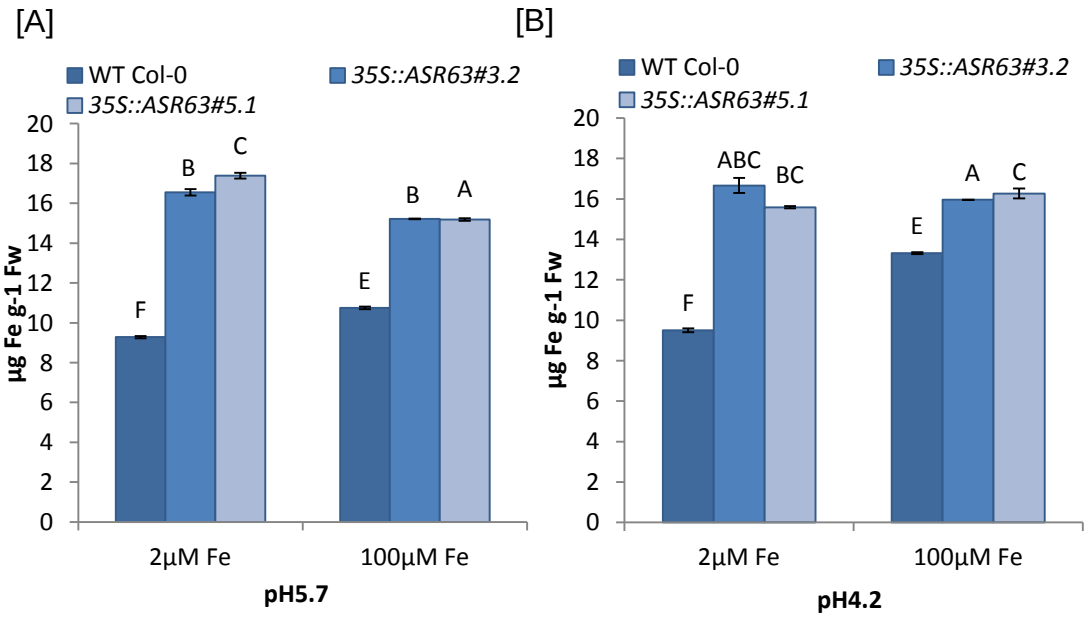


Figure 38: Measurement of total iron content in *Arabidopsis* shoots. Quantification of iron in shoot tissues as determined by ferrozine assay from seedlings grown at pH5.7 [A] and pH 4.2 [B]. Values are mean +/- SEM, ANOVA ($P<0.05$), levels not connected by same letter are significantly different.

Since iron content in seeds is of special importance to the human health, we evaluated the iron content in the *Arabidopsis* embryos by Perl's staining technique. Seeds collected from wild type and two independent *35S::ASR63* lines grown under normal conditions were used for this method. Seeds were imbibed in water for 48 hours and embryos were excised for staining. The Perl staining which can detect only Fe^{2+}, revealed elevated iron content in the *35S::ASR63* lines compared to wild type embryos (Fig. 39A, upper panel). In contrast, enhanced Perl's stain (Roschzttardtz et al., 2009) which can detect the presence of both iron forms, Fe^{2+} and Fe^{3+}, revealed a clear distinction in iron contents of wild type and *35S::ASR63* lines as compared to the wild-type embryos (Fig. 39A, lower panel). To further

quantify seed iron content, total iron was measured in whole seeds and embryos (n=60) of wild type and *35S::ASR63* lines using ferrozine assay. The total iron content in embryos of *35S::ASR63* lines was found to be about 3-5 fold higher than the wild type control while it was 5-7 fold for the whole seeds (Fig. 39B).

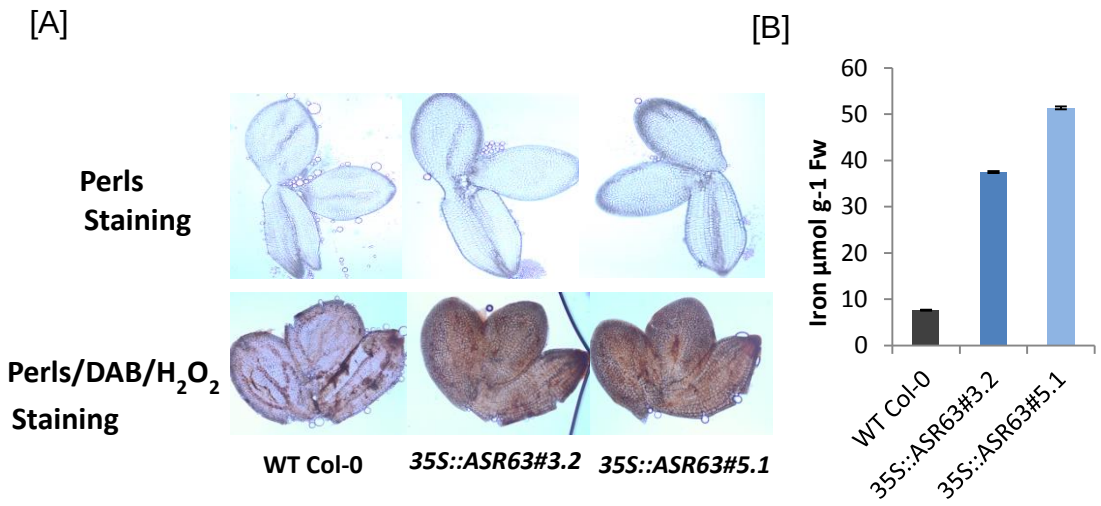


Figure 39 : Quantification of iron in *Arabidopsis* seeds. [A] Qualitative assay of iron in seed embryos by perl's staining (upper panel) and enhanced perl's staining (lower panel). Mature seeds imbibed in water for 48 hours used for excising embryos. [B] Quantitative measurement of iron in seeds expressed in µmol/gram/fresh weight (n=60 seeds).

5.2.5 Increased activity of iron-sulphur (Fe-S) proteins indicate higher iron availability in *35S::ASR63* lines under iron deficiency

To further characterize the observed phenotype biochemically we analysed the activity of two Fe-S proteins, aconitase and sulphite reductase under iron deficient conditions. Since most of the metabolically active iron is bound to Sulphur in Fe-S clusters (Forieri et al., 2013), measuring the activity of Fe-S proteins reflects on the availability of iron in the plant. Aconitase is an enzyme containing an Fe-S cluster that catalyzes the reversible isomerization of citrate to isocitrate via cis-aconitate in the tricarboxylic acid cycle (Peyret et al., 1995). Since aconitase acitivity depends on iron content in the cell, aconitase activity can be used as marker for iron accumulation.

Another Fe-S protein sulphite reductase (SiR) analysed is involved in assimilitary reduction of sulfite to sulphide in the sulphate assimilation pathway (Khan et al., 2010).

To assess the aconitase activity, 10 day old seedlings were transplanted to iron replete or iron depleted condition at pH 4.2 and pH 5.7 and grown for a further four days. Total protein was extracted from shoot tissue from seedlings grown on iron replete or iron depleted conditions at pH 5.7 and pH 4.2. Protein concentration was measured and equal amount of protein was loaded on the polyacrylamide gel. The in gel activity assay was done according to the protocol described previously (Bernard et al., 2009). Results showed higher aconitase activity in *35S::ASR63* lines under 0 µM Fe at both pH 5.7 and 4.2 compared to wild type control. The aconitase activity in *35S::ASR63* lines at 2 µM Fe was slightly higher compared to wild type control at both pH 4.2 and 5.7 (Fig. 40).

SiR activity was found to be higher in *35S::ASR63* lines at 0 µM Fe and 100 µM Fe under pH 4.2 but no obvious difference was found at 2 µM Fe treatment compared to wild type control. Under pH 5.7 there was a clear difference in activity of SiR between wild type and *35S::ASR63* lines under iron depleted conditions (0 and 2 µM Fe) while no obvious difference was seen under iron replete condition (Fig. 40). Collectively, our results suggested that *35S::ASR63* lines accumulate more iron under iron depleted conditions and the accumulated iron is bioavailable as enzyme cofactor.

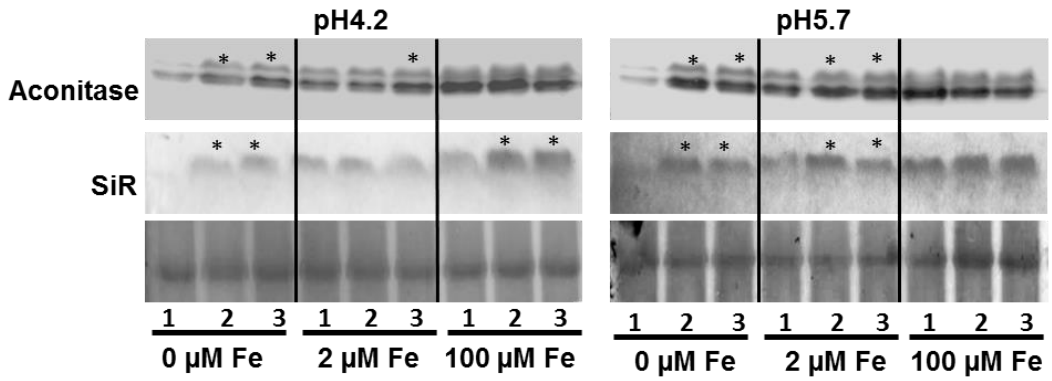


Figure 40: Activity of Fe-S proteins in *35S::ASR63* lines. Protein extracted from 4 day treated seedlings was run on native PAGE and analysed for aconitase and SiR activity. Equal quantity of protein was loaded into each well. Lower panel shows the RubisCo protein band as loading control. 1 - WT Col-0; 2 -*35S::ASR63#3.2* ; 3 - *35S::ASR63#5.1*; SiR- Sulfate reductase. Higher activity is marked in asterisks.

5.2.6 Measurement of reactive oxygen species (ROS) in *35S::ASR63* lines under iron deficiency

iron deficiency is known to make the plants susceptible to oxidative damage by ROS as iron is required for ROS scavenging enzymes associated with detoxification of hydrogen peroxide (catalase, peroxidase, ascorbate peroxidase) and dismutation of superoxide (Fe-superoxide dismutase). In order to check whether *35S::ASR63* lines are efficient in scavenging ROS under iron deficiency, we analysed production of hydrogen peroxide (H_2O_2) and superoxide ($O_2^{\bullet-}$) under iron replete (Fe+, 100 µM Fe) and iron deplete (Fe-, 0 µM Fe) conditions at pH 4.2 and pH 5.7. Seven day old seedlings were transferred to Fe(+) and Fe(-) media and grown for further four days before collecting the shoot tissue for analysis. H_2O_2 measurement showed that *35S::ASR63* lines have significantly lower amount of H_2O_2 under Fe(-) condition at pH 4.2 as well as pH 5.7 compared to wild type control suggesting that *35S::ASR63* lines efficiently scavenge H_2O_2 under iron deficiency conditions (Fig.41).

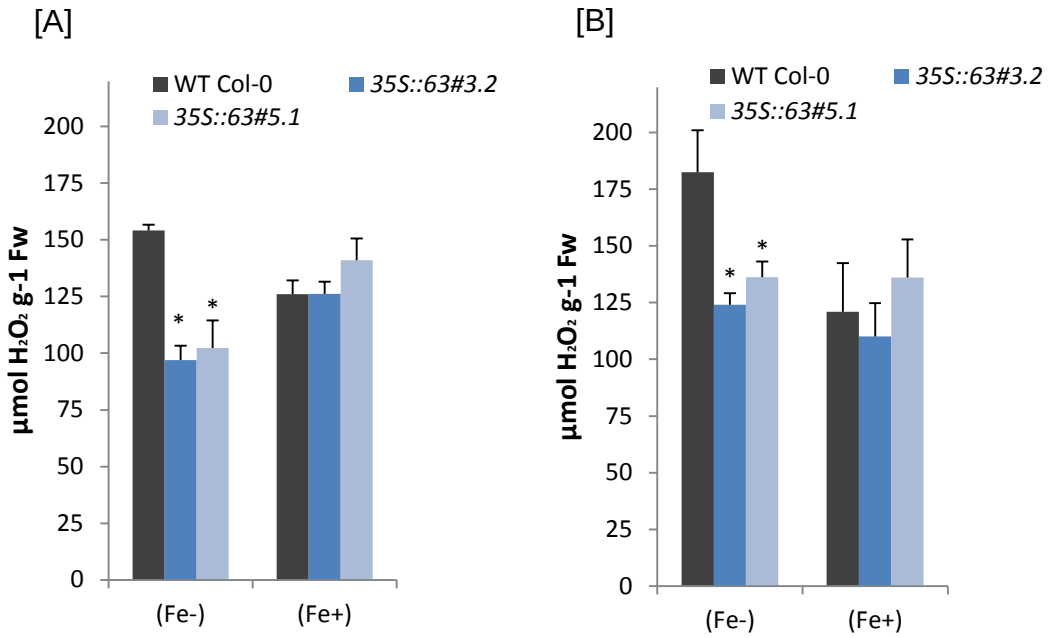


Figure 41: Hydrogen peroxide production in *35S::ASR63* lines under iron deficiency.Hydrogen peroxide production in *35S::ASR63* lines under iron sufficient (Fe+) and iron deficient (Fe-) conditions at pH 4.2 (A) and pH 5.7 (B), expressed in µmol per gram fresh weight. Values are mean +/- SEM, students t test (P<0.05), means compared with respective wild types.

Interestingly the concentration of superoxide was found to be significantly higher in *35S::ASR63* lines at normal conditions, 100 µM Fe at pH 5.7 (Fig.42B). The same result was seen in Fe(-) condition also where superoxide levels were similar to Fe(+) conditions. The same trend was observed at pH 4.2 where *35S::ASR63* lines had higher superoxide levels compared to wild type both at Fe(+) and Fe(-) conditions. However, at pH 4.2 superoxide levels of wild type and *35S::ASR63* lines were reduced under Fe(-) compared to Fe(+) conditions (Fig. 42A).

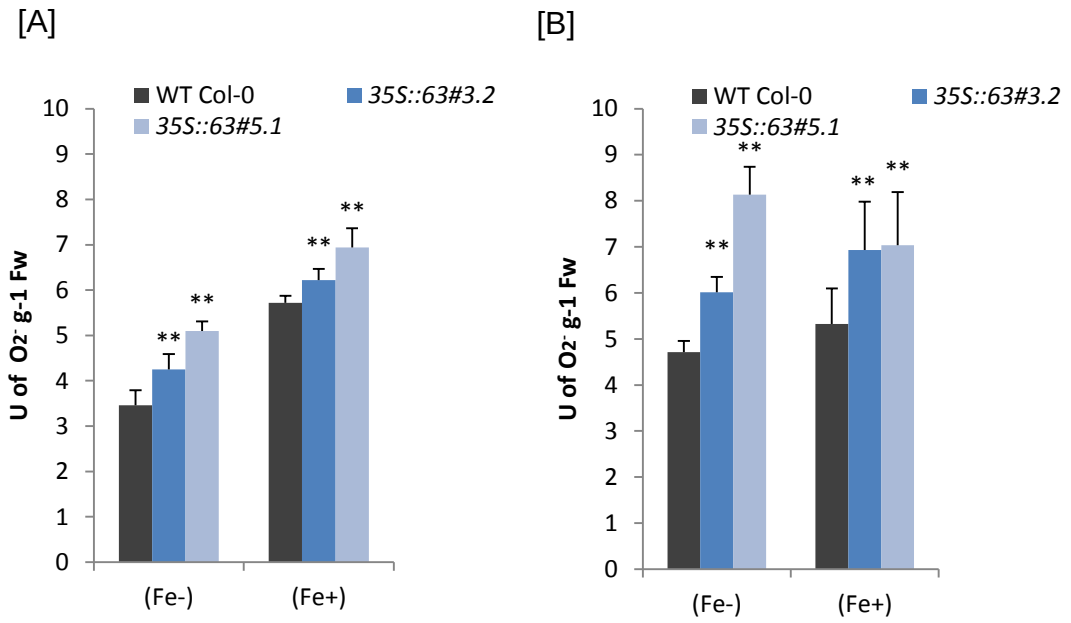


Figure 42: Measurement of superoxide in *35S::ASR63* lines under iron deficiency.Superoxide levels in *35S::ASR63* lines under iron sufficient (Fe+) and iron deficient (Fe-) conditions at pH 4.2 (A) and pH 5.7 (B), expressed in units per gram fresh weight. Values are mean +/- SEM, students t test (P<0.01), means compared with respective wild types.

5.2.7 The antioxidants Glutathione and Ascorbic acid are increased in *35S::ASR63* lines under iron deficiency

Glutathione (GSH) and ascorbic acid are considered to be the most important antioxidants and play a major role in detoxification of reactive oxygen species under different stress conditions. It has been shown recently that GSH and ascorbic acid prevents chlorosis and accumulation of reactive oxygen species under iron deficiency in *Arabidopsis* (Ramirez et al., 2013). Hence we checked the levels of GSH and ascorbic acid under iron depleted and iron replete conditions in order to see whether GSH and ascorbic acid levels are increased in *35S::ASR63* lines. Seven days old seedlings were transplanted to iron depleted or iron replete conditions at pH 4.2 and pH 5.7 and grown further 4 days before collecting the shoot tissue for GSH estimation. The total GSH content at pH 4.2 was increased in *35S::ASR63* lines under iron depleted conditions (2 µM Fe) compared to wild type whereas no significant difference was observed under iron replete conditions (100 µm Fe). At 2 µM Fe the total GSH content was increased by 40-75 nmol

per gram fresh weight compared to wild type control (Fig.43A). The total GSH was also increased at pH 5.7 under iron depleted conditions by 80-90 nmol per gram fresh weight compared to wild type. However under iron replete condition at pH 5.7, total GSH was higher in wild type compared to *35S::ASR63* lines (Fig.43B).

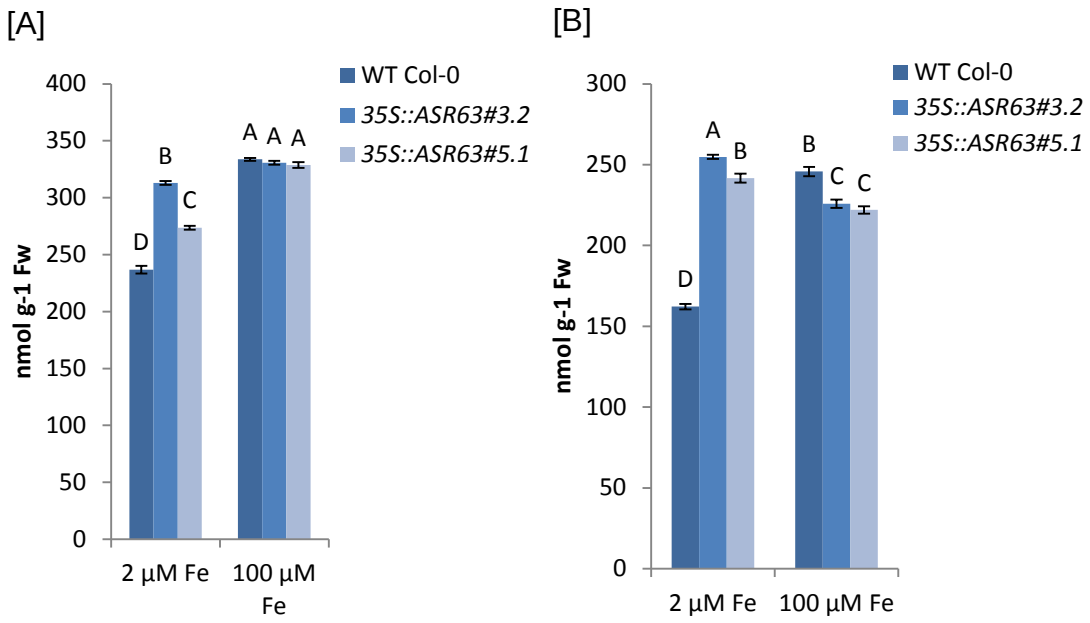


Figure 43: Total GSH content in *35S::ASR63* lines under iron deficiency. Total GSH content from shoot tissue compared between wild type and *35S::ASR63* lines at pH 4.2 [A] and pH 5.7 [B] under iron deficiency. Values are mean +/- SEM, one way ANOVA ($P<0.05$), levels not connected by same letter are significantly different.

Similarly, total ascorbic acid levels were also found to be increased in *35S::ASR63* lines compared to wild type under both iron depleted and iron replete conditions irrespective of pH difference (Fig.44). Ascorbic acid is a multifunctional metabolite involved in redox balance in plants and also gives protection under iron deficiency conditions. These results collectively suggest that the *ASR63* overexpression leads to increased accumulation of antioxidants GSH and ascorbic acid thereby preventing chlorosis in *35S::ASR63* lines under iron starvation.

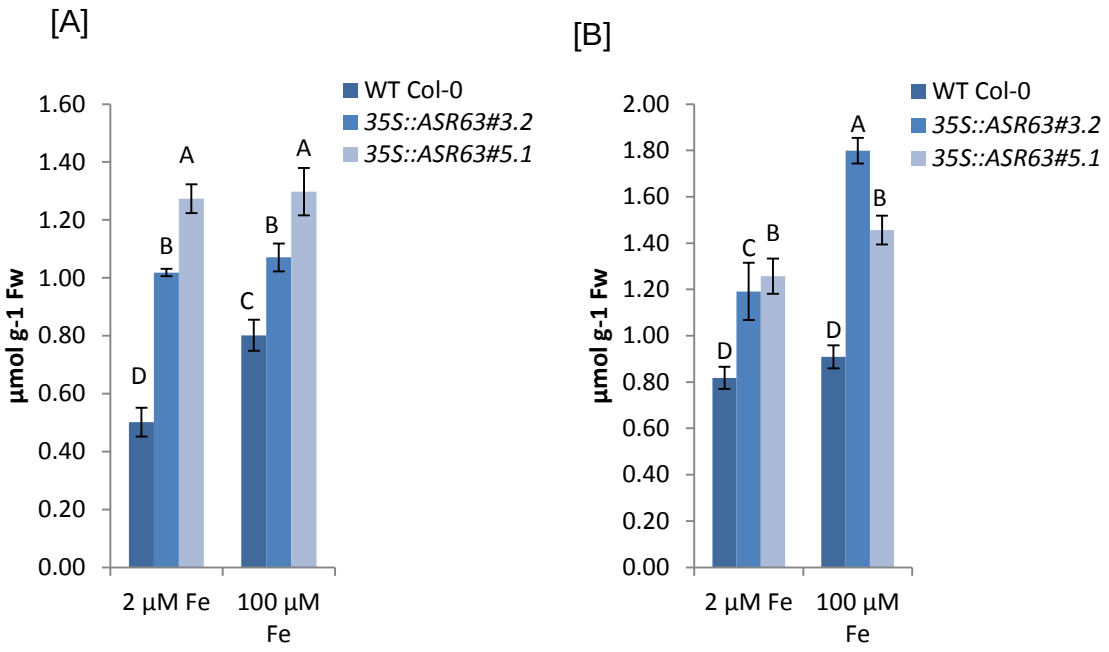


Figure 44: Total ascorbic acid levels in *35S::ASR63* lines under iron deficiency.Total ascorbic acid content from shoot tissue compared between wild type and *35S::ASR63* lines at pH 4.2 [A] and pH 5.7 [B] under iron deficiency. Values are mean +/- SEM, one way ANOVA (P<0.05), levels not connected by same letter are significantly different.

5.2.8 *35S::ASR63* lines accumulate higher bioavailable iron compared to wild type plants

Since *35S::ASR63* lines accumulated higher iron, we analysed the bioavailability iron in wild type and *35S::ASR63* lines. Because this would also help to assess the potential utilization of the gene for biotechnological applications as iron deficiency is one of the major causes of concern in majority of the developing countries. The iron bioavailability was assessed according to standard method using Caco-2 cells as described previously (Glahn et al., 1998) with a slight modification as described in the methods section. The rosette leaves were used for the analysis. The result suggested that about 63 to 73% of the total iron accumulated in *35S::ASR63* lines was in fact bioavailable where as it was only 34% in wild type control (Fig.45). These results suggested that overexpression of *ASR63* leads to accumulation of more absorbable iron in the plant tissues.

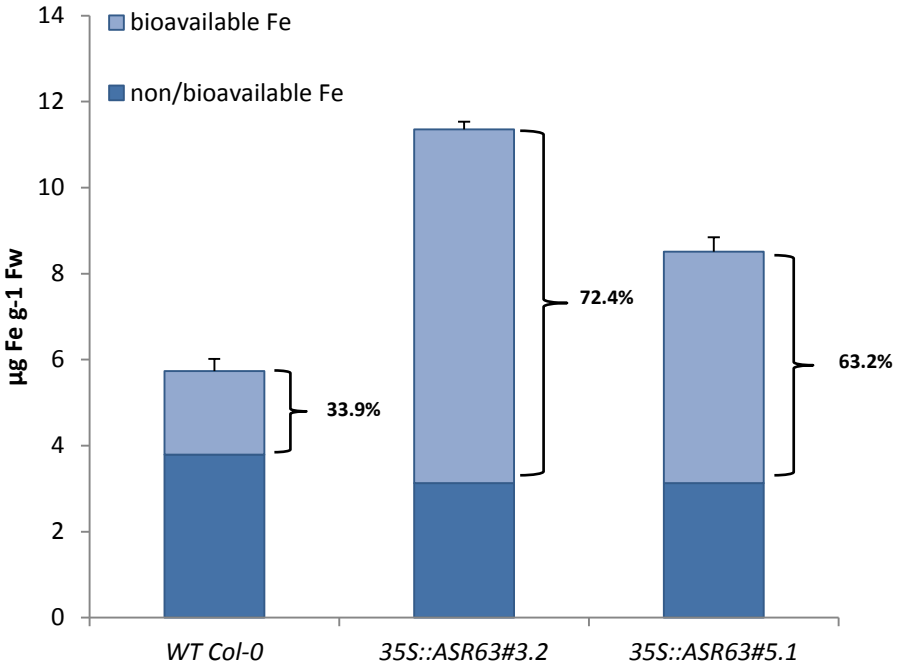


Figure 45: Iron bioavailability assay. Four week old rosette leaves were used for the iron bioavailability analysis. Digested leaf samples were treated to Caco-2 cells for four hours before analysis. Iron content is expressed in micrograms iron per gram fresh weight of leaves. Values are mean +/- SEM from three replicates. The experiment was repeated three times with similar results. The percentage values represent the bioavailable iron calculated from total iron in the treated sample.

