DNA LEVEL CHARACTERISTICS AND PHENOTYPIC FEATURES REVEALED BY KNOCKOUT, OVEREXPRESSION AND ADDITION MUTANTS FOR DUPLICATION-RESISTANT GENES

by

CHENGBO ZHOU

(Under the Direction of Andrew H. Paterson)

ABSTRACT

Gene duplication provides raw material for evolution in the genomes of higher eukaryotes. The propensity for genes to return to single copy dosage after duplication varies, and is partly reflected by their copy numbers in other species. Arabidopsis single copy genes can be divided into subsets with different characteristics, including different likelihoods of showing knockout (KO) phenotypes. We are particularly interested in the subset that was consistently restored to single copy dosage after independent whole-genome duplications in multiple lineages, known as duplication-resistant (DR) genes. The recurring removal of duplicated gene copies is associated with gene function, as implied by GO terms enriched in DR genes, making functional study of DR genes an approach to decipher the basis of duplication resistance. In this study, we assessed Arabidopsis DR gene functions via their KO phenotypes. Compared to non-DR singletons, T-DNA insertion mutants of DR genes were not prone to exhibit phenotypes under stress conditions, showing a low occurrence of ABA and cold sensitive phenotypes and an overrepresentation of salt-related phenotypes. On the other hand, severe phenotypes, including visible seedling alternations and phenotypes indicative of homozygous lethality, occurred more

frequently in DR genes than other singletons. In addition, Arabidopsis genotypes with artificial duplication (Addition, AD) (OE, by coupling to 35S promoters) of DR genes were made, to examine the consequences of having multiple copies of these genes. Visible duplication effects, as reflected by AD phenotypes, were more likely to occur in DR genes than in other singletons. Seedling phenotypes, absent in AD lines, were found in OE lines at a higher percentage than KO mutants. As for other genes, duplication and overexpression of DR genes can also result in stress tolerance. Genes with both OE and KO phenotypes constitute approximately one third of DR genes with OE lines, indicating that dosage sensitivity may be responsible for duplication resistance of some DR genes.

INDEX WORDS: duplication resistance, Arabidopsis, mutant, stress, phenotype

DNA LEVEL CHARACTERISTICS AND PHENOTYPIC FEATURES REVEALED BY KNOCKOUT, OVEREXPRESSION AND ADDITION MUTANTS FOR DUPLICATION-RESISTANT GENES

by

CHENGBO ZHOU

BS, Huazhong Agricultural University, China, 2009

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2016

DNA LEVEL CHARACTERISTICS AND PHENOTYPIC FEATURES REVEALED BY KNOCKOUT, OVEREXPRESSION AND ADDITION MUTANTS FOR DUPLICATION-RESISTANT GENES

by

CHENGBO ZHOU

Major Professor:	Andrew H. Paterson
Committee:	CJ Tsai
	James H. Leebens-Mack
	Rodney Mauricio
	Xiaoyu Zhang

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2016

ACKNOWLEDGEMENTS

To my advisor Professor Andrew Paterson, thank you for always being patient, supportive and awesome.

To Professor Kenneth Feldmann, thank you for making this project possible.

To my committee members, thank you all for joining my committee and for your help. I learnt a lot from the classes you taught.

To Ms. Katy Millward, thank you for teaching me all the techniques and for making the transgenic lines.

To Mr. Mitchell Feldmann, thank you for helping with the phenotype screen and taking the photographs.

To Ms. Susan Watkins, thank you for helping me with all the complicated forms. You are the best.

To Professor Brigitte Bruns, thank you for always encouraging me, for all your help and for being a truly amazing person.

To Ms. Carla Feldmann, thank you for taking good care of me while I was working on this project in Arizona.

To people in PBIO department and PGML, thank you for being nice, friendly and helpful.

To my friends and family, thank you for your love, support and always being there for me. I love you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Gene duplication at various scales is an inevitable part of genome evolution [1]. Duplications of individual genes and small genomic segments served as a continuous source of duplicated genes [2, 3] and were responsible for the expansion of gene families involved in response to environmental stimuli [4, 5]. Whole genome duplications, followed by massive gene loss [6], restructured the genome, brought dramatic changes to metabolic networks, and sometimes led to the born of new species [7]. Many higher eukaryotes have experienced whole genome duplication or triplication [8, 9], which occurred most frequently in flowering plants [10-16]. The abundance of genome duplication events makes angiosperms an outstanding model to study fate after gene duplication [17], a part as important as duplication itself during genome evolution.

The various evolutionary fates after gene duplication [1] can be summarized into two, retention and loss, according to the current snapshot of genomes: some genes have multiple copies and others do not. Duplicates retention, accompanied usually by sequence and functional divergence, occurred to a small number of genes, thus was considered rare and non-random [6]. Consequently, factors contributing to duplication retention have been extensively studied and became relatively well-understood [18-20]. In contrast, duplicates removal via 'non-functionalization' [21], as a fate awaiting most genes [6], seems to be random and received less attention.

While many singletons may randomly revert back to single copy status after duplication, evidence supporting non-random loss of duplicates abound. In Arabidopsis, some genes were repeatedly restored to 'singleton' after subsequent whole genome duplications [22]. Moreover, the number of single copy genes conserved among multiple species is much higher than expected assuming random gene loss [23, 24], and only decreases slightly as more genomes are sampled [25]. In addition, the chance of staying single copy seems to be associated with gene function. Genes in specific functional categories show more gene loss than genome average [26, 27], and certain protein domains are enriched in singletons of Arabidopsis and Oryza [28].

The preference for single copy is most evident in duplication-resistant (DR) genes, which tend to be singletons or have low copy number in multiple plant species they were conserved in. Given that all angiosperms are paleo-polyploids [17], DR genes were capable of surviving multiple and independent genome duplications as single copy, thus revealed strong resistance to duplication. As a result, DR genes became a favorite research target to study non-random gene loss.

Several groups of DR genes fitting the above definition have been identified, and their characteristics described. Single copy genes shared between Arabidopsis and rice have shorter proteins and more introns than the genome averages, along with distinctive promoter sequence features indicative of housekeeping function [23]. A subset of these genes conserved as single copy in moss and alga has low dN/dS value, indicating that they were under purifying selection thus not a group of random genes [23]. The shared single copy genes among Arabidopsis, Populus, Vitis and Oryza have more exons, fewer known domains, and are overrepresented in specific functional categories such as chloroplast, plastid and DNA or RNA metabolism, compared to the rest of the genome [25]. The functional bias remained similar when the number

of species increased to twenty [29]. Genes conserved as single copy among all or most of 20 flowering plants are well conserved in non-plant species as well, suggesting that they tend to function in basic processes important to probably all living creatures [29]. Their presumable involvement in housekeeping functions is further supported by their high expression breadth [29]. Additional characteristics of these DR genes include high expression level and biased codon use [29].

While the distinctive features of DR genes offered no clear explanation for duplication resistance, some inspired hypothesis about possible contributors to the persistent single copy status. The enrichment of chloroplast-targeting genes, for example, led to 'Select single copy gene hypothesis', stating that the duplication of chloroplast-targeting DR genes might disturb the dosage balance with their chloroplast genome encoded partners [30]. The 'Dominant-negative mutation hypothesis', on the other hand, springs from the observation that many DR genes are components of protein complexes [29]. According to this hypothesis, a mutated duplicate copy may encode proteins that are capable of binding other components but form inactive complex instead. Experimental confirmation of these hypotheses would be valuable for future research on duplication resistance, thus motivates further exploration of DR genes' characteristics.

The overrepresented GO terms in DR genes associated duplication resistance with gene function. Being perhaps an indispensable piece to solve the duplication resistance puzzle, the function of DR genes is far from being fully explored. More definitive than predicted functions is empirical observation of phenotypic changes resulting from silencing a gene. For example, gene importance is more directly reflected on the severity of its loss-of-function phenotypes than the annotated function. Meanwhile, the presence of knockout (KO) phenotype may serve as a strong proof and present a visual outcome of predicted dosage balance perturbation.

Thanks to the abundance of mutant resources in Arabidopsis [31-35], KO phenotypes were available for many genes. Using the published phenotype datasets [36-38], features regarding to KO phenotypes of DR genes and other subsets of Arabidopsis single copy genes grouped by their copy numbers in four other plant species, were described in Chapter 1. A majority of DR genes were in lack of any reported KO phenotype, and mutants for some may have never been studied. Therefore, to complete DR genes' KO phenotype dataset, a platform for in-depth functional study, their T-DNA insertional mutants were ordered and phenotyped. Several stress conditions were applied for phenotyping, as conditional phenotypes were extremely diverse and normally invisible thus remain largely uninvestigated. Due to the similar reason, root phenotypes were searched for as well. The stress and root phenotypes from DR genes, along with striking seedling phenotypes shown under normal condition, were described and compared to other single copy genes in Chapter 2.

The single copy status of DR genes among plant species with varying evolutionary speed suggests a relatively fast removal of their duplicated copies, the presence of which may exert negative effect thus was selected against. To assess the direct outcome of gene duplication, addition (AD) lines were made, each containing a transgenic copy of a target gene. As dosage increase is usually the cause of duplication effect [39-41], overexpression (OE) lines, where target genes were overexpressed by the CaMV 35S promoter, were also used to visualize intensified versions of duplication effect, which may be too mild to detect sometimes. The two kinds of mutants were phenotyped similarly as the KO mutants. Phenotypes from AD and OE lines were presented and discussed in Chapter 3.

REFERENCES

1. Zhang J: Evolution by gene duplication: an update. Trends in ecology & evolution 2003, 18:292-298.

2. Gu X, Wang YF, Gu JY: Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nat Genet 2002, 31:205-209.

3. Rizzon C, Ponger L, Gaut BS: Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLoS Comput Biol 2006, 2:e115.

4. Leister D: Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. Trends Genet 2004, 20:116-122.

5. Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH: Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiology 2008, 148:993-1003.

6. Lynch M: The Evolutionary Fate and Consequences of Duplicate Genes. Science 2000, 290:1151-1155.

7. Levy AA, Feldman M: The impact of polyploidy on grass genome evolution. Plant Physiology 2002, 130:1587-1593.

8. Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y: The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol 2006, 7:R43.

9. Dehal P, Boore JL: Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 2005, 3:e314.

10. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al: Ancestral polyploidy in seed plants and angiosperms. Nature 2011, 473:97-100.

11. Bowers JE, Chapman BA, Rong J, Paterson AH: Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 2003, 422:433-438.

12. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al: The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 2007, 449:463-467.

13. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al: The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 2006, 313:1596-1604.

14. Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC: Altered patterns of fractionation and exon deletions in Brassica rapa support a two-step model of paleohexaploidy. Genetics 2012, 190:1563-1574.

15. Blanc G, Wolfe KH: Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 2004, 16:1667-1678.

16. Cui LY, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, et al: Widespread genome duplications throughout the history of flowering plants. Genome Res 2006, 16:738-749.

17. Wang Y, Wang X, Paterson AH: Genome and gene duplications and gene expression divergence: a view from plants. Ann N Y Acad Sci 2012, 1256:1-14.

18. Birchler JA, Riddle NC, Auger DL, Veitia RA: Dosage balance in gene regulation: biological implications. Trends Genet 2005, 21:219-226.

19. Conant GC, Wolfe KH: Turning a hobby into a job: How duplicated genes find new functions. Nature Reviews Genetics 2008, 9:938-950.

20. Freeling M: Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol 2009, 60:433-453.

21. Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M: Gene Loss and Evolutionary Rates Following Whole Genome Duplication in Teleost Fishes Molecular Biology and Evolution 2006, 23:1808-1816.

22. Chapman BA, Bowers JE, Feltus FA, Paterson AH: Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. Proc Natl Acad Sci U S A 2006, 103:2730-2735.

23. Armisen D, Lecharny A, Aubourg S: Unique genes in plants: specificities and conserved features throughout evolution. BMC Evol Biol 2008, 8:280.

24. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH: Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 2008, 18:1944-1954.

25. Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J, dePamphilis CW: Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. BMC Evol Biol 2010, 10:61.

26. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y: Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci U S A 2005, 102:5454-5459.

27. Blanc G, Wolfe KH: Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 2004, 16:1679-1691.

28. Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC: Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. Trends Genet 2006, 22:597-602.

29. De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y: Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci U S A 2013, 110:2898-2903.

30. Edger PP, Pires JC: Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res 2009, 17:699-717.

31. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al: Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 2003, 301:653-657.

32. Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C, et al: A high-throughput Arabidopsis reverse genetics system. Plant Cell 2002, 14:2985-2994.

33. Rosso MG, Li Y, Strizhov N, Reiss B, Dekker K, Weisshaar B: An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol 2003, 53:247-259.

34. Woody ST, Austin-Phillips S, Amasino RM, Krysan PJ: The WiscDsLox T-DNA collection: an arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. Journal of Plant Research 2007, 120:157-165.

35. Feldmann KA: T-DNA Insertion Mutagenesis in Arabidopsis - Mutational Spectrum. Plant Journal 1991, 1:71-82.

36. Kuromori T, Wada T, Kamiya A, Yuguchi M, Yokouchi T, Imura Y, Takabe H, Sakurai T, Akiyama K, Hirayama T, et al: A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of Arabidopsis. Plant Journal 2006, 47:640-651.

37. Hanada K, Kuromori T, Myouga F, Toyoda T, Li WH, Shinozaki K: Evolutionary persistence of functional compensation by duplicate genes in Arabidopsis. Genome Biol Evol 2009, 1:409-414.

38. Lloyd J, Meinke D: A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. Plant Physiology 2012, 158:1115-1129.

39. Murakami T, Garcia CA, Reiter LT, Lupski JR: Charcot-Marie-Tooth disease and related inherited neuropathies. Medicine (Baltimore) 1996, 75:233-250.

40. Singleton A, Gwinn-Hardy K: Parkinson's disease and dementia with Lewy bodies: a difference in dose? Lancet 2004, 364:1105-1107.

41. Hardy J: Amyloid double trouble. Nat Genet 2006, 38:11-12.

CHAPTER 2

LITERATURE REVIEW

BACKGROUND

Gene duplications in the genomes of higher eukaryotes

Duplication of DNA happens on various scales. Whole genome duplication (WGD), followed by gene loss and genome rearrangement, restructures a genome, provides raw material for evolution, and often precedes the emergence of new species. Many higher eukaryotes have experienced whole genome duplication or triplication. There was a WGD in a yeast ancestor before the divergence of S. cerevisiae and K. waltii [1]. Two whole genome duplications are thought to have happened in early vertebrate evolution [2], followed by an independent WGD in the fish lineage [1].

Genome duplications have been remarkably abundant in flowering plants, with WGDs in a seed plant common ancestor [3], an angiosperm common ancestor [3], both monocot and dicot lineages [4-6], and numerous lineage specific events [7-10]. Even 'model' plants with small genomes, such as Arabidopsis and Oryza, are paleo-polyploids, with the footprints of past WGD events evident in their genomes [11, 12]. The prevalence of polyploidy events in plants is thought to contribute to their evolutionary success [13, 14] and to their attributes as crops [15-21], making them suitable, if not the best, models for study of gene fates after duplication.

'Small scale duplications' (SSDs), i.e., of individual genes or small genomic segments, may collectively contribute to the expansion of duplicated genes, thus to the restructuring of metabolic networks. In human, 30%-52% of gene duplication events arise from SSDs [22],

comprising 48% of all human protein coding genes [23]. Tandem duplicates, one type of SSDs, account for 16% and 14% of Arabidopsis and rice genes, respectively [24]. SSDs were considered responsible for the expansion of gene families involved in response to environmental stimuli, such as immune-related families (human), disease-resistance genes (NBS-LRR) [25], and receptor-like kinases [26]. Indeed, their more or less continuous occurrence may make SSDs better suited than (infrequent and episodic) WGDs to provide raw material for a genome to evolve environmental adaptations. For example, the evolution of C4 photosynthesis in cereals largely involved SSDs, although the entire required gene set had been duplicated previously in a WGD with most of the genes subsequently lost.

Following duplication, gene pairs experience one of three general fates

Most gene functional groups show post-duplication gene preservation/loss rates that are indistinguishable from the genome-wide average. Such 'neutral' loss of duplicated genes presumably involves mutations that were not strongly selected against thus were able to accumulate and eventually inactivate the genes [27], closely resembling the 'nonfunctionalization' described [e.g. [28]] as the fate of the vast majority of duplicated genes. Population genetic models suggest that loss of most duplicated genes may happen during the first few million years following duplication [29]. Intrinsic or extrinsic factors may influence gene retention rates. Duplicated genes are presumably identical at 'birth' (duplication), however in a newly polyploid nucleus they may have as much as several million years of independent evolution (allopolyploidy), and one copy usually has a higher probability of loss than the other. For example, different yeast species lose the corresponding copy of their shared single copy genes in most cases [30]. Homeologous chromosomes may be epigenetically 'marked', and lose

genes (or mutate) at different rates [31]. Losses of chromosomes or large segments could result in gene loss that is largely if not wholly independent of gene function.