5.2.9 *ASR63* overexpression improves the growth of *aft1* mutant in *Saccharomyces cerevisiae* in a copper dependent manner

Functional complementation of yeast mutants using plant genes is a routine method to identify the gene functions. In order to find the possible pathway for *ASR63* in iron homeostasis we used range of yeast mutants defective in iron uptake and homeostasis and overexpressed *ASR63* in those mutants under GPD promoter. We found that *ASR63* improved the growth of yeast mutant *aft1*. Aft1p in yeast is an iron-dependent transcription factor involved in transcriptional activation of a set of genes involved in iron acquisition, mobilization of stored iron and metabolic perturbations during growth under iron limited conditions (Philpott and Protchenko, 2008). Yeast *aft1* grows poorly and requires excess iron for normal growth and also exhibit

reduced iron uptake (Rutherford et al., 2003). The growth of *aft1* transformed with *ASR63* was compared to the wild type (YKB779) transformed with empty vector. *aft1* transformed with empty vector was used as negative control. Drop test result suggested that *ASR63* improves growth of *aft1* mutant on YPD media (Fig. 46). However, *ASR63* overexpression did not improve growth of *aft1* when YPD media was supplemented with iron chelator Ferrozine (FRZ - 300 µM). Increased copper is known to improve the growth of *aft1* and *fet3* mutants (Bleackley et al., 2011). In order to analyse whether excess copper improves the growth of *aft1* overexpressing *ASR63,* excess copper and low iron media (300 µM FRZ and 2.5 mM $CuSO_4$) was used. The results showed that when YPD + ferrozine media was supplemented with excess of copper (2.5 mM $CuSO_4$), the growth was improved in both *aft1* and *aft1/ASR63.* This suggested that the *ASR63* improves growth of *aft1* through a copper dependent pathway.

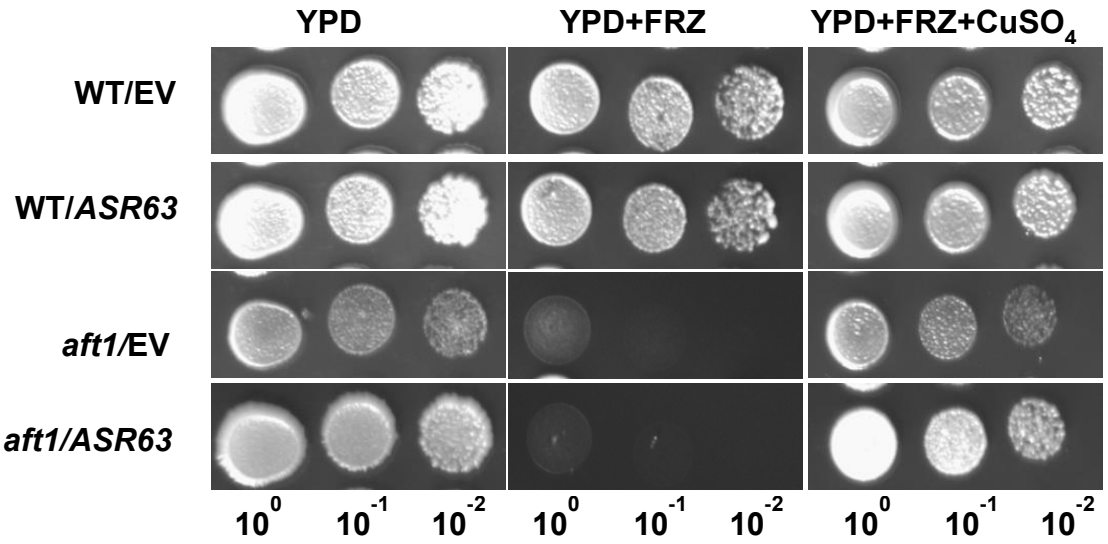


Figure 46: The copper dependent complementation of *aft1* mutant by *ASR63.* Overnight grown cultures were used for drop test. Cultures were adjusted to OD 2 and two additional serial dilutions were spotted on to the media as indicated. WT-Wild type (YKB779); EV-empty vector; YPD-yeast peptone dextrose media; FRZ-ferrozine. Images were taken 3 days after incubation at 28 °C.

5.2.10 Overexpression of *ASR63* in *aft1* mutant leads to increased accumulation of Iron

It has been shown previosly that *aft1* mutant cells show reduced iron uptake under iron deficient conditions (Berthelet et al., 2010; Yamaguchi-Iwai et al., 1995). In order to check whether *ASR63* overexpression leads to increased uptake of iron in *aft1* we analysed the uptake of iron under low iron (2 µM Fe) and iron sufficient (30 µM Fe) conditions in *aft1* overexpressing *ASR63*, wild type (YKB779) and *aft1* transformed with empty vector (*EV/aft1*). The result showed that *aft1* overexpressing *ASR63* accumulating significantly higher amount of iron compared to wild type and *aft1* transformed with empty vector under low iron conditions. However under iron sufficient conditions *aft1* overexpressing *ASR63* showed lower accumulation of iron compared to wild type control but significantly higher compared to *aft1* transformed with empty vector (Fig.47).

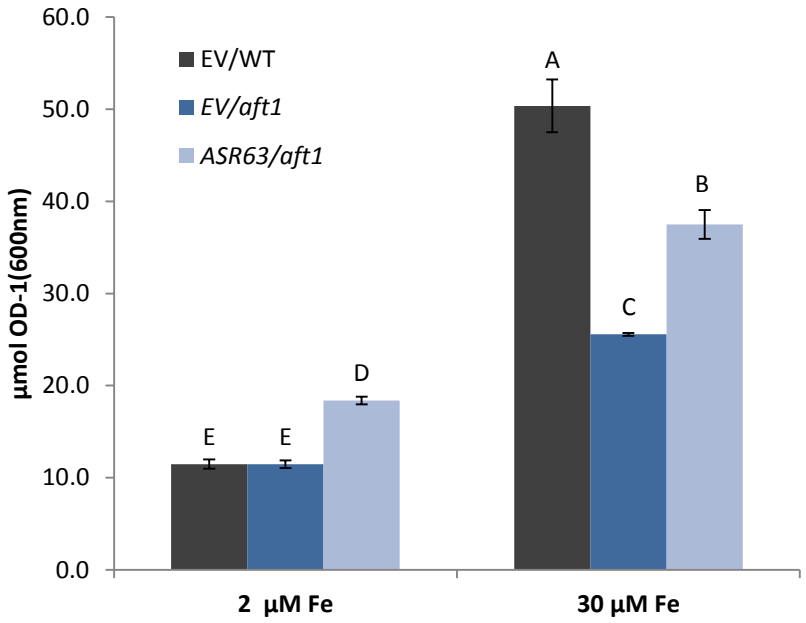


Figure 47: Measurement of iron accumulation in *aft1* mutant overexpressing *ASR63.* Values are +/- SEM, one way ANOVA (P<0.05), levels not connected by same letter are significantly different. EV/*aft1*- *aft1* transformed with empty vector; EV/WT - Wild type transformed with empty vector; *ASR63*/*aft1* - *aft1* transformed with *ASR63.*

5.2.11 Activity of [Fe-S] enzyme aconitase support higher Fe availability in *aft1* overexpressing *ASR63*

To further confirm the higher iron accumulation, the Fe-S protein aconitase activity was analysed in wild type, *aft1* and *ASR63/aft1* (*aft1* transformed with *ASR63*) under iron sufficient (100 µM Fe) and low iron conditions (2 µM Fe). The overnight grown cultures in the SOD media with indicated concentration of iron were used for the aconitase assay. The aconitase activity was determined using a NADP-coupled assay in the presence of excess of isocitrate dehydrogenase as described previously (Drapier and Hibbs, 1996). The results show that the aconitase activity in *ASR63/aft1* cells was similar to that of wild type under both iron sufficient and iron deficient conditions while in *aft1*/EV aconitase activity was significantly reduced under low iron as well as iron sufficient conditions (Fig. 48). The results further confirm that *ASR63* overexpression leads to increased uptake of iron in *aft1* cells complementing the *aft1* phenotype.

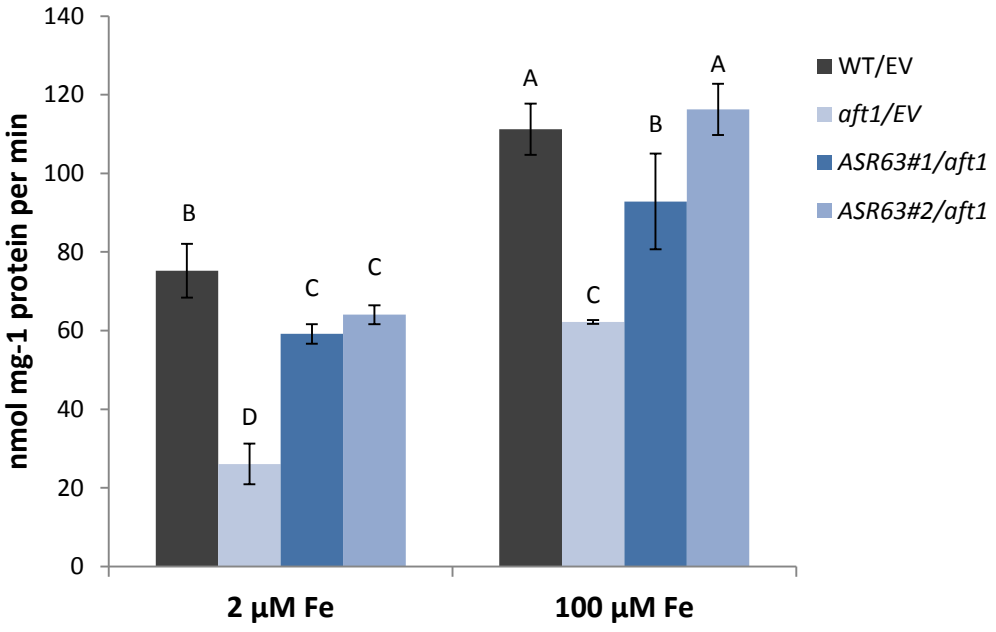


Figure 48: Measurement of aconitase activity under iron deficiency in Yeast. Aconitase activity under iron sufficient (100 µM Fe) and low iron (2 µM Fe) conditions were measured. Values are +/- SEM, one way ANOVA ($P<0.05$), levels not connected by same letter are significantly different.

5.3 Discussion

5.3.1 *ASR63* is a Brassicaceae specific gene likely to have originated *de novo*

Our analysis of *ASR63* indicated that it is found only in the Brassicaceae lineage. This was supported by lack of blastp and tblastn hits in sequence databases. To further support this we analysed the syntenic blocks in Brassicaceae and the closest outgroup species in Cleomaceae. If the flanking regions of a pair of sequence fragment are in an orthologous relationship in two related species then the sequence fragments in question are expected to be in a orthologous status even though they do not have a high BLAST match (Cai et al., 2008). From the syntenic analysis it is clear that, *ASR63* orthologous sequences are present in all the six Brassicaceae species analysed but missing in the outgroup species *T. hassleriana* suggesting that the *ASR63* is likely to have originated only in Brassicaceae. This was further supported by the absence of conserved gene families related to *ASR63* while its flanking genes have orthologs in many species outside of Brassicaceae including grasses (Fig. 35). This suggested that *ASR63* has originated *de novo* after the split of Brassicaceae, and further the *ASR63* locus has diverged so much that it was not possible to find the orthologus non-coding sequence outside of Brassicaceae.

The evolution of new genes (or orphan genes) has always been an essential part of evolutionary genetic studies. The emergence of new genes is suggested to bring lineage specific evolutionary innovations to the organism through lineage specific adaptations. The two major mechanisms of new gene origination considered are the duplication divergence model and *de novo* emergence model. The *de novo* emergence of genes has been shown in several cases in Drosophila, yeast, humans as well as in plants (discussed in section 1.2.4). Recently it has been shown in humans that bidirectional promoters could be a source of emergence of new genes in a genome (Gotea et al., 2013). In fact many of the active promoters from yeast to mammals are found to be bidirectional and known to initiate transcription on both orientation which is known as ‘divergent transcription’ and is

observed for the majority of mammalian promoters (Seila et al., 2008). In plants, there are only a few examples of studies on individual bidirectional gene pairs, this may be because genome wide studies on bidirectional promoters are scarce. Here we propose that *ASR63* has originated by bidirectionality of the promoter of the upstream gene AT4G18590. The bidirectionality of the promoter likely resulted in producing cryptic unstable transcript of the *ASR63* into non-coding RNA which over time gained the necessary splice sites and ORF to translate a protein which possibly gained function in *A. thaliana*. Recently it has been shown in yeast that non genic sequences are frequently transcribed (called 'protogenes') and go through a non-coding phase which gets translated at low frequencies to form short peptides. Some of the peptides overtime may confer advantage to the organism and become fixed in the population forming a new gene (Carvunis et al., 2012).

Studies on *de novo* gene origination by bidirectional promoters have just begun. A recent study identified 24 hominoid-specific *de novo* genes in human (Xie et al., 2012), of which five genes are derived from bidirectional promoters, suggesting that the divergent transcription through bidirectional promoter is an important source of *de novo* gene origination. In plants there are only few genome wide studies on bidirectional promoters in *A. thaliana*, Rice and *Populus* (Dhadi et al., 2009; Kourmpetli et al., 2013; Wang et al., 2009). A recent study identified 1696 bidirectional transcript pairs in maize and analysed the promoter activity of most of the transcript pairs expressed in maize embryo (Liu et al., 2014). The same study also compared the bidirectional transcript pairs with other monocot (*Zea mays*, *Sorghum bicolor, Oryza sativa*) and eudicot (*Arabidopsis thaliana* and *Glycine max)* plant species and show that bidirectional gene pairs are abundant in those species. To further understand the role of the promoter in the evolution of *ASR63*, a comparison and analysis of promoter activities of *ASR63* and its orthologous regions in other Brassicaceae species needs to be done. This would give a clearer understanding about the evolutionary origin and functionality of *ASR63* in the Brassicaceae lineage.

5.3.2 Ectopic overexpression of *ASR63* confers tolerance to iron deficiency condition in *A. thaliana* possibly through maintaining optimum redox status

Iron is an essential micronutrient for plant growth and also for majority of all forms of life. The deficiency of iron drastically affects the plant growth and particularly in crop plants reduces the yield and quality of the produce. In plants, the majority of the genes involved in iron acquisition and homeostasis have been discovered. The major genes involved in iron acquisition are known to be transcriptionally upregulated under iron deficiency to enhance iron uptake to cope up with the iron requirement of the plant (Hindt and Guerinot, 2012; Kobayashi and Nishizawa, 2012). *ASR63* shows expression in many parts of the plant such as leaves and roots, mainly in vasculature and also in flower petals. The ectopic expression of *ASR63* led to increased tolerance to iron deficiency condition. Since, under low pH, iron is more soluble and readily available for uptake, the phenotype was also analysed at pH 4.2 along with pH 5.7. The tolerant phenotype was observed at both the pH levels analysed suggesting pH has no influence on the phenotype. The very obvious phenotype due to iron deficiency is chlorosis which occurs due to reduced biosynthesis of chlorophyll. The overexpression lines appeared to accumulate enough iron since they did not show chlorotic phenotype. This was supported by measurement of total iron content which indicated that *35S::ASR63* lines can indeed accumulate more iron in shoot tissue under iron deficiency conditions which is almost the same as that of normal conditions (100 µM Fe). Interestingly iron accumulation was higher in shoots compared to wild type control even under normal conditions.

The knockdown of *Arabidopsis OPT3*, an oligopeptide transporter family gene, leads to higher accumulation of iron in the leaves while the *opt3* null mutant is embryo lethal (Stacey et al., 2008). However constitutive expression of *AtOPT3* did not show accumulation of iron. The overexpression of nicotianamine synthase gene (*NAS*) in *Arabidopsis* resulted in accumulation of iron in both roots and shoots under iron deficiency but the plants remained chlorotic compared to wild type (Cassin et al., 2009). However *35S::ASR63* lines showed normal growth under iron

deficiency and also showed higher accumulation of iron compared to wild type control under normal growth conditions (Fig.39). The higher activity of Fe-S proteins under iron deficiency conditions (Fig.40) further support that *35S::ASR63* lines accumulate more iron which possibly helps the plant to function normally under iron depletion. This suggests that *ASR63* might be involved in facilitating the uptake and transport of iron in *A. thaliana*. Further many of the proteins involved in iron transport within the plant are expressed in the vascular tissues. Hence investigating the transcriptome changes upon overexpression of *ASR63* will help to understand the relation between *ASR63* function and iron uptake/transport.

Reactive oxygen species (ROS) are the normal products of plant metabolism but any stress condition leads to excessive production of ROS which can cause oxidative stress and ultimately results in oxidative damage to cells/tissues if not effectively scavenged. Hence ROS production was quantified (H_2O_2 and superoxide anion) in *35S::ASR63* lines to see the effect of *ASR63* overexpression on ROS under iron depletion. The results suggested that *35S::ASR63* lines indeed effectively scavenged ROS under iron depletion. Although superoxide content was higher in *35S::ASR63* lines (both under normal and iron depleted condition), superoxide is a moderately reactive and short lived ROS (half-life of approx. 1 µs.) and its detoxification involves conversion to H_2O_2 enzymatically (by dismutases) or non-enzymatically. H_2O_2 was found to be effectively detoxified in *35S::ASR63* lines under iron depletion. However it may not be the direct effect of *ASR63* overexpression. Because iron deficiency induced oxidative stress is due to less availability of iron which is a constituent or cofactor for many of the antioxidant enzymes. It has been shown that in sunflower the iron deficiency causes reduced production of peroxidases specifically ascorbate peroxidase (Ranieri et al., 2001) which is involved in detoxifying peroxides using ascorbate as substrate. A recent study in *Arabidopsis* showed that exogenous application of glutathione (GSH) and ascorbic acid (ASC) prevent chlorosis and the accumulation of ROS under iron deficiency in the leaves (Ramirez et al., 2013). The quantification of GSH and ASC in *35S::ASR63*

lines suggested that the absence of chlorosis phenotype in the transgenic *ASR63* lines is likely due to the increased GSH and ASC.

The antioxidants GSH and ASC are the major metabolites keeping the cell redox status under control. Increase in GSH and ASC also increase the activity of ascorbate peroxidase (Ramirez et al., 2013) and hence *35S::ASR63* lines effectively scavenge H_2O_2 thereby preventing oxidative stress caused due to the iron depletion. However Ramirez et al. (2013) also found that supplementing GSH and ASC protected *Arabidopsis* seedlings under iron deficiency without altering the total iron content in the treated seedlings. Hence accumulation of iron in *35S::ASR63* lines might be independent of increase in GSH and ASC content under iron deficiency and needs to be addressed in future studies.

5.3.3 *Arabidopsis* orphan gene *ASR63* exhibits a functional phenotype in the distant species *Saccharomyces cerevisiae*

Identifying the functionality or genetic pathways of plant genes using the yeast system (e.g. *S. cerevisiae*) is a popular and proven method. This methodology has been used to identify the functionality of large number of plant genes (discussed in chapter 4) including some of the genes involved in plant iron homeostasis. In order to evaluate the iron uptake phenotype exhibited by *ASR63* in *Arabidopsis,* a range of yeast mutants defective in iron homeostasis were transformed with the *ASR63* coding sequence under *GPD* promoter to see whether any mutants can be rescued. Surprisingly, *ASR63* rescued the mutant for Aft1p (*aft1*) which is a major iron dependent transcription factor in yeast. *ASR63* improves the growth of *aft1* in normal YPD media. However when YPD media was supplemented with the iron chelator ferrozine (FRZ), *ASR63* has no effect on growth of the *aft1* mutant. In yeast, iron and copper metabolism are very closely linked as copper is required for high affinity iron transport cuproenzymes Fet3p and Fet5p (van Bakel et al., 2005). Fet3p is a ferroxidase that converts ferrous into ferric iron and the permease, Ftr1p, transports ferric iron into cytoplasm. Fet3p-Ftr1p constitutes the high affinity iron transporter in the plasma membrane. In order

to investigate whether copper has any effect on improving the growth of *aft1* by *ASR63* overexpression, excess copper (2.5 mM) was used. The excess copper of upto 2.5 mM was used in a previous study and above 3 mM is considered as toxic (van Bakel et al., 2005). Indeed excess copper supplementation to YPD+ferrozine media rescued the phenotype suggesting the growth improvement of *aft1* by *ASR63* is copper dependent. Although ferrozine is thought to be iron specific chelator, it was previously reported that ferrozine can form coloured complexes with monovalent copper (Stookey, 1970). This is possibly the reason for excess copper requirement for *ASR63* when ferrozine was used in the media. It needs to be further investigated whether *ASR63* can rescue *aft1* in the absence of copper in the media which can further confirm copper dependent rescue of *aft1* growth phenotype by *ASR63* in *S. cerevisiae*.

Most importantly *aft1* transformed with *ASR63* accumulates more iron compared to the mutant and also wild type under iron deficiency conditions, a phenotype also observed in *35S::ASR63* lines in *Arabidopsis* which suggest a role for *ASR63* in iron uptake and homeostasis. Higher activity of Fe-S protein aconitase further supports iron availability in *aft1* due to *ASR63* overexpression. Since the phenotype conferred by *ASR63* on *aft1* is linked with the availability of copper in the media, investigating the effect of *ASR63* in the mutants for cuproenzymes Fet3p (*fet3)* and Fet5p (*fet5*), seem to be the obvious next steps to dissect the pathway of *ASR63* function in yeast. Overall, the functionality of the orphan gene *ASR63* in yeast suggests that orphan genes can be used in a distant species as well to confer novel traits.

5.3.4 *ASR63* could be a potential target gene for biotechnological application for crop improvement

Apart from characterizing the functionality of orphan genes, the long term goal of the present work is to identify whether the genes conferring abiotic stress tolerance can be utilized for biotechnological application for example, in crop improvement. Since *35::ASR63* lines accumulated more iron and iron is not only an essential element for plants but also is an

important micronutrient for human health. For human consumption whatever the amount of iron accumulated in the plant product, has to be bioavailable. Hence we analysed the amount of bioavailable iron in the *35S::ASR63* lines. Interestingly, the *35S::ASR63* lines accumulated about 35 to 40% more bioavailable iron in the rosette tissue (Fig.45). Further bioavailability of the iron that is accumulated in the seed needs to be analysed as seed is the edible part in major food crops (cereals and pulses). It is important to note that about 12% of the world's population is suffering from iron deficient anaemia and iron deficiency is considered to be one of the most important disease precursors (world health organization report). Hence iron biofortification in major food crops is in urgent need. However it is important to understand the mechanism of action of *ASR63* which would help to design strategies for use of this gene in other crop plants for biofortification. For now it can be considered as a potential gene for utilization in nutritional enhancement of crop plants.

5.4 Conclusions

In the genomics era, large scale sequencing efforts combined with comparative genomics has led to the identification of lineage specific novel genes called orphan genes (*de novo* genes). The earlier assumption that *de novo* evolution of genes is a rare event has been questioned by the increasing evidence for *de novo* gene emergence in many of the species from bacteria to fungi to humans. This study further adds to the existing evidences for *de novo* gene emergence and a possible mechanism for the origin that is through divergent transcription. Divergent transcription has been shown to be a common process in many organisms from yeast to humans, however there are not many studies on this. From this study it can be stated that in plants *de novo* gene evolution from bidirectional promoters is an important mechanism of gene evolution which can contribute to the evolution of novel species specific traits.

Identification of *ASR63* in iron deficiency tolerance in *Arabidopsis* will further help to increase our understanding of regulatory network of iron deficiency responses. Iron deficiency is a major problem in agricultural crops which drastically reduces the crop yield and productivity (particularly in calcareous soils). In addition, the nutrient status of plants is also very important which can directly affect human health. *ASR63,* a potential new player in iron homeostasis pathway holds promise for not only improving plant growth under iron deficiency but also human health. The future work should be focussed on dissecting the functional pathway of *ASR63* which will help understand how novel genes become functional and get integrated into the conserved pathways.

Chapter 6: A novel function for the Brassicaceae specific orphan gene *ASR63* (AT4G18580) in seed germination in *Arabidopsis thaliana*

6.1 Introduction

Seeds are important organ in higher plants which are not only required for the reproductive success of the plant but also for the dispersal of the species. In the seed plant life cycle, seed germination is a crucial process. Seed germination is a complex process which is highly regulated by endogenous hormones and external environmental factors particularly water, light and low temperature (Finkelstein et al., 2008; Penfield et al., 2005; Seo et al., 2009). Moreover the optimum conditions for germination are often species or lineage specific which means seed dormancy played a significant role in adaptation and evolution of seed plants (Seo et al., 2009). The role of plant hormones in seed germination has been well studied. It is known that the antagonistic balance of the abscisic acid (ABA) and gibberellic acid (GA) in promoting dormancy and germination constitutes the basic genetic framework of seed dormancy (Nonogaki, 2014; Seo et al., 2009). Indeed, high ABA/GA ratio prevents seed germination; ABA deficiency during seed development is associated with absence of primary dormancy (vivipary), and ABA overexpression is associated with enhanced dormancy (Graeber et al., 2012; Rajjou et al., 2012; Yamaguchi et al., 2007).

Recently it was shown that increased ABA biosynthesis is required to establish seed dormancy during embryo maturation not only in *Arabidopsis*, but also in wheat plants (Schramm et al., 2013). However, genetic studies revealed that ABA produced by maternal tissues or supplied exogenously is not sufficient to induce dormancy (Finch-Savage and Leubner-Metzger, 2006; Finkelstein et al., 2008). As a result, to induce seed dormancy the ABA synthesis should occur in the embryo and/or endosperm. Some genes like *NCED6* and *NCED9* essential for ABA synthesis in both the embryo and endosperm during dormancy induction were identified. Indeed, *Arabidopsis* plants over expressing *NCED6* and *NCED9* had 3-4 fold increase in ABA content in the seeds compared to the wild-type seeds, while *nced6* and *nced9* T-DNA mutants had 30-50% ABA reduction and level of their expression correlated with seed dormancy and germination (Lefebvre et al.,

2006). Additional studies have identified specific genes correlated with dormancy maintenance (such as *ZEP,* encoding zeaxanthin epoxidase) and dormancy release via ABA catabolism (*CYP707A2*) (Cadman et al., 2006; Millar et al., 2006). *ABI3*, *ABI4* and *ABI5* encode different types of transcriptional factors that regulate the expression of ABA-responsive genes in seeds (Giraudat et al., 1992) and reduce the sensitivity of seed germination to exogenous ABA as well as exhibit additional defects in various aspects of seed maturation but do not seem to be altered in vegetative responses to ABA (Finkelstein et al., 2008; Koornneef et al., 1984). Conversely, mutations in *ERA1* (*ENHANCED RESPONSE TO ABA 1*), encoding the β-subunit of a farnesyl transferase, increase the sensitivity of seed germination to applied ABA, prolong seed dormancy and improved ability to withstand drought (Leung et al., 1997; Pei et al., 1998).

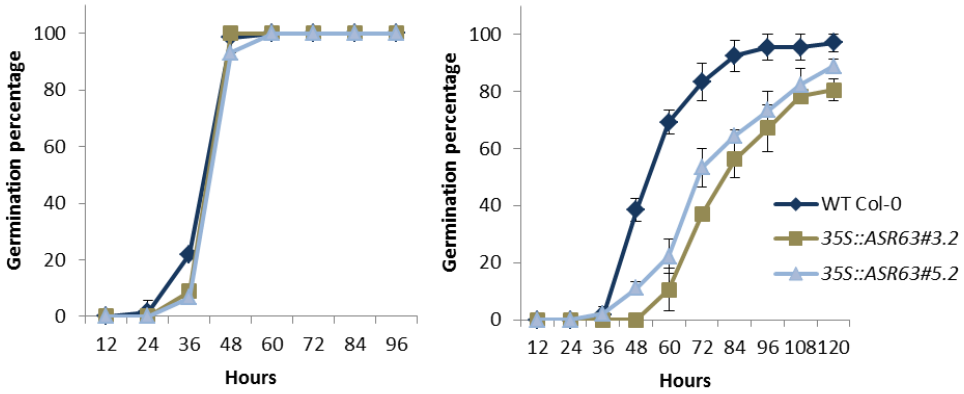
On the other hand the inhibitory effect of ABA is counteracted by GA and ethylene. GA counteracts ABA in the early and late phase of germination whereas ethylene is important in the late phase (Linkies et al., 2009). It has been found that dry seeds of *Arabidopsis* contains physiologically relevant amounts of bioactive GA_4 which further increases during the late phase of germination (Ogawa et al., 2003). Further, transcripts of GA_{20} and GA_3 oxidases, key enzymes regulating GA biosynthesis, accumulate during the early phase of germination. GA/ABA ratio determines the process of germination. During the early stages GA/ABA ratio increased up to 3 fold and in the late phase it further increased up to 10 fold, possibly due to the degradation of ABA at this stage. Further GA_{20} and GA_3 oxidases are induced by red light and cold temperature (Yamauchi et al., 2004). Cold stratification of seeds in *Arabidopsis* is associated with accumulation of bioactive GAs, GA_{20} and GA_3 oxidases which helps to break dormancy. Among other factors involved in seed dormancy establishment and release, the role of ROS (reactive oxygen species, such as hydrogen peroxide and superoxide) as ABA regulators has been revealed (El-Maarouf-Bouteau and Bailly, 2008; Liu et al., 2010).

This study identifies a novel function for the *de novo* gene *ASR63* in seed germination. The Brassicaceae specific gene *ASR63* confers tolerance to iron deficiency upon constitutive expression in *A. thaliana*. Here we show that *ASR63* overexpression also leads to sensitivity to ABA during seed germination. Furthermore, the germinating seeds accumulate more ROS suggesting a possible involvement in ROS mediated signalling during germination in *A. thaliana*.

6.2 Results

6.2.1 Seed germination in *ASR63* overexpression lines is affected by ABA and Paclobutrazol

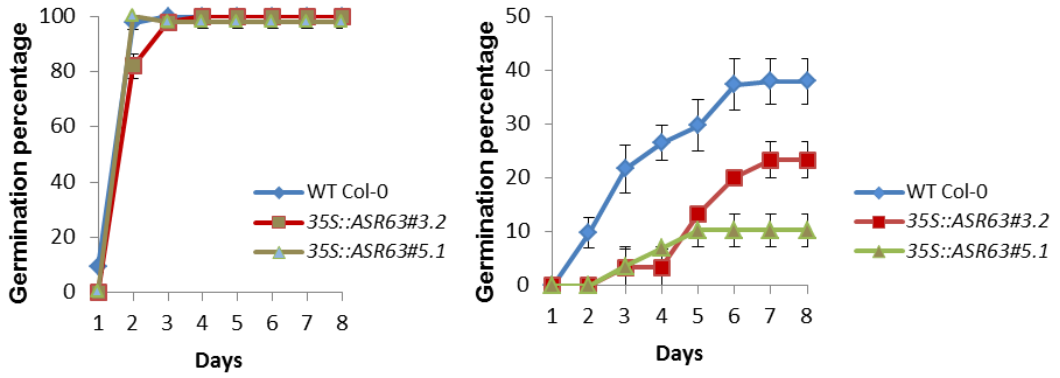
Seed germination is tightly controlled by the abscisic acid (ABA) and gibberellic acid (GA). High ABA and low GA levels prevent seeds germination. To find out if there is any link with ABA in the germination phenotype, the *35S::ASR63* lines were germinated on ABA containing media (n=~120). The seeds of *35S::ASR63* lines did not have any obvious seed phenotypes compared to control. Seed germination was tested on 1.5 µM ABA on MS media and radicle emergence was scored for ABA response from every 12 hours after 3 days of stratification at 4 $^{\circ}$C. *35S::ASR63* lines showed slower rates of radical emergence on 1.5 µM ABA compared to wild type control. Although *35S::ASR63* lines showed slower rates of radical emergence initially, *35S::ASR63* lines reached 100% germination after 7 days (156 hrs) while wild type reached 100% germination within 5 days (Fig.49).



[A] [B]

Figure 49: Effect of ABA on germination of *35S::ASR63* lines. Effect of abscisic acid on seed germination in *35S::ASR63* lines. Radicle emergence was calculated every 12 hrs after 3 day-stratification at 4 °C on 1X MS plates. Significant decrease of germination rate was observed in *35S::ASR63* lines at 1.5 µM ABA [B] while no difference was seen under 0 µM ABA [A]. Values are mean +/- SEM (student's t test, P<0.05).

To investigate any link with GA in germination phenotype of *35S::ASR63* lines, paclobutrazol (PAC), a GA inhibitor was used. The seeds of wild type and *35S::ASR63* lines were germinated (n=~120) on MS media supplemented with 4 µM PAC and the germination was monitored on daily basis for up to one week. *35S::ASR63* lines showed significantly reduced germination rate compared to wild type. After 6 days 35 % of the wild type seeds had germinated whereas only 10 -15% of the *35S::ASR63* lines seeds had germinated when seeds were treated with 4 µM paclobutrazol (Fig.50).



[A] [B]

Figure 50: Effect of paclobutrazol on germination of *35S::ASR63* lines. Effect of GA inhibitor (PAC) on seed germination of *35S::ASR63* lines. Radical emergence was calculated every day after 3 day-stratification at 4 °C on 1X MS plates. Significant decrease of germination rate was observed in *35S::ASR63* lines at 4 µM PAC [B] while no difference was seen under 0 µM PAC [A]. Values are mean +/- SEM (student's t test, P<0.05).

6.2.2 Superoxide accumulation in germinating seeds of *35S::ASR63* lines

Reactive oxygen species (ROS, such as hydrogen peroxide and superoxide) also play a major role as ABA regulators during germination. Therefore, we investigated the level of ROS in 24h and 48h-imbibed embryos of wild type and *35S::ASR63* lines. No difference was observed in the hydrogen peroxide level between embryos of wild-type and *35S::ASR63* lines, stained with 3 mM 3,3'-diaminobenzidine which reacts with H_2O_2 to give a dark-brown colour. However, superoxide content, visualized by 1mM nitroblue-tetrazolium, showed higher amount in 24 hrs imbibed seeds and was approximately 4-5 fold higher in the embryos of *35S::ASR63* lines imbibed for 48 hrs after stratification at 4 °C (Fig.51). It indicates that *ASR63* may directly or indirectly be involved in the superoxide balance in the germinating seeds of *35S::ASR63* lines. Based on these results, it is possible to consider that *ASR63* is involved in the regulation of ABA balance either directly or indirectly by superoxide induction, and thus affect seed dormancy.

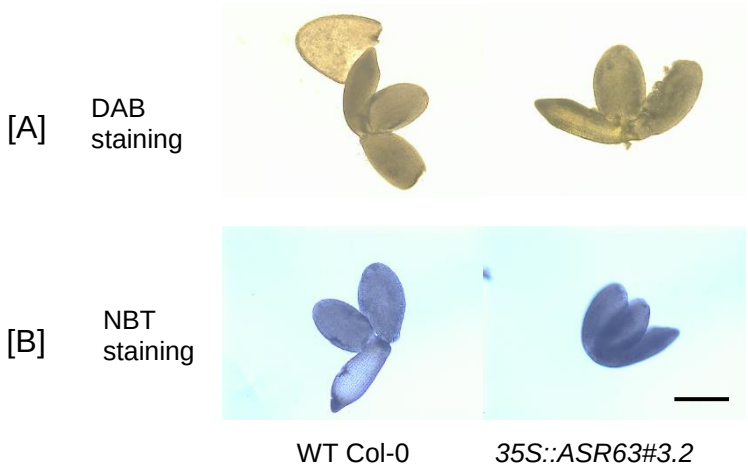


Figure 51: Qualitative analysis of Superoxide accumulation in germinating seeds. Reactive oxygen species (ROS) formation in wild-type and *35S::ASR63* germinating seeds. Embryos were extracted from 48 hrs stratified seeds (approx. 40 seeds per line) followed by 30 min staining with DAB for hydrogen peroxide detection (A) and with NBT for superoxide detection (B).

6.3 Discussion

Seed germination is one of the most complex and tightly regulated processes in plants. Plant hormones, particularly ABA and GA, govern the seed germination process although other hormones such as ethylene, cytokinin, auxin and brassinosteroids are also involved in cross talk with other components of hormonal signalling. ABA has been defined as a negative regulator of germination while GA positively regulates germination. In our analysis *ASR63* overexpression caused the seeds to become sensitive to ABA. Increased ABA sensitivity is usually associated with increased ABA accumulation, which is directly linked with increased seed dormancy and increased pre-harvest sprouting tolerance (Finkelstein et al., 2008). Thus it can be assumed that delayed seed germination of the stored seeds is due to the ABA accumulation or also due to alteration of GA content. Moreover *35::ASR63* lines also showed sensitivity to GA inhibitor paclobutrazol (PAC). The increased sensitivity to ABA and PAC is usually

detected in seeds with a higher ABA level. Therefore, the sensitivity of *35::ASR63* lines to ABA and PAC is likely due to altered seed ABA content.

Although ROS production in seeds is considered detrimental, recent evidence suggests that ROS might be playing a key role in signalling during seed germination or dormancy release (El-Maarouf-Bouteau and Bailly, 2008). The analysis of ROS in germinating seeds indicated that *35S::ASR63* lines accumulate more superoxide after imbibition compared to controls. The ROS production is known to increase after seed imbibition due to the resumption of respiration and is known to play a positive role in germination (Bailly, 2004; Leymarie et al., 2012). However we did not see any difference in H_2O_2 accumulation. It has been shown that ABA suppresses the production of ROS during imbibition in rice seeds and that reduced ROS content is involved with inhibition of germination by ABA (Ye et al., 2012b). Hence, the ROS content under ABA treatment need to be measured to see whether there is suppression of ROS which would explain the sensitive phenotype observed under ABA treatment.

Further, it is interesting to note that ascorbic acid, apart from its antioxidant property, can also act as substrate (as proton donor) for some dioxygenases involved in plant hormone biosynthesis including GA. Ye et al. (2012b) also showed that a high dose of ascorbic acid inhibits germination of rice seeds. We have shown that *35S::ASR63* lines accumulate more ascorbic acid in their shoots under normal conditions (Fig. 43, chapter 5). So, the analysis of ascorbic acid content in the germinating seeds of *35S::ASR63* lines could answer why they show sensitivity to PAC. Based on our result, ectopic expression of *ASR63* might potentially be interfering with ABA and ROS signalling during germination. Whether *ASR63* is directly involved in affecting seed germination or it is involved indirectly through production of antioxidant ascorbic acid, needs further investigation.

6.4 Conclusions

In this chapter, an additional phenotype for *35S::ASR63* lines in seed germination was identified. The sensitive phenotypes exhibited by *35S::ASR63* lines have generated important leads regarding possible next experiments to perform in order to identify any link with ABA and ROS signalling in seed germination. The process of seed germination is highly complex involving tightly regulated signalling pathways that include several phytohormones and molecular cross-talks between them, in addition to environmental cues influencing the germination. The importance of ROS in seed germination is just beginning to be uncovered and *ASR63* could be a new player involved in seed germination in *A. thaliana*.

Chapter 7: Material and Methods

7.1 Chemicals and reagents

All the chemicals and reagents used in the experiments are from Sigma Aldrich unless otherwise mentioned.

7.2 Plant materials and growth conditions

Seeds of *Arabidopsis* mutant lines were obtained from European Arabidopsis Stock Centre (NASC, Nottingham, UK), INRA (France) and RIKEN bioresource centre (Japan). Seeds were surface-sterilized in 70% ethanol for 1 minute and then treated with sterilization solution [5% (v/v) sodium hypochlorite solution + 0.05% (v/v) TritonX-100] for 10 min, washed 5 times with sterilized water, plated on to half strength Murashige and Skoog (0.5X MS) medium [MS salts with vitamins (Sigma,M5519) 2.2 g/L, 0.7% (w/v) agar, pH 5.7, 0.5% (w/v) sucrose], stratified at 4 $^{\circ}$C for 3 days in the dark, and grown in a controlled plant growth chamber [21/18 $^{\circ}$C, 16/8 h light/dark, with light intensity 150 µE m^{-2} s^{-1}.

For transplanting to soil, seedlings were transferred to pots with growth medium (compost:perlite:vermiculite::4:1:1) and maintained under controlled environment growth chamber (16 h photoperiod with 21 °C and 8 h dark at 18 $^{\circ}$C) with ligh intensity 120-150 µE m^{-2} s^{-1}

7.3 DNA isolation and genotyping

Genomic DNA was isolated with the CTAB method (Doyle and Doyle, 1987). Modifications are described. Approximately 100 mg of leaf tissue was ground in 200 µl 2xCTAB buffer (2 % w/v Cetyltriethylammoniumbromide; 100 mM Tris-HCl (pH 8.0); 20 mM EDTA; 1.4 M NaCl; 1 % w/v Polyvinylpyrolidon) mixed thoroughly and incubated for at least one hour under agitation at 65 °C. 200 µl of chloroform:isoamylalcohol mixture (Chloroform:Isoamylalcohol::24:1) was added, mixed thoroughly and centrifuged for 15 minutes. 150 µl of the upper phase was transferred to a new 1.5 ml reaction tube, mixed with 200 µl of 2-propanol and incubated for five minutes at RT. DNA was precipitated at 13000 x g, washed in 70%

ethanol (EtOH) for 5 minutes at 13000 xg and air dried until no traces of EtOH are left. DNA was resuspended in 30µl of sterile double distilled water. The genotype of the mutant lines were verified by PCR as described (http://signal.salk.edu/tdnaprimers.2.html) and the T-DNA insertion was confirmed by sequencing. All the sequencing was done with GATC Biotech, Germany.

7.4 Abiotic stress screening (Chapter 2)

For the screening of segregating lines, half strength MS media with agar supplemented by 150 mM sodium chloride (NaCl) for salt stress, 375 mM mannitol for osmotic stress, 1.5 µM ABA or 20% PEG for surrogate drought screen was used. For the germination growth assay the seeds of wild type and mutant lines were directly germinated on the stress media followed by 4 days stratification at 4 $^{\circ}$C. The germination and growth was monitored on a daily basis starting 5th day. The phenotype was scored as ratio of resistant to sensitive seedlings or vice versa. Seedlings with true leaves were scored as resistant and seedlings with only cotyledonary leaves or ungerminated seeds were scored as sensitive. For genotoxic stress, seeds were germinated on half strength MS media, grown for 7 days. After 7 days seedlings were transferred to half strength MS media supplemented with 150 ppm methyl methane sulfonate (MMS). Phenotypes were scored as ratio of resistant (green): sensitive (dead) seedlings. .Lines showing segregation for resistant and sensitive phenotype were considered as promising candidates for further analysis (Table-4).

For salt stress screening (with homozygous lines), seeds of control and mutant lines were sown on 0.5X MS agar medium supplemented with 0 mM or 150 mM NaCl, stratified for 3 days before transferring to growth cabinet. Green seedlings were counted 16 days after germination.

For root length measurements, seeds of control and mutant lines were sown on 0.5X MS agar medium, grown vertically for 7 days and then transplanted to 0.5X MS agar medium supplemented with stress agent (ABA,

Mannitol or PEG) and grown vertically for another 7 days before taking root length measurements.

7.5 Chlorophyll quantification and photochemical activity of PSII

Chlorophyll assay was carriedout as described by Eun-Deok et al. (2009). Approximately 300 mg of fresh leaves from 4-week old seedlings collected, weighed and frozen in liquid nitrogen followed by lysis in tissue lyser (Qiagen). The powdered tissues were transferred to 15 ml Falcon tube containing 5 ml of 80% acetone, mixed in dark for 30 min. The tubes were centrifuged at 4 °C for 15 min (3,000 rpm), the supernatant transferred to a new centrifuge tube and absorbance (A) of chlorophyll content measured using spectrophotometer. The total chlorophyll (a+b) concentration was calculated using the following formula;

C_{a+b} (mg/g) = [8.02×A663+20.20xA645]×V/1000×W

Where A= absorbance; volume of the extract; W = Weight of fresh leaves (g).

Photochemical activity of PSII was scored using a PAM-2000 device (Walz). Fv/Fm ratios were measured on at least 15 plants using 3 leaves per plant after dark acclimation of 30 minutes.

7.6 RNA isolation and qRT PCR

RNA was extracted using plant RNA mini kit (Bioline,USA) according to the manufacturer's protocol. cDNA synthesis was performed with 500 ng to1ug of RNA using the RevertAid First Strand cDNA Synthesis Kit™ (Fermentas). qRT-PCR reactions were performed in triplicates using a Bio-Rad CFX™ thermocycler and SYBR Green Jumpstart Ready mix™ (Sigma). The PCR profile was 94 °C for 10 minutes followed by 40 cycles of 94 °C for 15 sec denaturation and 60 °C for 60 seconds (annealing and extension). Primers were designed with QuantPrime online tool. (http://www.quantprime.de/main.php?page=home)

7.7 Cloning and transformation

For making overexpression constructs, the open reading frame (ORF) of each orphan gene was amplified from cDNA using primers (Table-3) with *attB* gateway site. The *attB*-PCR product was then recombined into the pDONR207 entry vector (Invitrogen) using BP Clonase II Enzyme mix. The entry clones were sequenced to confirm the ORF before proceeding further. The entry clone was then recombined into binary vectors pMDC32 (Fig. 52) for plant expression (Curtis and Grossniklaus, 2003) using LR Clonase II Enzyme mix (Invitrogen).

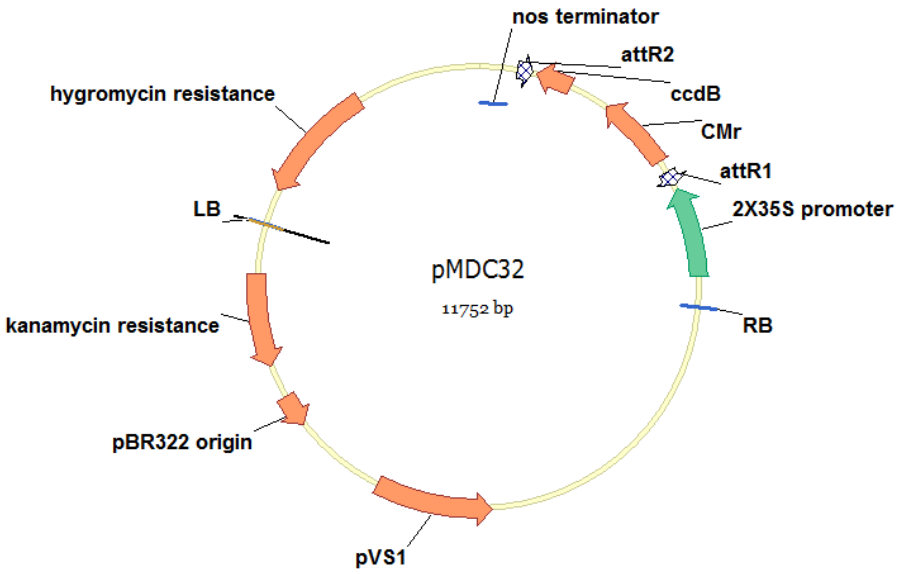


Figure 52: Vector map of plant expression vector pMDC32.

Cloning and transformation reactions were done according to manufacturer's protocol (Invitrogen, USA) except that the amount of BP and LR enzyme mix used was half of the recommended volume. The expression constructs transformed to *Agrobacterium* strain GV3101 pMP90 (a kind gift from Ueli Grossniklaus lab, Switzerland) by electroporation. *Arabidopsis* wild type Columbia-0 plants were transformed using the floral dip method (Clough and Bent, 1998). Transformed seeds were selected on MS medium supplemented with 25 mg/L hygromycin as described by (Harrison et al., 2006).

Similarly for yeast expression, the orphan ORFs from pDONR207 were cloned into yeast episomal vector pAG426GPD-ccdB (Addgene, USA) (Fig 53) through LR cloning reaction as described above.

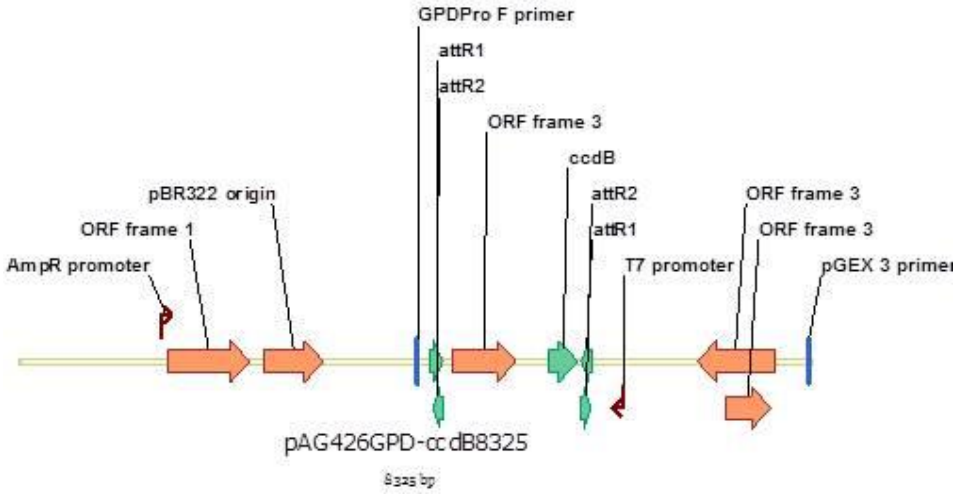


Figure 53: Vector map of yeast episomal vector pAG426GPD-ccdB.

For promoter GUS construct (chapter 5), *ASR63* gene was amplified along with 980 bp upstream of translation start site using following primers

ASR63pGUS-F-
GGGGACAAGTTTGTACAAAAAAGCAGGCTTAACATTCATTCGTCTTTCTCAC
ASR63pGUS-R- GGGGACCACTTTGTACAAGAAAGCTGGGTACACAACAAACCACTCATC

The amplicon was then cloned into pDONR207 and then transferred to pMDC163 (Curtis and Grossniklaus, 2003) (Fig.54). Cloning and transformations were done as described above.

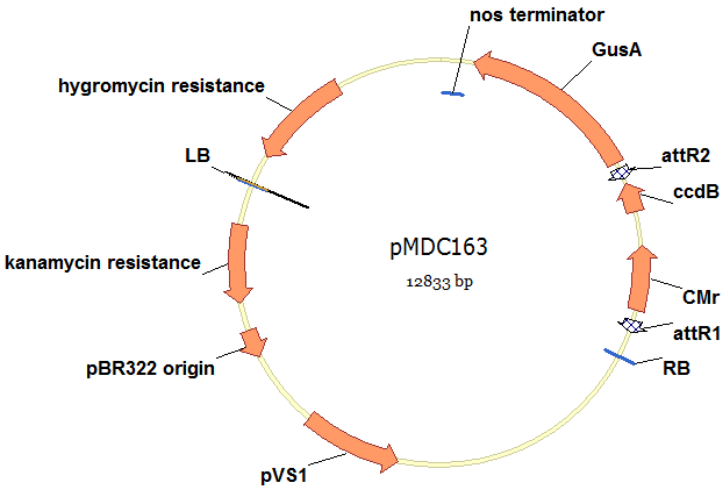


Figure 54: Vector map of plant expression vector pMDC163.

7.8 Abiotic stress screening (Chapter 3)

Seed sterilisation protocol and growth conditions were same as mentioned in section 7.2, unless otherwise mentioned. For germination tests, seeds of wild type and transgenic lines were surface sterilised, stratified at 4°C and plated on 1X MS media supplemented with different concentrations of stress agents (Table 10 and 11): The specific concentrations of stress agents were selected after an extensive literature search for abiotic stress screens on *A. thaliana* Col-0 ecotype.

Table 10: Abiotic stress treatments with concentrations used for germination tests

Stress agent	Concentration (unit)
NaCl	0,75,100,150 (mM)
mannitol	0,200,300,400 (mM)
ABA	0,1.5,2,3 (µM)
IAA	0,0.5,1,2 (µM)
ACC	0,0.1,0.5,1 (µM)
methyl jasmonate	0,0.5,1,5 (µM)

The germination was monitored on daily basis and scored on the basis of hypocotyl emergence and expressed in percentage as germination ratio.

For the root growth assays under stress, wild type and transgenic lines were grown on vertically on MS plates for 5-7 days and then transferred to stress media. The concentrations of stress agents used for root growth mentioned below.

Table 11: Abiotic stress treatments with concentrations used for root growth assays

Stress agent	Concentrations used (concentration unit)
NaCl	0,120,150,200 (mM)
mannitol	0,200,300,400 (mM)
ABA	0,1.5,2,3 (µM)
IAA	0,0.5,1,2 (µM)
ACC	0,0.1,0.5,1 (µM)
methyl jasmonate	0,1,10,50 (µM)

The position of the root tips were marked immediately after transplant and the root length was measured after 5-7 days from the previous mark to the new position of the root tips. Measurements were done using the image processing tool, imageJ (Abràmoff et al., 2004).

7.9 DNA extraction and real time PCR

The DNA extraction was done using the CTAB method as described previously (section 7.3). The hygromycin resistance gene (*HygR*) in the construct pMDC32 was used as a target gene and the *bZIP1* gene was used as internal calibrator. The primers for both genes were designed using Primer-BLAST tool (Ye et al., 2012a). The primer specificity and efficiency were tested by constructing standard curve with five serial 5-fold dilutions. The template DNA used was a mixture of DNA (approx. 50 ng) from ten transgenic lines. The amplification reaction was carried out in a volume of 10 µL using SensiMix^{TM} SYBR No-ROX kit (Bioline, USA). The cycling conditions were as recommended - initial step of 95 °C for 10 min for polymerase activation, PCR (40 cycles), 95 °C, 15 sec for melting, 60 °C, 30 sec for annealing and 72 °C, 30 sec for extending, followed by melt curve analysis, 0.5 °C increment from 70 °C to 95 °C. The reactions were run on Bio-Rad CFX96 real time system. The standard curve analysis gives slope, intercept and correlation coefficients which are used for copy number calculation. The standard curve analysis was run two times independently to confirm the primer efficiency and specificity. Both HygR and bZIP1 primers had the primer efficiency of 92.4%.with correlation coefficient's of 0.99. PCR specificity was confirmed by melt curve analysis.