Genes in some specific functional categories duplicate and reduplicate. Several gene functional groups are preferentially preserved in duplicate [32-37]. Coding regions of genes preserved in duplicate tend to be functionally complex [32], under purifying selection [28, 32], and evolve in concert [38, 39], with tendencies to retain duplicate genes involved in signal transduction and transcription, and lose DNA repair genes [34, 35]. Pfam domain-based groupings reveal heterogeneity in the broad GO categories used in many studies for analysis of gene retention/loss patterns, for example showing one abundant protein-protein interaction domain (LRR) to be usually preserved in duplicate while two less-abundant domains (SET, TPR) are usually restored to singleton state [40]. Regulatory divergence between members of preserved gene pairs may contribute much to morphological complexity [41], perhaps offering important benefits to polyploidized lineages [42]. Classical ideas about one gene copy diverging to new function [neofunctionalization [43, 44]] have more recently been tempered by findings that many duplicated genes may subdivide ancestral functions [subfunctionalization [45]]. Subfunctionalization may be a stepping-stone to neofunctionalization [46]. Retention of some duplicated genes may be an indirect consequence of the fate(s) of nearby genes. In human, for example, pre-existing fixed duplicates may facilitate the fixation of neighboring duplicates [47], resulting in less than average copy number variation for genes adjacent to WGD-duplicates [48].

Other specific genes and gene functional groups show more extensive loss of duplicate copies than the genome-wide average. Duplicated gene loss may not be ‘neutral’ when increased dosage of gene products reduces fitness, such as in defense reactions [49, 50] and stress responses [51, 52], or when genes in dosage balance with each other are not duplicated at the

same time (in the case of SSDs) [53]. Accordingly, some gene functional groups are preserved in duplicate significantly less frequently than the genome-wide average. This observation alone might be viewed as noise – among thousands of functional groups, some must incur more gene loss than others due to random factors. However, the gene functional groups that have incurred greatest loss of duplicated copies are closely correlated following independent duplications in Arabidopsis and rice that are separated by more than 100 million years of evolution, at statistical probabilities that essentially rule out false positives [37]. Multiple genome alignments likewise show some individual genes to have been repeatedly restored to single-copy status following many different genome duplications in independent angiosperm lineages [36, 54]. Repeated restoration of certain genes to singleton status at a greater-than random frequency suggests that an underlying set of principles of molecular evolution may contribute to the fates of gene and genome duplications [37].

Functional importance of single copy vs. duplicated genes

The functional importance of genes can be measured through different approaches, perhaps the most direct one being to evaluate the phenotypic effect caused by rendering a gene nonfunctional. Traditionally, one would think that single copy genes are more prone than duplicated genes to revealing 'knockout' (KO) phenotypes because of a lack of genetic redundancy. KOs of duplicated genes may be masked at least somewhat by one another and less likely to reveal a phenotype, reasoning that is generally supported by large-scale KO mutant screens in yeast [55] and nematode [56]. However, most, if not all, duplicated genes experience DNA sequence divergence over time [57], presumably eroding their ability to provide functional compensation for one another.

Indeed, the ability of duplicate genes to functionally compensate for each other is highly variable among different species [58]. In mouse, 'essential' genes, defined as those for which deletion causes lethality or sterility, occur in proportions that are not significantly different in duplicates versus singletons, based on a sample of genes with phenotypic data [59, 60]. With more data collected from individual experiments, mouse KO mutants for duplicated genes may prove more likely to be lethal than those for single copy genes [58].

Though a number of large-scale mutant screens have been done in Arabidopsis, few have addressed the relationship between phenotypes and gene duplication level. Analysis of phenotypic data from 3871 Arabidopsis KO mutants with transposon or T-DNA insertions plus 1489 published KO phenotypes, found KOs to cause a significantly lower proportion of phenotypic changes in duplicated genes than single copy genes [61]. In another study of 2400 published KO phenotypes, no significant difference was found in the proportions of phenotypes among unique genes, somewhat redundant genes, and highly redundant genes, grouped according to Blastp results [4].

While literature surveys may be inherently biased (with genes showing no KO phenotype perhaps reported less frequently), Hanada et al. [61] found no significant difference between de novo and published evidence. Further, as discussed by Hanada et al [61], the percentage and severity of gene KO phenotypes are related to the sequence divergence of duplicated genes, indicating that duplicated genes could be divided into subgroups with distinct likelihoods of finding phenotypes.

Single copy genes are not a homogenous population

While a broad generalization seems intuitive that single copy genes are more prone to show KO phenotypes than duplicated genes, a variety of subgroups of single copy genes can be

distinguished, for example, lineage-specific singletons (LSSs) and duplication-resistant (DR) genes.

LSSs are defined as genes that have no homolog identified in other species. Though LSSs receive little attention in the literature, they constitute a majority of lineage-specific genes (LSGs) [62]. LSGs have received more attention, and for example are believed to be fast-evolving as indicated by significantly more SNP alleles and higher non-synonymous to synonymous mutation ratios in Arabidopsis [62]. Other characteristics of LSGs include short protein length [63-65], fewer than average exons [62, 63], lack of functional annotation [62, 63], low expression level, limited expression breadth, and stress-related expression patterns [66]. Arabidopsis LSGs with annotation are enriched in secretory proteins [62, 63] and mitochondrion-targeting proteins [62], and are underrepresented in chloroplast-targeting genes. Their expression patterns [64] along with the GO terms of their co-expressed genes [62] associate LSGs with reproduction-related functions and stress-response.

Species specific WGDs, different evolutionary rates and functional divergence might reasonably be expected to impart variation among different plant species in the copy numbers of conserved genes [67]. The distinguishing characteristic of DR genes is that they have little variation in copy number across taxa, being single copy in most plant species [67]. While different DR lists have been published by different research groups based on various criteria, some of their characteristics are remarkably consistent, including more exons/introns than average genes [63, 68], conservation in metazoans [63, 69], a higher than average proportion of plastid-targeting genes and genes functioning in DNA/RNA metabolism [68, 69], and increased purifying selection [63, 69].

RESULTS

Classification of Arabidopsis single copy genes

The number of Arabidopsis single copy genes, as determined by different approaches, varies from 2570 [63] to over 14993 [69] including estimates of 4863 [70], 5460 [68], 5980 [71] and 8499 [4] , each with its own advantage for a particular study. Yet, common features were observed for LSGs (most being LSSs) and DR genes among independent studies [62-65] [68, 69], despite the difference in gene lists. Both being single copy in Arabidopsis, LSSs and DR genes were distinguished by their copy numbers in other plant species, which could serve as a parameter to divide Arabidopsis single copy genes into different groups.

Here we define as single copy those genes that have no non-self blastp hit in the same genome with e-value less than or equal to 1e-10. By this criterion, there are 4136 single copy genes in Arabidopsis, near the low end of the range of other estimates [68, 70, 71]. To determine the copy number of homologs in other species, we considered a relatively slowly evolving monocot genome (Oryza) [72], a eurosid outgroup thought to resemble the ancestral eudicot genome (Vitis)[6, 67], and eudicot genomes with and without recent WGD (Eurosid Ⅰ and Eurosid Ⅱ)(Populus [7] and Carica [73], respectively) [74].

Based on the copy number of their homologs in other species, Arabidopsis single copy genes can be divided into LSSs, with no homologs in all four species; and conserved singletons, with one or more homolog(s) in at least one species. Further, conserved singletons can be divided into DR genes, with only single copy homologs in all four species; singletons with their homologs duplicated in all four species (SD); and others (OT), including singletons remaining fully conserved (OT-F, with homolog(s) in every species) and partly conserved (OT-P, having homolog(s) in at least one other but not all four species) (Table 2.1).

To compare the general features of our singletons to others reported, the GO percentage (=the number of genes with certain GO term/total number of genes) was calculated and compared between each subset and the whole genome (Table S2.1, S2.2). The enrichment of DR genes in DNA or RNA metabolism, plastid and chloroplast as well as the overrepresentation of LSS genes in mitochondrion and unknown proteins (unknown biological processes/molecular functions/cellular components; Table S2.2), were generally consistent with prior reports [68, 69].

The likelihood of showing knockout phenotypes is different among subsets of singletons

To investigate levels and patterns of functional importance among different subsets of single copy genes, we assessed the percentage of KO mutants showing a phenotype in each subset, using three datasets that included a wide range of phenotypes. Two datasets [4, 61] introduced under the section "Functional importance of Single copy vs. Duplicated genes" were utilized. For Hanada et al.'s dataset [61], only the 3871 genes with KO mutants were used in the following analysis, because the 1489 genes with published phenotypes were redundant with Lloyd and Meinke's dataset [4]. In addition, we included Kuromori et al's screen of 4000 DS-insertional mutants [75], containing 3800 genes.

The statistic that we adopted for comparing the likelihood of knockout phenotypes in different gene subsets is 'phenotype percentage' (PP), calculated as the number of genes in a group for which knockout mutants show discernible phenotypes, divided by the total number of sampled genes in the group. For example, in the 4000DS dataset [75], there are 46 DR genes, of which 2 showed mutant phenotypes, thus the phenotype percentage is 4.35% (2/46) (Table S2.3). In the Lloyd and Meinke dataset [4], containing only 2400 phenotypes/genes without any background mutant population, all the genes in each subset are considered 'sampled genes', as it is a genome-wide collection of phenotype data. For example, the phenotype percentage for the

DR subset in Lloyd and Meinke's dataset is 12%, with 49 DR knockouts showing phenotypes among 409 sampled (Table S2.3).

Datasets composed of previously published phenotypes, such as Lloyd and Meinke, tend to have a higher PP, because they incorporate efforts from multiple experiments by a number of research groups, increasing the chance of phenotype discovery. For example, the PP of 4000DS (3%) is lower than those of the other two data sets (6.5% for Hanada, 9% for Lloyd and Meinke), though within the range of results for other Arabidopsis studies [76]. Moreover, there may be a bias in favor of mutants showing phenotypes being published more frequently than those that lack phenotypes.

The overall pattern of PPs among the four subsets of Arabidopsis single-copy genes is similar in the three datasets (Figure 2.1), except for SD in 4000DS: on average (Table S2.3), SD [singletons with their homologs duplicated in four other species] has the highest PP (16.1%), DR shows slightly lower (8.6%) but similar PP to OT (10.6%), and LSS has the lowest PP (0.1%).

SD stands out with the highest average PP, having the highest PP in Hanada and Lloyd and Meinke datasets but not in 4000DS. Only 7 SDs are in the 4000DS population, and none showed a phenotype. However, three have phenotypes in the other two datasets, indicating that the low PP for SD in 4000DS may be partly due to the quality of the mutants (e.g., genes are not entirely knocked out) rather than the small sample size.

LSS also stands out, having the lowest PP across all datasets. LSSs are suspected to have lineage specific functions that are not identified yet, or involved in stress responses (above). Therefore, their low PP could reflect phenotypes that are too subtle to be easily detected, or failure to assess the conditions under which their phenotypes are expressed. LSSs show a striking deficiency of KO mutants, with less than 50% having available homozygous mutants in the

SALK collection of homozygous KO lines (http://signal.salk.edu/cgi-bin/homozygotes.cgi), versus more than 80% for the other three subsets (Figure S2.1). This raises the tantalizing hypothesis that some LSS genes are 'essential', rendering homozygous KO plants inviable.

It was somewhat unexpected that DR genes had phenotype percentages slightly lower than the overall averages for OT and much lower than that of OT-F (Table S2.3). This seems to suggest that copy number regulation (or lack thereof) in other species does not predict the likelihood of finding a KO phenotype in Arabidopsis.

The proportions of essential genes differ among subsets of singletons

In addition to presence/absence, severity of a phenotype is another criterion in evaluating gene importance. Most studies in other organisms (bacteria, yeast, fruit fly, mouse and others) use the proportion of essential genes, i.e. for which KO is lethal or sterile, as a measure of importance for a group of genes [58]. However, in Arabidopsis, most large scale mutant screens have not looked for lethality, as it is not obvious (e.g., requiring silique or seed check through microscopes [77, 78]; or segregation ratio analysis [79]). The absence of homozygous knockout mutants is an indicator of essentiality of genes, but further confirmation is required to rule out alternative reasons.

No lethal phenotype was reported in the 4000DS dataset. The "seed" Category of Hanada et al [61] includes embryo-defective mutants, one major subgroup of lethal mutants, yet these are not separated from other seed phenotypes in this dataset. The Lloyd and Meinke dataset makes lethal phenotypes a discrete category, showing the proportion of essential genes to be distributed similarly to the overall phenotype percentage among the four subsets (Figure 2.2).

Vegetative, reproductive and seed phenotype percentages varied among datasets but were similar to overall phenotype percentages among the four subsets of single copy genes

Phenotypes can be grouped in different ways. For example, there are eight major phenotype groups (seedling, leaves, flowering and growth, stems, branching, flowers (morphology) siliques, seed yield) in the 4000DS dataset, three (Vegetative, reproductive, seed) in Hanada's dataset, and four (essential, morphological, cellular and biochemical, conditional) in Lloyd and Meinke's dataset.

The PPs of particular gene groups may shed light on function as well as functional importance of the underlying genes. For example, a group of genes with most KO phenotypes being seed defective might tend to function in seed development, and may be viewed as more important than a group consisting mainly of leaf phenotypes [61].

For comparison among the three datasets, phenotypes were regrouped according to Hanada et al. [61], the one with the fewest groups (Table 2.2). During regrouping, the main groups of phenotypes in each dataset were not broken into more detailed subsets. While some discrepancies remain in the contents of phenotype groups, several common features are clear.
DRs and OTs have similar PPs in most cases. The exception is again the 4000DS dataset, in which most DR phenotypes are contributed by "seed", while a high portion of OT phenotypes belong to "vegetative" (Figure 2.3). This might imply that DR KO mutant phenotypes are somehow more serious than OT mutants. However, such a conclusion is constrained by limited phenotypic data.

The variation of SD PP among different datasets is the largest in each phenotypic category (Figure 2.3, Figure S2.2). Two SD mutants with vegetative phenotypes in the Hanada dataset also have phenotypes in the Lloyd and Meinke dataset, but one of them became "seed".

Though one gene having multiple phenotype categories is not rare [4, 75], “seed” in Lloyd and Meinke’s dataset corresponds to lethal phenotypes caused by defective embryo/seed, leaving no viable seedling to show any vegetative phenotype. Such conflicting phenotype data might be caused by the lack of confirmation, for example, via complementation test, different mutant allele(s), and gene product measurement. Despite the variations, the SD group still holds the highest PP in at least one dataset in “Seed” and “Vegetative”. In contrast, no reproductive phenotype is found in SD.

Low reproductive PP is a common phenomenon for other subsets and among datasets (Figure 2.3), and even for a much earlier summary of Arabidopsis phenotypes [80]. A possible cause is that most gametophytic defects are lethal and thus are classified as “essential” instead of “reproductive” [4]. Compared to vegetative phenotypes, reproductive phenotypes are considered more severe because of their direct effect on the next generation [61]. The survival of a lineage may depend in part on having only some minimal number of genes for which deletion confers a severe defect. The abundance of “seed” phenotypes, arguably the most severe among the three categories, could possibly be explained by multiple screens targeting them as part of a long-term endeavor to identify all essential genes [78, 81].

As noted above, the PPs of LSSs were still the lowest in all comparisons. LSS phenotypes are only found in Lloyd and Meinke’s dataset. There are three vegetative phenotypes and one seed phenotype for LSS, following the general trend that vegetative phenotypes are the most abundant.

In summary, the PP patterns in these specific phenotype groups resemble those of all phenotypes and lethal phenotypes (above), tending to be lowest in LSS, highest in SD, and

similar between OT and DR classes. Therefore, it seems that such patterns are not caused by the over- or under-representation of a specific type of phenotype in a certain subset.

Though the three datasets studied here have phenotype data for useful numbers of genes in each subset, there are still many genes completely uninvestigated. Among those with their mutants studied, some were merely screened for visually-evident morphology alternations, while others were additionally scrutinized for non-obvious phenotypes such as cellular level abnormalities. Therefore, the absence of KO phenotypes in many cases may be for reasons unrelated to gene function: KO lines might never be assessed under a certain condition or using a particular method. Moreover, the inconsistency in phenotype data among different datasets reflects the problem of mutant quality: phenotypes from different mutants for one gene might be different or even conflicting (eg. lethal or non-lethal), indicating that not all mutants were real knockouts. Collectively, if the scientific community would like to further investigate these subsets of Arabidopsis single copy genes, more needs to be done to improve and expand the phenotype data.