7.10 Transgene copy number calculation

To calculate the copy number of the transgene, qRT-PCR based method was followed as described previously (Chen and Lin, 2010). In brief, this method calculates copy numbers by obtaining the value of X_0/R_0 =

$10^{[(Ct,X - IX)/SX] - [(Ct, R - IR)/SR]}$ where I_X and I_R are intercepts of the standard curves of target and internal control genes, respectively, Sx and S_R are slopes of the standard curves of target and internal control genes, respectively and $Ct,_X$ and $Ct,_R$ are the detected threshold cycles of amplification of the target and internal control genes to a tested sample (transgenic line DNA) and the values are reported by the Bio-Rad CFX manager software. If the copy number of the internal control gene (R_0) is well defined (copy number one for bZIP1), the copy number of the target gene (X_0) can easily be deduced from the I_X , I_R , S_X , S_R , $Ct,_X$ and $Ct,_R$ in the tested sample. The slopes of both target gene (S_X) and internal control (S_R) was -3.519 and intercepts (I_X and I_R) 29.8 and 30.5 respectively. For each plant DNA sample, triplicate sets of PCR reactions with primers of the HygR and bZIP1 and one negative control (reaction samples without DNA templates), were prepared and run in a 96-well plate. Because the transgenic lines in the T1 generation are heterozygotes (or hemizygous) for the transgenes (*HygR*) and homozygotes for the endogenous gene (*bZIP1*), the value of X_0/R_0 has to be doubled to arrive at the copy number of the target gene (*HygR* in this case)

7.11 Yeast strains, transformation and stress screening

The *Saccharomyces cerevisiae* strains W303-1A, JM26 and JM162 were obtained from JM Mulet, IBMCP, Valencia, Spain. Both JM26 and JM162 strains were in the *Saccharomyces cerevisiae* strain W303-1A background. The strains JM162 was used for drought, JM26 for salt and W303-1A for oxidative stress and heat stress (Table 12). W303-1A was used as wild type control in salt and drought screens. The yeast transformations were performed as described previously (Gietz and Schiestl, 2007). The transformed colonies were selected on synthetic dropout media without uracil (SD-URA) and verified by colony PCR before continuing further. The genotypes of the strains used are given below. The media recipe for SOD-URA is as follows:

For 100 ml
Yeast nitrogen based – 0.67 g
Yeast synthetic drop-out media – 0.14g
Glucose – 2g
Histidine – 20 mg
Leucine – 100mg
Tryptophan – 20 mg
L-Adenine - 20 mg
Distilled water till 100 ml, pH -5.6,
Add agar, 2g for 100 ml solution.

Table 12: Yeast strains and their genotypes used in abiotic stress screening

Yeast strain	Stress conditions analysed	Genotype of strains
W303-1A	Heat and oxidative stress	*MATa* [*leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15*]
JM26	Salt and LiCl	*MATa* [*leu2-3,112 ura 3-1 trp1-1, ade 2-1 his3-11,15 can 1-100, ena 1-4::HIS3, nha1::TRP1*]
JM162	Drought (Sorbitol)	MATa [*his4 leu2 ura3-52 jnm1::LEU2 ade2 ade3, gpd1:trp1*]

YPD medium was used for growing yeast strains which contained 1% (w/v) yeast extract (Difco), 2% (w/v) peptone (Fluka) and 2% (w/v) glucose (Sigma). Solid media contained 2% bacteriological-grade agar. For stress screens, YPD media was supplemented with 1.7 M sorbitol (for drought), 5 mM H_2O_2 (for oxidative stress) or 300 mM NaCl (for salt stress). The growth of yeast cells under different stress conditions was assayed by spotting serial dilutions (1:10, 1:100 and 1:1000) of saturated cultures as described (Kanhonou et al., 2001). The stress screening plates were incubated for 2-3 days at 28 $^{\circ}$C before taking observations. For heat stress plates were incubated at 37 $^{\circ}$C. Pictures were taken 3 days after plating using the G:BOX gel imaging system (Syngene, UK).

7.12 Measurement of intracellular sodium content in Yeast

The ion content was measured as described previously (Rios et al., 2013). Cells were grown in YPD to an absorbance at 660 nm of 0.6–0.7, centrifuged for 5 min at 1900 *g*, resuspended at the same concentration in YPD containing 0.3 M NaCl, and incubated at 30 °C for 90 min. Aliquots were taken, centrifuged in plastic tubes for 5 min at 700 g and 4 °C, and washed twice with 10 mL of ice-cold solution of 20 mM $MgCl_2$. The cell pellets were resuspended in 0.5 mL of 20 mM $MgCl_2$. Ions were extracted by heating the cells for 15 min at 95 °C. After centrifugation, aliquots of the supernatant were analyzed with an atomic absorption spectrometer (SensAA) in flame emission mode (this experiment was done by collaborating lab in IBMCP, Valencia).

7.14 Iron deficiency screening

For the iron depletion assay, 7 day old seedlings grown on 1X MS agar media (supplemented with 0.5% sucrose) were transplanted to normal and low iron media and grown for another two weeks before scoring the phenotypes. Low iron media consisted of 1X MS basal salts (Sigma,M0654), 1X microsalts (recipe below), 1X vitamins (Sigma, M7150), EDTA and $FeSO_4$ to a final concentration of 0, 2 or 100 µM with pH 4.2 and 5.7 (Table 13 and 14). For screening, seeds from wild type and *35S*::*ASR63* lines were surface sterilised and stratified for 3 days at 4 $^{\circ}$C before plating them on 1X MS media. Plates were kept vertically for germination and growth of the seedlings. After 7 days of germination seedlings were transferred to different iron media with 0, 2 or 100 µM Fe with pH 4.2 and 5.7. Seedlings were further grown for another 14 days to score the phenotype.

Table 13: MS media recipe

Component (mg/L)	Macrosalts (10x)	Microsalts (100x)	EDTA (100x)	FeSO4 (100x) (Filter sterilization)
Ammonium nitrate	8250.0/500 ml			
Calcium chloride anhydrous	1661.0/500 ml			
Magnesium sulfate	903.0/500 ml			
Potassium nitrate	9500.0/500 ml			
Potassium phosphate monobasic	850.0/500 ml			
Boric acid		155 mg/250 ml		
Cobalt chloride • $6H_2O$		0.625 mg/250 ml		
Cupric sulfate • $5H_2O$		0.625 mg/250 ml		
Zinc sulfate • $7H_2O$		215 mg/250 ml		
Manganese sulfate • H_2O		422.5 mg/250 ml		
Molybdic acid (sodium salt) • $2H_2O$		6.25 mg/250 ml		
Potassium iodide		20.75 mg/250 ml		
Na_2-EDTA			373 mg/100 ml	
Ferrous sulfate • $7H_2O$				278 mg/100 ml
	Macrosalts (10x)	**Microsalts 100x**	**EDTA (100x)**	**FeSO4 (100x)**
Amount to prepare 1 L	100 ml	10 ml	10 ml	10 ml
pH at room temperature	4.3	n/a		

Table 14: MS Media for different Fe treatments

MS (Fe treatment)	100 µM Fe	5 µM Fe	2 µM Fe	0 µM Fe
Macrosalts (10x)	100 ml/L	100 ml/L	100 ml/L	100 ml/L
Microsalts (100x)	10ml/L	10ml/L	10ml/L	10ml/L
Vitamins (Sigma M7150) (100x)	10ml/L	10ml/L	10ml/L	10ml/L
EDTA (100x)	10ml/L	0.5 ml/L	0.2 ml/L	-----
FeSO4 (100x)	10ml/L	0.5 ml/L	0.2 ml/L	-----
Sucrose	5g/L	5g/L	5g/L	5g/L
pH	5.7	5.7	5.7	5.7
Plant agar	8g/L	8g/L	8g/L	8g/L

7.15 Measurement of iron content

Iron content was measured as described previously (Riemer et al., 2004). In brief, 100 mg of shoot tissue was collected from seedlings grown on iron sufficient or iron depleted media for 14 days and placed in eppendorf tubes and mixed with 100µL of 10 mM HCl (the solvent for the iron standard $FeCl_3$), 100 µL 50 mM NaOH, 100 µL of the iron-releasing reagent (a freshly mixed solution of equal volumes of 1.4M HCl and 4.5% (w/v) KMnO4 in H2O). These mixtures were incubated for 2 hrs at 60 °C within a fume hood, since chlorine gas is produced during the reaction. The mixture was cooled to room temperature and 30 µL of iron-detection reagent (6.5mM ferrozine, 6.5mM neocuproine, 2.5M ammonium acetate, and 1M ascorbic acid dissolved in water) was added to each tube. The mixture was incubated for 30 min. 280 µL of the solution was transferred into a 96-well plate and the absorbance was measured at 550 nm on a microplate reader (Sarstedt, Ireland). The iron content of the sample was calculated by comparing its absorbance to that of a range of standard concentrations of equal volume that had been prepared in a way similar to that of the sample (mixture of 100 µL of $FeCl_3$ standards (0–300 µM) in 10mM HCl, 100 µL 50mM NaOH, 100 µL releasing reagent, and 30 µL detection reagent). Total iron content from the seeds was measured using 60 seeds with or without seed coat after imbibing in water for 48 hours. Same assay was used for detection of iron content in yeast cells wherein yeast cultures with optical density (OD) 1 were used and the iron content was expressed as micromole per OD of iron.

7.16 Perl's staining for iron

For Perl's staining, embryos were excised from seeds previously imbibed in distilled water for 3 hrs using a binocular magnifying lens. The isolated embryos were vacuum infiltrated with equal volumes of 4% (v/v) HCl and 4% (w/v) K-ferrocyanide (Perl's stain solution) for 15 min and incubated for 30 min at room temperature (Stacey et al., 2008). Enhanced Perl's staining was done as described by (Meguro et al., 2007). After washing with distilled water, the embryos were incubated in a methanol solution containing 0.01 M NaN_3 and 0.3% (v/v) H_2O_2 for 1 h, and then washed with 0.1 M

phosphate buffer (pH 7.4). For the intensification reaction the embryos were incubated between 10 to 30 min in a 0.1 M phosphate buffer (pH 7.4) solution containing 0.025% (w/v) DAB (Sigma), 0.005% (v/v) H_2O_2, and 0.005% (w/v) $CoCl_2$ (intensification solution). The reaction was stopped by rinsing with distilled water.

7.17 In-gel protein assay

For in-gel analysis of aconitase activities, proteins were extracted by grinding 50 to 100 mg of leaf tissue with 1.5 volumes of 50 mM Tris-HCl, pH 8.0, 50 mM KCl, 0.2% (v/v) Triton X-100, 2 mM sodium citrate, and 1 mM dithiothreitol. Samples were centrifuged for 15 min at 4 °C at maximum speed in a microfuge. The supernatant was mixed with 0.25 volume of loading buffer (20 mM Tris-HCl 8.0, 80% [v/v] glycerol, and 0.1% [w/v] bromophenol blue), and 80 µg of protein was loaded per lane. The proteins were separated on a minigel with a stacking gel of 4% (w/v) polyacrylamide (PA) and a running gel of 7.5% (w/v) PA in 0.13 M Tris-borate buffer, pH 8.6, and 3.6 mM sodium citrate. Gels were run in 0.025 M Tris-Gly buffer, pH 8.3, and 3.6 mM sodium citrate at 120 V for 3.5 h. The activity of aconitases was visualized by incubating the gels in 0.1 M Tris-HCl, pH 8.0, 2.5 mM $MgCl_2$, 0.5 mM NADP, 2.5 mM cis-aconitase, 5 mg mL^{-1} thiazolyl blue tetrazolium bromide, 50 µg mL^{-1} phenazine methosulfate, and 0.13 units mL^{-1} porcine isocitrate dehydrogenase for 10 to 30 min at room temperature.(Bernard et al., 2009).

For in-gel SiR activity, *Arabidopsis* leaf samples (10 mg) weighed and crushed with a mortar and pestle in liquid nitrogen. Buffer containing 0.25 M sucrose, 50 mM Tris–HCl (pH 7.5), 1 mM EDTA, 0.025% Triton X-100 and a protein inhibitor mix (aprotinin, leupeptin and pepstatin, 10 µg/ml each) was added to each sample in a 1:20 ratio (w/v). The crushed plant material was centrifuged (Eppendorf 5417R) at 2,200 rpm for 10 min, mixed by vortexing, incubated on ice for 30 min and centrifuged at 14,000 rpm for 20 min. The resulting supernatant was heated at 55 °C for 90 sec. and subjected to centrifugation at 14,000 rpm for 5 min. The total soluble protein content in the resulting supernatant was determined according to the Bradford method, with

BSA as the standard (Bradford, 1976). 100 µg of protein was loaded per lane on the gel. PAGE was performed as described above. After the run, the gel was incubated with reaction solution which consisted of 0.05 M Tris–HCl, pH 7.5, 50 mM β-mercaptoethanol, 6 mM sodium dithionite (dissolved in 150 mM NaHCO3), 0.7 mM MVH, 0.05 mM NADPH and 0.4 mM lead acetate. When required, the reaction time was extended by adding fresh reaction solution. The reaction was stopped by immersion of the gel in double-distilled water and the bands were visualised.

7.18 Measurement of reactive oxygen species

The 7 days old seedlings (WT, *35S::ASR63#3.2, 35S::ASR63#5.1*) grown on 1X MS media were transplanted to Fe(+) or Fe(-) media and grown for further 4 days. The shoot tissue was collected and weighed before extracting the sample with phosphate buffer (pH 7.5). The samples were centrifuged at least two times to get rid of any debris by transferring to new tube each time.

Superoxide measurement was done using epinephrine. The reaction mixture consisted of 50 µL epinephrine (4 mM), 100 µL 100 mM Tris-HCl, pH 7.8,10 µL H_2O and 40 µL plant extract. A blank was used which consisted of 10 µL CuZn-SOD (2100 U/mL) instead of H_2O. Absorbance was measured at 480 nm after 5 and 10 minutes of incubation at room temperature and expressed as Units of superoxide/g/FW where unit is OD@ 480 nm.

Total H_2O_2 content was measured using Amplex Red reagent (10-acetyl-3,7-dihydroxyphenoxazine from Invitrogen). Amplex Red reagent reacts with H_2O_2 in a 1:1 stoichiometry to produce the red-fluorescent oxidation product, resorufin. Amplex red was dissolved in DMSO to make 10 mM stock solution. The amplex red staining solution consisted of 10 mM amplex red, 10 U/mL HRP (horse radish peroxidase) and 50 mM sodium phosphate buffer pH 7.4. 50 µL staining solution was added to 50 µL sample and 50 µL standard in a 96 well plate. Each sample was used in triplicates and incubated for 30 minutes at room temperature. After 30 min. absorbance

was measured at 560 nm. H_2O_2 standards were prepared using 1 mM H_2O_2 (0, 5, 10, 20, 50, 70, 100, 200 µM).

7.19 Measurement of Glutathione

Glutathione (reduced and oxidized forms) was assayed by an enzymatic recycling procedure (Griffith, 1980) in which it is sequentially oxidized by 5,5'-dithiobis (2 nitrobenzoic acid/DTNB) and reduced by NADPH in the presence of GR. The extent of 2-nitro-5-thiobenzoic acid formation is monitored at 412 nm. For specific assay of GSSG, GSH was masked by derivatization with 2-vinylpyridine. Leaves and roots (1g) were homogenized in 5 ml 5% (w/v) metaphosphoric acid and centrifuged for 10 min at 14,000*g*. Samples (100 µl) were mixed with 250 µl of 50 mM HEPES-KOH (pH 7.5) and 150 µl of 10% (w/v) sulphosalicylic acid. Samples were centrifuged for 10 min at 14,000*g* and the supernatant was divided into two, one half for total glutathione determination and the other for GSSG determination. Samples used for GSSG determination were treated with 2-vinilpyridine (final concentration 2%) and incubated for 1 h at 25 °C. For both, total glutathione and GSSG, 10-50 µl of the samples were mixed with: 100 mM potassium phosphate buffer, (pH 7.5), 0.2 mM NADPH, 1 mM DTNB, 2.5 mM EDTA. The reaction was initiated by the addition of 0.25 Units of GR and the increase in absorbance was followed for 4 min at 412 nm. GSH and GSSG contents were derived from GSH standard curve, since reduced and oxidized forms of glutathione have been shown to produce the same standard curves under these assay conditions (Griffith, 1980). The total GSH was expressed as nanomoles per gram fresh weight of the tissue.

7.20 Determination of total ascorbic acid

Ascorbate (ASC) and dehydroascorbate (DHA) were determined according to (Law et al., 1983). Plant tissue was homogenized in 5% meta-phosphoric acid in a ratio of 1:5 (w/v). The extract was centrifuged at 14000 g for 20 minutes. Supernatants were stored at –70 °C until used for analysis. Reaction mixture for ascorbate determination contained, 100 µL plant extract, 100 µL of 5 mM EDTA (in potassium phosphate buffer 150 mM, pH

7.4), and 200 µL water. For total ascorbate (ASC+DHA) determination reaction mixture contained, 100 µL plant extract, 100 µL of 5 mM EDTA (in 150 mM potassium phosphate buffer pH 7.4), 50 µL of 10 mM DTT. The mixture was incubated for 15 min at room temperature and excess DTT was oxidized by the addition of 50 µL of 0.5 % (w/v) N-ethylmaleimide. In both reaction mixtures colour was developed by adding 200 µL of 10 % TCA, 200 µL of 44 % (v/v) o-phosphoric acid, 200 µL of 4 % (w/v) α, α dipyridil in 70% EtOH, and 100 µL of 3 % (w/v) ferric chloride ($FeCl_3$). After mixing thoroughly, the samples were incubated at 37 $^{\circ}$C for 1 hr and absorbance at 525 nm was measured. Total ascorbate and ascorbate content were calculated on the basis of a standard curve in a range of 0-40 nmol ascorbate. DHA content was calculated as the difference between total ascorbate and ASC. The ascorbate content was expressed in nano moles per gram fresh weight of the tissue.

7.21 Iron bioavailability assay

Iron bioavailability assay was done as described previously (Garrett et al., 2000; Glahn et al., 1998) with slight modifications. In brief, 300 mg of leaf tissue was homogenised with 1ml of saline/BHT solution (100 mM saline/150µM BHT). The homogenized meal was acidified (pH 2 with HCl) before addition of porcine pepsin (prepared with) to a final concentration of 1.8 mg/mL and incubated at 37 °C in a water bath for 1 hr. After incubation time the pH of the gastric digest was increased to pH 6 by drop wise addition of 1M $NAHCO_3$. About 250 µl of pancreatine-bile extract mixture per 1 ml of original sample volume (5x dilution till 2.4 mg porcine bile extract per 1 mL digest and 0.4mg pancreatin per mL digest) was added and the pH was adjusted to 7 with 1M NaOH. To the mixture 250 µl of 120mM NaCl/5 mM KCl solution was added and incubated in a water bath at 37 °C for 2 hrs. The mixture was then centrifuged at room temperature for 10 min. at maximum speed, and aqueous fraction was collected and centrifuged again to collect the clear supernatant which was used for treating the cells.

Caco-2 cells used in the experiment were a kind gift from Dr. Aoife Boyd, department of microbiology, NUI Galway. Cell culture was done as described previously (Glahn et al., 1998). For the bioavailability assay 1 mL of cells were seeded at a density of 20,000 cells/mL per well in 24-well plates (24-well cell culture cluster dishes, Sarstedt). The cells were grown in Dulbecco's modified Eagle's medium (GIBCO) with 10% v/v fetal calf serum (GIBCO), 25 mmol/L HEPES and antibiotic solution (100X penicillin (5000 units)-streptomycin (5 mg/L) solution, Sigma). The cells were maintained at 37 °C in an incubator with a 5% CO_2 - 95% atmospheric air at constant humidity; the medium was changed every 2 days. The cells were used in the iron uptake experiments at 14-16 days post seeding.

Before treating the cells with the digested sample, cells in the 24-well plate were washed twice with minimum essential medium (MEM). Cells were treated with 1 mL of test medium containing 0.75 mL MEM (OPTI-MEM, GIBCO) plus 0.25 mL iron fraction or saline (as control) in triplicate wells for each sample. 1 mL of test sample was used as time zero sample (initial Fe concentration). Cells were then incubated at 37 °C for 4 hrs. 1 mL of test sample was collected from the wells as treated sample (for Fe concentration after treatment). The Fe concentration was measured in both time zero and 4 hours treated samples and the difference in iron concentration indicate the amount of iron taken up by cells and is expressed as percentage bioavailable iron per gram fresh weight of tissue analysed. The iron content was measured by ferrozine assay as described in section 7.15.

7.22 Yeast strains and transformations (chapter 5)

Yeast strains YKB779 (*MATa* [*ade2-101 his3-Δ200 lys2-801 leu2-801 lue2-Δ1 ura3-52 trp1-Δ63*]) and *aft1* (*MATa* [*aft1Δ::natMX4, can1Δ::STE2pr-SP-his5, lyp1Δ his3D1 leu2Δ0 ura3Δ0 met15Δ0 LYS2+*]) are a kind gift from Dr. Kristin Baetz, Univ. of Ottawa, Canada. The yeast transformations were done as described in section 4.5.1. Yeast was grown in YPAD medium (1% (w/v) Bacto yeast extract, 2% (w/v) Bacto peptone, adenine hemisulfate 80 mg/l) (for plates, 18 g/l Bacto Agar added). The yeast cells to be transformed

are usually regrown for two generations in liquid YPAD medium (2% (w/v) Bacto yeast extract, 4% (w/v) Bacto peptone, 4% (w/v) glucose and adenine hemisulfate 80 mg/l). The transformed colonies were selected on synthetic complete drop out (SOD) media without uracil. List of all the yeast strains used I this study are given in Table- 15.

7.23 Kinetic assay for aconitase activity in Yeast

Aconitase activity was measured by the NADP-coupled aconitase assay according to Drapier and Hibbs (1996). Yeast cells grown under iron deficient or iron sufficient conditions were collected from 20 mL of cultures with OD 10 and total protein was extracted using 10 ml extraction buffer (0.25 M sucrose in 50 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES), pH 7.4). 50 µL of protein (100 µg/mL) was added to the reaction mixture (1 mM NADP, 1 mM sodium citrate, 1 U of aconitase-free isocitrate dehydrogenase in 150 µL). The reaction kinetic was read at 340 nm at 37 $^{\circ}$C for 40 min. at 10 min. intervals. A sample without citrate and isocitrate dehydrogenase was used as blank. The activity was expressed as nmol/mg of protein per minute.

7.24 Germination analysis and hormone treatments (chapter 6)

For measuring germination rates, seeds of wild type and *35S::ASR63* lines were germinated on 1X MS medium, stratified at 4 $^{\circ}$C for 4 days and transferred to long day regime. The germination was scored in terms of radical emergence. Cotyledon emergence was scored after 3 days. For hormone treatments, 1X MS media was supplemented with 0, 0.5 or 1.5 µM ABA or 0, 4 or 10 µM PAC. Germination was scored in terms of cotyledon emergence on a daily basis and expressed as a percentage.

Table 15: List of Yeast mutants used in the study

Sl.No	Name	Mating type	Mutations	auxotrophies	Background	Obtained from
#1	*fet3/fet4*	a/l diploid	fet3-2::HIS3/fet3-2::HIS3 fet4-1::LEU2/fet4-1::LEU2		W303	David Eide, University of Wisconsin-Madison
#2	*aft2*	a	aft2Δ::nat	can1Δ::STE2pr-SP-his5 lyp1Δ his3D1 leu2Δ0 ura3Δ0 met15Δ0 LYS2 +	YKB779	Kristin Baetz
#3	*fet3*	a	fet3Δ::nat	can1Δ::STE2pr-SP-his5 lyp1Δ his3D1 leu2Δ0 ura3Δ0 met15Δ0 LYS2 +	YKB779	Kristin Baetz
#4	*rim101*	a	rim101Δ::nat	can1Δ::STE2pr-SP-his5 lyp1Δ his3D1 leu2Δ0 ura3Δ0 met15Δ0 LYS2 +	YKB779	Kristin Baetz
#5	*aft1*	a	aft1Δ::natMX4	can1Δ::STE2pr-SP-his5 lyp1Δ his3D1 leu2Δ0 ura3Δ0 met15Δ0 LYS2 +	YKB779	Kristin Baetz
#6	wt-YKB779	a	wild-type	ade2-101 his3-Δ200 lys2-801 leu2-801 lue2-Δ1 ura3-52 trp1-Δ63	YKB779	Kristin Baetz
#7	*Δaft1Δrim101*	a	rim101Δ::nat// aft1Δ::natMX4	ade2-101 his3-Δ200 lys2-801 leu2-801 lue2-Δ1 ura3-52 trp1-Δ63	YKB779	Kristin Baetz
#8	wt-EUROSCARF	a	wild-type	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF	Prof. Hana Sychrová,
#9	*tok1/ ena1-5*	a	tok1Δ ena1-5Δ	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF	Prof. Hana Sychrová,
#10	*nha1/ ena1-5*	a	nha1Δ ena1-5Δ	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF	Prof. Hana Sychrová,
#11	*tok1/ nha1/ ena1-*	a	tok1Δ nha1Δ ena1-5Δ	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF	Prof. Hana Sychrová,
#12	*trk1/trk2*	a	trk1Δ trk2Δ	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF	Prof. Hana Sychrová,
#13	*trk1/ trk2/ tok1/*	a	trk1Δ trk2Δ tok1Δ	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF	Prof. Hana Sychrová,
#14	*trk1/ trk2/nha1*	a	trk1Δ trk2Δ nha1Δ	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF	Prof. Hana Sychrová,
#15	*trk1/ trk2/ ena1-5*	a	trk1Δ trk2Δ ena1-5Δ	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF	Prof. Hana Sychrová,
#16	*trk1/ trk2/ tok1/ n*	a	trk1Δ trk2Δ tok1Δ nha1Δ ena1-5Δ	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF	Prof. Hana Sychrová,

Sl.No	Name	Mating type	Mutations	auxotrophies	Background	Obtained from
#17	wt-BY4741	a	wild-type	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	BY4741	Prof. Amparo Pascual-Ahuir // Markus Proft, Valencia, Spain
#18	*sco1*	a	Δsco1	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	BY4741	Prof. Amparo Pascual-Ahuir // Markus Proft, Valencia, Spain
#19	*rcs1*	a	Δrcs1	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	BY4741	Prof. Amparo Pascual-Ahuir // Markus Proft, Valencia, Spain
#20	*cox19*	a	Δcox19	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	BY4741	Prof. Amparo Pascual-Ahuir // Markus Proft, Valencia, Spain
#21	*isa1*	a	Δisa1	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	BY4741	Prof. Amparo Pascual-Ahuir // Markus Proft, Valencia, Spain
#22	*ftr1*	a	Δftr1	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	BY4741	Prof. Amparo Pascual-Ahuir // Markus Proft, Valencia, Spain
#23	*fet3*	a	Δfet3	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	BY4741	Prof. Amparo Pascual-Ahuir // Markus Proft, Valencia, Spain
#24	*isa2*	a	Δisa2	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	BY4741	Prof. Amparo Pascual-Ahuir // Markus Proft, Valencia, Spain
#25	*mrs4*	a	Δmrs4	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	BY4741	Prof. Amparo Pascual-Ahuir // Markus Proft, Valencia, Spain
#26	WT-DY150	a	wild-type	ura3-52 leu2-3 112 trp1-1 his3-11 ade2-1 can1-100 (oc)	DY150	Prof. Kaplan, US
#27	*ccc1*	a	Δccc1	ura3 leu2 trp1 his3 ade2 can1 Δccc1::HIS3	DY150	Prof. Kaplan, US
#28	*mrs3/mrs4*	a	Δmrs3::kanMax, Δmrs4::kanMax	ura3-52 leu2-3 112 trp1-1 his3-11 ade2-1 can1-100 (oc)	DY150	Prof. Kaplan, US
#29	*mrs3/mrs4/rim2*	a	Δmrs3::kanMax, Δmrs4::kanMax, Δrim2::HIS3	ura3 leu2 trp1 his3 ade2 can1 Δrim2::HIS3	DY150	Prof. Kaplan, US
#30	*fet3/fet4*	a	Δfet3::kHIS, Δfet4::LEU2	ade2 can1 trp1 ura3 probably correct	DY150	Prof. Kaplan, US

7.25 Detection of reactive oxygen species in germinating seeds

For the detection of hydrogen peroxide, 48 hrs imbibed seeds were excised to remove seed coat and the embryos were incubated in 1 mg/mL solution of DAB (3′ 3′ diaminobenzidine, Sigma) for 1 hour. Embryos were rinsed with deionized water three times. The presence of hydrogen peroxide was visualised by dark brown precipitate. Superoxide anion accumulation was detected using nitroblue tetrazolium (NBT) staining as described previously (Leymarie et al., 2012). In short, embryos were dissected from 48 hrs. imbibed seeds and incubated in 6 mM NBT (Sigma) in 10 mM Tris–HCl buffer (pH 7.4) at room temperature for 1 hr. Embryos were rinsed three times with deionized water, and superoxide anion was visualized as precipitates of dark blue insoluble formazan compounds.