DISCUSSION

Similarly to duplicated genes, single copy genes are a diverse group. Their characteristics can be related to their conservation level as well as the copy number of their homologs in other plant species, according to which we divided Arabidopsis single copy genes into four subsets: LSS, OT, DR, and SD. For subsets DR and LSS, their descriptors (protein size, gene structure, expression pattern and others) have been investigated in detail in other studies. Yet, phenotypes, which are most directly related to gene function, have not previously been summarized and compared among those subsets. We compared phenotype datasets for Arabidopsis gene knockout

mutants to investigate the proportion of genes with phenotypes (phenotype percentage--PP) in each subset.

Despite differences in the three Arabidopsis phenotype datasets with regard to the numbers of genes and phenotypes, it remains an accurate generalization that the conserved groups (OT, DR, SD) carry more important functions than the non-conserved group (LSS), as reflected by higher proportions of knockout mutant phenotypes as well as essential genes. This is consistent with findings in bacteria [82], yeast [83][84] and mouse [60]. Further, the PP is positively correlated with the number of species the genes are conserved in (Figure S2.3), using the most up-to-date Lloyd and Meinke dataset. A gene preserved across long evolutionary time (i.e., in many species) is likely to function in fundamental metabolic processes such as protein synthesis or DNA replication. The disturbance of these processes, through removing such genes from a genome, usually has negative consequences [85].

The most consistent results are found for LSSs, in which the phenotype percentage is consistently low, with only 6 (0.4%) published phenotypes for the 1669 genes through 2012. In contrast to conserved genes, LSSs are most simply explained by recent origin, and thus have not been 'tested' by natural selection over long evolutionary times. An alternate, less parsimonious hypothesis is that LSSs have not survived the test of time, and have been eliminated from most lineages. The lack of phenotypes for their knockout mutants in all three datasets implies that they may not have important functions. Yet, the lack of HM SALK lines (Figure S2.1) compared to other subsets suggests that they might have essential functions. A systematic effort to isolate HM mutants for those without HM SALK lines would test whether this possibility is true. It is also worth investigating whether LSSs frequently have conditional phenotypes, given their possible role in stress response (section "Single copy genes are not a homogenous population").

OT is the largest subset of the four classes of single copy genes, containing 2011 (48.6% of) genes , and can be further divided into ‘partly’ (790) and ‘fully’ conserved single copy genes (1221, without DR and SD). The PP is much higher in fully-conserved (13.5%) than partly-conserved OTs (4.8%) (Table S2.3), making fully-conserved OTs the main contributor to the high PP of OT. This group represents single copy genes with reasonable copy number variation among their homologs.

SD is one extreme subgroup of fully conserved single copy genes, with duplicated homologs in all four plant species, making them a small yet unique population. With high PP and high proportions of annotated function in almost every category (Table S2.1), SD involves relatively well-known genes that are duplicated in many taxa but not in Arabidopsis. The retention of multiple gene copies in many other taxa implies either subfunctionalization or some advantage of multiple gene doses, with the single Arabidopsis copy perhaps being more prone than an average gene to show a KO phenotype. To confirm this assumption, functional analysis of their homologs in other plant species is necessary.

Particularly perplexing are the DR class of fully conserved genes, with little copy number variation. The DR class has lower average PP (8.6%) than other fully conserved single copy genes (13.5%) in the three datasets (Table S2.3). The fact that following many independent genome duplications these genes are recurrently restored to low copy number but not lost [37], seems to imply that they serve some important function(s). Increased knowledge of the functions of DR genes would shed light on the underlying mechanisms that favor their repeated return to single-copy status after duplication.

REFERENCES

1. Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y: The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol 2006, 7:R43.

2. Dehal P, Boore JL: Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 2005, 3:e314.

3. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al: Ancestral polyploidy in seed plants and angiosperms. Nature 2011, 473:97-100.

4. Lloyd J, Meinke D: A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. Plant Physiology 2012, 158:1115-1129.

5. Bowers JE, Chapman BA, Rong J, Paterson AH: Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 2003, 422:433-438.

6. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al: The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 2007, 449:463-467.

7. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al: The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 2006, 313:1596-1604.

8. Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC: Altered patterns of fractionation and exon deletions in Brassica rapa support a two-step model of paleohexaploidy. Genetics 2012, 190:1563-1574.

9. Blanc G, Wolfe KH: Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 2004, 16:1667-1678.

10. Cui LY, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, et al: Widespread genome duplications throughout the history of flowering plants. Genome Res 2006, 16:738-749.

11. Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, et al: The Genomes of Oryza sativa: a history of duplications. PLoS Biol 2005, 3:e38.

12. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 2000, 408:796-815.

13. Fawcett JA, Maere S, Van de Peer Y: Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc Natl Acad Sci U S A 2009, 106:5737-5742.

14. GUO Xin-yi XG-h, ZHANG Yang, HU Wei-min, FAN Long-jiang: Small-Scale Duplications Play a Significant Role in Rice Genome Evolution. RICE SCIENCE 2005, 12:173-178.

15. Renny-Byfield S, Gong L, Gallagher JP, Wendel JF: Persistence of sub-genomes in paleopolyploid cotton after 60 million years of evolution. Mol Biol Evol 2015.

16. Renny-Byfield S, Wendel JF: Doubling down on genomes: polyploidy and crop plants. Am J Bot 2014, 101:1711-1725.

17. Feldman M, Levy AA: Genome Evolution Due to Allopolyploidization in Wheat. Genetics 2012, 192:763-774.

18. Jiang C, Wright RJ, El-Zik KM, Paterson AH: Polyploid formation created unique avenues for response to selection in Gossypium (cotton). Proc Natl Acad Sci U S A 1998, 95:4419-4424.

19. Wright R, Thaxton P, Paterson A, El-Zik K: Polyploid formation in Gossypium has created novel avenues for response to selection for disease resistance. Genetics 1998, 149:1987-1996.

20. Ming R, Liu SC, Moore PH, Irvine JE, Paterson AH: QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res 2001, 11:2075-2084.

21. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J: Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature 2012, 492:423-427.

22. Gu X, Wang YF, Gu JY: Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nat Genet 2002, 31:205-209.

23. Singh PP, Affeldt S, Malaguti G, Isambert H: Human dominant disease genes are enriched in paralogs originating from whole genome duplication. PLoS Comput Biol 2014, 10:e1003754.

24. Rizzon C, Ponger L, Gaut BS: Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLoS Comput Biol 2006, 2:e115.

25. Leister D: Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. Trends Genet 2004, 20:116-122.

26. Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH: Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiology 2008, 148:993-1003.

27. Haldane JBS: The part played by recurrent mutation in evolution. The American Naturalist 1933, 67:5–19.

28. Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M: Gene Loss and Evolutionary Rates Following Whole Genome Duplication in Teleost Fishes Molecular Biology and Evolution 2006, 23:1808-1816.

29. Lynch M, Conery JS: The evolutionary fate and consequences of duplicate genes. Science 2000, 290:1151-1155.

30. Makino T, McLysaght A: Positionally biased gene loss after whole genome duplication: Evidence from human, yeast, and plant. Genome Res 2012, 22:2427-2435.

31. Thomas BC, Pedersen B, Freeling M: Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Research 2006, 16:934-946.

32. Chapman BA, Bowers JE, Feltus FA, Paterson AH: Buffering crucial functions by paleologous duplicated genes may impart cyclicality to angiosperm genome duplication. Proc Natl Acad Sci U S A 2006, 103:2730-2735.

33. Seoighe C, Gehring C: Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. Trends in Genetics 2004, 20:461-464.

34. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y: Modeling gene and genome duplications in eukaryotes. Proceedings of the National Academy of Sciences of the United States of America 2005, 102:5454-5459.

35. Blanc G, Wolfe KH: Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 2004, 16:1679-1691.

36. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH: Unraveling Ancient Hexaploidy through Multiply-aligned Angiosperm Gene Maps. Genome Research 2008, 18:1944-1954.

37. Paterson AH, Chapman BA, Kissinger J, Bowers JE, Feltus FA, Estill J, Marler BS: Convergent retention or loss of gene/domain families following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces, and Tetraodon. Trends in Genetics 2006, 22:597-602.

38. Gao LZ, Innan H: Very low gene duplication rate in the yeast genome. Science 2004, 306:1367-1370.

39. Wang X, Tang H, Bowers JE, Feltus FA, Paterson AH: Extensive concerted evolution of rice paralogs and the road to regaining independence. Genetics 2007, 177:1753-1763.

40. Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC: Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. Trends Genet 2006, 22:597-602.

41. Freeling M, Thomas BC: Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Research 2006, 16:805-814.

42. Comai L: The advantages and disadvantages of being polyploid. Nature Reviews Genetics 2005, 6:836-846.

43. Ohno S: Evolution by gene duplication. Berlin: Springer; 1970.

44. Stephens S: Possible significance of duplications in evolution. Advances in Genetics 1951, 4:247-265.

45. Lynch M, Force A: The probability of duplicate gene preservation by subfunctionalization. Genetics 2000, 154:459-473.

46. He XL, Zhang JZ: Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 2005, 169:1157-1164.

47. Kondrashov FA, Kondrashov AS: Role of selection in fixation of gene duplications. J Theor Biol 2006, 239:141-151.

48. Makino T, McLysaght A, Kawata M: Genome-wide deserts for copy number variation in vertebrates. Nat Commun 2013, 4:2283.

49. Felton GW, Korth KL: Trade-offs between pathogen and herbivore resistance. Curr Opin Plant Biol 2000, 3:309-314.

50. Baldwin IT: Jasmonate-induced responses are costly but benefit plants under attack in native populations. Proc Natl Acad Sci U S A 1998, 95:8113-8118.

51. Yoshida T, Mogami J, Yamaguchi-Shinozaki K: ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. Curr Opin Plant Biol 2014, 21:133-139.

52. Atkinson NJ, Urwin PE: The interaction of plant biotic and abiotic stresses: from genes to the field. J Exp Bot 2012, 63:3523-3543.

53. Edger PP, Pires JC: Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res 2009, 17:699-717.

54. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH: Synteny and colinearity in plant genomes. Science 2008, 320:486-488.

55. Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH: Role of duplicate genes in genetic robustness against null mutations. Nature 2003, 421:63-66.

56. Conant GC, Wagner A: Duplicate genes and robustness to transient gene knock-downs in Caenorhabditis elegans. Proc Biol Sci 2004, 271:89-96.

57. Woollard A: Gene duplications and genetic redundancy in C. elegans. WormBook 2005:1-6.

58. Hannay K, Marcotte EM, Vogel C: Buffering by gene duplicates: an analysis of molecular correlates and evolutionary conservation. BMC Genomics 2008, 9:609.

59. Liang H, Li WH: Gene essentiality, gene duplicability and protein connectivity in human and mouse. Trends Genet 2007, 23:375-378.

60. Liao BY, Zhang JZ: Mouse duplicate genes are as essential as singletons. Trends in Genetics 2007, 23:378-381.

61. Hanada K, Kuromori T, Myouga F, Toyoda T, Li WH, Shinozaki K: Evolutionary persistence of functional compensation by duplicate genes in Arabidopsis. Genome Biol Evol 2009, 1:409-414.

62. Lin H, Moghe G, Ouyang S, Iezzoni A, Shiu SH, Gu X, Buell CR: Comparative analyses reveal distinct sets of lineage-specific genes within Arabidopsis thaliana. BMC Evol Biol 2010, 10:41.

63. Armisen D, Lecharny A, Aubourg S: Unique genes in plants: specificities and conserved features throughout evolution. BMC Evol Biol 2008, 8:280.

64. Yang X, Jawdy S, Tschaplinski TJ, Tuskan GA: Genome-wide identification of lineage-specific genes in Arabidopsis, Oryza and Populus. Genomics 2009, 93:473-480.

65. Campbell MA, Zhu W, Jiang N, Lin H, Ouyang S, Childs KL, Haas BJ, Hamilton JP, Buell CR: Identification and characterization of lineage-specific genes within the Poaceae. Plant Physiology 2007, 145:1311-1322.

66. Donoghue MT, Keshavaiah C, Swamidatta SH, Spillane C: Evolutionary origins of Brassicaceae specific genes in Arabidopsis thaliana. BMC Evol Biol 2011, 11:47.

67. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH: Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 2008, 18:1944-1954.

68. Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J, dePamphilis CW: Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. BMC Evol Biol 2010, 10:61.

69. De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y: Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci U S A 2013, 110:2898-2903.

70. Han F, Peng Y, Xu L, Xiao P: Identification, characterization, and utilization of single copy genes in 29 angiosperm genomes. BMC Genomics 2014, 15:504.

71. Guo YL: Gene family evolution in green plants with emphasis on the origination and evolution of Arabidopsis thaliana genes. Plant Journal 2013, 73:941-951.

72. Matsumoto T, Wu JZ, Kanamori H, Katayose Y, Fujisawa M, Namiki N, Mizuno H, Yamamoto K, Antonio BA, Baba T, et al: The map-based sequence of the rice genome. Nature 2005, 436:793-800.

73. Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, et al: The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature 2008, 452:991-996.

74. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH: Synteny and collinearity in plant genomes. Science 2008, 320:486-488.

75. Kuromori T, Wada T, Kamiya A, Yuguchi M, Yokouchi T, Imura Y, Takabe H, Sakurai T, Akiyama K, Hirayama T, et al: A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of Arabidopsis. Plant Journal 2006, 47:640-651.

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

77. Meinke DW, Sussex IM: Embryo-lethal mutants of Arabidopsis thaliana: a model system for genetic analysis of plant embryo development. Dev Biol 1979, 72:50-61.

78. McElver J, Tzafrir I, Aux G, Rogers R, Ashby C, Smith K, Thomas C, Schetter A, Zhou Q, Cushman MA, et al: Insertional mutagenesis of genes required for seed development in Arabidopsis thaliana. Genetics 2001, 159:1751-1763.

79. Bryant N, Lloyd J, Sweeney C, Myouga F, Meinke D: Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. Plant Physiology 2011, 155:1678-1689.

80. Meinke DW, Meinke LK, Showalter TC, Schissel AM, Mueller LA, Tzafrir I: A sequence-based map of Arabidopsis genes with mutant phenotypes. Plant Physiology 2003, 131:409-418.

81. Meinke D, Muralla R, Sweeney C, Dickerman A: Identifying essential genes in Arabidopsis thaliana. Trends Plant Sci 2008, 13:483-491.

82. Jordan IK, Rogozin IB, Wolf YI, Koonin EV: Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome Res 2002, 12:962-968.

83. Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, Eisen MB, Feldman MW: Functional genomic analysis of the rates of protein evolution. Proc Natl Acad Sci U S A 2005, 102:5483-5488.

84. Zhang JZ, He XL: Significant impact of protein dispensability on the instantaneous rate of protein evolution. Mol Biol Evol 2005, 22:1147-1155.

85. Tsukaya H, Byrne ME, Horiguchi G, Sugiyama M, Van Lijsebettens M, Lenhard M: How do 'housekeeping' genes control organogenesis?-unexpected new findings on the role of housekeeping genes in cell and organ differentiation. Journal of Plant Research 2013, 126:3-15.

Table 2.1 Classification of Arabidopsis single copy genes.

Groups	Subsets	Copy number in five plant species				
		Arabidopsis	Oryza	Vitis	Carica	Populus
Non-conserved	LSS	1	0	0	0	0
Partially conserved	OT-P*	1	≥0	≥0	≥0	≥0
Fully-conserved	DR	1	1	1	1	1
	SD	>1	>1	>1	>1	>1
	OT-F*	1	>0	>0	>0	>0

*OT-P and OT-F can be combined together as OT

Table 2.2 Phenotype re-grouping in the three datasets.

Common grouping	Hanada	4000DS[1]	Lloyd and Meinke[2]
vegetative	altered germination, seedling, root, rosette, or transition to flowering [83]	seedling (agar plate),leaves (soil), flowering and growth (bolting time, plant height, growth) , stems (color/size/others), branching	Class "V" (germination, cotyledon, hypocotyl, pigment, growth rate, size, root, leaf, stem, shoot structure, miscellaneous shoot)
reproductive	abnormal flower, silique, seed coat, or gamete [83]	flowers (structure), siliques (shape), seed yield (yield)	Class "R" (flower, silique morphology, ovule, pollen, sterility, seed, seed coat)
seed	embryo- or endosperm-defective or seed pigment mutant [83]	seed yield (seed shape)	Class "S" (embryo defective, seed defective)

1. Genes in more than one category were re-grouped according to the most severe category, based on phenotype severity: Seed > Reproductive > Vegetative.