8. General conclusions and future directions

Every organism contains unique set of genes called orphan genes which are absent in other species/lineage. The widespread occurrence of orphan genes has become more evident as genomes of many species have been sequenced in the recent years. Some orphan genes may acquire functions that allow improved adaptations to changes in the environment. When this occurs, some orphan genes are expected to convey evolutionary innovations in the organism. We defined orphan genes in Brassicaceae lineage in our previous work along with evolutionary mechanisms of their origin (Donoghue et al., 2011). We also identified that orphan genes are enriched for various abiotic stress responses which became the basis for the study taken up in this thesis. With the aim of ascribing functions to the abiotic stress responsive orphan genes in *A. thaliana*, three different approaches were followed to screen for orphan genes under abiotic stress conditions namely, drought, salt, temperature stress (heat and cold) and nutrient deficiency. The approaches followed were; loss-of-function screening, gain-of-function screening and gain-of-function screening through heterologous expression in yeast.

The loss of function screen was performed using T-DNA insertion mutant lines for individual orphan genes to identify any phenotypes under abiotic stress when the gene function is abolished. This screen identified one promising candidate for salt stress where the loss of the gene conferred tolerance to salt stress (Chapter 2). The orphan gene, *ASR50*, identified in the loss of function screen is proposed to be involved in salt tolerance and will be subjected to further investigation in the lab. The complementation of the mutant and expression of *ASR50* under a constitutive promoter will help to decipher its function under salinity stress. In addition another gene *ASR70* displayed ABA sensitive phenotype during seed germination and it was found to be Brassicales specific gene. Further investigation of *ASR70* by complementation and overexpression under a constitutive promoter is underway. ABA sensitivity is closely related to drought tolerance (Fujita et al.,

2005) and hence *asr70* mutant lines will be analysed for drought stress under transpiring conditions. However it is also necessary to investigate whether *ASR70* is involved in seed germination as ABA is a major hormone regulating seed germination.

The gain of function screen described in chapter 3 involved overexpression of the orphan genes under CaMV 35S constitutive promoter. The overexpression screen identified two promising orphan candidates, namely *ASR35* and *ASR63*, for salt and iron deficiency tolerance respectively. In addition, two genes namely *ASR12* conferring salt tolerance and *ASR51* causing morphological defect were also identified in this study. *ASR63* was further analysed in depth to investigate its role in iron deficiency tolerance in chapter 5 and 6. However *ASR35*, *ASR12* and *ASR51* are also promising candidates for functional analysis and will be investigated further in the lab. *ASR35* is an orphan gene and is found to be associated with plant cell wall. Further it was shown that *ASR35* displays ecotype-specific expression patterns (Aufsatz et al., 1998). Recently it has been shown for one *Arabidopsis de novo* gene that it is under tight epigenetic control and display different expression patterns across *Arabidopsis* ecotypes (Silveira et al., 2013). It has been inferred that such epigenetic regulation will allow the new gene to adjust optimum expression pattern which possibly play a role in adapatation. In this perspective, *ASR35* is an interesting candidate to investigate further to elucidate its functionality. *ASR51* is another interesting candidate identified in this study which causes severe morphological defect upon ectopic expression. *ASR51* could be directly involved in morphogenesis in *Arabidopsis*. Overall, both loss-of-function and gain-of-function screens are demonstrated as successful for identifying functionality of orphan genes and both approaches used together can help to better understand the functionality of orphan genes in *A. thaliana*.

The third approach followed in this study, heterologous expression in yeast system, provides an important tool for functional characterization of plant genes. Surprisingly, Brassicaceae specific orphan genes conferred different degrees of stress related phenotypes (tolerance or sensitivity) in

yeast mutants. Since the plant lineage split from yeast approximately 1600 million years ago, plant orphan genes showing functional phenotypes in yeast was remarkable. However, considering that yeast and plants share conserved pathways for many of the cellular responses, some of the orphan gene functions evolved in plant lineages can serve as lineage-trapped innovations and will be able to function in yeast. A recent study has shown that an *A. thaliana* specific orphan gene can function in another plant species, soybean, which diverged from *Arabidopsis* approximately 100 million years ago. (Li and Wurtele, 2014).

Although gain of function sometimes may lead to neomorphic phenotypes (Prelich, 2012), the results from characterization of *ASR63* (in chapter 5) indicate that its phenotype may not be due to a neomorphic effect because *ASR63* function seems to be linked with iron homeostasis in both plant and yeast. One obvious possibility for a protein to function in another species or lineage is that it has a conserved structure or motif to function in a pathway. Some evidences from metagenomic studies do suggest that proteins with no sequence similarity to any known proteins show structural similarity to conserved protein domains/motifs (Das et al., 2009; Godzik, 2011). Another possibility is that the protein is intrinsically disordered. Intrinsically disordered proteins, proteins with no proper structure, are known to be involved in various functions including cell signalling and regulation (Oldfield and Dunker, 2014). Hence some of the orphan genes might be coding for intrinsically disordered proteins that can carry out biological function without gaining a proper conformation. Although there is not enough evidence yet to support this hypothesis, the possibility of intrinsic disorderness cannot be ruled out. However, detailed characterisation of more orphan genes showing functional phenotypes in yeast or other species will shed some more light on this aspect.

A major finding in this work is the identification of the orphan gene *ASR63* which confers tolerance to iron deficiency in *A. thaliana* when it is ectopically expressed. The results in yeast also suggest that *ASR63* is most likely acting in iron homeostasis pathway. However, more experimentation is

needed to determine the precise mechanism of action of *ASR63* in iron deficiency responses. Transcriptome analysis of *35S::ASR63* lines under iron deficiency will help to identify the potential pathway. Analysis of *ASR63* knockout or knock down lines (which is underway) will reveal its role in iron homeostasis in Arabidopsis. Further, analysis of the effect of *ASR63* overexpression in ascorbic acid or GSH deficient mutants will help to identify any link with biosynthesis of antioxidants. The loss of function screen for *ASR63* in *A. thaliana* combined with yeast genetic analyses (using iron and copper homeostasis mutants), will further help in understanding the *ASR63* function. The use of yeast iron and copper homeostasis mutants will help to identify possible direct link with the pathway and also genetic ordering (epistasis experiment) of *ASR63* in the respective pathway.

In addition, the evolutionary origin of orphan genes is another area of active investigation in the field of evolutionary genetics. In this context we propose that the divergent transcription by the bidirectional promoter led to the origin of *ASR63* as a *de novo* gene in Brassicaceae. The comparison of *ASR63* locus with other Brassicaceae species has confirmed the presence of the gene in those species. However it is not known whether *ASR63* is functional in all those species and whether the promoter indeed is bidirectional in them. Hence a detailed comparison of the *ASR63* locus and the functional analysis of promoter region with its orthologs in sister taxa (which is underway) will help us to understand how *ASR63* gained a proper gene structure (specifically intron and splice junctions) during diversification of Brassicaceae. The bidirectionality of the *ASR63* promoter in Arabidopsis and in the sister species will be confirmed by making promoter:reporter constructs and will be analysed by transient expression assay (protoplasts/tobacco leaf infiltration). In addition splice site variations (if any) between *ASR63* and its Brassicaceae orthologs will be analysed computationally. This will further help to define the *ASR63* evolutionary origin. Finally, *ASR63* overexpression lines also show increased sensitivity to phytohormone ABA and a GA inhibitor PAC. This is possibly due to the increased production of antioxidants and superoxide in *35S::ASR63* lines.

Moreover the loss of function screen for *ASR63* should help to identify whether *ASR63* has any direct effect on germination in *A. thaliana*.

Overall, this work has analysed a set of Brassicaceae specific orphan genes of *A. thaliana* by considering abiotic stress responsiveness as the basis for evolutionary novelty. This PhD thesis has identified a *de novo* gene *ASR63*, which has a potential to be used for enhancing iron content in crop plants. *De novo* origin of genes, once considered a rare mechanism of origin, is now considered as an important route to genome novelty. Most importantly, this study also showed that the orphan gene functionality can be transferred across species suggesting the possibility of exploiting lineage trapped innovations. Although some pieces of evidence suggest that orphan genes can form conserved motifs at the protein level, more studies on orphan protein structures will be needed to find how orphan genes can function in distant species. Moreover, this study highlights the importance of orphan genes and gives an evidence of a *de novo* gene which confers a novel trait to the organism. We consider that this study will set a stage for comprehensive study of orphan genes in plant species in general, which could help in identifying genes with novel traits. Hence, from the applied research point of view, identifying orphan genes conferring traits like stress tolerance would be beneficial and such genes can be used in improving crop plants through biotechnological approaches.

References

Abràmoff, M.D., Magalhães, P.J., and Ram, S.J. (2004). Image processing with ImageJ. Biophotonics International *11*, 36-43.

Abrusan, G. (2013). Integration of new genes into cellular networks, and their structural maturation. Genetics *195*, 1407-1417.

Adachi, N., and Lieber, M.R. (2002). Bidirectional gene organization: a common architectural feature of the human genome. Cell *109*, 807-809.

Aharoni, A., Dixit, S., Jetter, R., Thoenes, E., van Arkel, G., and Pereira, A. (2004). The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis*. The Plant cell *16*, 2463-2480.

Alba, M.M., and Castresana, J. (2005). Inverse relationship between evolutionary rate and age of mammalian genes (vol 22, pg 598, 2005). Molecular Biology and Evolution *22*, 1159-1159.

Alderson, A., Sabelli, P.A., Dickinson, J.R., Cole, D., Richardson, M., Kreis, M., Shewry, P.R., and Halford, N.G. (1991). Complementation of snf1, a mutation affecting global regulation of carbon metabolism in yeast, by a plant protein kinase cDNA. Proc Natl Acad Sci USA *88*, 8602-8605.

Alonso, J.M., and Ecker, J.R. (2006). Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. Nat Rev Genet *7*, 524-536.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., and Cheuk, R. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science *301*, 653-657.

Alvarez, J.P., Pekker, I., Goldshmidt, A., Blum, E., Amsellem, Z., and Eshed, Y. (2006). Endogenous and synthetic microRNAs stimulate simultaneous, efficient, and localized regulation of multiple targets in diverse species. The Plant cell *18*, 1134-1151.

Arendsee, Z.W., Li, L., and Wurtele, E.S. (2014). Coming of age: orphan genes in plants. Trends in plant science *19*, 698-708.

Aufsatz, W., Amry, D., and Grimm, C. (1998). The ECS1 gene of *Arabidopsis* encodes a plant cell wall-associated protein and is potentially linked to a locus influencing resistance to *Xanthomonas campestris*. Plant Mol Biol *38*, 965-976.

Aufsatz, W., and Grimm, C. (1994). A new, pathogen-inducible gene of Arabidopsis is expressed in an ecotype-specific manner. Plant Mol Biol *25*, 229-239.

Bai, Y., Casola, C., and Betran, E. (2008). Evolutionary origin of regulatory regions of retrogenes in Drosophila. BMC Genomics *9*, 241.

Bailly, C. (2004). Active oxygen species and antioxidants in seed biology. Seed Science Research *14*, 93-107.

Banno, H., Ikeda, Y., Niu, Q.W., and Chua, N.H. (2001). Overexpression of *Arabidopsis* ESR1 induces initiation of shoot regeneration. The Plant cell *13*, 2609-2618.

Bardwell, L. (2005). A walk-through of the yeast mating pheromone response pathway. Peptides *26*, 339-350.

Bartel, P.L., and Fields, S. (1997). The yeast two-hybrid system (Oxford University Press).

Beaudoin, N., Serizet, C., Gosti, F., and Giraudat, J. (2000). Interactions between Abscisic Acid and Ethylene Signaling Cascades. The Plant cell *12*, 1103-1115.

Beggs, J.D. (1978). Transformation of yeast by a replicating hybrid plasmid. Nature *275*, 104-109.

Begun, D.J., Lindfors, H.A., Kern, A.D., and Jones, C.D. (2007). Evidence for de novo evolution of testis-expressed genes in the *Drosophila* yakuba/Drosophila erecta clade. Genetics *176*, 1131-1137.

Bernard, D.G., Cheng, Y., Zhao, Y., and Balk, J. (2009). An allelic mutant series of *ATM3* reveals its key role in the biogenesis of cytosolic iron-sulfur proteins in *Arabidopsis*. Plant Physiol *151*, 590-602.

Berthelet, S., Usher, J., Shulist, K., Hamza, A., Maltez, N., Johnston, A., Fong, Y., Harris, L.J., and Baetz, K. (2010). Functional genomics analysis of the *Saccharomyces cerevisiae* iron responsive transcription factor Aft1 reveals iron-independent functions. Genetics *185*, 1111-1128.

Blanc, G., Hokamp, K., and Wolfe, K.H. (2003). A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome research *13*, 137-144.

Bleackley, M.R., Young, B.P., Loewen, C.J., and MacGillivray, R.T. (2011). High density array screening to identify the genetic requirements for

transition metal tolerance in Saccharomyces cerevisiae. Metallomics : integrated biometal science *3*, 195-205.

Bolle, C., Schneider, A., and Leister, D. (2011). Perspectives on Systematic Analyses of Gene Function in *Arabidopsis thaliana*: New Tools, Topics and Trends. Curr Genomics *12*, 1-14.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. The Plant cell *12*, 2383-2394.

Borsani, O., Cuartero, J., Valpuesta, V., and Botella, M.A. (2002). Tomato tos1 mutation identifies a gene essential for osmotic tolerance and abscisic acid sensitivity. Plant J *32*, 905-914.

Bosch, T.C., Augustin, R., Anton-Erxleben, F., Fraune, S., Hemmrich, G., Zill, H., Rosenstiel, P., Jacobs, G., Schreiber, S., Leippe, M.*, et al.* (2009). Uncovering the evolutionary history of innate immunity: the simple metazoan Hydra uses epithelial cells for host defence. Dev Comp Immunol *33*, 559-569.

Botstein, D., and Fink, G.R. (2011). Yeast: an experimental organism for 21st Century biology. Genetics *189*, 695-704.

Bouche, N., and Bouchez, D. (2001). *Arabidopsis* gene knockout: phenotypes wanted. Curr Opin Plant Biol *4*, 111-117.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry *72*, 248-254.

Cadman, C.S., Toorop, P.E., Hilhorst, H.W., and Finch-Savage, W.E. (2006). Gene expression profiles of Arabidopsis Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. Plant J *46*, 805-822.

Cai, J., Zhao, R., Jiang, H., and Wang, W. (2008). *De novo* origination of a new protein-coding gene in *Saccharomyces cerevisiae*. Genetics *179*, 487-496.

Cai, J.J., Woo, P.C.Y., Lau, S.K.P., Smith, D.K., and Yuen, K.Y. (2006). Accelerated evolutionary rate may be responsible for the emergence of lineage-specific genes in Ascomycota. Journal of Molecular Evolution *63*, 1-11.

Carvunis, A.R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charloteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B.*, et al.* (2012). Proto-genes and de novo gene birth. Nature *487*, 370-374.

Cassin, G., Mari, S., Curie, C., Briat, J.F., and Czernic, P. (2009). Increased sensitivity to iron deficiency in *Arabidopsis thaliana* overaccumulating nicotianamine. J Exp Bot *60*, 1249-1259.

Chaerle, L., Leinonen, I., Jones, H.G., and Van Der Straeten, D. (2007). Monitoring and screening plant populations with combined thermal and chlorophyll fluorescence imaging. J Exp Bot *58*, 773-784.

Chen, G.Q., and Lin, J.T. (2010). Use of Quantitative Polymerase Chain Reaction for Determining Copy Numbers of Transgenes in *Lesquerella fendleri*. American Journal of Agricultural and Biological Sciences *5*, 415.

Chen, S., Ni, X., Krinsky, B.H., Zhang, Y.E., Vibranovski, M.D., White, K.P., and Long, M. (2012). Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene. EMBO J *31*, 2798-2809.

Chen, S., Zhang, Y.E., and Long, M. (2010). New genes in Drosophila quickly become essential. Science *330*, 1682-1685.

Cheng, Y., Dai, X., and Zhao, Y. (2006). Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. Genes Dev *20*, 1790-1799.

Chuang, C.F., and Meyerowitz, E.M. (2000). Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. Proc Natl Acad Sci USA *97*, 4985-4990.

Clark-Adams, C.D., Norris, D., Osley, M.A., Fassler, J.S., and Winston, F. (1988). Changes in histone gene dosage alter transcription in yeast. Genes Dev *2*, 150-159.

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

Clouse, S.D., Langford, M., and McMorris, T.C. (1996). A brassinosteroid-insensitive mutant in *Arabidopsis thaliana* exhibits multiple defects in growth and development. Plant Physiol *111*, 671-678.

Colangelo, E.P., and Guerinot, M.L. (2004). The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. The Plant cell *16*, 3400-3412.

Conant, G.C., and Wolfe, K.H. (2008). Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet *9*, 938-950.

Cooley, L., Kelley, R., and Spradling, A. (1988). Insertional Mutagenesis of the *Drosophila*.

Cordaux, R., Udit, S., Batzer, M.A., and Feschotte, C. (2006). Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proc Natl Acad Sci USA *103*, 8101-8106.

Curie, C., Panaviene, Z., Loulergue, C., Dellaporta, S.L., Briat, J.F., and Walker, E.L. (2001). Maize yellow stripe1 encodes a membrane protein directly involved in Fe(III) uptake. Nature *409*, 346-349.

Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol *133*, 462-469.

Darbani, B., Briat, J.F., Holm, P.B., Husted, S., Noeparvar, S., and Borg, S. (2013). Dissecting plant iron homeostasis under short and long-term iron fluctuations. Biotechnol Adv *31*, 1292-1307.

Das, D., Kozbial, P., Axelrod, H.L., Miller, M.D., McMullan, D., Krishna, S.S., Abdubek, P., Acosta, C., Astakhova, T., Burra, P.*, et al.* (2009). Crystal structure of a novel Sm-like protein of putative cyanophage origin at 2.60 A resolution. Proteins *75*, 296-307.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.

Decreux, A., and Messiaen, J. (2005). Wall-associated kinase WAK1 interacts with cell wall pectins in a calcium-induced conformation. Plant Cell Physiol *46*, 268-278.

Demuth, J.P., De Bie, T., Stajich, J.E., Cristianini, N., and Hahn, M.W. (2006). The evolution of mammalian gene families. PLoS One *1*, e85.

Dhadi, S.R., Krom, N., and Ramakrishna, W. (2009). Genome-wide comparative analysis of putative bidirectional promoters from rice, Arabidopsis and Populus. Gene *429*, 65-73.

Ding, Y., Zhao, L., Yang, S., Jiang, Y., Chen, Y., Zhao, R., Zhang, Y., Zhang, G., Dong, Y., Yu, H.*, et al.* (2010). A young Drosophila duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes. PLoS Genet *6*, e1001255.

Ding, Y., Zhou, Q., and Wang, W. (2012). Origins of new genes and evolution of their novel functions. Annual Review of Ecology, Evolution, and Systematics *43*, 345-363.

Domazet-Loso, T., and Tautz, D. (2003). An evolutionary analysis of orphan genes in Drosophila. Genome research *13*, 2213-2219.

Donoghue, M.T., Keshavaiah, C., Swamidatta, S.H., and Spillane, C. (2011). Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana*. BMC Evol Biol *11*, 47.

Doyle, J.J., and Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin *19*, 11-15.

Drapier, J.C., and Hibbs, J.B., Jr. (1996). Aconitases: a class of metalloproteins highly sensitive to nitric oxide synthesis. Methods Enzymol *269*, 26-36.

Dujon, B. (1996). The yeast genome project: what did we learn? Trends in Genetics *12*, 263-270.

Durrett, T.P., Gassmann, W., and Rogers, E.E. (2007). The FRD3-mediated efflux of citrate into the root vasculature is necessary for efficient iron translocation. Plant Physiol *144*, 197-205.

Eide, D., Broderius, M., Fett, J., and Guerinot, M.L. (1996). A novel iron-regulated metal transporter from plants identified by functional expression in yeast. Proc Natl Acad Sci USA *93*, 5624-5628.

El-Maarouf-Bouteau, H., and Bailly, C. (2008). Oxidative signaling in seed germination and dormancy. Plant Signal Behav *3*, 175-182.

Emerson, J.J., Kaessmann, H., Betrán, E., and Long, M. (2004). Extensive Gene Traffic on the Mammalian X Chromosome. Science *303*, 537-540.

Eswaran, N., Parameswaran, S., Sathram, B., Anantharaman, B., Kumar, G.R., and Tangirala, S.J. (2010). Yeast functional screen to identify genetic determinants capable of conferring abiotic stress tolerance in *Jatropha curcas*. BMC Biotechnol *10*, 23.

Eun-Deok, K., Z. Jeffrey, C., and Zhongfu, N. (2009). Chlorophyll and starch assays.

Feldmann, K.A., Malmberg, R.L., and Dean, C. (1994). Mutagenesis in *Arabidopsis*. Cold Spring Harbor Monograph Archive *27*, 137-172.

Feldmann, K.A., and Marks, M.D. (1987). Agrobacterium-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach. Molecular and General Genetics MGG *208*, 1-9.

Felippes, F.F., Schneeberger, K., Dezulian, T., Huson, D.H., and Weigel, D. (2008). Evolution of Arabidopsis thaliana microRNAs from random sequences. RNA *14*, 2455-2459.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y.*, et al.* (2013). Efficient genome editing in plants using a CRISPR/Cas system. Cell Res *23*, 1229-1232.

Finch-Savage, W.E., and Leubner-Metzger, G. (2006). Seed dormancy and the control of germination. New Phytol *171*, 501-523.

Finkelstein, R., Reeves, W., Ariizumi, T., and Steber, C. (2008). Molecular aspects of seed dormancy. Annu Rev Plant Biol *59*, 387-415.

Finkelstein, R.R., Gampala, S.S., and Rock, C.D. (2002). Abscisic acid signaling in seeds and seedlings. The Plant cell *14 Suppl*, S15-45.

Finkelstein, R.R., and Zeevaart, J.A.D. (1994). Gibberellin and Abscisic Acid Biosynthesis and Response. Cold Spring Harbor Monograph Archive *27*, 523-553.

Fischer, D., and Eisenberg, D. (1999). Finding families for genomic ORFans. Bioinformatics *15*, 759-762.

Forieri, I., Wirtz, M., and Hell, R. (2013). Toward new perspectives on the interaction of iron and sulfur metabolism in plants. Front Plant Sci *4*, 357.

Forsburg, S.L. (2001). The art and design of genetic screens: yeast. Nat Rev Genet *2*, 659-668.

Foyer, C.H., and Shigeoka, S. (2011). Understanding oxidative stress and antioxidant functions to enhance photosynthesis. Plant Physiol *155*, 93-100.

Fujita, Y., Fujita, M., Satoh, R., Maruyama, K., Parvez, M.M., Seki, M., Hiratsu, K., Ohme-Takagi, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2005). AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in *Arabidopsis*. The Plant cell *17*, 3470-3488.

Gaber, R., Dreyer, I., Horeau, C., Lemaillet, G., Zimmermann, S., Bush, D., Rodriguez-Navarro, A., Schachtman, D., Spalding, E., and Sentenac, H.

(1999). Identification and characterization of plant transporters using heterologous expression systems. Journal of experimental botany *1073*, 15.

Gao, H., Zheng, X.M., Fei, G., Chen, J., Jin, M., Ren, Y., Wu, W., Zhou, K., Sheng, P., Zhou, F.*, et al.* (2013). *Ehd4* encodes a novel and *Oryza*-genus-specific regulator of photoperiodic flowering in rice. PLoS Genet *9*, e1003281.

Garrett, D.A., Failla, M.L., and Sarama, R.J. (2000). Estimation of carotenoid bioavailability from fresh stir-fried vegetables using an in vitro digestion/Caco-2 cell culture model. J Nutr Biochem *11*, 574-580.

Ghassemian, M., Nambara, E., Cutler, S., Kawaide, H., Kamiya, Y., and McCourt, P. (2000). Regulation of abscisic acid signaling by the ethylene response pathway in *Arabidopsis*. The Plant cell *12*, 1117-1126.

Gietz, R.D., and Schiestl, R.H. (2007). Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc *2*, 35-37.

Gilmour, S.J., Sebolt, A.M., Salazar, M.P., Everard, J.D., and Thomashow, M.F. (2000). Overexpression of the *Arabidopsis CBF3* transcriptional activator mimics multiple biochemical changes associated with cold acclimation. Plant Physiol *124*, 1854-1865.

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F., and Goodman, H.M. (1992). Isolation of the *Arabidopsis ABI3* gene by positional cloning. The Plant cell *4*, 1251-1261.

Glahn, R.P., Lee, O.A., Yeung, A., Goldman, M.I., and Miller, D.D. (1998). Caco-2 cell ferritin formation predicts nonradiolabeled food iron availability in an in vitro digestion/Caco-2 cell culture model. J Nutr *128*, 1555-1561.

Godzik, A. (2011). Metagenomics and the protein universe. Curr Opin Struct Biol *21*, 398-403.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H.*, et al.* (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science *296*, 92-100.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M.*, et al.* (1996). Life with 6000 genes. Science *274*, 546, 563-547.

Golling, G., Amsterdam, A., Sun, Z., Antonelli, M., Maldonado, E., Chen, W., Burgess, S., Haldi, M., Artzt, K., and Farrington, S. (2002). Insertional

mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. Nature genetics *31*, 135-140.

Gong, W., Shen, Y.P., Ma, L.G., Pan, Y., Du, Y.L., Wang, D.H., Yang, J.Y., Hu, L.D., Liu, X.F., Dong, C.X.*, et al.* (2004). Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. Plant Physiol *135*, 773-782.

Gotea, V., Petrykowska, H.M., and Elnitski, L. (2013). Bidirectional promoters as important drivers for the emergence of species-specific transcripts. PLoS One *8*, e57323.

Graeber, K., Nakabayashi, K., Miatton, E., Leubner-Metzger, G., and Soppe, W.J. (2012). Molecular mechanisms of seed dormancy. Plant Cell Environ *35*, 1769-1786.

Greene, E.A., Codomo, C.A., Taylor, N.E., Henikoff, J.G., Till, B.J., Reynolds, S.H., Enns, L.C., Burtner, C., Johnson, J.E., and Odden, A.R. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. Genetics *164*, 731-740.

Griffith, O.W. (1980). Determination of glutathione and glutathione disulfide using glutathione reductase and 2-vinylpyridine. Anal Biochem *106*, 207-212.

Grotz, N., and Guerinot, M.L. (2006). Molecular aspects of Cu, Fe and Zn homeostasis in plants. Biochem Biophys Acta *1763*, 595-608.

Guo, W.J., Li, P., Ling, J., and Ye, S.P. (2007). Significant comparative characteristics between orphan and nonorphan genes in the rice (*Oryza sativa* L.) genome. Comparative and functional genomics *2007*.

Guo, Y.L. (2013). Gene family evolution in green plants with emphasis on the origination and evolution of Arabidopsis thaliana genes. Plant J *73*, 941-951.

Halder, G., Callaerts, P., and Gehring, W.J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. Science *267*, 1788-1792.

Hamilton, A.J., Bouzayen, M., and Grierson, D. (1991). Identification of a tomato gene for the ethylene-forming enzyme by expression in yeast. Proc Natl Acad Sci USA *88*, 7434-7437.

Hara, K., Kajita, R., Torii, K.U., Bergmann, D.C., and Kakimoto, T. (2007). The secretory peptide gene *EPF1* enforces the stomatal one-cell-spacing rule. Genes Dev *21*, 1720-1725.

Harrison, S.J., Mott, E.K., Parsley, K., Aspinall, S., Gray, J.C., and Cottage, A. (2006). A rapid and robust method of identifying transformed *Arabidopsis thaliana* seedlings following floral dip transformation. Plant Methods *2*, 19.

Haydon, M.J., and Cobbett, C.S. (2007). A novel major facilitator superfamily protein at the tonoplast influences zinc tolerance and accumulation in Arabidopsis. Plant Physiol *143*, 1705-1719.

Hedges, S.B. (2002). The origin and evolution of model organisms. Nat Rev Genet *3*, 838-849.

Heinen, T.J., Staubach, F., Haming, D., and Tautz, D. (2009). Emergence of a new gene from an intergenic region. Curr Biol *19*, 1527-1531.

Henikoff, S., Tatchell, K., Hall, B.D., and Nasmyth, K.A. (1981). Isolation of a gene from Drosophila by complementation in yeast. Nature *289*, 33-37.

Herrero, E., Ros, J., Belli, G., and Cabiscol, E. (2008). Redox control and oxidative stress in yeast cells. Biochimica et biophysica acta *1780*, 1217-1235.

Hilson, P., Allemeersch, J., Altmann, T., Aubourg, S., Avon, A., Beynon, J., Bhalerao, R.P., Bitton, F., Caboche, M., Cannoot, B.*, et al.* (2004). Versatile gene-specific sequence tags for *Arabidopsis* functional genomics: transcript profiling and reverse genetics applications. Genome research *14*, 2176-2189.

Himmelbach, A., Yang, Y., and Grill, E. (2003). Relay and control of abscisic acid signaling. Curr Opin Plant Biol *6*, 470-479.

Hindt, M.N., and Guerinot, M.L. (2012). Getting a sense for signals: regulation of the plant iron deficiency response. Biochimica et biophysica acta *1823*, 1521-1530.

Hinnen, A., Hicks, J.B., and Fink, G.R. (1978). Transformation of yeast. Proc Natl Acad Sci USA *75*, 1929-1933.

Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Penagaricano, F., Lindquist, E., Pedraza, M.A., Barry, K.*, et al.* (2014). Insights into the maize pan-genome and pan-transcriptome. The Plant cell *26*, 121-135.

Huang, G.T., Ma, S.L., Bai, L.P., Zhang, L., Ma, H., Jia, P., Liu, J., Zhong, M., and Guo, Z.F. (2011). Signal transduction during cold, salt, and drought stresses in plants. Mol Biol Rep.

Ichikawa, T., Nakazawa, M., Kawashima, M., Iizumi, H., Kuroda, H., Kondou, Y., Tsuhara, Y., Suzuki, K., Ishikawa, A., Seki, M.*, et al.* (2006). The FOX hunting system: an alternative gain-of-function gene hunting technique. Plant J *48*, 974-985.

Ichikawa, T., Nakazawa, M., Kawashima, M., Muto, S., Gohda, K., Suzuki, K., Ishikawa, A., Kobayashi, H., Yoshizumi, T., Tsumoto, Y.*, et al.* (2003). Sequence database of 1172 T-DNA insertion sites in Arabidopsis activation-tagging lines that showed phenotypes in T1 generation. Plant J *36*, 421-429.

Inoue, H., Kobayashi, T., Nozoye, T., Takahashi, M., Kakei, Y., Suzuki, K., Nakazono, M., Nakanishi, H., Mori, S., and Nishizawa, N.K. (2009). Rice OsYSL15 is an iron-regulated iron(III)-deoxymugineic acid transporter expressed in the roots and is essential for iron uptake in early growth of the seedlings. J Biol Chem *284*, 3470-3479.

Izawa, S., Inoue, Y., and Kimura, A. (1995). Oxidative stress response in yeast: effect of glutathione on adaptation to hydrogen peroxide stress in *Saccharomyces cerevisiae*. FEBS letters *368*, 73-76.

Jaglo-Ottosen, K.R., Gilmour, S.J., Zarka, D.G., Schabenberger, O., and Thomashow, M.F. (1998). *Arabidopsis CBF1* overexpression induces *COR* genes and enhances freezing tolerance. Science *280*, 104-106.

Jamieson, D.J. (1998). Oxidative stress responses of the yeast *Saccharomyces cerevisiae*. Yeast *14*, 1511-1527.

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.H., Jeong, D.H., Lee, J., Kim, C., Jang, S., Lee, S., and Yang, K. (2000). T-DNA insertional mutagenesis for functional genomics in rice. Plant J *22*, 561-570.

Jeong, D.H., An, S., Park, S., Kang, H.G., Park, G.G., Kim, S.R., Sim, J., Kim, Y.O., Kim, M.K., Kim, S.R.*, et al.* (2006). Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. Plant J *45*, 123-132.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S.*, et al.* (2011). Ancestral polyploidy in seed plants and angiosperms. Nature *473*, 97-100.

Johnson, Z.I., and Chisholm, S.W. (2004). Properties of overlapping genes are conserved across microbial genomes. Genome research *14*, 2268-2272.

Joung, J.K., and Sander, J.D. (2013). TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol *14*, 49-55.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome research *20*, 1313-1326.

Kaessmann, H., Vinckenbosch, N., and Long, M. (2009). RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet *10*, 19-31.

Kakimoto, T. (1996). CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. Science *274*, 982-985.

Kandlbinder, A., Finkemeier, I., Wormuth, D., Hanitzsch, M., and Dietz, K.J. (2004). The antioxidant status of photosynthesizing leaves under nutrient deficiency: redox regulation, gene expression and antioxidant activity in *Arabidopsis thaliana*. Physiol Plant *120*, 63-73.

Kanhonou, R., Serrano, R., and Palau, R.R. (2001). A catalytic subunit of the sugar beet protein kinase CK2 is induced by salt stress and increases NaCl tolerance in *Saccharomyces cerevisiae*. Plant Mol Biol *47*, 571-579.

Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat Biotechnol *17*, 287-291.

Kato, S., Endoh, H., Masuhiro, Y., Kitamoto, T., Uchiyama, S., Sasaki, H., Masushige, S., Gotoh, Y., Nishida, E., Kawashima, H.*, et al.* (1995). Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. Science *270*, 1491-1494.

Khalturin, K., Anton-Erxleben, F., Sassmann, S., Wittlieb, J., Hemmrich, G., and Bosch, T.C. (2008). A novel gene family controls species-specific morphological traits in *Hydra*. PLoS Biol *6*, e278.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., and Bosch, T.C. (2009). More than just orphans: are taxonomically-restricted genes important in evolution? Trends Genet *25*, 404-413.

Khan, M.S., Haas, F.H., Samami, A.A., Gholami, A.M., Bauer, A., Fellenberg, K., Reichelt, M., Hansch, R., Mendel, R.R., Meyer, A.J.*, et al.* (2010). Sulfite reductase defines a newly discovered bottleneck for assimilatory sulfate reduction and is essential for growth and development in *Arabidopsis thaliana*. The Plant cell *22*, 1216-1231.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K. (2007). The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant J *50*, 347-363.

Kim, J.H., Park, S.M., Kang, M.R., Oh, S.Y., Lee, T.H., Muller, M.T., and Chung, I.K. (2005). Ubiquitin ligase MKRN1 modulates telomere length homeostasis through a proteolysis of hTERT. Genes Dev *19*, 776-781.

Kim, J.I., Sharkhuu, A., Jin, J.B., Li, P., Jeong, J.C., Baek, D., Lee, S.Y., Blakeslee, J.J., Murphy, A.S., Bohnert, H.J.*, et al.* (2007). *yucca6*, a dominant mutation in *Arabidopsis*, affects auxin accumulation and auxin-related phenotypes. Plant Physiol *145*, 722-735.

Kim, S.A., Punshon, T., Lanzirotti, A., Li, L., Alonso, J.M., Ecker, J.R., Kaplan, J., and Guerinot, M.L. (2006). Localization of iron in Arabidopsis seed requires the vacuolar membrane transporter VIT1. Science *314*, 1295-1298.

Klatte, M., Schuler, M., Wirtz, M., Fink-Straube, C., Hell, R., and Bauer, P. (2009). The analysis of *Arabidopsis* nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. Plant Physiol *150*, 257-271.

Knowles, D.G., and McLysaght, A. (2009). Recent de novo origin of human protein-coding genes. Genome research *19*, 1752-1759.

Kobayashi, T., and Nishizawa, N.K. (2012). Iron uptake, translocation, and regulation in higher plants. Annu Rev Plant Biol *63*, 131-152.

Kondou, Y., Higuchi, M., and Matsui, M. (2010). High-throughput characterization of plant gene functions by using gain-of-function technology. Annu Rev Plant Biol *61*, 373-393.

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

Koornneef, M., Reuling, G., and Karssen, C. (1984). The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. Physiologia plantarum *61*, 377-383.

Kourmpetli, S., Lee, K., Hemsley, R., Rossignol, P., Papageorgiou, T., and Drea, S. (2013). Bidirectional promoters in seed development and related hormone/stress responses. BMC plant biology *13*, 187.

Krull, M., Brosius, J., and Schmitz, J. (2005). Alu-SINE exonization: en route to protein-coding function. Mol Biol Evol *22*, 1702-1711.

Kuhn, J.M., Boisson-Dernier, A., Dizon, M.B., Maktabi, M.H., and Schroeder, J.I. (2006). The protein phosphatase *AtPP2CA* negatively regulates abscisic acid signal transduction in *Arabidopsis*, and effects of *abh1* on *AtPP2CA* mRNA. Plant Physiol *140*, 127-139.

Kuo, C.H., and Kissinger, J.C. (2008). Consistent and contrasting properties of lineage-specific genes in the apicomplexan parasites *Plasmodium* and *Theileria*. BMC evolutionary biology *8*, 108.

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

Lanquar, V., Lelievre, F., Bolte, S., Hames, C., Alcon, C., Neumann, D., Vansuyt, G., Curie, C., Schroder, A., Kramer, U*., et al.* (2005). Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. The EMBO journal *24*, 4041-4051.

Law, M.Y., Charles, S.A., and Halliwell, B. (1983). Glutathione and ascorbic acid in spinach (*Spinacia oleracea*) chloroplasts. The effect of hydrogen peroxide and of Paraquat. The Biochemical journal *210*, 899-903.

LeClere, S., and Bartel, B. (2001). A library of *Arabidopsis* 35S-cDNA lines for identifying novel mutants. Plant molecular biology *46*, 695-703.

Lefebvre, V., North, H., Frey, A., Sotta, B., Seo, M., Okamoto, M., Nambara, E., and Marion-Poll, A. (2006). Functional analysis of *Arabidopsis* NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. Plant J *45*, 309-319.

Leonelli, S. (2007). *Arabidopsis*, the botanical *Drosophila*: from mouse cress to model organism. Endeavour *31*, 34-38.

Leung, J., Merlot, S., and Giraudat, J. (1997). The *Arabidopsis ABSCISIC ACID-INSENSITIVE2 (ABI2)* and *ABI1* genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. The Plant cell *9*, 759-771.

Levine, M.T., Jones, C.D., Kern, A.D., Lindfors, H.A., and Begun, D.J. (2006). Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression. Proc Natl Acad Sci U S A *103*, 9935-9939.

Leymarie, J., Vitkauskaite, G., Hoang, H.H., Gendreau, E., Chazoule, V., Meimoun, P., Corbineau, F., El-Maarouf-Bouteau, H., and Bailly, C. (2012). Role of reactive oxygen species in the regulation of *Arabidopsis* seed dormancy. Plant & cell physiology *53*, 96-106.

Li, C.Y., Zhang, Y., Wang, Z., Cao, C., Zhang, P.W., Lu, S.J., Li, X.M., Yu, Q., Zheng, X., Du, Q*., et al.* (2010a). A human-specific *de novo* protein-coding gene associated with human brain functions. PLoS Comput Biol *6*, e1000734.

Li, D., Dong, Y., Jiang, Y., Jiang, H., Cai, J., and Wang, W. (2010b). A de novo originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. Cell Res *20*, 408-420.

Li, J., Nam, K.H., Vafeados, D., and Chory, J. (2001a). BIN2, a new brassinosteroid-insensitive locus in *Arabidopsis*. Plant Physiol *127*, 14-22.

Li, L., Chen, O.S., McVey Ward, D., and Kaplan, J. (2001b). CCC1 is a transporter that mediates vacuolar iron storage in yeast. The Journal of biological chemistry *276*, 29515-29519.

Li, L., Foster, C.M., Gan, Q., Nettleton, D., James, M.G., Myers, A.M., and Wurtele, E.S. (2009). Identification of the novel protein QQS as a component of the starch metabolic network in *Arabidopsis* leaves. Plant J *58*, 485-498.

Li, L., and Wurtele, E.S. (2014). The QQS orphan gene of Arabidopsis modulates carbon and nitrogen allocation in soybean. Plant biotechnology journal.

Li, R., Li, Y., Zheng, H., Luo, R., Zhu, H., Li, Q., Qian, W., Ren, Y., Tian, G., Li, J.*, et al.* (2010c). Building the sequence map of the human pan-genome. Nat Biotechnol *28*, 57-63.

Lin, H., Moghe, G., Ouyang, S., Iezzoni, A., Shiu, S.H., Gu, X., and Buell, C.R. (2010). Comparative analyses reveal distinct sets of lineage-specific genes within *Arabidopsis thaliana*. BMC evolutionary biology *10*, 41.

Lingam, S., Mohrbacher, J., Brumbarova, T., Potuschak, T., Fink-Straube, C., Blondet, E., Genschik, P., and Bauer, P. (2011). Interaction between the bHLH transcription factor FIT and ETHYLENE INSENSITIVE3/ETHYLENE INSENSITIVE3-LIKE1 reveals molecular linkage between the regulation of iron acquisition and ethylene signaling in Arabidopsis. The Plant cell *23*, 1815-1829.

Linkies, A., Muller, K., Morris, K., Tureckova, V., Wenk, M., Cadman, C.S., Corbineau, F., Strnad, M., Lynn, J.R., Finch-Savage, W.E.*, et al.* (2009). Ethylene interacts with abscisic acid to regulate endosperm rupture during germination: a comparative approach using Lepidium sativum and Arabidopsis thaliana. The Plant cell *21*, 3803-3822.

Lipman, D.J., Souvorov, A., Koonin, E.V., Panchenko, A.R., and Tatusova, T.A. (2002). The relationship of protein conservation and sequence length. BMC evolutionary biology *2*.

Lisch, D. (2013). How important are transposons for plant evolution? Nature Reviews Genetics *14*, 49-61.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. The Plant cell *10*, 1391-1406.

Liu, X., Zhou, X., Li, Y., Tian, J., Zhang, Q., Li, S., Wang, L., Zhao, J., Chen, R., and Fan, Y. (2014). Identification and functional characterization of bidirectional gene pairs and their intergenic regions in maize. BMC Genomics *15*, 338.

Liu, Y., Ye, N., Liu, R., Chen, M., and Zhang, J. (2010). H2O2 mediates the regulation of ABA catabolism and GA biosynthesis in *Arabidopsis* seed dormancy and germination. Journal of experimental botany *61*, 2979-2990.

Liu, Y.G., and Whittier, R.F. (1995). Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics *25*, 674-681.

Lockton, S., and Gaut, B.S. (2009). The contribution of transposable elements to expressed coding sequence in *Arabidopsis thaliana*. J Mol Evol *68*, 80-89.

Long, D., Martin, M., Sundberg, E., Swinburne, J., Puangsomlee, P., and Coupland, G. (1993). The maize transposable element system Ac/Ds as a mutagen in *Arabidopsis*: identification of an albino mutation induced by Ds insertion. Proceedings of the National Academy of Sciences *90*, 10370-10374.

Long, M., Betran, E., Thornton, K., and Wang, W. (2003). The origin of new genes: glimpses from the young and old. Nat Rev Genet *4*, 865-875.

Long, M., and Langley, C.H. (1993). Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila. Science *260*, 91-95.

Long, M., Rosenberg, C., and Gilbert, W. (1995). Intron phase correlations and the evolution of the intron/exon structure of genes. Proc Natl Acad Sci U S A *92*, 12495-12499.

Long, T.A., Tsukagoshi, H., Busch, W., Lahner, B., Salt, D.E., and Benfey, P.N. (2010). The bHLH transcription factor POPEYE regulates response to iron deficiency in *Arabidopsis* roots. The Plant cell *22*, 2219-2236.

López-Madrigal, S., Latorre, A., Porcar, M., Moya, A., and Gil, R. (2011). Complete genome sequence of “Candidatus Tremblaya princeps” strain

PCVAL, an intriguing translational machine below the living-cell status. Journal of bacteriology *193*, 5587-5588.

Luikenhuis, S., Perrone, G., Dawes, I.W., and Grant, C.M. (1998). The yeast *Saccharomyces cerevisiae* contains two glutaredoxin genes that are required for protection against reactive oxygen species. Molecular biology of the cell *9*, 1081-1091.

Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science *290*, 1151-1155.

Maeda, T., Wurgler-Murphy, S.M., and Saito, H. (1994). A two-component system that regulates an osmosensing MAP kinase cascade in yeast. Nature *369*, 242-245.

Maguin, E., Prevost, H., Ehrlich, S., and Gruss, A. (1996). Efficient insertional mutagenesis in lactococci and other gram-positive bacteria. Journal of Bacteriology *178*, 931-935.

Marquez, J.A., and Serrano, R. (1996). Multiple transduction pathways regulate the sodium-extrusion gene PMR2/ENA1 during salt stress in yeast. FEBS letters *382*, 89-92.

Marroni, F., Pinosio, S., and Morgante, M. (2014). Structural variation and genome complexity: is dispensable really dispensable? Curr Opin Plant Biol *18*, 31-36.

Mathews, H., Clendennen, S.K., Caldwell, C.G., Liu, X.L., Connors, K., Matheis, N., Schuster, D.K., Menasco, D.J., Wagoner, W., Lightner, J.*, et al.* (2003). Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. The Plant cell *15*, 1689-1703.

Matsuno, M., Compagnon, V., Schoch, G.A., Schmitt, M., Debayle, D., Bassard, J.E., Pollet, B., Hehn, A., Heintz, D., Ullmann, P.*, et al.* (2009). Evolution of a novel phenolic pathway for pollen development. Science *325*, 1688-1692.

Maxwell, K., and Johnson, G.N. (2000). Chlorophyll fluorescence--a practical guide. Journal of experimental botany *51*, 659-668.

McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeted screening for induced mutations. Nature biotechnology *18*, 455-457.

McCarrey, J.R., and Thomas, K. (1987). Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. Nature *326*, 501-505.

McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots. Nature *411*, 709-713.

Meeks-Wagner, D., and Hartwell, L.H. (1986). Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. Cell *44*, 43-52.

Meguro, R., Asano, Y., Odagiri, S., Li, C., Iwatsuki, H., and Shoumura, K. (2007). Nonheme-iron histochemistry for light and electron microscopy: a historical, theoretical and technical review. Arch Histol Cytol *70*, 1-19.

Meier, S., Ruzvidzo, O., Morse, M., Donaldson, L., Kwezi, L., and Gehring, C. (2010). The Arabidopsis *wall associated kinase-like 10* gene encodes a functional guanylyl cyclase and is co-expressed with pathogen defense related genes. PLoS One *5*, e8904.

Mell, J.C., and Burgess, S.M. (2001). Yeast as a Model Genetic Organism. In eLS (John Wiley & Sons, Ltd).

Mendoza-Cozatl, D.G., Xie, Q., Akmakjian, G.Z., Jobe, T.O., Patel, A., Stacey, M.G., Song, L., Demoin, D.W., Jurisson, S.S., Stacey, G.*, et al.* (2014). OPT3 Is a Component of the Iron-Signaling Network between Leaves and Roots and Misregulation of OPT3 Leads to an Over-Accumulation of Cadmium in Seeds. Molecular plant.

Meunier, J., Lemoine, F., Soumillon, M., Liechti, A., Weier, M., Guschanski, K., Hu, H., Khaitovich, P., and Kaessmann, H. (2013). Birth and expression evolution of mammalian microRNA genes. Genome Res *23*, 34-45.

Miethke, M., and Marahiel, M.A. (2007). Siderophore-based iron acquisition and pathogen control. Microbiology and molecular biology reviews : MMBR *71*, 413-451.

Millar, A.A., Jacobsen, J.V., Ross, J.J., Helliwell, C.A., Poole, A.T., Scofield, G., Reid, J.B., and Gubler, F. (2006). Seed dormancy and ABA metabolism in *Arabidopsis* and barley: the role of ABA 8'-hydroxylase. Plant J *45*, 942-954.

Morgante, M., De Paoli, E., and Radovic, S. (2007). Transposable elements and the plant pan-genomes. Curr Opin Plant Biol *10*, 149-155.

Morrissey, J., Baxter, I.R., Lee, J., Li, L., Lahner, B., Grotz, N., Kaplan, J., Salt, D.E., and Guerinot, M.L. (2009). The ferroportin metal efflux proteins function in iron and cobalt homeostasis in Arabidopsis. The Plant cell *21*, 3326-3338.

Motamayor, J.C., Mockaitis, K., Schmutz, J., Haiminen, N., Donald III, L., Cornejo, O., Findley, S.D., Zheng, P., Utro, F., and Royaert, S. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome biology *14*, r53.

Mulet, J.M., Alemany, B., Ros, R., Calvete, J.J., and Serrano, R. (2004). Expression of a plant serine O-acetyltransferase in *Saccharomyces cerevisiae* confers osmotic tolerance and creates an alternative pathway for cysteine biosynthesis. Yeast *21*, 303-312.

Mulet, J.M., Leube, M.P., Kron, S.J., Rios, G., Fink, G.R., and Serrano, R. (1999). A novel mechanism of ion homeostasis and salt tolerance in yeast: the Hal4 and Hal5 protein kinases modulate the Trk1-Trk2 potassium transporter. Molecular and cellular biology *19*, 3328-3337.

Muller, H.J. (1932). Further studies on the nature and causes of gene mutations. Paper presented at: Proc 6th Int Congr Genet.

Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. Annual review of plant biology *59*, 651-681.

Mussig, C., and Altmann, T. (2001). Brassinosteroid signaling in plants. Trends in endocrinology and metabolism: TEM *12*, 398-402.

Nakamura, H., Hakata, M., Amano, K., Miyao, A., Toki, N., Kajikawa, M., Pang, J., Higashi, N., Ando, S., Toki, S.*, et al.* (2007). A genome-wide gain-of function analysis of rice genes using the FOX-hunting system. Plant molecular biology *65*, 357-371.

Nakazawa, M., Ichikawa, T., Ishikawa, A., Kobayashi, H., Tsuhara, Y., Kawashima, M., Suzuki, K., Muto, S., and Matsui, M. (2003). Activation tagging, a novel tool to dissect the functions of a gene family. Plant J *34*, 741-750.

Nanjo, T., Sakurai, T., Totoki, Y., Toyoda, A., Nishiguchi, M., Kado, T., Igasaki, T., Futamura, N., Seki, M., Sakaki, Y.*, et al.* (2007). Functional annotation of 19,841 *Populus nigra* full-length enriched cDNA clones. BMC Genomics *8*, 448.

Neill, S., Desikan, R., and Hancock, J. (2002). Hydrogen peroxide signalling. Curr Opin Plant Biol *5*, 388-395.

Neme, R., and Tautz, D. (2014). Evolution: dynamics of de novo gene emergence. Curr Biol *24*, R238-240.

Nogi, Y., Shimada, H., Matsuzaki, Y., Hashimoto, H., and Fukasawa, T. (1984). Regulation of expression of the galactose gene cluster in

Saccharomyces cerevisiae. II. The isolation and dosage effect of the regulatory gene GAL80. Molecular & general genetics : MGG *195*, 29-34.

Nonogaki, H. (2014). Seed dormancy and germination-emerging mechanisms and new hypotheses. Frontiers in plant science *5*, 233.

Ochman, H. (2001). Lateral and oblique gene transfer. Curr Opin Genet Dev *11*, 616-619.

Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y., and Yamaguchi, S. (2003). Gibberellin biosynthesis and response during Arabidopsis seed germination. The Plant cell *15*, 1591-1604.

Ogihara, Y., Mochida, K., Kawaura, K., Murai, K., Seki, M., Kamiya, A., Shinozaki, K., Carninci, P., Hayashizaki, Y., Shin, I.T.*, et al.* (2004). Construction of a full-length cDNA library from young spikelets of hexaploid wheat and its characterization by large-scale sequencing of expressed sequence tags. Genes & genetic systems *79*, 227-232.

Ohno, S. (1970). Evolution by gene duplication (London: George Alien & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag.).

Oldfield, C.J., and Dunker, A.K. (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. Annual review of biochemistry *83*, 553-584.

Osborn, T.C., Pires, J.C., Birchler, J.A., Auger, D.L., Chen, Z.J., Lee, H.S., Comai, L., Madlung, A., Doerge, R.W., Colot, V.*, et al.* (2003). Understanding mechanisms of novel gene expression in polyploids. Trends in genetics : TIG *19*, 141-147.

Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. Plant Journal *53*, 674-690.

Ostergaard, L., and Yanofsky, M.F. (2004). Establishing gene function by mutagenesis in *Arabidopsis thaliana*. Plant J *39*, 682-696.

Page, D.R., and Grossniklaus, U. (2002). The art and design of genetic screens: Arabidopsis thaliana. Nat Rev Genet *3*, 124-136.

Parinov, S., and Sundaresan, V. (2000). Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. Curr Opin Biotechnol *11*, 157-161.

Pei, Z.M., Ghassemian, M., Kwak, C.M., McCourt, P., and Schroeder, J.I. (1998). Role of farnesyltransferase in ABA regulation of guard cell anion channels and plant water loss. Science *282*, 287-290.

Penfield, S., Josse, E.M., Kannangara, R., Gilday, A.D., Halliday, K.J., and Graham, I.A. (2005). Cold and light control seed germination through the bHLH transcription factor *SPATULA*. Curr Biol *15*, 1998-2006.