2. The class and description is from Lloyd and Meinke (2012) supplemental table S1.

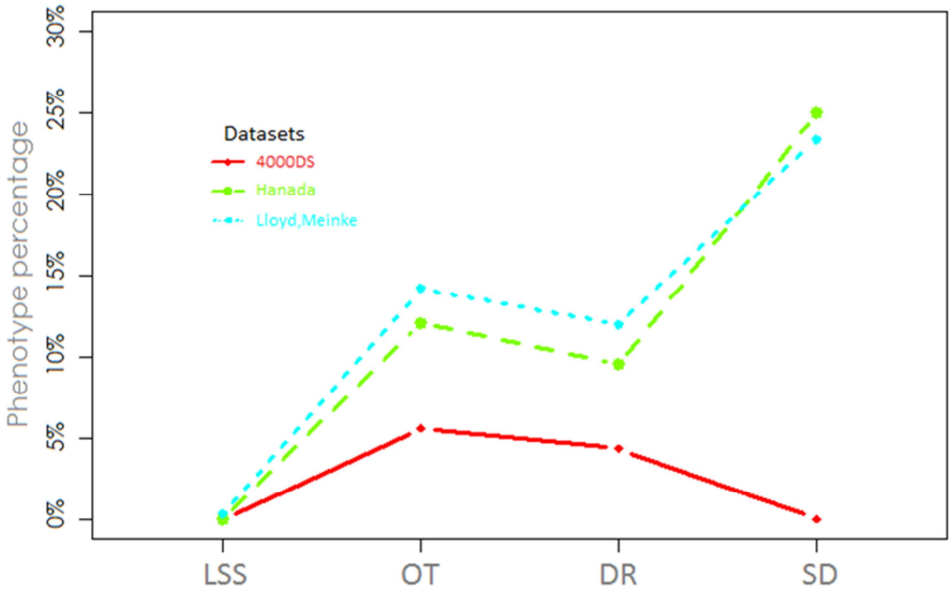

Figure 2.1 Phenotype percentages of LSS, OT, DR and SD genes in three datasets.

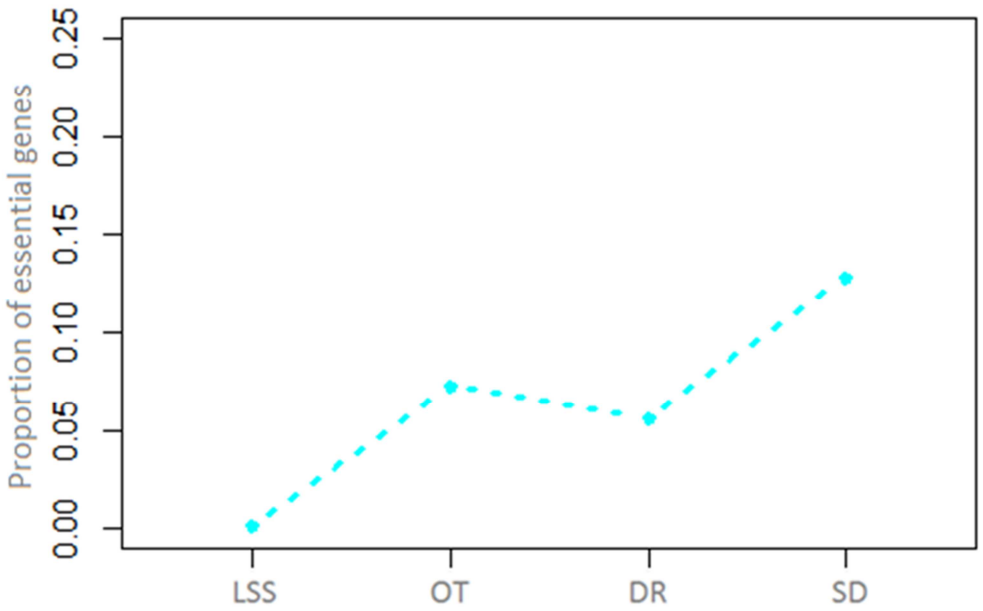

Figure 2.2 The proportion of essential LSS, OT, DR and SD genes in the Lloyd and Meinke dataset.

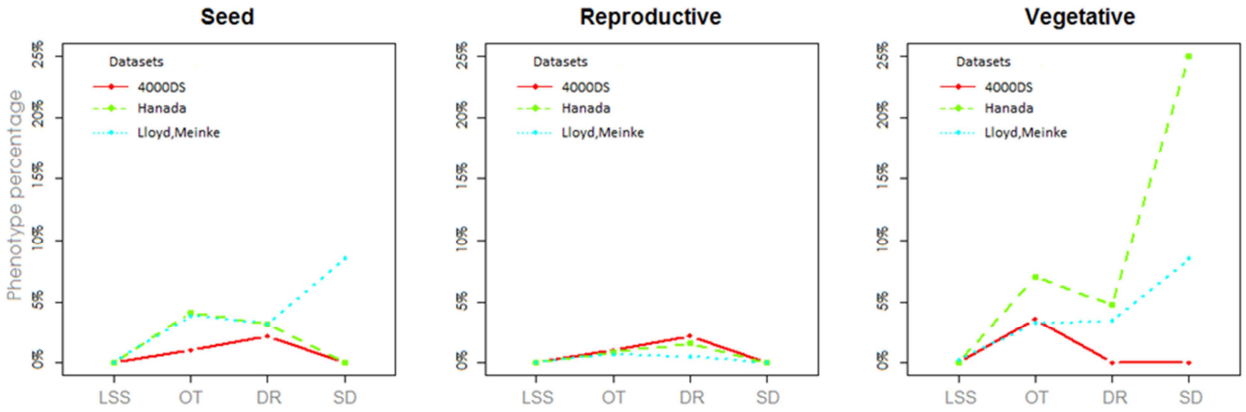

Figure 2.3 Comparison of seed, reproductive and vegetative phenotype percentages.

CHAPTER 3

PHENOME ANALYSIS OF DUPLICATION-RESISTANT GENES USING ARABIDOPSIS T-DNA INSERTIONAL MUTANTS

Chengbo Zhou, Kenneth Feldmann, Kathryn Millward and Andrew H. Paterson. To be submitted to Physiologia Plantarum.

ABSTRACT

Genes conserved among multiple plant species are expected to have basic and important functions, thus being a vital component of plant genomes. Among conserved plant genes, 'duplication-resistant' (DR) genes stand out because of their strong tendency to return to single-copy dosage despite recurrent, independent duplication events. Their dosage sensitivity strongly suggests that DR genes experience different selective constraints than other conserved genes. The recurring removal of duplicated gene copies is a rare and as yet not-well-explained fate of gene duplication. This fate shows non-random association with gene function, as implied by enriched GO terms in DR genes, making functional study of DR genes a path by which to decipher the basis of duplication resistance. In this study, we assessed Arabidopsis DR genes' function via their knockout phenotypes. T-DNA insertional mutants of 179 DR genes and 64 non-DR singletons were screened for root and stress-triggered phenotypes. Compared to single-copy-random (SCR) genes, most with duplicated homologs in other plant species, T-DNA insertional mutants of DR genes were less likely to exhibit phenotypes than SCR genes under all tested stress conditions except salt. The overrepresentation of salt phenotypes, together with underrepresentation of ABA and cold sensitive phenotypes in the DR group were also seen in comparison with other non-DR singletons (nDR), which had been early DR candidates later found to be non-DR by analysis of more genomes. On the other hand, severe phenotypes, including visible seedling alternations as well as phenotypes indicative of homozygous lethality, occurred more frequently in the DR group than the controls. Distinctive phenotypic features of DR genes may provide further information toward learning the selective constraints that repeatedly return these genes to single-copy dosage. The phenotypes discovered and mutants characterized herein, serve as a foundation for in-depth functional study. Better understanding of

DR genes' function would contribute greatly to plant functional genomic analysis, and provide insight into the evolutionary basis of their persistent single copy status.

ABBREVIATIONS

WGD (whole genome duplication); HM (homozygous); HT (heterozygous); WT (wildtype); PP (phenotype percentage)

INTRODUCTION

The evolutionary history of plant genomes is characterized by often-recurrent whole genome duplications (WGDs) as well as numerous small-scale duplications of individual genes or small genomic regions (SSDs) [1, 2]. In angiosperms, the original set of genes in the last universal common ancestor (LUCA) has been at least duplicated (without exception) and re-duplicated in most species. Most duplicated gene copies were either non-functionalized due to accumulated mutations [3], or became new genes without recognizable sequence resemblance to the original copy [3]. In either case, these once-duplicated genes reverted to single copy dosage. In contrast, multiple copies were retained for a minority of genes, forming gene families of various sizes.

Due to diverse processes and some random factors, the copy numbers of conserved genes varies among species. Yet, there exist appreciable numbers of genes that are widely conserved only in single copies or low copy numbers. These genes serve as valuable markers for phylogenic study [4, 5]. Moreover, their recurring return to low copy number following recurring duplication events, which has been shown to be non-random [6], may be meaningful for fitness of a genotype, though the underlying selective mechanisms remain unclear.

Genes with persistent single copy status have been assessed by several research groups [4, 7, 8]. The names given to these genes vary, thus to avoid confusion, they are referred to

herein as ‘duplication-resistant (DR) genes’. The lists of DR candidates identified by each research group differ from one another, due to the study of different species and use of different analytical methods. Nevertheless, much congruence was found in their characteristics. The number of DR genes is much higher than expected under the assumption of random loss, indicating that selective forces contribute to their single copy status. Indeed, purifying selection acting on these genes is implied by evidence including low dN/dS rates [7] and low Ka values [8]. Moreover, DR genes are structurally complex [4, 7], containing greater than average numbers of exons/introns; and functionally biased, being over-represented in GO terms such as DNA or RNA metabolism, chloroplast and plastid [4, 8].

Insights into causes of ‘duplication resistance’ (copy number regulation) have been inspired by functional features of DR genes. For example, enrichment for chloroplast-located DR proteins raised the hypothesis that dosage balance between these DR proteins and chloroplast-genome-encoded interaction partners would be disturbed by nuclear but not chloroplast genome duplication, resulting in selection against unilateral duplication [2]. Reasonable as it is, the theory is limited to a small subset of DR genes. Duplication resistance of genes encoding subunits of macromolecular complexes, for example, is not well explained by the dosage balance hypothesis [9], according to which these genes should be WGD duplicates instead of persistent singletons [2]. In this case, it may not be increased dosage that causes trouble, but rather mutated proteins produced by one of the duplicated copies that compete with the wild type protein in binding to other components, and form inactive complexes [8]. Experimental confirmation of hypotheses about causes of duplication resistance would be valuable, and motivates functional study in deciphering DR genes’ copy number syndromes.

More definitive than predicted functions is empirical observation of phenotypic changes resulting from silencing a gene, directly reflecting its function. For example, a plant lacking proteins functioning in a critical metabolic pathway may experience premature death, while the absence of enzymes involved in pigment synthesis may merely change color intensity or pattern. Therefore, to study gene function, mutants with disturbance of few, preferably one, gene(s) per line were made and their phenotypes searched [10].

Arabidopsis mutant resources [10-14] and their phenotypic descriptions [15-17] are abundant, and from the published knockout (KO) phenotypes some phenotypic features of DR genes were extracted (Chapter 2). Compared to other fully conserved singletons (OT-F), Arabidopsis DR genes were less likely to either reveal knockout phenotypes or be essential, indicating that functions carried by DR genes might not be as important as those of OT-F genes (Chapter 2). The relatively low phenotype percentage (PP) of DR genes was seen in all three phenotype categories, namely, Seed, Vegetative and Reproductive (Chapter 2). The results suggest that duplication resistance might not be directly related to functional importance, nor correlated with a particular phenotype category. However, these observations may change with new phenotype data, especially given that as many as 88% (360/409) of DR genes lacked reported knockout phenotypes (Chapter 2). Among those without phenotypes are genes whose mutants have not been investigated, implying the possibility of uncovering new phenotypes.

The potential for phenotype discovery is particularly large in conditional phenotypes, which require a specific environment to discern, because of the vast range of conditions that plants must adapt to in nature. For example, different concentrations of salt as well as a variety of measurements (germination rate, root length, leaf numbers and others) could be applied in a salt screen. Environmental conditions affecting crop production have been of continuing interest,

including drought [18], salinity [19], and cold [20]. Candidate genes conferring stress tolerance have been important targets for development of molecular tools for crop improvement [21]. Though numerous stress screens involving thousands of genes have been performed in Arabidopsis [22-27], none have focused on DR genes, implying that many DR mutants have not been tested under stress yet.

In this study, T-DNA insertional mutants (SALK lines) [10] of Arabidopsis DR genes, and non-DR but conserved singletons as the control group, were assessed in a variety of phenotype screens, focusing on four stress conditions that sought to emulate major stresses threatening crop productivity, including drought (ABA, mannitol), salt, cold and heat, and high sucrose condition. The original DR candidates include those defined in Chapter 2, as well as conserved single copy genes (between Arabidopsis and Rice) with DR domains, while the original control group (SCR) contains mostly SD and OT-F genes. Later adjustment (detailed in methods) resulted in re-classifying some genes from the DR group to the 'nDR' control group. The final sample was limited to qualified DR candidates with available SALK lines, preferably with insertions in exons or introns. The five stresses were chosen because they were relatively well studied [28-32], thus easier to connect DR function with known stress response pathways or networks. Together with the highly conserved nature of the test genes, knowledge gained through this study may contribute to molecular breeding of stress-tolerant crops.

In addition to stress-response candidates, root length of mutants compared to wildtype (WT) Col0 plants was scored on vertical plates. Though not intentionally screened, abnormal seedlings were noted for plants grown in growth chambers and referred to as soil growth phenotypes. Phenotypes indicating lethality for lines that lacked homozygous (HM) plants for T-DNA insertion, such as dying seedlings and abnormal seeds in HT siliques, were classified as

potentially lethal (PL) phenotypes. Together, these phenotypes exhibited a relatively complete picture of the outcome of DR genes' knockout. Through comparing the phenotype data between DR genes and other singletons, phenotypic features that may be associated with duplication resistance were identified.

MATERIALS AND METHODS

DR, SCR and nDR gene lists

Arabidopsis genes were blasted against one another as well as those of four other plant genomes (*Carica papaya, Populus trichocarpa, Vitis vinifera, and Oryza sativa*). Genes with exactly one blastp hit (e value less than or equal to 1e-10) in all five plant genomes were considered DR candidates, forming one DR subgroup named 'strict singleton (SS)', which was further divided into non-syntenic and syntenic classes according to gene location [33, 34]. The other subgroup, 'protein domain (PD)', contains genes conserved as single copy in both Arabidopsis and Oryza and with protein functional (Pfam) domains that are significantly enriched in singletons [6].

Copy numbers of the DR candidates in 21 genomes (*Manihot esculenta, Ricinus communis, Populus trichocarpa, Medicago truncatula, Glycine max, Cucumis sativus, Prunus persica, Arabidopsis thaliana|Arabidopsis lyrata, Carica papaya, Citrus sinensis|Citrus clementina, Eucalyptus grandis, Vitis vinifera, Mimulus guttatus, Aquilegia coerulea, Sorghum bicolor|Setaria italica|Oryza sativa|Brachypodium distachyon, Zea mays*) were assessed on Phytozome 7.0. To reduce redundancy and make the dataset easier to analyze, multiple counts were 'collapsed' into a single count for species that experienced the same sets of WGDs or whole genome triplications (WGTs). Particularly, average counts were taken for two arabidopsis, four diploid grass, and two citrus species, with the only exception that the count cannot be zero

as long as one lineage is nonzero. For example, if the count was '1' in *Arabidopsis thaliana* and '0' in *Arabidopsis lyrata*, the collapsed count would be '1' which assumes that *A. lyrata* contains incomplete data. After this collapse, 16 lineages remain. DR candidates without homologs in two or more genomes were dropped from the initial list, and were named 'nDR', while the remaining 303 genes were considered DR candidates (Table S3.1).

Other Arabidopsis singletons, which were not single copy in at least one of the other four species, were named SCR (single copy random) genes, as their variable single copy status suggested random factors [2]. The original control group (SCR) contained 24 randomly selected Arabidopsis single copy genes conserved in all four other species ('SD' or 'OT-F' in Chapter 2), and three Arabidopsis single copy genes conserved in at least one but not all four other species ('OT-P' in Chapter 2) (Table S3.1).

Genotyping

Seeds ordered from Arabidopsis Biological Resource Center (ABRC) were sown on soil in sets of six, from which a pooled leaf sample was collected for DNA extraction. Plant genome DNA was extracted by cetyltrimethylammonium bromide (CTAB)-based Method [35] with modifications.

PCR was done on the pooled sample for each mutant line. Primers were found on http://signal.salk.edu/tdnaprimers.2.html, where the genotype scoring method could be found. Instead of conducting the standard protocol, which in our case often resulted in nonspecific amplification, two PCR reactions were done with the primer combination (LP+RP) and (LB+RP) (http://signal.salk.edu/tdnaprimers.2.html), respectively. The genotypes corresponding to different banding patterns were shown in Figure M3.1, with WT standing for 'wildtype', HM representing 'homozygous' and HT for 'heterozygous' for the T-DNA insertion.