Peyret, P., Perez, P., and Alric, M. (1995). Structure, genomic organization, and expression of the Arabidopsis thaliana aconitase gene. Plant aconitase show significant homology with mammalian iron-responsive element-binding protein. J Biol Chem *270*, 8131-8137.

Phillips, P.C. (2008). Epistasis-the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet *9*, 855-867.

Philpott, C.C., and Protchenko, O. (2008). Response to iron deprivation in *Saccharomyces cerevisiae*. Eukaryotic cell *7*, 20-27.

Popping, B., Gibbons, T., and Watson, M.D. (1996). The *Pisum sativum* MAP kinase homologue (PsMAPK) rescues the *Saccharomyces cerevisiae hog1* deletion mutant under conditions of high osmotic stress. Plant molecular biology *31*, 355-363.

Prelich, G. (2012). Gene overexpression: uses, mechanisms, and interpretation. Genetics *190*, 841-854.

Prince, V.E., and Pickett, F.B. (2002). Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet *3*, 827-837.

Priyanka, B., Sekhar, K., Reddy, V.D., and Rao, K.V. (2010). Expression of pigeonpea hybrid-proline-rich protein encoding gene (CcHyPRP) in yeast and *Arabidopsis* affords multiple abiotic stress tolerance. Plant biotechnology journal *8*, 76-87.

Quintero, F.J., and Blatt, M.R. (1997). A new family of K+ transporters from Arabidopsis that are conserved across phyla. FEBS letters *415*, 206-211.

Quintero, F.J., Garciadeblas, B., and Rodriguez-Navarro, A. (1996). The SAL1 gene of *Arabidopsis*, encoding an enzyme with 3'(2'),5'-bisphosphate nucleotidase and inositol polyphosphate 1-phosphatase activities, increases salt tolerance in yeast. The Plant cell *8*, 529-537.

Rajjou, L., Duval, M., Gallardo, K., Catusse, J., Bally, J., Job, C., and Job, D. (2012). Seed germination and vigor. Annual review of plant biology *63*, 507-533.

Ramirez, L., Bartoli, C.G., and Lamattina, L. (2013). Glutathione and ascorbic acid protect *Arabidopsis* plants against detrimental effects of iron deficiency. Journal of experimental botany *64*, 3169-3178.

Rampey, R.A., Woodward, A.W., Hobbs, B.N., Tierney, M.P., Lahner, B., Salt, D.E., and Bartel, B. (2006). An Arabidopsis basic helix-loop-helix leucine zipper protein modulates metal homeostasis and auxin conjugate responsiveness. Genetics *174*, 1841-1857.

Ranieri, A., Castagna, A., Baldan, B., and Soldatini, G.F. (2001). Iron deficiency differently affects peroxidase isoforms in sunflower. Journal of experimental botany *52*, 25-35.

Ranz, J.M., and Parsch, J. (2012). Newly evolved genes: moving from comparative genomics to functional studies in model systems. How important is genetic novelty for species adaptation and diversification? Bioessays *34*, 477-483.

Ratzkin, B., and Carbon, J. (1977). Functional expression of cloned yeast DNA in Escherichia coli. Proc Natl Acad Sci U S A *74*, 487-491.

Rausell, A., Kanhonou, R., Yenush, L., Serrano, R., and Ros, R. (2003). The translation initiation factor eIF1A is an important determinant in the tolerance to NaCl stress in yeast and plants. The Plant Journal *34*, 257-267.

Reddy, A.S., Ali, G.S., Celesnik, H., and Day, I.S. (2011). Coping with stresses: roles of calcium- and calcium/calmodulin-regulated gene expression. The Plant cell *23*, 2010-2032.

Reinhardt, J.A., Wanjiru, B.M., Brant, A.T., Saelao, P., Begun, D.J., and Jones, C.D. (2013). *De novo* ORFs in Drosophila are important to organismal fitness and evolved rapidly from previously non-coding sequences. PLoS Genet *9*, e1003860.

Resource, T.A.I. The Arabidopsis Information Resource.

Rice Full-Length c, D.N.A.C., National Institute of Agrobiological Sciences Rice Full-Length c, D.N.A.P.T., Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M.*, et al.* (2003). Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science *301*, 376-379.

Riemer, J., Hoepken, H.H., Czerwinska, H., Robinson, S.R., and Dringen, R. (2004). Colorimetric ferrozine-based assay for the quantitation of iron in cultured cells. Anal Biochem *331*, 370-375.

Rine, J., Hansen, W., Hardeman, E., and Davis, R.W. (1983). Targeted selection of recombinant clones through gene dosage effects. Proc Natl Acad Sci U S A *80*, 6750-6754.

Rios, G., Cabedo, M., Rull, B., Yenush, L., Serrano, R., and Mulet, J.M. (2013). Role of the yeast multidrug transporter Qdr2 in cation homeostasis and the oxidative stress response. FEMS yeast research *13*, 97-106.

Rizzon, C., Ponger, L., and Gaut, B.S. (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLoS Comput Biol *2*, e115.

Robinson, N.J., Procter, C.M., Connolly, E.L., and Guerinot, M.L. (1999). A ferric-chelate reductase for iron uptake from soils. Nature *397*, 694-697.

Rogers, E.E., and Guerinot, M.L. (2002). FRD3, a member of the multidrug and toxin efflux family, controls iron deficiency responses in Arabidopsis. The Plant cell *14*, 1787-1799.

Rosado, A., Schapire, A.L., Bressan, R.A., Harfouche, A.L., Hasegawa, P.M., Valpuesta, V., and Botella, M.A. (2006). The *Arabidopsis* Tetratricopeptide Repeat-Containing Protein TTL1 Is Required for Osmotic Stress Responses and Abscisic Acid Sensitivity. Plant Physiology *142*, 1113-1126.

Roschzttardtz, H., Conéjéro, G., Curie, C., and Mari, S. (2009). Identification of the Endodermal Vacuole as the Iron Storage Compartment in the Arabidopsis Embryo. Plant Physiology *151*, 1329-1338.

Rutherford, J.C., Jaron, S., and Winge, D.R. (2003). Aft1p and Aft2p mediate iron-responsive gene expression in yeast through related promoter elements. The Journal of biological chemistry *278*, 27636-27643.

Ryu, M.Y., Cho, S.K., and Kim, W.T. (2010). The *Arabidopsis* C3H2C3-Type RING E3 Ubiquitin Ligase AtAIRP1 Is a Positive Regulator of an Abscisic Acid-Dependent Response to Drought Stress. Plant Physiology *154*, 1983-1997.

Sabath, N., Wagner, A., and Karlin, D. (2012). Evolution of viral proteins originated de novo by overprinting. Mol Biol Evol *29*, 3767-3780.

Sakurai, T., Satou, M., Akiyama, K., Iida, K., Seki, M., Kuromori, T., Ito, T., Konagaya, A., Toyoda, T., and Shinozaki, K. (2005). RARGE: a large-scale database of RIKEN Arabidopsis resources ranging from transcriptome to phenome. Nucleic acids research *33*, D647-650.

Sassone-Corsi, P. (2002). Unique chromatin remodeling and transcriptional regulation in spermatogenesis. Science *296*, 2176-2178.

Sato, K., Shin, I.T., Seki, M., Shinozaki, K., Yoshida, H., Takeda, K., Yamazaki, Y., Conte, M., and Kohara, Y. (2009). Development of 5006 full-length CDNAs in barley: a tool for accessing cereal genomics resources. DNA research : an international journal for rapid publication of reports on genes and genomes *16*, 81-89.

Schaaf, G., Honsbein, A., Meda, A.R., Kirchner, S., Wipf, D., and von Wiren, N. (2006). AtIREG2 encodes a tonoplast transport protein involved in iron-dependent nickel detoxification in Arabidopsis thaliana roots. The Journal of biological chemistry *281*, 25532-25540.

Schachtman, D.P., and Schroeder, J.I. (1994). Structure and transport mechanism of a high-affinity potassium uptake transporter from higher plants. Nature *370*, 655-658.

Schmidt, W., and Buckhout, T.J. (2011). A hitchhiker's guide to the *Arabidopsis* ferrome. Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale *49*, 462-470.

Schramm, E.C., Nelson, S.K., Kidwell, K.K., and Steber, C.M. (2013). Increased ABA sensitivity results in higher seed dormancy in soft white spring wheat cultivar 'Zak'. TAG Theoretical and applied genetics Theoretische und angewandte Genetik *126*, 791-803.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Nakajima, M., Enju, A., Akiyama, K., Oono, Y.*, et al.* (2002). Functional annotation of a full-length Arabidopsis cDNA collection. Science *296*, 141-145.

Seo, M., Nambara, E., Choi, G., and Yamaguchi, S. (2009). Interaction of light and hormone signals in germinating seeds. Plant molecular biology *69*, 463-472.

Serrano, R. (1996). Salt tolerance in plants and microorganisms: toxicity targets and defense responses. Int Rev Cytol *165*, 1-52.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., and Qiu, J.-L. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. Nature biotechnology *31*, 686-688.

She, X., Horvath, J.E., Jiang, Z., Liu, G., Furey, T.S., Christ, L., Clark, R., Graves, T., Gulden, C.L., Alkan, C.*, et al.* (2004). The structure and evolution of centromeric transition regions within the human genome. Nature *430*, 857-864.

Silveira, A.B., Trontin, C., Cortijo, S., Barau, J., Del Bem, L.E., Loudet, O., Colot, V., and Vincentz, M. (2013). Extensive natural epigenetic variation at a *de novo* originated gene. PLoS Genet *9*, e1003437.

Sivitz, A.B., Hermand, V., Curie, C., and Vert, G. (2012). Arabidopsis bHLH100 and bHLH101 control iron homeostasis via a FIT-independent pathway. PLoS One *7*, e44843.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Depamphilis, C.W., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. American journal of botany *96*, 336-348.

Somerville, C. (2000). The twentieth century trajectory of plant biology. Cell *100*, 13-25.

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

Stacey, M.G., Koh, S., Becker, J., and Stacey, G. (2002). AtOPT3, a member of the oligopeptide transporter family, is essential for embryo development in Arabidopsis. The Plant cell *14*, 2799-2811.

Stacey, M.G., Patel, A., McClain, W.E., Mathieu, M., Remley, M., Rogers, E.E., Gassmann, W., Blevins, D.G., and Stacey, G. (2008). The *Arabidopsis* AtOPT3 protein functions in metal homeostasis and movement of iron to developing seeds. Plant Physiol *146*, 589-601.

Stephen, D.W., and Jamieson, D.J. (1996). Glutathione is an important antioxidant molecule in the yeast *Saccharomyces cerevisiae*. FEMS microbiology letters *141*, 207-212.

Stookey, L.L. (1970). Ferrozine- A new spectrophotometric reagent for iron. Analytical chemistry *42*, 779-781.

Struhl, K., and Davis, R.W. (1977). Production of a functional eukaryotic enzyme in Escherichia coli: cloning and expression of the yeast structural gene for imidazole-glycerolphosphate dehydratase (his3). Proc Natl Acad Sci U S A *74*, 5255-5259.

Suzuki-Fujimoto, T., Fukuma, M., Yano, K.I., Sakurai, H., Vonika, A., Johnston, S.A., and Fukasawa, T. (1996). Analysis of the galactose signal

transduction pathway in *Saccharomyces cerevisiae*: interaction between Gal3p and Gal80p. Molecular and cellular biology *16*, 2504-2508.

Sychrova, H. (2004). Yeast as a model organism to study transport and homeostasis of alkali metal cations. Physiological research / Academia Scientiarum Bohemoslovaca *53 Suppl 1*, S91-98.

Tautz, D., and Domazet-Loso, T. (2011). The evolutionary origin of orphan genes. Nat Rev Genet *12*, 692-702.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S.*, et al.* (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A *102*, 13950-13955.

Toll-Riera, M., Bosch, N., Bellora, N., Castelo, R., Armengol, L., Estivill, X., and Alba, M.M. (2009). Origin of primate orphan genes: a comparative genomics approach. Mol Biol Evol *26*, 603-612.

van Bakel, H., Strengman, E., Wijmenga, C., and Holstege, F.C. (2005). Gene expression profiling and phenotype analyses of *S. cerevisiae* in response to changing copper reveals six genes with new roles in copper and iron metabolism. Physiological genomics *22*, 356-367.

Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. Nat Rev Genet *10*, 725-732.

van der Fits, L., and Memelink, J. (2000). *ORCA3*, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. Science *289*, 295-297.

Verghese, J., Abrams, J., Wang, Y., and Morano, K.A. (2012). Biology of the heat shock response and protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system. Microbiology and molecular biology reviews : MMBR *76*, 115-158.

Verslues, P.E., Agarwal, M., Katiyar-Agarwal, S., Zhu, J., and Zhu, J.K. (2006). Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. Plant J *45*, 523-539.

Vert, G., Grotz, N., Dedaldechamp, F., Gaymard, F., Guerinot, M.L., Briat, J.F., and Curie, C. (2002). IRT1, an *Arabidopsis* transporter essential for iron uptake from the soil and for plant growth. The Plant cell *14*, 1223-1233.

Vinckenbosch, N., Dupanloup, I., and Kaessmann, H. (2006). Evolutionary fate of retroposed gene copies in the human genome. Proc Natl Acad Sci U S A *103*, 3220-3225.

Wang, Q., Wan, L., Li, D., Zhu, L., Qian, M., and Deng, M. (2009). Searching for bidirectional promoters in *Arabidopsis thaliana*. BMC bioinformatics *10 Suppl 1*, S29.

Wang, W., Zheng, H., Fan, C., Li, J., Shi, J., Cai, Z., Zhang, G., Liu, D., Zhang, J., Vang, S.*, et al.* (2006). High rate of chimeric gene origination by retroposition in plant genomes. The Plant cell *18*, 1791-1802.

Waters, B.M., Chu, H.H., Didonato, R.J., Roberts, L.A., Eisley, R.B., Lahner, B., Salt, D.E., and Walker, E.L. (2006). Mutations in *Arabidopsis* yellow stripe-like1 and yellow stripe-like3 reveal their roles in metal ion homeostasis and loading of metal ions in seeds. Plant Physiol *141*, 1446-1458.

Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malancharuvil, E.J., Neff, M.M.*, et al.* (2000). Activation tagging in Arabidopsis. Plant Physiol *122*, 1003-1013.

Wilson, G., Bertrand, N., Patel, Y., Hughes, J., Feil, E., and Field, D. (2005). Orphans as taxonomically restricted and ecologically important genes. Microbiology *151*, 2499-2501.

Wilson, G.A., Feil, E.J., Lilley, A.K., and Field, D. (2007). Large-scale comparative genomic ranking of taxonomically restricted genes (TRGs) in bacterial and archaeal genomes. PLoS One *2*, e324.

Winkler, A., Arkind, C., Mattison, C.P., Burkholder, A., Knoche, K., and Ota, I. (2002). Heat stress activates the yeast high-osmolarity glycerol mitogen-activated protein kinase pathway, and protein tyrosine phosphatases are essential under heat stress. Eukaryotic cell *1*, 163-173.

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., and Rieseberg, L.H. (2009). The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci U S A *106*, 13875-13879.

Woodward, C., Bemis, S.M., Hill, E.J., Sawa, S., Koshiba, T., and Torii, K.U. (2005). Interaction of auxin and *ERECTA* in elaborating *Arabidopsis* inflorescence architecture revealed by the activation tagging of a new member of the *YUCCA* family putative flavin monooxygenases. Plant Physiol *139*, 192-203.

Wu, D.-D., Irwin, D.M., and Zhang, Y.-P. (2011). De novo origin of human protein-coding genes. PLoS genetics *7*, e1002379.

Wu, D.D., and Zhang, Y.P. (2013). Evolution and function of de novo originated genes. Mol Phylogenet Evol *67*, 541-545.

Xiao, W., Liu, H., Li, Y., Li, X., Xu, C., Long, M., and Wang, S. (2009). A rice gene of de novo origin negatively regulates pathogen-induced defense response. PLoS One *4*, e4603.

Xie, C., Zhang, Y.E., Chen, J.Y., Liu, C.J., Zhou, W.Z., Li, Y., Zhang, M., Zhang, R., Wei, L., and Li, C.Y. (2012). Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs. PLoS Genet *8*, e1002942.

Yamaguchi-Iwai, Y., Dancis, A., and Klausner, R.D. (1995). AFT1: a mediator of iron regulated transcriptional control in *Saccharomyces cerevisiae*. The EMBO journal *14*, 1231-1239.

Yamaguchi, S., Kamiya, Y., and Nambara, E. (2007). Regulation of ABA and GA levels during seed development and germination in Arabidopsis. Annual Plant Reviews Volume 27: Seed Development, Dormancy and Germination, 224-247.

Yamanaka, T., Miyama, M., and Tada, Y. (2009). Transcriptome profiling of the mangrove plant *Bruguiera gymnorhiza* and identification of salt tolerance genes by *Agrobacterium* functional screening. Biosci Biotechnol Biochem *73*, 304-310.

Yamauchi, Y., Ogawa, M., Kuwahara, A., Hanada, A., Kamiya, Y., and Yamaguchi, S. (2004). Activation of gibberellin biosynthesis and response pathways by low temperature during imbibition of *Arabidopsis thaliana* seeds. The Plant cell *16*, 367-378.

Yan, J., Tsuichihara, N., Etoh, T., and Iwai, S. (2007). Reactive oxygen species and nitric oxide are involved in ABA inhibition of stomatal opening. Plant, cell & environment *30*, 1320-1325.

Yang, L., Zou, M., Fu, B., and He, S. (2013). Genome-wide identification, characterization, and expression analysis of lineage-specific genes within zebrafish. BMC Genomics *14*, 65.

Yang, X., Jawdy, S., Tschaplinski, T.J., and Tuskan, G.A. (2009). Genome-wide identification of lineage-specific genes in *Arabidopsis, Oryza* and *Populus*. Genomics *93*, 473-480.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012a). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC bioinformatics *13*, 134.

Ye, N., Zhu, G., Liu, Y., Zhang, A., Li, Y., Liu, R., Shi, L., Jia, L., and Zhang, J. (2012b). Ascorbic acid and reactive oxygen species are involved in the inhibition of seed germination by abscisic acid in rice seeds. Journal of experimental botany *63*, 1809-1822.

Yesilirmak, F., and Sayers, Z. (2009). Heterelogous expression of plant genes. International journal of plant genomics *2009*, 296482.

Yu, H., Chen, X., Hong, Y.Y., Wang, Y., Xu, P., Ke, S.D., Liu, H.Y., Zhu, J.K., Oliver, D.J., and Xiang, C.B. (2008). Activated expression of an Arabidopsis HD-START protein confers drought tolerance with improved root system and reduced stomatal density. The Plant cell *20*, 1134-1151.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X.*, et al.* (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science *296*, 79-92.

Yuan, Y., Wu, H., Wang, N., Li, J., Zhao, W., Du, J., Wang, D., and Ling, H.Q. (2008). FIT interacts with AtbHLH38 and AtbHLH39 in regulating iron uptake gene expression for iron homeostasis in *Arabidopsis*. Cell Res *18*, 385-397.

Zemojtel, T., Penzkofer, T., Schultz, J., Dandekar, T., Badge, R., and Vingron, M. (2007). Exonization of active mouse L1s: a driver of transcriptome evolution? BMC Genomics *8*, 392.

Zhai, Z., Gayomba, S.R., Jung, H.I., Vimalakumari, N.K., Pineros, M., Craft, E., Rutzke, M.A., Danku, J., Lahner, B., Punshon, T.*, et al.* (2014). OPT3 Is a Phloem-Specific Iron Transporter That Is Essential for Systemic Iron Signaling and Redistribution of Iron and Cadmium in Arabidopsis. The Plant cell *26*, 2249-2264.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., Li, X., Pierick, C.J., Dobbs, D., Peterson, T.*, et al.* (2010). High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. Proceedings of the National Academy of Sciences *107*, 12028-12033.

Zhao, L., Saelao, P., Jones, C.D., and Begun, D.J. (2014). Origin and spread of de novo genes in *Drosophila melanogaster* populations. Science *343*, 769-772.

Zhao, Y., Christensen, S.K., Fankhauser, C., Cashman, J.R., Cohen, J.D., Weigel, D., and Chory, J. (2001). A role for flavin monooxygenase-like enzymes in auxin biosynthesis. Science *291*, 306-309.

Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and Wang, W. (2008). On the origin of new genes in *Drosophila*. Genome Res *18*, 1446-1455.

Zhu, J.K. (2002). Salt and drought stress signal transduction in plants. Annual review of plant biology *53*, 247-273.

Conferences and Presentations

- **Sandesh Rao**, Channa Keshavaiah, Wei Li, Mark Donoghue and Charles Spillane, Identification and characterisation of a novel mutant *sf7* displaying fasciation in *Arabidopsis thaliana*. Oral presentation at the Irish Plant Scientists' Association Meeting (IPSAM), May 16-17, 2013, Galway, Ireland.

- **Sandesh Rao**, Channa Keshavaiah, Wei Li, Mark Donoghue and Charles Spillane, Characterisation of a novel mutant *sf7* affecting fasciation and vascular development in *Arabidopsis thaliana.* Oral presentation at the Irish Plant Scientists' Association Meeting (IPSAM), Aril 3-5, 2012, Dublin, Ireland.

- Mark T.A. Donoghue, Channa Keshavaiah, **Sandesh H Swamidatta**, Charles Spillane, Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana.* Poster presentation at International Plant and Animal Genome-XIX Conference, January 15-19,2011, San Diego, California,USA.

Publications

- Mark T.A. Donoghue, Channa Keshavaiah, **Sandesh H. Swamidatta** and Charles Spillane,(2011), Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana. BMC Evolutionary Biology,* **11**:47.

Invention disclosure

- Charles Spillane, **Sandesh Rao** and Galina Brychkova. 'Iron-fortification under normal and iron-deprived conditions by metabolic engineering', submitted to NUI Galway. **IDF-2014-003**.

Appendices

A1. A fasciation phenotype caused by a T-DNA insertion at the *WUSCHEL* locus shows a paramutation like phenomenon in *Arabidopsis thaliana*

Introduction

Paramutation is a epigenetic phenomenon wherein trans interaction between alleles or homologous sequences (which could be two alleles of the same gene, or two transgenes, or one transgene and one endogenous gene) leads to establishment of distinct gene expression states that are meiotically heritable for generations (Arteaga-Vazquez and Chandler, 2010). Paramutation resembles genetic mutation as it is heritable but do not cause a change in DNA sequence. It is often associated with changes in DNA methylation and chromatin structure and hence considered as an epigenetic phenomenon. The phenomenon was first observed in Pea (Bateson and Pellew, 1915) and extensively studied in Maize. Paramutation has fascinated geneticists from a long time because it defies the genetic models of heredity and evolution. Paramutation like interactions are also reported in other species including Drosophila (de Vanssay et al., 2012), mammals and humans (Suter and Martin, 2010).

The best examples of paramutation in plants are the genes involved in colour formation which are easy to identify. Some of the classical examples include, maize *r1* (*colored 1*), *p1* (*pericarp color 1*), *b1* (*booster 1*) and *pl1* (*purple plant 1*) loci (Brink, 1958; Coe Jr, 1959; Hollick et al., 1995; Sidorenko and Peterson, 2001), *Petunia an3* (van Houwelingen et al., 1999) and transgenic *A1* locus (Meyer et al., 1993) which determine the red and purple pigments in various tissues. The paramutation involves three types of alleles; a paramutagenic allele which transfers its epigenetic state to another allele called paramutable allele and the allele that does not participate in the paramutation is termed as neutral allele. The paramutable allele becomes paramutagenic once it acquires epigenetic state of the paramutagenic allele. Some paramutated alleles may exhibit secondary paramutation in some cases and can paramutate new paramutable allele in subsequent interactions. The general principle of paramutation is given in Fig.1.

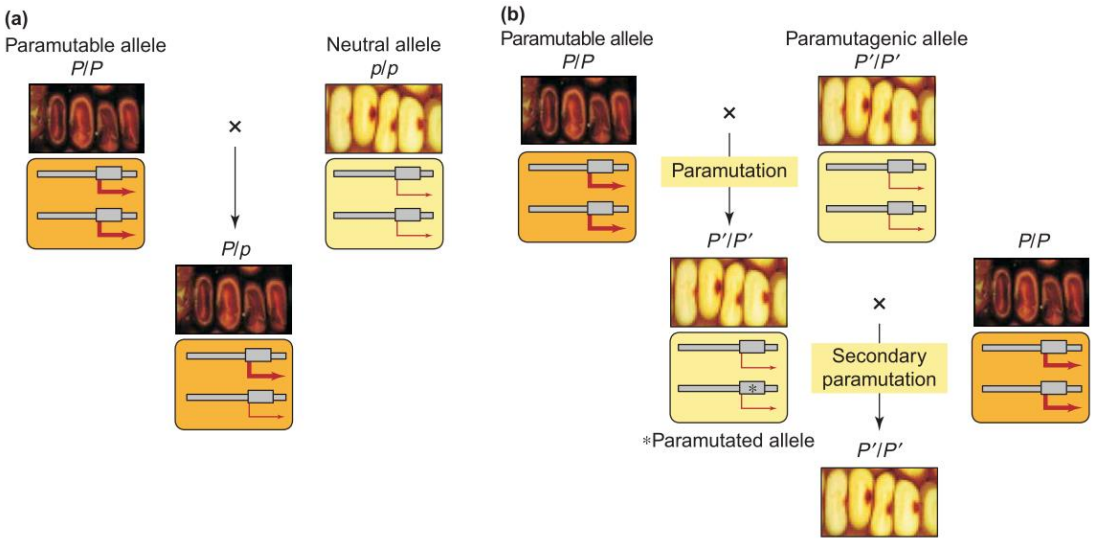


Figure 1. The principle of paramutation.

(a) A paramutable, highly expressed allele (*P*) maintains the active state when combined with a neutral, non- or weakly expressed allele (*p*), with both alleles separating unchanged afterwards. Thickness of arrows denotes expression levels, thick for high expression and thin arrow for low expression level.
(b) When combined with a paramutagenic allele (*P'*), the same paramutable allele is paramutated and expression is reduced. This change in expression is heritable and the modified allele can now paramutate a newly introduced paramutable allele (secondary paramutation). The illustrations show the effects of *p1* paramutation on the phenotype of maize kernels (*p1* controls kernel pigmentation).
Figure adapted from Stam and Mittelsten Scheid (2005).

From various studies of paramutation in plants and animals, it has been found that paramutation is governed by two mechanisms; one is by cis-acting signals, signals physically associated with the gene that they regulate and the other by trans-acting signals. Cis-acting signals include DNA methylation or histone modifications leading to changes in chromatin structure which ultimately leads to change in gene expression status. On the other hand trans-acting signals involve transcription factors or small RNAs (sRNAs) which regulate their own expression levels by feedback loops (Bonasio et al., 2010; Pilu, 2011). The involvement of DNA methylation in establishing paramutation is well known (Pilu et al., 2009; Sidorenko and Peterson, 2001; Walker, 1998). Furthermore, repeated sequences (direct as well as in inverted orientation) have been found to be associated with

paramutation in few examples of maize (Kermicle et al., 1995; Sidorenko and Peterson, 2001; Stam et al., 2002). In maize *b1* locus, the transition from B-I to B' (the two *b1* alleles) correlates with hypermethylation of tandem repeats located 100 Kb upstream of coding region. Recently it has been shown that these tandem repeats in *b1* locus are sufficient for paramutation induced trans-generational silencing through small interfering RNAs (siRNAs) (Belele et al., 2013). The involvement of RNA molecules in paramutation has been described in various examples. It has been established that the transfer of information between paramutable and paramutagenic alleles is mediated by RNA molecules (Alleman et al., 2006; Erhard et al., 2009; Rassoulzadegan et al., 2006; Sidorenko et al., 2009).

Arabidopsis T-DNA lines are a valuable resource for finding gene functions using reverse genetic approach. A phenomenon of gene silencing by trans-inactivation has been observed in many of the transgenic lines containing T-DNA wherein one T-DNA exerts a dominant epigenetic silencing on another T-DNA (unlinked) in trans (Matzke et al., 1989; Stam et al., 1997). The transcriptional gene silencing in *Arabidopsis* T-DNA lines was triggered by the promoter sequences that are common to T-DNA insertions, usually CaMV 35S promoter (Daxinger et al., 2008). Xue et al. (2012) found a SALK T-DNA triggered epigenetic change which led to the enhanced expression of the gene in which another T-DNA was residing. The second T-DNA became paramutagenic which involved DNA methylation modulated by MET1 and ROS1. In this study we report a SALK T-DNA residing in the promoter region of *WUSCHEL* displaying severe fasciated phenotype. Interestingly the phenotype is segregated as a dominant phenotype and transmitted only in selfed heterozygous progenies. However homozygous selfed progenies reverted to wild type phenotype in subsequent generations suggesting a possible paramutation like phenomenon.

Results and discussion

ORFAS mutant displays a strong fasciation phenotype

The gene AT2G17960 was identified as an orphan gene in our previous work (Donoghue et al., 2011). In order to analyse the gene function, a T-DNA insertion line (SALK_018310) for this gene was analysed identifying homozygous and heterozygous lines and the morphological phenotypes were scored. Most evident phenotype of homozygous plants is the severe fasciation of the stem which is characterised by flattened stem with severe twisting (Fig.2B) at some places. Hence we called the locus as *ORFAS* (for ORphan FASciated). The *ORFAS* mutant plants often produced bunchy type of inflorescences with terminal clustering of flower and flower buds. Surprisingly heterozygous (*ORFAS/orfas*) lines also showed varying degrees of fasciation. Further we analysed two additional insertion lines (SAIL_51_F07 and GABI_830G06), none of which showed any phenotypic defect. Overall the T-DNA insertion at the *ORFAS* locus is a dominant mutation since heterozygotes also show mutant phenotype.

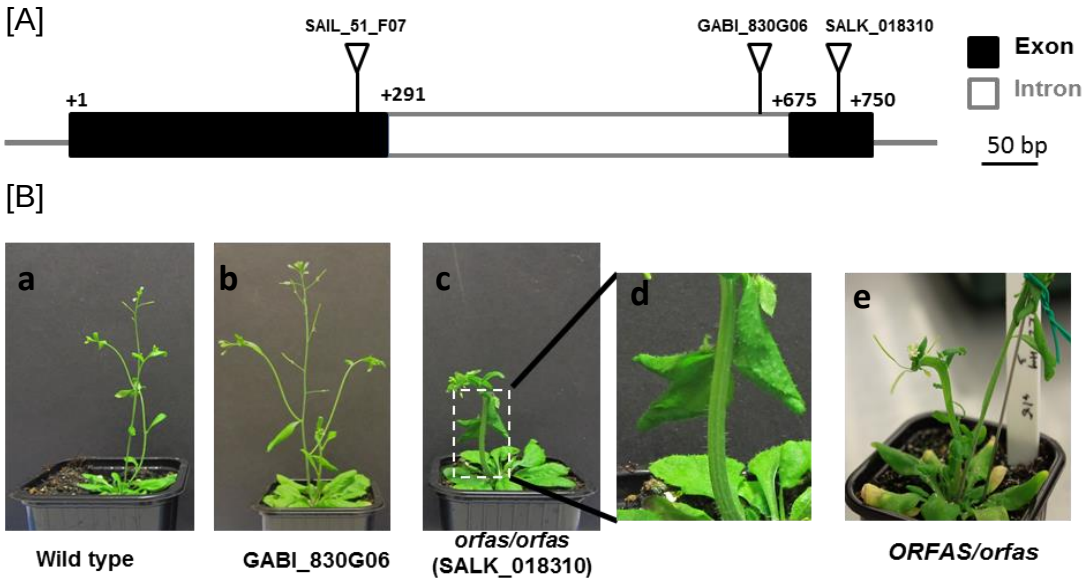


Fig.2: Phenotypic analysis of *ORFAS* mutant.

[A] Schematic representation of *ORFAS* locus (annotation as per TAIR10) labelled with the T-DNA insertions analysed in the study.

[B] Fasciated phenotype of *orfas/orfas* (c). The other T-DNA line (b) is normal as wild type (a). Close up of fasciated stem of *orfas/orfas* in (d) and the fasciated phenotype in *ORFAS/orfas* (e).

ORFAS locus lies on the *WUSCHEL* promoter

Since only one T-DNA insertion showed the fasciated phenotype we closely looked at the *ORFAS* locus and found that *ORFAS* is situated 1708 bp upstream of the start codon (ATG sequence) of *WUSCHEL* (*WUS*), a well characterized homeodomain transcription factor involved in meristem cell maintenance (Laux, 2003, 2009; Laux and Mayer, 1998; Laux et al., 1996). The *WUS* promoter and its regulatory regions extend up to the 5 kb upstream of the *WUS* gene from its start position+1 (Baurle and Laux, 2005). Based on Baurle and Laux (2005) findings, we found that *ORFAS* locus lies on a region required for enhanced expression levels of *WUS* (Fig.3). This suggested that T-DNA insertion in *orfas* is disrupting *WUS* promoter and the phenotype manifested by *orfas* mutant might be due to disruption of *WUS* rather than *ORFAS* itself.

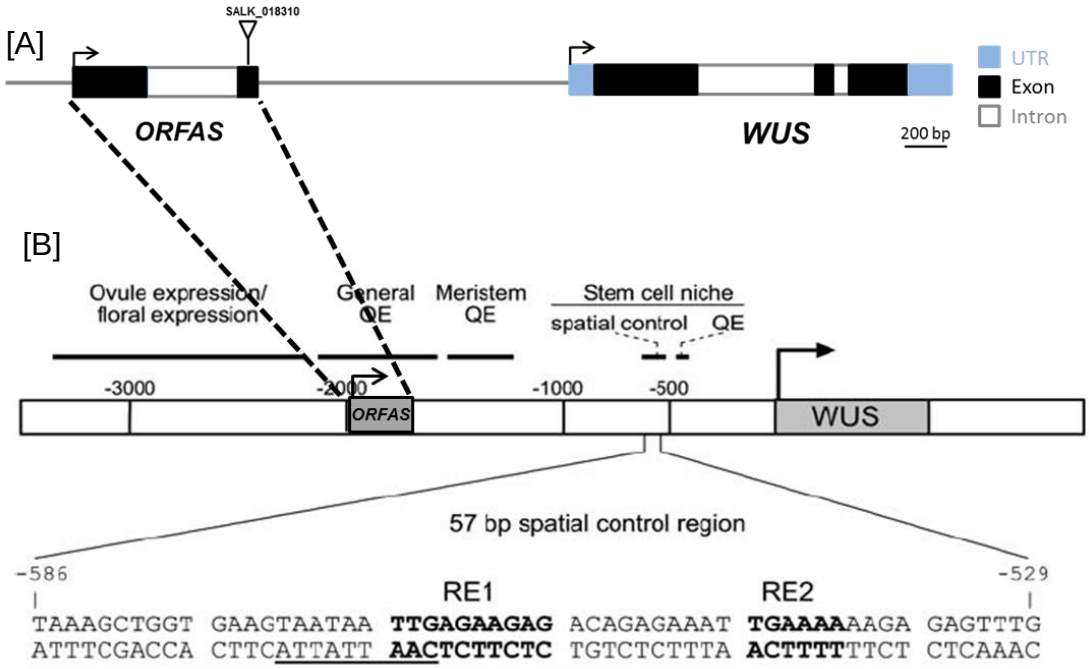


Fig 3. *ORFAS* locus and its link with *WUS.*

[A] Schematic representation of *ORFAS* and *WUS* locus. T-DNA insertion on the annotated *ORFAS* locus is shown (annotations as per TAIR10).
[B] The approximate positions of regulatory domains are indicated. *ORFAS* locus lies on the general quantitative element (QE) of *WUS*. At bottom, nucleotides within the 57-bp spatial control region essential for promoter activity in the stem cell niche of the inflorescence meristem (RE1 and RE2) are indicated in boldface letters. Figure modified from Baurle and Laux (2005).

orfas/*orfas* causes ectopic expression of *WUSCHEL*

Since we hypothesised that *orfas* disrupts *WUS* regulatory region, we analysed the expression of *WUS* mRNA in *orfas/orfas* in rosette leaf, root, shoot apex, and flower buds to see any expression changes in *WUS,* compared to wild type. The expression was analysed using semi-quantitative RT PCR. In wild type *WUS* expression was seen in shoot apex and flower buds but not in leaf and root tissues. However in *orfas* mutant, in addition to shoot apex and flower buds *WUS* expression was also seen in leaf and roots. This confirmed that T-DNA in *orfas* disrupts the *WUS* regulatory region and this leads to ectopic expression of *WUS* in *orfas* mutant. Further it has been reported previously that ectopic expression of WUS causes multiple morphological defects such as fasciated inflorescence stem and siliques, elongated pistils and ectopic floral buds on the inflorescence stem (Xu et al., 2005). This further suggested that the fasciated phenotype in *orfas* is due to ectopic *WUS* expression.

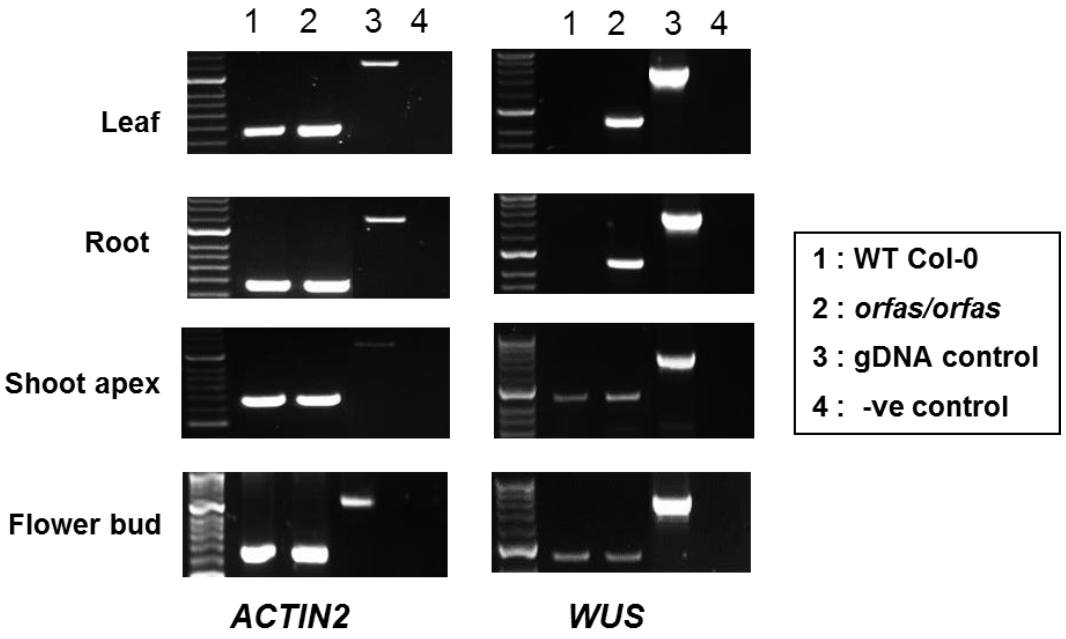


Fig.4: Ectopic expression of *WUS* in *orfas* mutant.

The expression analysis was done using semi-quantitative RT PCR (reaction cycles 28). *Actin2* was used as an internal control (left panel).

Segregation of *orfas* shows paramutation like inheritance

The dominant fasciation phenotype of *ORFAS/orfas* segregated in Mendelian pattern. The selfed heterozygous progenies of *ORFAS/orfas* segregated in 1:2:1 (A/a:A/a:a/a) genotypic ratio or 1:3 (normal:fasciated) phenotypic ratio (Fig 4). However we found that the selfed homozygous progenies of *orfas* displayed normal phenotype instead of expected fasciation in homozygous dominant condition. The homozygous lines did not show the fasciation phenotype in subsequent generations (up to 4 generations analysed). This suggested that homozygous lines somehow rectify the mutation which could be directly involved with *WUS* regulation.

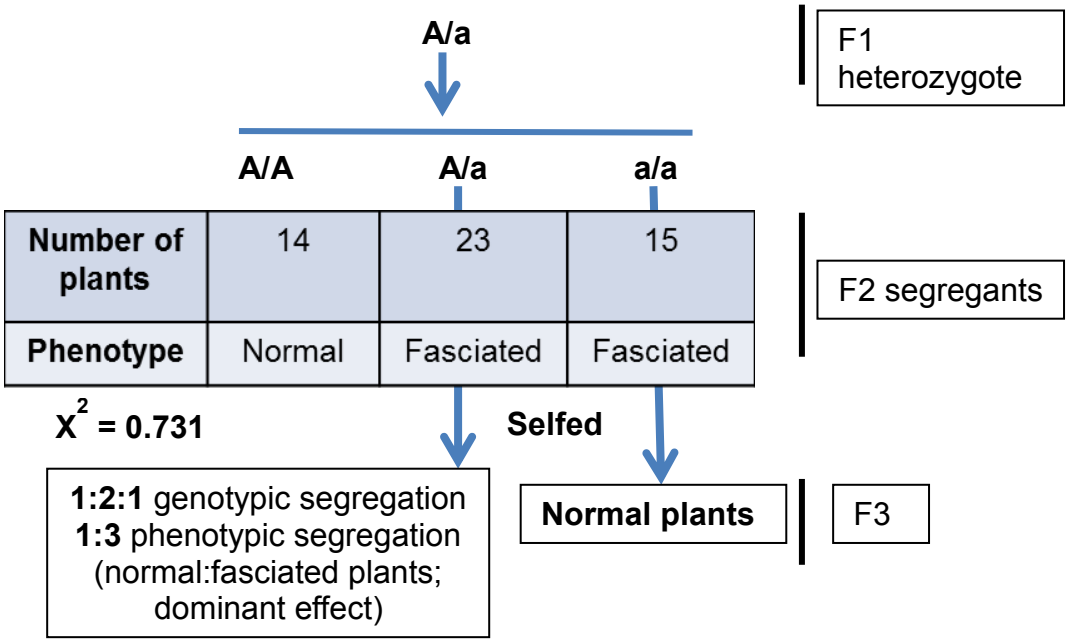


Fig. 5. Segregation pattern of *orfas* lines.

Segregation pattern of *orfas* mutant lines is depicted. Selfed *orfas/orfas* progenies revert back to the normal phenotype. However selfed *ORFS/orfas* shows Mendelian segregation for fasciation as a dominant phenotype.

We then analysed the transmission of fasciation phenotype by performing crosses with all possible combinations using wild type (A/A), *orfas/orfas* (a/a) and *ORFAS/orfas* (A/a). When fasciated heterozygote crossed with wild type, 50% of the F1 progenies show fasciation as expected. However a cross between homozygous fasciated plant (a/a) and

wild type (A/A) result in 50% of the progenies showing fasciation as opposed to 100% expected for the dominant phenotype in the F1 progeny. When homozygous *orfas*/*orfas* (a/a) plant which is not displaying fasciation crossed with wild type plants, the resulting F1 progenies displayed normal phenotype. Finally, heterozygous plants crossed with either homozygous *orfas* with fasciation or without fasciation resulted in 50% or 20% F1 progenies showing fasciation respectively.

Crossing genotypes	Parental phenotypes	Ratio of normal: fasciated plants	Expected genotypes
A/A x A/a	A/a with fasciation	**1:1**	**A/A : A/a**
A/A x a/a	a/a not showing phenotype	**All normal**	**A/a**
A/A x a/a	a/a with fasciation	**1:1**	**A/a**
A/a x a/a	a/a not showing phenotype	**4:1**	**A/a : a/a**
A/a x a/a	a/a with fasciation	**1:1**	**A/a : a/a**

Table-1. Transmission of fasciation phenotype in different cross directions.
For simplification dominant and recessive alleles were denoted as ‘A’ and ‘a’ respectively.

The data presented here (table-1) is just a phenotypic data from the ongoing experimentation and needs to be compared with genotypic data of the F1 progenies of respective crosses. This will help to arrive at a possible hypothesis for the unusual segregation pattern of the *orfas* T-DNA line. The reversion of fasciation to wild type phenotype in selfed homozygous *orfas* (fasciated) progenies and subsequent stable inheritance indicate a paramutation like epigenetic phenomenon. The trans-inactivation or epigenetic suppression of T-DNA insertion mutants has been observed in several instances (Daxinger et al., 2008; Gao and Zhao, 2013; Sandhu et al., 2013; Xue et al., 2012). However in *orfas* lines, we did not find any additional

T-DNA insertions (as identified by qRT-PCR based copy number detection). We suspect that T-DNA may be triggering small RNA mediated epigenetic changes at the *WUS* locus resulting in suppression of the fasciation phenotype. In maize, it has been shown that paramutation is associated with SiRNA biogenesis (Arteaga-Vazquez and Chandler, 2010). Hence, the plausible next step experiments are (a) to cross *ORFAS/orfas* and *orfas/orfas* lines with epigenetic modifier mutants such as *drm2*, *rdr2*, *dcl3* (involved in RNA directed DNA methylation) (b) methylation analysis at the *WUS* locus in *ORFAS/orfas*, *orfas/orfas*, *orfas/orfas* selfed and *ORFAS/ORFAS* and (c) northern blot for small RNAs at the *WUS* locus in *ORFAS/orfas*, *orfas/orfas*, *orfas/orfas* selfed and *ORFAS/ORFAS*.

Conclusions

The epigenetic phenomenon of paramutation has intrigued geneticists for a long time. Although paramutation in plants has been intensively studied, the exact requirements for establishing paramutation are still not clear. The phenomenon observed in this study indicates a paramutation like effect and we propose that it is likely mediated by small RNAs. Although there are a lot of unanswered questions regarding the inheritance of the fasciation phenotype and the phenomenon observed in this study, the proposed next experiments will help in answering those questions.

References

Alleman, M., Sidorenko, L., McGinnis, K., Seshadri, V., Dorweiler, J.E., White, J., Sikkink, K., and Chandler, V.L. (2006). An RNA-dependent RNA polymerase is required for paramutation in maize. Nature *442*, 295-298.

Arteaga-Vazquez, M.A., and Chandler, V.L. (2010). Paramutation in maize: RNA mediated trans-generational gene silencing. Curr Opin Genet Dev *20*, 156-163.

Bateson, W., and Pellew, C. (1915). On the genetics of “rogues” among culinary peas (*Pisum sativum*). Journal of Genetics *5*, 13-36.

Baurle, I., and Laux, T. (2005). Regulation of *WUSCHEL* transcription in the stem cell niche of the Arabidopsis shoot meristem. The Plant cell *17*, 2271-2280.

Belele, C.L., Sidorenko, L., Stam, M., Bader, R., Arteaga-Vazquez, M.A., and Chandler, V.L. (2013). Specific tandem repeats are sufficient for paramutation-induced trans-generational silencing. PLoS Genet *9*, e1003773.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. Science *330*, 612-616.

Brink, R.A. (1958). Paramutation at the *R* locus in maize. Paper presented at: Cold Spring Harbor symposia on quantitative biology (Cold Spring Harbor Laboratory Press).

Coe Jr, E.H. (1959). A regular and continuing conversion-type phenomenon at the *B* locus in maize. P Natl Acad Sci USA *45*, 828.

Daxinger, L., Hunter, B., Sheikh, M., Jauvion, V., Gasciolli, V., Vaucheret, H., Matzke, M., and Furner, I. (2008). Unexpected silencing effects from T-DNA tags in Arabidopsis. Trends in plant science *13*, 4-6.

de Vanssay, A., Bouge, A.L., Boivin, A., Hermant, C., Teysset, L., Delmarre, V., Antoniewski, C., and Ronsseray, S. (2012). Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature *490*, 112-115.

Donoghue, M.T., Keshavaiah, C., Swamidatta, S.H., and Spillane, C. (2011). Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana*. BMC evolutionary biology *11*, 47.

Erhard, K.F., Jr., Stonaker, J.L., Parkinson, S.E., Lim, J.P., Hale, C.J., and Hollick, J.B. (2009). RNA polymerase IV functions in paramutation in Zea mays. Science *323*, 1201-1205.

Gao, Y., and Zhao, Y. (2013). Epigenetic suppression of T-DNA insertion mutants in Arabidopsis. Molecular plant *6*, 539-545.

Hollick, J.B., Patterson, G.I., Coe-Jr, E.H., Cone, K.C., and Chandler, V.L. (1995). Allelic interactions heritably alter the activity of a metastable maize pl allele. Genetics *141*, 709.

Kermicle, J.L., Eggleston, W.B., and Alleman, M. (1995). Organization of paramutagenicity in *R-stippled* maize. Genetics *141*, 361-372.

Laux, T. (2003). The stem cell concept in plants: a matter of debate. Cell *113*, 281-283.

Laux, T. (2009). Stem cells serve as a source of new cells. Semin Cell Dev Biol *20*, 1082.

Laux, T., and Mayer, K.F. (1998). Cell fate regulation in the shoot meristem. Semin Cell Dev Biol *9*, 195-200.

Laux, T., Mayer, K.F., Berger, J., and Jurgens, G. (1996). The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*. Development *122*, 87-96.

Matzke, M.A., Primig, M., Trnovsky, J., and Matzke, A.J. (1989). Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. The EMBO journal *8*, 643-649.

Meyer, P., Heidmann, I., and Niedenhof, I. (1993). Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. The Plant Journal *4*, 89-100.

Pilu, R. (2011). Paramutation: just a curiosity or fine tuning of gene expression in the next generation? Curr Genomics *12*, 298-306.

Pilu, R., Panzeri, D., Cassani, E., Cerino Badone, F., Landoni, M., and Nielsen, E. (2009). A paramutation phenomenon is involved in the genetics of maize low phytic acid1-241 (lpa1-241) trait. Heredity *102*, 236-245.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., and Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature *441*, 469-474.

Sandhu, K.S., Koirala, P.S., and Neff, M.M. (2013). The *ben1-1* brassinosteroid-catabolism mutation is unstable due to epigenetic modifications of the intronic T-DNA insertion. G3 *3*, 1587-1595.

Sidorenko, L., Dorweiler, J.E., Cigan, A.M., Arteaga-Vazquez, M., Vyas, M., Kermicle, J., Jurcin, D., Brzeski, J., Cai, Y., and Chandler, V.L. (2009). A dominant mutation in mediator of paramutation2, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. PLoS Genet *5*, e1000725.

Sidorenko, L.V., and Peterson, T. (2001). Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize *p1* gene. The Plant cell *13*, 319-335.

Stam, M., Belele, C., Dorweiler, J.E., and Chandler, V.L. (2002). Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. Genes & development *16*, 1906-1918.

Stam, M., and Mittelsten Scheid, O. (2005). Paramutation: an encounter leaving a lasting impression. Trends in plant science *10*, 283-290.

Stam, M., Mol, J.N.M., and Kooter, J.M. (1997). Review article: the silence of genes in transgenic plants. Annals of Botany *79*, 3-12.

Suter, C.M., and Martin, D.I. (2010). Paramutation: the tip of an epigenetic iceberg? Trends in genetics : TIG *26*, 9-14.

van Houwelingen, A., Souer, E., Mol, J., and Koes, R. (1999). Epigenetic Interactions among Three dTph1 Transposons in Two Homologous Chromosomes Activate a New Excision–Repair Mechanism in Petunia. The Plant Cell Online *11*, 1319-1336.

Walker, E.L. (1998). Paramutation of the *r1* locus of maize is associated with increased cytosine methylation. Genetics *148*, 1973-1981.

Xu, Y.Y., Wang, X.M., Li, J., Li, J.H., Wu, J.S., Walker, J.C., Xu, Z.H., and Chong, K. (2005). Activation of the *WUS* gene induces ectopic initiation of floral meristems on mature stem surface in *Arabidopsis thaliana*. Plant molecular biology *57*, 773-784.

Xue, W., Ruprecht, C., Street, N., Hematy, K., Chang, C., Frommer, W.B., Persson, S., and Niittyla, T. (2012). Paramutation-like interaction of T-DNA loci in Arabidopsis. PLoS One *7*, e51651.

A2. Publication plans for the genes identified in this thesis

Orphan gene ID	Gene Name	Findings to date	Experiments to do	Target journal
AT2G25510	*ASR50*	Salt tolerant phenotype - KO line	(a) Identification and screening of additional loss of function alleles; (b) Complemntation of the *asr50-1* phenotype (c) Determining whether overexpression of *ASR50* increases the salt sensitivity (i.e. a dosage response test); and (d) Determining whether *CBF3* overexpression contributes to the *asr50-1* phenotype through generation and analysis of an *asr50 cbf3* double mutant line.	Plant Cell & Environment
AT1G31580	*ASR35*	Salt tolerant phenotype - o/e line	(a) Elucidation of the evolutionary origins and distribution; (b) Comparative analysis under salt stress of seedings of knockout lines, wild type and *ASR35* overexpression lines on MS plates; (c) Comparative analysis under salt stress of mature plants of knockouts, wild type and *ASR35* overexpression lines on soil; (d) Use of yeast heterologous system to identify possible pathways the *ASR35* gene may be interacting with	Plant science
AT4G39675	*ASR70*	ABA sensitive phenotype - Ko line	(a) Complementation of the *asr70* mutant lines with intact *ASR70* (b) Analysis of ASR70 overexpression lines to determine any dosage response to ABA (c) To investigate the *asr70* mutant lines and *ASR70* overepression lines under drought stress in transpiring conditions i.e on soil	Plant molecular biology
AT3G56260	*ASR12*	Osmotic tolerance- o/e line	(a) Elucidation of its evolutionary origins and distribution; (b) Comparative analysis under salt stress of seedings of knockouts (-/-), wild type (+/+) and *ASR12* o/e lines on MS plates; (c) Comparative analysis under salt stress of mature plants of knockouts, wild type and *ASR12* o/e lines on soil; (d) Use of yeast heterologous system to identify possible pathways the *ASR12* gene may be interacting with.	Gene
AT2G28570	*ASR51*	Morphological defect- o/e line	(a) To investigate response of hormones -gibberellic acid and brassinosteroid to identify any potential link with these hormone biosynthetic pathways (b) To analyse microarray datsets to find potential leads towards a developmental pathway (c) To generate and analyse knock out or knock down lines for *ASR51* to see if it shows any morphological defect (d) To generate transgenic *ASR51* with its own promoter to analyse any dosage effects of *ASR51* on plant morphology.	Gene