Seeds harvested from homozygous (HM) plants, with identical T-DNA insertion in both alleles of the target gene, were directly used in phenotype screens. For heterozygous (HT) samples, with one knockout allele and one functional allele of the target gene, 12 seeds were planted and the resulting seedlings were genotyped to seek one or more HM plant(s), from which homozygous (selfed) seeds were collected for phenotype screens.

Lethality test

For lines lacking HM plants, young fruits of HT parents (if available) were assayed under a dissecting microscope to identify abnormally developed seeds.

Planting

Soil was put into pots sitting in flats and totally saturated with water before planting, with 2-3 cm water remaining in the flat. Once the seeds were put in pots, each flat was covered with a transparent plastic lid, and kept at 4o C for 3 days to reduce dormancy and improve synchrony of germination. After moving the flats to a growth chamber (22°C, 16:8 light: dark cycle, 100µ Einsteins), the lid was kept on for several days. After removing the lids, flats were bottom watered when the soil became noticeably dry.

Stress screens, root screens and Kanamycin test

Stress screens include those involving exposure to ABA, mannitol, cold growth, cold germination, heat, heat recovery, salt and sucrose. These were done on petri dishes with specific media. All media contained exogenous sucrose (5g/L) except for that used in the sucrose screen. The basic medium used for root, etiolation, cold, heat and Kanamycin (20mg/L and 50mg/L) screens was 0.5xMS (Murashige and Skoog [36]). The basic medium used for mannitol (375mM), ABA (1.5µM), sucrose (300mM) and salt (125mM) was 1xMS. Both ABA and

mannitol media were supplemented with MES (2-(N-morpholino)ethanesulfonic acid) hydrate (5g/L), which serves as a buffer to stabilize pH.

Seed Sterilization

Seeds used in stress, root and kanamycin screens were sterilized by a mix of clorox and water (volume 1:1) with 2 drops of Triton X-100 per 100 ml solution. In a 1.5ml Eppendorf tube, seeds were mixed well with the clorox solution and submerged for 8 minutes, then rinsed with 1.5 ml sterile water three times. Sterile water (1ml) was added after the rinses to enable planting seeds on the plates with a pipette. Soft agar (0.1%) was added instead if the seeds needed to be stored (at 4°C) for more than three days.

Plate arrangement

ABA, mannitol, sucrose, salt, cold growth, heat/heat recovery screens: For each mutant line and controls (Col0), 10-15 seeds were plated evenly in a straight line parallel to one side of a square plate (15 x 100 x 100mm). Six to eight lines (that may include the WT) were screened per plate (Figure M3.2).

Root screens: For each mutant line as well as the control (Col0), 10-15 seeds were plated evenly in a straight line parallel to one side of a square plate (15 x 100 x 100mm). The seeds were about 1 cm from the top of the plate when the plate was positioned vertically. Two mutant lines were plated side by side, each taking one half of the plate's side length (Figure M3.2).

Cold germination screen: Square plates (15 x 100 x 100mm) with 6x6 grids on the bottoms were used. On each plate, seeds from two replicates of 17 mutant lines as well as controls were plated randomly, with each replicate occupying one grid. While plating, one drop of seed suspension was placed in each grid, containing 20-30 seeds.

Heat/Heat recovery and cold growth screens (early trials): Some of the earliest screens were done on round plates (100 x 15 mm). Three mutant lines or two mutant lines with one control were screened per plate, each occupying a 120 degree sector, where 20-30 seeds were plated evenly (Figure M3.3).

Kanamycin screens: Four lines (or seeds from four individual plants) were inoculated per plate, each taking ¼ of the plate area. Around 20~30 seeds were plated evenly for each line (or individual plant) (Figure M3.3).

Phenotyping

If not noted otherwise below, plates were first photographed at the 14th day after plating, and again every week for 2~4 weeks; growth conditions were 22°C, 100 µEinsteins light intensity, 16:8 hour light: dark cycle.

After plating seeds, ABA and sucrose plates were photographed at the 7th day; Root plates were placed at 4o C for 3 days before being kept vertical in a growth chamber; Cold germination plates were put in a 10o C growth chamber; Cold growth plates were put in a 22°C growth chamber for 7 days before transfer to a 10°C growth chamber; Heat plates were wrapped with vent tape instead of parafilm (which would melt at 34°C) and put into a 34°C growth chamber; Heat recovery plates stayed at 34°C for 7 days before being transferred to a 22°C growth chamber, where they were photographed after 7 days; Kanamycin scoring started at the 10th~12th day in a 22°C growth chamber, and scores were directly recorded in a notebook (not photographed).

Phenotype scoring

Phenotype scoring was done using photos of the plates for all screens except Kanamycin. One score was assigned to each mutant line per plate, based on the average performance of 10-30

individuals in that mutant line compared to the control. In most cases, the score is an integer ranging from 0 to 5, with 0 representing ‘no germination’, 1 for ‘very sensitive’, 2 for ‘somewhat sensitive’, 3 for ‘normal’ (similar to wild type), 4 for ‘somewhat tolerant’, and 5 for ‘very tolerant’. Sensitive mutants are smaller than the control while tolerant mutants are bigger (Figure M3.4). For a few cases in which the most striking difference was color, yellow was considered to be ‘sensitive’ while ‘green’ was ‘tolerant’. When only a few (<50%) individuals within a mutant line showed phenotypes, the score ‘seg’ (for segregating) was assigned (Figure M3.4).

In addition to depicting the performance of seedlings, root scores mainly represented root length, with 1 being ‘short’, 2 being ‘slightly short’, 3 being normal, 4 being ‘slightly long’, and 5 being ‘long’ (Figure M3.5). The 0 still stands for ‘no germination’ and ‘seg’ means only a few plants within a mutant line showed a phenotype.

The average of all non-seg scores of a particular mutant was calculated per trial, representing the overall performance during the 2-4 week growth period. If there was only one non-seg score, it would be recorded as ‘single read’ instead of a number. The ‘single read’ was not considered a valid phenotypic score if no other trial was done on this mutant or no other score was assigned because it raised the possibility of various artifactual outcomes, for example, that the plate was contaminated. For a final score of one trial, mutants with average scores from 2.5-3.5 were re-assigned with score 3; mutants with average scores lower than 2.5 were re-assigned with 1 (if there was no 3) or 2 (if there was one 3); those with average scores higher than 3.5 were re-assigned with 5 (no 3) or 4 (one 3) (Figure M3.6).

Trial scores were combined together to create a final score. Similar scores were interpreted as an increase in the level of confidence with which sensitivity or tolerance was

inferred. For example, if a mutant was assigned score 2 in both trials, then the final score would be 1. In contrast, divergent scores were interpreted as a decrease in the level of confidence with which sensitivity or tolerance was inferred. For example, if a mutant was 1 in one trial and 3 in the other, the final score would be 2. The same rule applied when there were three or more trials. For example, if a mutant was 1 in one trial and 3 in the other two trials, then the final score would be 3 (Figure M3.6).

No numeric score was assigned in the cold germination screen. Instead, if the two replicates on the same plates (Figure M3.2) had similar phenotypes that were each different from WT, the mutant was scored according to their phenotypes, including 'CS' (cold sensitive), 'CT' (cold tolerant), and 'YG' (yellow green). Mutants tested in only one trial received the score 'SR' (single read).

For the Kanamycin screen, green plants with visible roots were scored as 'R', representing 'resistant', while pale-green plants with very shallow roots were scored as 'S', representing 'sensitive'. If both 'R' and 'S' appeared in one line, the mutant was scored as 'seg'.

Validation of mild phenotypes

To validate the phenotypes we observed, the sizes of some randomly chosen mutants with sensitive or tolerant phenotypes as well as their corresponding controls were measured and compared. As we were using photos for measurement, the relative length/width compared to a fixed reference length was necessary to enable comparisons among different plates. As most of the plates we used (especially in the later stage of this study) have the 6x6 grid patterned bottom, the length of one grid was chosen as the reference length. Size of a single plant was calculated by the product of the relative length and relative width. T-tests were applied to assess statistical significance of the difference in size between the chosen mutant and the control.

RESULTS

Genotyping and genotype groups

A total of 250 SALK lines were genotyped for T-DNA insertions, including 181 DR, 37 nDR and 32 SCR lines (Table 3.1). There are two different SALK lines for two DR genes, one nDR gene and three SCR genes (Table S3.2; green cells in column 'gene'), only one of which were involved in the root and stress screens though all were grown for seed collection. The SALK lines were divided into three genotype groups (Table 3.1). One or more homozygous (HM) knockout plant(s) were identified in 185 (74%) SALK lines, forming the HM group. Both heterozygous (HT) and wild type (WT) plants were present in 27 (11%) SALK lines without homozygous plants, constituting the potentially lethal (PL) group [for which homozygous knockout mutants may be inviable]. The wild type (WT) group contains 34 (14%) SALK lines in which all plants were wild type. Others (OT) include 4 lines (2%) with only HT plants or plants with unknown genotypes.

For WT and OT groups, phenotypes may not be associated with target genes. In WT lines, the T-DNA insertions were not at their expected places, with their real locations unknown. Both DR and control groups included WT lines, which were probably caused by incorrect mapping of T-DNA locations [37]. Two WT lines (SALK_081322 and SALK_040220C) were also recorded as WT in a chloroplast function database (http://rarge-v2.psc.riken.jp/chloroplast/), agreeing with our results. DR knockout mutants have the highest percentage of WT (16%), similar to what was reported in the chloroplast functional database [37], while the percentage is lower in nDR (8%) and SCR (9%).

Lines with only HT plants may indicate abnormal insertion events. For example, perhaps only a short segment of the T-DNA left border was inserted instead of the whole region, and

therefore did not block the (LP+RP) reaction. In this case, the banding patterns of HT and HM would be indistinguishable. The small insertion might exert no influence on transcription thus a full-length mRNA could still be produced. Therefore, the HT lines were not considered knockout mutants.

One (Mutant 21) of the two lines with unknown genotype did not have amplification of the wild type allele in the control (wildtype Col0 DNA). This failure of amplification might indicate a problem with the LP primer. The (LB+RP) primer combination amplified in the pooled sample, indicating that nothing was wrong with LB or RP primers and the sample DNA. The control DNA worked well for (LP+RP) primers of other SALK lines using the same PCR mixture and PCR machine, eliminating the potential problem of the control DNA and other PCR-related factors. Mutant 21 was kanamycin-resistant and not segregating, thus could possibly be HM. As to the other line (Mutant 3_30; SALK_067058C), which was HM in a leaf phenomics study [38], no PCR product was detected.

Potentially lethal (PL) SALK lines

DR genes have a higher percentage of potentially lethal SALK lines (12%) than both SD (9%) and nDR (8%). Abnormal seedlings were detected in 63% of PL lines grown on soil (Figure 3.1c,d), and defective seeds were found in siliques of heterozygous plants in 41% of PL lines (Figure 3.1a,b; Table S3.3), supporting the lethality of the absent HM genotype. About 30% of PL genes were confirmed as essential to plant survival by other research groups, and 15% have published morphological phenotypes (Table S3.3).

Evidence of HM knockout lethality was found in root and other screens as well. Mutant 45 segregated for small, red seedlings in the sucrose screen (Figure 3.2). For mutant 56, small

grey seedlings were found in the ABA screen, and a few tiny white seedlings occurred in the mannitol screen, in agreement with its reported lethal phenotype [39](Figure 3.2).

Of the 20 PL lines tested in root or stress screens, most were scored as 3 (no difference from WT), agreeing with the fact that these plants were HT or WT for T-DNA insertions. The expected 'seg' phenotype rarely appeared, perhaps because the number of seeds for each line was too small to be confident of including HM seed. Though two or three of 10-14 seeds from selfed HT plants were expected to be HM, the frequency of HM plants might be less than 25%. Consistent with this possibility, no line showed 'seg' phenotypes in more than two screens, even for the confirmed lethal mutants (eg. Mutant 56).

In two PL lines, phenotypes were observed in most but not all plants, while only one quarter or fewer of seedlings were expected to be HM. The observed phenotype in one mutant (mutant 3_23; SALK_080169C) is similar to its reported phenotype [40], yet it is unknown why so many seedlings showed phenotypes. These lines perhaps segregated at atypical ratios or the T-DNA insertion caused a dominant phenotype, with their HT plants revealing phenotypes as well.

Interestingly, a few tolerant phenotypes were found, especially in the ABA screen. Most of these phenotypes were mild (score 4) except for one in sucrose (Mutant 149). Rather than being a mutant effect, the slightly larger size might be within the normal range of wild type Arabidopsis plants, or due to unknown environmental factors.

Homozygous (HM) SALK lines: Soil growth phenotypes

As many as 14% (18/130) of DR HM mutants, but no SCR or nDR mutants, showed phenotypes under normal growth conditions while growing plants for fresh seeds or genotyping (Table S3.4)(Table 3.2). Most of the phenotypes were described as 'small' (Table S3.4), and

could be associated with many factors including late germination, retarded growth, dwarfism, seedling lethality, watering or seed quality. Therefore, they were grouped based upon more detailed descriptions in addition to 'small'. For example, a mutant with description 'small and yellow green' would be put in the 'pigment' category. There are seven categories in total for soil growth phenotypes of DR HM mutants (Table 3.2).

Phenotypes in the 'lethal' category may not derive from the target genes. Small, dying plants did not survive to the 6-8 leaf stage when leaves were collected for DNA extraction, therefore HM plants were not identified in PCR. Instead, lethal phenotypes might be caused by T-DNA insertion in non-target genes, or by environmental factors. Interestingly, one of the five genes in the 'lethal' category was reported to be essential: Its knockout mutant has abnormal chloroplasts thus cannot survive without exogenous sucrose [41]. The reported phenotype belongs to a mutant (Feldmann T-DNA line, no. 2755) different from ours (SALK_021962C), which experienced PCR problems, therefore the only putative HM plant might not be truly HM.

The distribution of phenotypes among other categories was not even, with the largest non-lethal category being 'pigment', containing four lines (Table 3.2). The enrichment of pigment mutants was not due to a high percentage (32.3%; 98/303) of chloroplast-targeting proteins, the elimination of which frequently results in a pigment defect [42], as only one out of the four proteins in 'pigment' is located in the chloroplast.

All three DR genes in the 'leaf' category were related to RNA metabolism (rRNA processing, tRNA/rRNA methyltransferase, RNA modification), a functional term occurring at a low frequency (4.2%, as 17 of the 409 genes with HM SALK lines in PhenoLeaf belonged to the GO category 'DNA or RNA metabolism') in leaf phenotypes [38]. Two of the three proteins were chloroplast-targeting, and both were in PhenoLeaf [38]. SALK_005531C (Mutant 40,

AT5G15390) had reduced trichomes in our study, yet was no different from WT in PhenoLeaf, where trichomes may not be scored. For gene AT4G21770, SALK_039518C (Mutant 1_49) was reported to have no phenotype under normal conditions, as we also observed. Small, curled leaves were found in another SALK line (SALK_149232C) for the same gene (Figure 3.3d). The T-DNA insertions were at different locations in the two mutants, with SALK_149232C in an intron and SALK_039518C in the 5'UTR, presumably contributing to their morphological difference.

Though most soil growth phenotypes were manifested in a few rather than all plants in a particular mutant line, 'seg' depicts those with about 25% of seedlings showing phenotypes. Two of the three 'seg' mutants (135 and 177) experienced PCR problems: amplification was poor in mutant 135, whereas in mutant 177 the (LP+RP) reaction failed in one trial. As a result, their HM plant(s) might in fact be HT. In addition, both mutants have published knockout phenotypes. Mutant 135 corresponds to a known essential gene (AT3G20070) [43, 44], so the soil growth phenotype (slightly smaller stature and fewer siliques compared to WT) may not be from HM plants. Instead, its segregation for lethality was reflected by the appearance of a few stunted plants in the root screen. As for mutant 177, a gamma-ray-induced mutant as well as a different SALK line of the same gene revealed blocked lateral root initiation, small bushy shoots, and male sterility [45], partly consistent with the small and abnormal seedlings we observed. In mutant 97, the segregating phenotype for tiny seedlings agreed with the light green seeds in HT siliques. Different from this observation, the published phenotypes include slightly early flowering and reduced responsiveness to GA and BR treatment during seed germination [46]. Therefore, the 'seg' phenotype of mutant 97 is likely to be unrelated to the target gene.

The ‘multiple’ group describes cases in which distinct phenotypes were observed in different individuals of the same mutant line. In mutant 115, one plant was late flowering, one had reduced fertility, and one was small, all being involved in the reported phenotypes [47]. In addition, the ‘flowering’ mutant (mutant 11) was early flowering, while the ‘seed’ mutant (mutant 131) had abnormal seed.

In summary, there are 12 DR HM mutants with soil growth phenotypes, after removing those likely not caused by the mutated target genes (all ‘lethal’ except mutant 99; all ‘seg’ except mutant 177). The soil growth PP in DR HM mutants thus became 9.2% (12/130).

Homozygous (HM) SALK lines: Root and stress screen phenotypes

Not all HM SALK lines were included in the root and stress phenotype dataset for various reasons. In early trials, plates were not photographed if nothing striking was noticed. In addition, a few plates were contaminated. All of these reasons led to ‘single read’ (Methods) results, which were not considered a valid phenotype score. One HM SALK line (SALK_017317C) was not tested because it was not yet in hand at the time of the screen. For gene AT4G21770 with two SALK lines, SALK_039518C was tested instead of SALK_149232C.

Segregating phenotypes

Segregating root or stress phenotypes were found in both DR and nDR. There were three ‘seg’ nDR knockout mutants (Figure 3.4), all with reliable genotyping results. Therefore the ‘seg’ appearance might be caused by factors other than scoring HT genotypes as HM, for example, phenotype penetrance being less than 100%.

In DR, six ‘seg’s were found in five mutants, with mutant 135 segregating in both root and sucrose screens (Figure 3.5). Three of the five mutants (Mutants 26, 177, 97) had soil growth

phenotypes, indicating that most 'seg's reiterated morphological changes expressed under normal conditions rather than reflecting root or stress specific phenotypes. As discussed above, genotyping error might account for the segregating phenotypes in mutants 177 and 135, while a background T-DNA insertion may be responsible for the 'seg' in mutant 97.

The non-segregating root phenotypes

A total of six (5%) DR, one (4%) nDR and one (5%) SCR HM mutant(s) showed root phenotypes (Table S3.5). One of these mutants had long roots, while the others had short roots, consistent with the observation that long roots appear much less frequently than short roots among published knockout phenotypes [17]. Four of the six DR root mutants had a striking phenotype (scored 1 or 5), whereas all nDR and SCR root phenotypes were mild (scored 2).

Most published root phenotypes co-occurred with other morphological changes, particularly dwarfism [17]. In contrast, only one root mutant (Mutant 51) had a soil growth phenotype, indicating that these genes tended to be involved in root-specific processes. Five out of eight root mutants correspond to proteins located in chloroplast or mitochondrion, suggesting over-representation of pathways regulating root development through signaling crosstalk between chloroplast/mitochondrion and nuclear genomes.

Hormone signaling played an important role in root development [48, 49]. ABA regulates root growth through interaction with other hormone signaling or synthesis pathways [50], promoting root growth at low concentration (<1μM) and inhibiting root growth at high concentration [51]. Therefore, mutants with defects in ABA signaling might have disturbed root development, and vice versa. The SCR root mutant was indeed sensitive to ABA, yet none of the DR mutants had ABA phenotypes, indicating that DR genes' functions differed from the control group in root growth regulation.

Among the six DR genes with root phenotypes, three might indirectly affect protein synthesis efficiency according to their annotation, including ribosomal protein S4, pseudouridine synthase and anthranilate phosphoribosyltransferase. Pseudouridine synthase is involved in RNA modification [52] and Anthranilate phosphoribosyl-transferase participates in tryptophan biosynthesis. Though no stress condition was applied in the root screen, two genes were implicated in stress response: NDF6 is likely to be a subunit of NDH, which helps to alleviate oxidative stress in chloroplasts [53]; ATTPR2, a putative co-chaperone of Hsp70/Hsp90, interacts with Hsp90, the expression of which is induced by a variety of stresses [54]. The remaining protein, a CAAX amino terminal protease family protein, might be involved in signal transduction, as CAAX proteins have essential roles in mammalian signaling pathways [55].

The non-segregating root phenotypes

DR genes with striking stress phenotype

About 7% of DR HM SALK lines, representing 9 DR genes, showed striking phenotypes (with a final score of 0, 1, or 5) in stress screen(s) (Table S3.6). The screens with the most striking-phenotype mutants are sucrose and cold germination, each containing three. These two screens were the only ones in which striking phenotypes were found in the nDR group.

Among the 9 DR mutants with striking stress phenotypes, two had phenotypes in more than one stress screen. Mutant 34 was sensitive to water stress, including ABA, salt and sucrose screens. Though it scored ‘single read’ in the mannitol screen, strong sensitivity to mannitol was also suggested (Figure S3.1). In contrast, it was indistinguishable from WT under temperature stress. Mutant 167 was particularly responsive to cold stress, showing tolerant phenotypes in both cold growth and cold germination screens. Whether it is involved in heat stress remains to be tested, as only single reads were obtained in both heat and heat recovery screen.

Two genes had published phenotypes for their knockdown or knockout mutants, namely, AT4G13670 (mutant 39) and AT2G31040 (mutant 65). The RNAi line of AT4G13670 was sensitive to heat stress (30 degree, 5 d) [56], yet its SALK line (SALK_096411C) had no phenotype in our heat screen, perhaps because we used a different condition (34 degree, 14 d). Instead, a phenotype (yellow green) similar to one published under heat stress [56] was detected in our cold germination screen (Figure S3.2), implying that the protein might be required for normal chloroplast development under both hot and cold conditions. The smaller size of mutant 65 reported by others (SALK_05729) [57] is not observed in our study, perhaps because the reported size difference is not striking. Similarly to mutant 39, mutant 65 was yellow green under cold stress, possibly due to chloroplast protein synthesis reduction [57], which might be aggravated under cold stress.

The two largest functional categories of DR genes with striking stress phenotypes were transcription/translation and signaling/stress response. Four out of nine genes were implicated in transcription or protein synthesis (113, 3_36, 39, 65), including a SET domain protein, a pseudouridine synthase family protein, PTAC5 (plastid transcriptionally active 5), and AtCGL160 (Arabidopsis CONSERVED ONLY IN THE GREEN LINEAGE160) [57]. SET domain proteins are involved in epigenetic control of transcription [58]. PTAC5 is a necessary component for plastid-encoded RNA polymerase (PEP)-dependent transcription under heat stress [56]. Though ATCGL160 is not directly involved in translation, its knockout mutant has less chloroplast protein than WT, presumably due to the reduced translation rate caused by insufficient ATP [57].

Three of the 9 DR genes with striking stress phenotypes were associated with signaling or stress response, according to their predicted functions. Mutants 34, 39 and 3_35 correspond to a

phosphoinositide binding protein [59], PTAC5, and a tetratricopeptide repeat (TPR) protein, respectively. Phosphoinositide binding proteins are the effectors of PtdIns(3)P [60], which plays important roles in diverse physiological processes, including ROS production in response to salt stress [61]. As mentioned above, PTAC5 is important for chloroplast transcription under heat stress. TPR-containing proteins may participate in hormone signaling [62] or function as co-chaperones of heat shock proteins [54].

Translation or stress response were also well represented in SCR genes with striking stress phenotypes, with one involved in defense response and the other in tRNA modification (Table S3.6). In contrast, none of the nDR genes with striking stress phenotypes belong to the two categories (Table S3.6): one is implicated in lipid metabolism and the other is unknown.

DR genes with mild stress phenotypes

There were 33 mutants of DR genes with mild phenotypes (scored 2 or 4) in stress screens. As discussed in the 'soil growth phenotypes' section, mutant 1_49 might not be a knockout and thus was removed from the list of DRs with mild stress phenotype.

Among the remaining 32 mutants, five also revealed striking phenotypes (including soil growth phenotypes) in other screens. Mutant 3_35 and 3_36 were sensitive in both cold growth and sucrose screens, while mutant 167 was tolerant in these two screens, indicating a positive correlation between cold growth and sucrose. In addition, two mutants had soil growth phenotypes: Mutant 26 had small silver leaves; Mutant 160 was small and yellow green. The basis for correlation between the soil growth phenotypes and stress-triggered phenotypes, sucrose sensitivity for mutant 26 and cold tolerance for mutant 160, was unclear.

Two of the 27 mutants with only mild phenotypes detected (Mutant 63 and 83) had phenotypes in multiple screens. Mutant 63 was sensitive to ABA, mannitol and sucrose, and was

tolerant of cold growth. The corresponding gene is involved in protein modification, therefore may affect a variety of proteins, perhaps accounting for the wide range of phenotypes. Mutant 83 was ABA tolerant and salt sensitive, consistent with its annotation as a desiccation-induced protein.

Five of the 27 mutants with only mild phenotypes have published loss-of-function phenotypes. Two were reported to have cellular biological (CLB) phenotypes [17] which were undetectable under our test conditions. One mannitol sensitive mutant (Mutant 67) had a conditional phenotype triggered by hrpA and DC3000 strains [63], suggesting that the protein might participate in crosstalk between biotic and abiotic stress signaling pathways. The other two genes (AT5G11450; AT4G31770) were reported to cause visible morphological changes [64] and lethality [65] upon deletion, respectively. The mutants we used for both genes (Mutant 78, SALK_122077C; Mutant 58, SALK_024527C) were different from those with published phenotypes (Mutant 78, an unknown T-DNA insertion line; Mutant 58, SALK_061118), perhaps explaining why we did not observe the same phenotypes. Instead, their stress phenotypes may reflect the knockdown effect.

None of the 27 genes with only mild phenotypes had functional annotations implicated in signal transduction or stress response, whereas one of the six SCR genes (AT5G62390) with a mild stress mutant phenotype was implicated in heat/cold/unfolded protein response. Instead, the largest category (8/27 genes, 30%) in the mild-phenotype DR group was ‘unknown’ (Table S3.7; empty cells in column ‘protein description’). The high proportion of unknown proteins among mutants with mild phenotypes was not unique to DR. Three of the four unknown SCR proteins were in the mild-phenotype group (Table S3.7, Table S3.5). Four of six nDR genes with mild

phenotypes were unknown (Table S3.7), while unknown proteins compromised only 11% (5/23) of the nDR genes with striking or no phenotypes (Table S3.5).

PP comparison between DR and the control groups among stress screens

As can be seen above, some genes have published severe knockout phenotypes which were absent in our mutants for the same genes. If no PCR problem occurred and the published phenotype was confirmed, the mutants used in our study were then considered to be knockdown (Table S3.9). Consequently, these presumably knockdown phenotypes were not used in calculating the PP (phenotype percentage) below.

The percentage of DR HM mutants showing phenotypes (29%) in at least one stress screen, hereafter referred to as the overall stress PP, was lower than SCR (40%), suggesting that DR HM mutants were not highly involved in water and temperature stress response. On the other hand, the overall stress PP of DR HM group (29%) is similar to nDR HM group (28%), consistent with their similar copy number (Methods). Relatively high SCR PP together with similar DR and nDR PPs were observed for striking phenotypes (Table 3.3), implying that this pattern is reliable and is not affected by possible misscoring of mild phenotypes (score 2 or 4).

Mutants with phenotypes in two or more stress screens may represent genes involved in crosstalk among different stress signaling pathways. The percentage of such mutants was lower in the DR group (5%) than both control groups (SCR, 10%; nDR, 7%) (Table S3.5). Therefore, DR genes were not highly involved in crosstalk between multiple stress response pathways, compared to other fully conserved singletons.

The percentage of DR HM mutants being sensitive or tolerant was lower than SCR HM mutants in seven stress screens (Table 3.3; Figure 3.6). Due to the small sample size (<30) of SCR HM SALKs, the differences in PP between DR and SCR groups were not statistically

significant (Table S3.8). Yet, some differences were obvious thus worth further testing using a larger SCR sample. The most striking difference was in cold germination, where the PP of the SCR group is almost four times as high as that of the DR group (Table 3.3). In contrast, DR mutants were more prone to reveal phenotypes in the salt screen than SCR mutants, none of which had salt phenotypes. High salt (NaCl) stress induces both osmotic stress and ion toxicity. As no mannitol phenotype was found in any of the salt mutants, the corresponding genes might be involved in detoxification or ion homeostasis rather than osmosis. Five of seven DR salt mutants had no ABA phenotype (Table S3.5), suggesting that these genes tended to participate in ABA-independent signaling pathways.

The PP of DR mutants was similar to that of nDR mutants in most screens, except for salt and ABA (Table 3.3; Figure 3.6). Similarly to SCR, there was no salt mutant in the nDR group. The difference of PP between DR and nDR groups in salt and other screens was not statistically significant, due to the small number of nDR genes with HM SALKs (<30). However, if the SCR and nDR were combined as a single control group, the difference between DR and the control for salt PP is slightly significant (p value=0.0434, Table S3.8). In the ABA screen, the nDR PP was higher than the DR PP, implying that DR genes were less involved in ABA-mediated processes during seed germination and early seedling development: stress response, germination [66], plant growth or water regulation [67].

There are more sensitive phenotypes (scored 0, 1, or 2) than tolerant phenotypes (scored 4 and 5) in all three groups, consistent with a prior study [26]. Sensitive PP exceeds tolerant PP in most screens except for ABA in DR and nDR groups (Tables 3.4, 3.5). The nDR group has more ABA tolerant phenotypes than sensitive ones, whereas in the DR group the numbers of tolerant and sensitive ABA phenotypes were equal.

Tolerant DR mutants were found in most screens except heat recovery and salt, whereas in the control groups (nDR and SCR) tolerant phenotypes were only detected in ABA and cold germination screens. Moreover, DR tolerant PPs in ABA, cold growth and sucrose were a bit higher than in other screens (Table 3.5). For sensitive phenotypes, DR had lower PP than both control groups in cold growth and cold germination screens (Table 3.4), perhaps reflecting a lack of positive regulators in response to low, non-freezing temperatures.

DR genes without observed phenotypes

There are 70 DR HM mutants with no observed phenotypes, after excluding the knockdown mutants (Table S3.9). Of these 70 DR genes, eight had published loss-of-function phenotypes that were visible under normal conditions and two were essential (Table S3.10). Several reasons might explain why we failed to observe the same phenotypes, the major one being that seven of the ten mutants we screened were different than those with reported phenotypes. Though different mutants should reveal the same phenotype(s) if the genes were truly knocked out, T-DNA insertions are not always able to eliminate the target gene product [68, 69]. As a result, 'leaky' products might attenuate or totally mask knockout phenotypes. In addition, non-striking soil growth phenotypes (eg. Mutant 120, SALK_007870) were not thoroughly/intentionally screened for in our study (Method). Further, four of the eight published phenotypes were about siliques or flowering (Table S3.10), reproductive stage phenotypes that were not covered in our root and stress screens. Finally, PCR problems may be responsible for the absence of a phenotype in one mutant (Mutant 111, SALK_093546C). Consequently, the two HM plants were not albino as reported [70]. The ten genes were removed from DR genes without observed phenotypes.

The largest functional category in the remaining 60 DR genes without observed phenotypes was ‘unknown proteins’, constituting 28% (17/60) of this DR genes’ subset (Table S3.11). The percentage was higher than DR genes with phenotypes (20%; 13/66) and the genome-wide average (18%, the percentage of protein-coding genes without known protein domains), but not as high as the mild phenotype DR group (30%; 8/27). Therefore, DR genes with unknown function were less likely to show phenotypes than other DR genes, and they tended to have mild, almost undetectable phenotypes. The high percentage of unknown proteins was also observed in nDR genes without observed phenotypes (24%, 5/21), but not in the SCR group (8.3%; 1/12) (Table S3.11).

Chloroplast targeting DR genes and their phenotypes

Of the sampled DR genes, 37% (66/179) encode chloroplast-located proteins. Removing those with SALK lines in WT or OT group (Table 3.1) or likely being knockdown, there are 55 chloroplast-targeting DR genes, 42% (23/55) of which had presumable knockout phenotypes (Table S3.12). The phenotype percentage of chloroplast-targeting DR genes was lower than other DR genes (47%; 43/92), suggesting that DR proteins located in chloroplast were not more likely to have knockout phenotypes than other DR proteins.

All four main types of observed phenotypes (PL, soil growth, root and stress) were found in DR chloroplast proteins, which occurred at a relative low frequency in soil growth (non-lethal) phenotypes and a relatively high frequency in root phenotypes (Table 3.6). Moreover, the proportion of DR proteins varied among stress screens as well and was high in genes with sucrose and cold germination phenotypes (Table S3.13).

DISCUSSION

All being paleo-polyploid, angiosperms serve as an outstanding model for studying fates of genes after duplication [71]. Among a variety of possible fates [72], the long-term fixation of duplicated copies is considered to be rare, non-random and important [3]. As a result, it has been heavily researched and relatively well explained by a number of hypotheses [9, 73, 74]. In contrast, neither the importance nor the cause of persistent single copy status is well understood, though such 'duplication-resistant' genes do not occur by chance [6, 34]. Intricate and diverse as the underlying mechanisms of duplication resistance may be, it remains a reasonable assumption that the biological roles of DR genes in plants hold an important piece of the answer. Indeed, several possible explanations for single-copy preference derive from features of annotated gene functions [4, 8].

Here we investigated DR genes' functions from the perspective of empirical phenotypes revealed by knockout mutants. There are four major phenotype categories in tested DR genes and control genes, namely, PL (potentially lethal), soil growth, root and stress, with the latter three being associated with homozygous (HM) T-DNA insertion (SALK) lines only. Compared to other single copy genes, DR genes were more prone to cause lethal or visible above-ground phenotypes upon T-DNA insertion, as was revealed by DR genes' higher PL and soil growth PPs. This finding was contrary to observations from published datasets, where other conserved singletons have higher PP than DR in each and every phenotype category (Chapter 2). Several factors could have caused this incongruence. For example, phenotypes from genes not involved in any previous dataset may elevate DR genes' PP. Meanwhile, the sampling of control genes happened not to cover genes with published phenotypes. No matter how high the PP might be in a gene group, mutants with phenotypes were still a minority. As a result, the chance of sampling

the corresponding genes through random selection was low. In addition, some phenotypes might have nothing to do with the target genes, resulting in false expansion of PP. Further validation of these observed phenotypes is important. Nevertheless, our result raises the possibility that DR genes might serve more important functions than other fully conserved singletons.

Despite the lack of above-ground phenotypes, the root PPs in control groups were similar to those of DR genes. Six DR mutants had short or long roots compared to wildtype Col0, constituting about 5% of the DR HM set. The corresponding DR genes may tend to function in root specific programs, reflected by the lack of other phenotypes: one mutant had a soil growth phenotype and one a stress phenotype. In particular, DR genes frequently appeared in root growth pathways regulated by chloroplast signals, as three of the six proteins (50%) were located in the chloroplast. In contrast, DR genes rarely participated in the portion of the root growth pathway intertwined with ABA signaling, compared to SCR genes.

DR genes were not highly involved in response to the tested stresses, with lower overall stress PP than SCR genes. This result was expected as no published evidence suggested overrepresentation of stress-related functions for DR genes. Indeed, growing evidence implicates single-gene duplications as key contributors to stress adaptation, consistent with the ongoing availability of this gene class (versus the episodic nature of whole genome duplications) and the need for continuous adaptation to environmental fluctuation. Rather than reacting to environmental stimulus, housekeeping functions seem to be a better fit for DR genes, as was suggested by frequent occurrence of TELO-box and scarcity of TATA-box containing promoters in the unique genes conserved between Arabidopsis and Oryza [7]. Under this hypothesis, severe phenotypes would be a more likely consequence of DR gene T-DNA insertion, supportive of the

higher occurrence of soil growth and potentially lethal phenotypes in DR mutants than the controls.

The relatively low PP of DR genes was seen in almost every stress screen, with the most striking difference relative to SCR in cold germination. Two SCR mutants and three DR mutants showed cold germination phenotypes. One of the SCR genes was pathogenesis-related. Proteins functioning in both cold and disease responses are not rare [75-77]. Moreover, disease resistance genes have generally high copy numbers and great copy number variation among species [78], presumably accounting for their low occurrence in DR genes. The lack of defense proteins, in turn, may contribute to the low cold germination PP in DR genes.

Salt is the only stress for which the PP is higher in the DR group than the SCR group. Based on their lack of osmotic stress phenotypes, DR genes with salt phenotypes tend to be involved in detoxification or ion homeostasis. Similar to DR genes with root phenotypes, they are also more prone to function in ABA-independent pathways. As no salt phenotype was found in the other control group (nDR) as well, the enrichment of salt pathway genes may have been associated with duplication resistance. Yet, as with any other stress, salt stress response is complicated, including the re-adjustment of numerous biological processes [28], making it difficult to clarify its correlation with duplication resistance. An attractive hypothesis is that these genes form a sub-network of salt signaling, within which duplication of any member may disturb overall network function. However, no co-expression or protein interaction was detected among these genes, indicating that they were probably not closely related.

The difference in stress PP between DR and SCR genes was not evident in comparing DR with nDR genes. The similar PPs of DR and nDR genes were expected based on their same copy numbers in the five plant species (Method). Despite the similarity in their proportions of

stress mutants, the nDR group has a lower salt PP and higher ABA PP than the DR group. In addition to drought mutants, the application of exogenous ABA could sort out candidates for various other stresses as well. Therefore, the low ABA PP of DR genes further supports the earlier conclusion that DR genes tend not to react to stress, especially to ABA-mediated parts.

Moreover, DR genes have a low proportion of positive regulators in cold signaling, as revealed by their much lower sensitive PP in the two cold screens than both control groups. Low temperature might favor survival of polyploid plants, as polyploidization sometimes increased plant cold tolerance [79]. During subsequent diploidization, it might confer a selective advantage for cold positive regulators to remain duplicated for a while, for example, favoring survival of future cold episodes. Therefore, they were less likely to occur in a gene group where the extra copies were quickly removed.

Of DR stress phenotypes, 25% (9) were striking and 75% (27) were mild. Predicted functions in stress signaling were only found in DR genes with striking phenotypes, not in those with mild phenotypes. Indeed, about a third of DR genes with mild stress phenotypes were of unknown function. Instead of being involved in the main stress pathways, mild-phenotype genes might have peripheral roles during stress response. On the other hand, the high proportion of unknown functions in the mild phenotype group indicates that unknown genes are prone to revealing non-striking knockout phenotypes. Phenotyping methods that are sufficiently sensitive to detect mild phenotypes may be important in studying genes of unknown function.

In addition to stress signaling, functions implicated in transcription or translation were also enriched in striking-stress-phenotype group. The stress phenotype might derive from reduction of key stress regulatory factors. Alternatively, a broad reduction in protein synthesis may also make plants more vulnerable to stresses. Other than stress phenotypes, disturbance of

genes which affect a group of other genes' production is prone to reveal phenotypes under normal conditions, as can be seen by their high occurrence in root and soil growth phenotype groups. Most of these genes belong to DNA or RNA metabolism, a GO category enriched in DR genes. Therefore, the enrichment of transcription/translation-related genes in multiple phenotype groups indicates that DNA/RNA metabolic DR genes have important functions.

Besides DNA/RNA metabolism, chloroplast is also an overrepresented GO term in DR genes. Chloroplast-targeting DR genes had a lower overall PP (42%; 23/55) to other DR genes (47%; 43/92) in our sample. If chloroplast-located DR proteins were indeed involved in dosage balance with their chloroplast-genome-encoded interactors, the disturbance of such balance via gene knockout is not reflected at the phenotype level. Duplication, on the other hand, might have a larger perturbing effect. Hence, it would be interesting to test mutants where chloroplast-targeting DR genes were duplicated. In addition, the result might imply that the effect of dosage balance disturbance is subtle and thus would not be visible during a few generations.

The proportion of chloroplast-targeting DR genes is distributed unevenly among the four major phenotype groups, being highest in root and lowest in above-ground phenotypes, indicating that they tend to regulate a particular part of, rather than broadly affect, plant growth/development. Similarly, these genes were strongly associated with particular stresses, such as sucrose and cold germination. The enrichment of chloroplast-targeting genes in sucrose mutants was observed in the control groups as well. Being the location where sucrose is produced, the chloroplast is expected to play important roles in sucrose signaling. The correlation between chloroplast-located proteins and cold germination was unique in DR: all DR cold germination mutant genes were chloroplast-targeting, while none were related to chloroplasts in either control group. In contrast, chloroplast DR genes were not highly

represented in cold growth mutants, indicating that these genes particularly respond to cold stress occurring before or during germination.

Agreement with published data in our observed phenotypes was found in some but not all genes with published loss-of-function phenotypes. PCR problems, phenotype mis-scoring, and background T-DNA insertion were among possible reasons, yet the major cause might be the use of different mutants. The phenotype of one mutant was frequently absent in another mutant of the same gene in published datasets (Chapter 2)(Table S3.14), implying that some transposon or T-DNA insertional mutants, SALK lines included, are not real knockouts. Updates on DR knockout phenotypes will occur with further validation. For example, the stress phenotype was merely a knockdown effect, whereas the real knockout mutant has a more severe phenotype. Nevertheless, a majority of our observations remained true for SALK lines used in this study.

The observed phenotypes, whether from knockout or knockdown mutant, serve as a starting point for further investigation of DR genes' functions, and also added new information to the known knockout/knockdown phenotypes. As the first thorough screen of stress-related phenotypes for DR mutants, these results depict a rough yet relatively complete picture of their phenotypic features, some of which appear to be related to duplication resistance. These features may inspire new ideas for future research, meanwhile taking us one step closer to demonstrating DR gene functions as well as clarifying the underlying force(s) favoring duplication resistance.

REFERENCES

1. Bowers JE, Chapman BA, Rong J, Paterson AH: Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 2003, 422:433-438.

2. Edger PP, Pires JC: Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res 2009, 17:699-717.

3. Lynch M: The Evolutionary Fate and Consequences of Duplicate Genes. Science 2000, 290:1151-1155.

4. Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J, dePamphilis CW: Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. BMC Evol Biol 2010, 10:61.

5. Han F, Peng Y, Xu L, Xiao P: Identification, characterization, and utilization of single copy genes in 29 angiosperm genomes. BMC Genomics 2014, 15:504.

6. Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC: Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. Trends Genet 2006, 22:597-602.

7. Armisen D, Lecharny A, Aubourg S: Unique genes in plants: specificities and conserved features throughout evolution. BMC Evol Biol 2008, 8:280.

8. De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y: Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci U S A 2013, 110:2898-2903.

9. Birchler JA, Riddle NC, Auger DL, Veitia RA: Dosage balance in gene regulation: biological implications. Trends Genet 2005, 21:219-226.

10. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al: Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 2003, 301:653-657.

11. Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C, et al: A high-throughput Arabidopsis reverse genetics system. Plant Cell 2002, 14:2985-2994.

12. Rosso MG, Li Y, Strizhov N, Reiss B, Dekker K, Weisshaar B: An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol 2003, 53:247-259.

13. Woody ST, Austin-Phillips S, Amasino RM, Krysan PJ: The WiscDsLox T-DNA collection: an arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. Journal of Plant Research 2007, 120:157-165.

14. Feldmann KA: T-DNA Insertion Mutagenesis in Arabidopsis - Mutational Spectrum. Plant Journal 1991, 1:71-82.

15. Kuromori T, Wada T, Kamiya A, Yuguchi M, Yokouchi T, Imura Y, Takabe H, Sakurai T, Akiyama K, Hirayama T, et al: A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of Arabidopsis. Plant Journal 2006, 47:640-651.

16. Hanada K, Kuromori T, Myouga F, Toyoda T, Li WH, Shinozaki K: Evolutionary persistence of functional compensation by duplicate genes in Arabidopsis. Genome Biol Evol 2009, 1:409-414.

17. Lloyd J, Meinke D: A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. Plant Physiology 2012, 158:1115-1129.

18. Jogaiah S, Govind SR, Tran LS: Systems biology-based approaches toward understanding drought tolerance in food crops. Crit Rev Biotechnol 2013, 33:23-39.

19. Gupta B, Huang B: Mechanism of salinity tolerance in plants: physiological, biochemical, and molecular characterization. Int J Genomics 2014, 2014:701596.

20. Yadav SK: Cold stress tolerance mechanisms in plants. A review. Agronomy for Sustainable Development 2010, 30:515-527.

21. Juan JX, Yu XH, Jiang XM, Gao Z, Zhang Y, Li W, Duan YD, Yang G: Agrobacterium-mediated transformation of tomato with the ICE1 transcription factor gene. Genetics and Molecular Research 2015, 14:597-608.

22. Zeng W, Brutus A, Kremer JM, Withers JC, Gao X, Jones AD, He SY: A genetic screen reveals Arabidopsis stomatal and/or apoplastic defenses against Pseudomonas syringae pv. tomato DC3000. PLoS Pathog 2011, 7:e1002291.

23. Quesada V, Ponce MR, Micol JL: Genetic analysis of salt-tolerant mutants in Arabidopsis thaliana. Genetics 2000, 154:421-436.

24. Koiwa H, Bressan RA, Hasegawa PM: Identification of plant stress-responsive determinants in arabidopsis by large-scale forward genetic screens. J Exp Bot 2006, 57:1119-1128.

25. Arenas-Huertero F, Arroyo A, Zhou L, Sheen J, Leon P: Analysis of Arabidopsis glucose insensitive mutants, gin5 and gin6, reveals a central role of the plant hormone ABA in the regulation of plant vegetative development by sugar. Genes Dev 2000, 14:2085-2096.

26. Song LH, Hegie A, Suzuki N, Shulaev E, Luo XZ, Cenariu D, Ma V, Kao S, Lim J, Gunay MB, et al: Linking genes of unknown function with abiotic stress responses by high-throughput phenotype screening. Physiologia Plantarum 2013, 148:322-333.

27. Warren G, McKown R, Marin A, Teutonico R: Isolation of mutations affecting the development of freezing tolerance in Arabidopsis thaliana (L.) Heynh. Plant Physiology 1996, 111:1011-1019.

28. Zhu JK: Salt and drought stress signal transduction in plants. Annu Rev Plant Biol 2002, 53:247-273.

29. Shinozaki K, Yamaguchi-Shinozaki K: Gene networks involved in drought stress response and tolerance. J Exp Bot 2007, 58:221-227.

30. Rolland F, Moore B, Sheen J: Sugar sensing and signaling in plants. Plant Cell 2002, 14 Suppl:S185-205.

31. Miura K, Furumoto T: Cold signaling and cold response in plants. Int J Mol Sci 2013, 14:5312-5337.

32. Saidi Y, Finka A, Goloubinoff P: Heat perception and signalling in plants: a tortuous path to thermotolerance. New Phytol 2011, 190:556-565.

33. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH: Synteny and collinearity in plant genomes. Science 2008, 320:486-488.

34. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH: Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 2008, 18:1944-1954.

35. Rogers SO, Bendich AJ: Extraction of DNA from plant tissues. In Plant molecular biology manual. Springer; 1989: 73-83

36. Murashige T, Skoog F: A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiologia Plantarum 1962, 15:473-497.

37. Myouga F, Akiyama K, Tomonaga Y, Kato A, Sato Y, Kobayashi M, Nagata N, Sakurai T, Shinozaki K: The Chloroplast Function Database II: a comprehensive collection of homozygous mutants and their phenotypic/genotypic traits for nuclear-encoded chloroplast proteins. Plant Cell Physiol 2013, 54:e2.

38. Wilson-Sanchez D, Rubio-Diaz S, Munoz-Viana R, Perez-Perez JM, Jover-Gil S, Ponce MR, Micol JL: Leaf phenomics: a systematic reverse genetic screen for Arabidopsis leaf mutants. Plant Journal 2014.

39. Qiao JW, Li J, Chu W, Luo MZ: PRDA1, a Novel Chloroplast Nucleoid Protein, is Required for Early Chloroplast Development and is Involved in the Regulation of Plastid Gene Expression in Arabidopsis (vol 54, pg 2071, 2013). Plant and Cell Physiology 2014, 55:467-467.

40. Quettier AL, Shaw E, Eastmond PJ: SUGAR-DEPENDENT6 encodes a mitochondrial flavin adenine dinucleotide-dependent glycerol-3-p dehydrogenase, which is required for glycerol catabolism and post germinative seedling growth in Arabidopsis. Plant Physiology 2008, 148:519-528.

41. Shimada H, Ohno R, Shibata M, Ikegami I, Onai K, Ohto MA, Takamiya K: Inactivation and deficiency of core proteins of photosystems I and II caused by genetical phylloquinone and plastoquinone deficiency but retained lamellar structure in a T-DNA mutant of Arabidopsis. Plant Journal 2005, 41:627-637.

42. Myouga F, Akiyama K, Motohashi R, Kuromori T, Ito T, Iizumi H, Ryusui R, Sakurai T, Shinozaki K: The Chloroplast Function Database: a large-scale collection of Arabidopsis Ds/Spm- or T-DNA-tagged homozygous lines for nuclear-encoded chloroplast proteins, and their systematic phenotype analysis. Plant Journal 2010, 61:529-542.

43. Tzafrir I, Pena-Muralla R, Dickerman A, Berg M, Rogers R, Hutchens S, Sweeney TC, McElver J, Aux G, Patton D, Meinke D: Identification of genes required for embryo development in Arabidopsis. Plant Physiology 2004, 135:1206-1220.

44. Tzafrir I, McElver JA, Liu Cm CM, Yang LJ, Wu JQ, Martinez A, Patton DA, Meinke DW: Diversity of TITAN functions in Arabidopsis seed development. Plant Physiology 2002, 128:38-51.

45. DiDonato RJ, Arbuckle E, Buker S, Sheets J, Tobar J, Totong R, Grisafi P, Fink GR, Celenza JL: Arabidopsis ALF4 encodes a nuclear-localized protein required for lateral root formation. Plant Journal 2004, 37:340-353.

46. Chen JG, Pandey S, Huang JR, Alonso JM, Ecker JR, Assmann SM, Jones AM: GCR1 can act independently of heterotrimeric G-protein in response to brassinosteroids and gibberellins in Arabidopsis seed germination. Plant Physiology 2004, 135:907-915.

47. Kim YJ, Zheng B, Yu Y, Won SY, Mo B, Chen X: The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. Embo Journal 2011, 30:814-822.

48. Overvoorde P, Fukaki H, Beeckman T: Auxin control of root development. Cold Spring Harb Perspect Biol 2010, 2:a001537.

49. Garay-Arroyo A, De La Paz Sanchez M, Garcia-Ponce B, Azpeitia E, Alvarez-Buylla ER: Hormone symphony during root growth and development. Dev Dyn 2012, 241:1867-1885.

50. Luo XJ, Chen ZZ, Gao JP, Gong ZZ: Abscisic acid inhibits root growth in Arabidopsis through ethylene biosynthesis. Plant Journal 2014, 79:44-55.

51. Rodrigues A, Santiago J, Rubio S, Saez A, Osmont KS, Gadea J, Hardtke CS, Rodriguez PL: The short-rooted phenotype of the brevis radix mutant partly reflects root abscisic acid hypersensitivity. Plant Physiology 2009, 149:1917-1928.

52. Hamma T, Ferre-D'Amare AR: Pseudouridine synthases. Chem Biol 2006, 13:1125-1135.

53. Wang P, Duan W, Takabayashi A, Endo T, Shikanai T, Ye JY, Mi HL: Chloroplastic NAD(P)H dehydrogenase in tobacco leaves functions in alleviation of oxidative damage caused by temperature stress. Plant Physiology 2006, 141:465-474.

54. Prasad BD, Goel S, Krishna P: In silico identification of carboxylate clamp type tetratricopeptide repeat proteins in Arabidopsis and rice as putative co-chaperones of Hsp90/Hsp70. PLoS One 2010, 5:e12761.

55. Manolaridis I, Kulkarni K, Dodd RB, Ogasawara S, Zhang Z, Bineva G, O'Reilly N, Hanrahan SJ, Thompson AJ, Cronin N, et al: Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1. Nature 2013, 504:301-305.

56. Zhong L, Zhou W, Wang H, Ding S, Lu Q, Wen X, Peng L, Zhang L, Lu C: Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in Arabidopsis under heat stress. Plant Cell 2013, 25:2925-2943.

57. Ruhle T, Razeghi JA, Vamvaka E, Viola S, Gandini C, Kleine T, Schunemann D, Barbato R, Jahns P, Leister D: The Arabidopsis protein CONSERVED ONLY IN THE GREEN LINEAGE160 promotes the assembly of the membranous part of the chloroplast ATP synthase. Plant Physiology 2014, 165:207-226.

58. Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R, Schulz I, Reuter G, Aalen RB: The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res 2001, 29:4319-4333.

59. Jensen RB, La Cour T, Albrethsen J, Nielsen M, Skriver K: FYVE zinc-finger proteins in the plant model Arabidopsis thaliana: identification of PtdIns3P-binding residues by comparison of classic and variant FYVE domains. Biochem J 2001, 359:165-173.

60. Wywial E, Singh SM: Identification and structural characterization of FYVE domain-containing proteins of Arabidopsis thaliana. Bmc Plant Biology 2010, 10:157.

61. Leshem Y, Seri L, Levine A: Induction of phosphatidylinositol 3-kinase-mediated endocytosis by salt stress leads to intracellular production of reactive oxygen species and salt tolerance. Plant Journal 2007, 51:185-197.

62. Schapire AL, Valpuesta V, Botella MA: TPR Proteins in Plant Hormone Signaling. Plant Signal Behav 2006, 1:229-230.

63. De-La-Pena C, Rangel-Cano A, Alvarez-Venegas R: Regulation of disease-responsive genes mediated by epigenetic factors: interaction of Arabidopsis-Pseudomonas. Mol Plant Pathol 2012, 13:388-398.

64. Roose JL, Frankel LK, Bricker TM: Developmental defects in mutants of the PsbP domain protein 5 in Arabidopsis thaliana. PLoS One 2011, 6:e28624.

65. Wang H, Hill K, Perry SE: An Arabidopsis RNA lariat debranching enzyme is essential for embryogenesis. Journal of Biological Chemistry 2004, 279:1468-1473.

66. Nakashima K, Yamaguchi-Shinozaki K: ABA signaling in stress-response and seed development. Plant Cell Rep 2013, 32:959-970.

67. Raghavendra AS, Gonugunta VK, Christmann A, Grill E: ABA perception and signalling. Trends Plant Sci 2010, 15:395-401.

68. Wang YH: How effective is T-DNA insertional mutagenesis in Arabidopsis? Journal of Biochemical Technology 2008, 1:11-20.

69. Rodriguez-Milla MA, Salinas J: Prefoldins 3 and 5 play an essential role in Arabidopsis tolerance to salt stress. Mol Plant 2009, 2:526-534.

70. Mudd EA, Sullivan S, Gisby MF, Mironov A, Kwon CS, Chung WI, Day A: A 125 kDa RNase E/G-like protein is present in plastids and is essential for chloroplast development and autotrophic growth in Arabidopsis. J Exp Bot 2008, 59:2597-2610.

71. Wang Y, Wang X, Paterson AH: Genome and gene duplications and gene expression divergence: a view from plants. Ann N Y Acad Sci 2012, 1256:1-14.

72. Zhang J: Evolution by gene duplication: an update. Trends in ecology & evolution 2003, 18:292-298.

73. Conant GC, Wolfe KH: Turning a hobby into a job: How duplicated genes find new functions. Nature Reviews Genetics 2008, 9:938-950.

74. Freeling M: Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol 2009, 60:433-453.

75. Yang HB, Shi YT, Liu JY, Guo L, Zhang XY, Yang SH: A mutant CHS3 protein with TIR-NB-LRR-LIM domains modulates growth, cell death and freezing tolerance in a temperature-dependent manner in Arabidopsis. Plant Journal 2010, 63:283-296.

76. Yeh S, Moffatt BA, Griffith M, Xiong F, Yang DS, Wiseman SB, Sarhan F, Danyluk J, Xue YQ, Hew CL, et al: Chitinase genes responsive to cold encode antifreeze proteins in winter cereals. Plant Physiology 2000, 124:1251-1264.

77. Seo PJ, Kim MJ, Park JY, Kim SY, Jeon J, Lee YH, Kim J, Park CM: Cold activation of a plasma membrane-tethered NAC transcription factor induces a pathogen resistance response in Arabidopsis. Plant Journal 2010, 61:661-671.

78. Zhang RZ, Murat F, Pont C, Langin T, Salse J: Paleo-evolutionary plasticity of plant disease resistance genes. BMC Genomics 2014, 15.

79. te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesova M, Pysek P: The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot 2012, 109:19-45.

Table 3.1 Number of DR, nDR and SCR SALK lines in each genotype group.

Group	**HM**	**PL**	**WT**	**OT**	**Total**
DR	130	21	29	1	181
nDR	30	3	3	1	37
SCR	24	3	3	2	32
Sum	184	27	35	4	250

Table 3.2 Soil growth phenotypes of DR HM SALK lines.

Category	Number of mutants	Phenotypes
flowering	1	early flowering
pigment	4	yellow green; dark green
leaf	3	silver leaves; small leaves; fewer trichomes
lethal	5	small, dying seedlings
seg	3	about 1/4 seedlings were abnormal
multiple	1	small, late flowering, reduced fertility
seed	1	abnormal seed

Table 3.3 Stress-related phenotype percentage for HM SALK lines. ‘Coldgro’ stands ‘cold growth’. ‘Heatrec’ stands for ‘Heat recovery’. ‘Coldger’ stands for ‘cold germination’. ‘Overall’ stands for ‘overall stress phenotype percentage’. ‘Striking’ stands for ‘striking phenotype’.

Group	**ABA**	**Mannitol**	**Salt**	**Sucrose**	**Heat**	**Heatrec**	**Coldgro**	**Coldger**	**Overall**	**Striking**
SCR	10%	5%	0%	10%	5%	5%	15%	11%	40%	10%
nDR	10%	3%	0%	7%	0%	0%	10%	4%	28%	7%
DR	6%	3%	6%	9%	2%	0%	8%	3%	29%	7%

Table 3.4 Percentage of mutants with sensitive phenotypes.

Group	ABA	Mannitol	Salt	Sucrose	Heat	Heatrec	Coldgro	Coldger
SCR	10%	5%	0%	10%	5%	5%	10%	11%
nDR	3%	3%	0%	7%	0%	0%	10%	4%
DR	3%	2%	6%	6%	2%	0%	5%	2%

Table 3.5 Percentage of mutants with tolerant phenotypes.

Group	ABA	Mannitol	Salt	Sucrose	Heat	Heatrec	Coldgro	Coldger
SCR	0%	0%	0%	0%	0%	0%	5%	0%
nDR	7%	0%	0%	0%	0%	0%	0%	0%
DR	3%	1%	0%	2%	1%	0%	3%	1%

Table 3.6 Percentage of chloroplast-targeting genes in DR phenotype groups. CT in the table stands for chloroplast-targeting. Mutant 99 was put into PL group instead of soil growth phenotype group. Mutant 135 was put into PL group.

Phenotype category	**Gene number**	**Number of CT genes**	**Percentage of CT genes**
PL	17	6	35%
soil growth	11	3	27%
root	6	3	50%
stress	34	11	32%

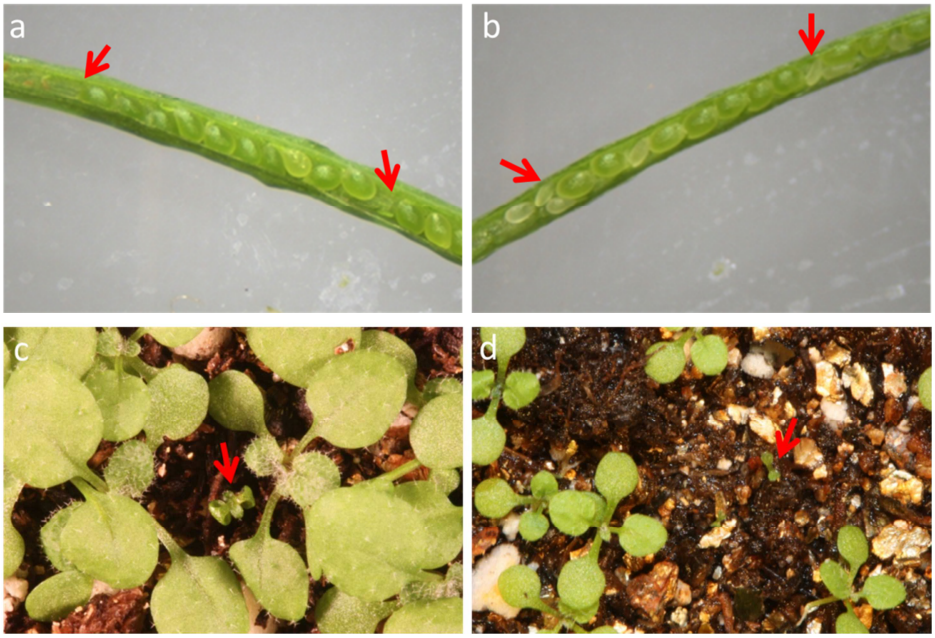

Figure 3.1 Phenotypes of SALK lines without HM knockout plant(s). The red arrows points to empty spaces (a), small seeds (b), and small/dying seedlings (c,d).

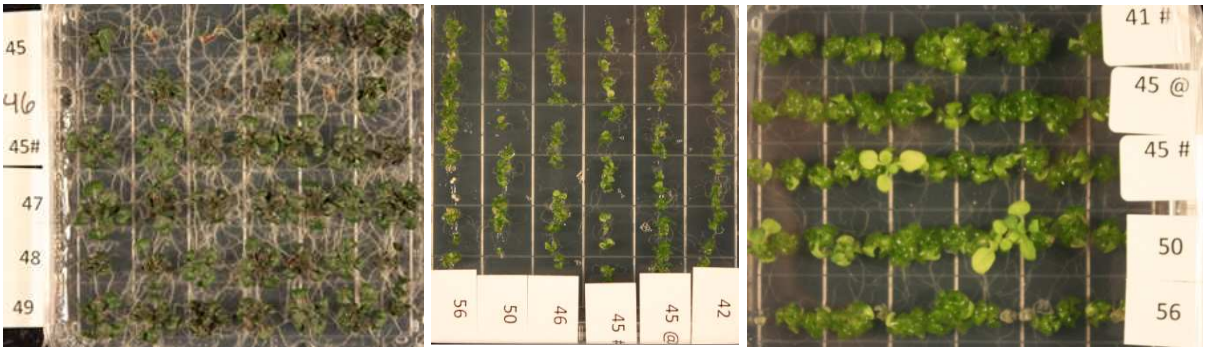

Figure 3.2 Segregating phenotypes for PL lines in stress screens (Left, mutant 45 in sucrose screen; middle, mutant 56 in mannitol screen; right, mutant 56 in ABA screen) The numbers represents SALK lines, which can be found in Table S3.2.

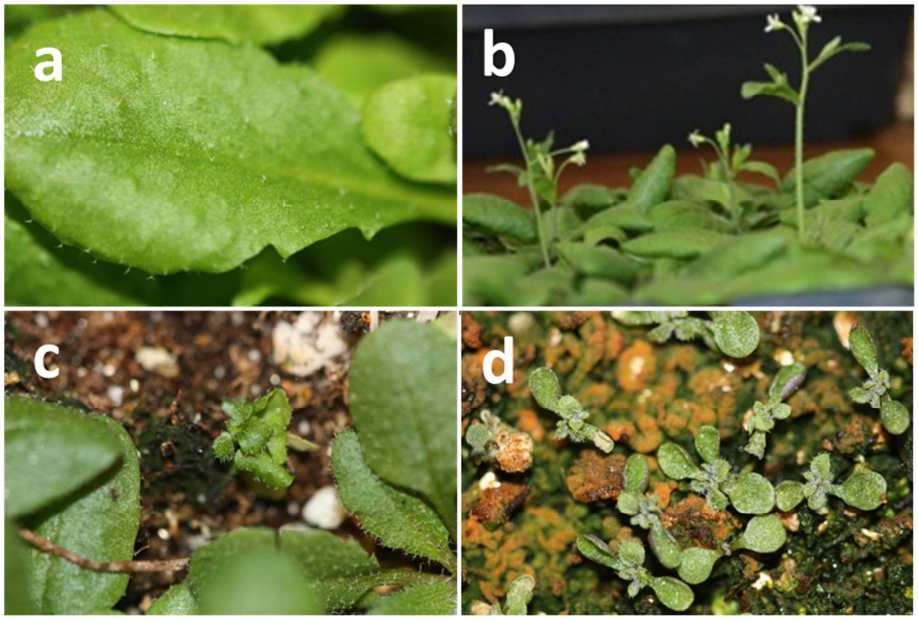

Figure 3.3 Phenotypes of the HM SALK lines. (a) Lack of trichomes; (b) early flowering; (c,d) small.

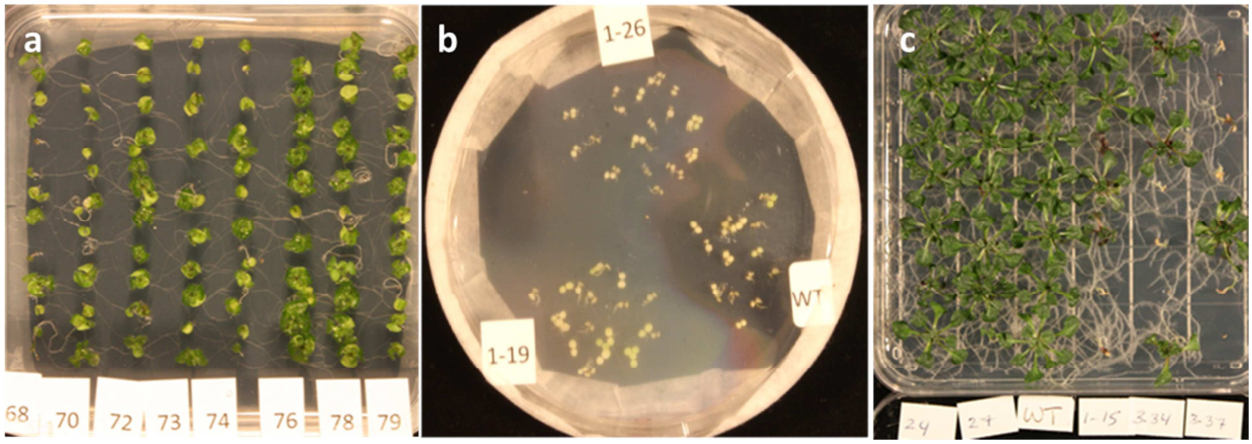

Figure 3.4 nDR HM segregating phenotypes. a. Mutant 68 in ABA b. Mutant 1_19 in heat c, d. Mutant 3_34 in sucrose. The numbers represent SALK lines, which can be found in Table S3.2. WT represents wild type Col 0.

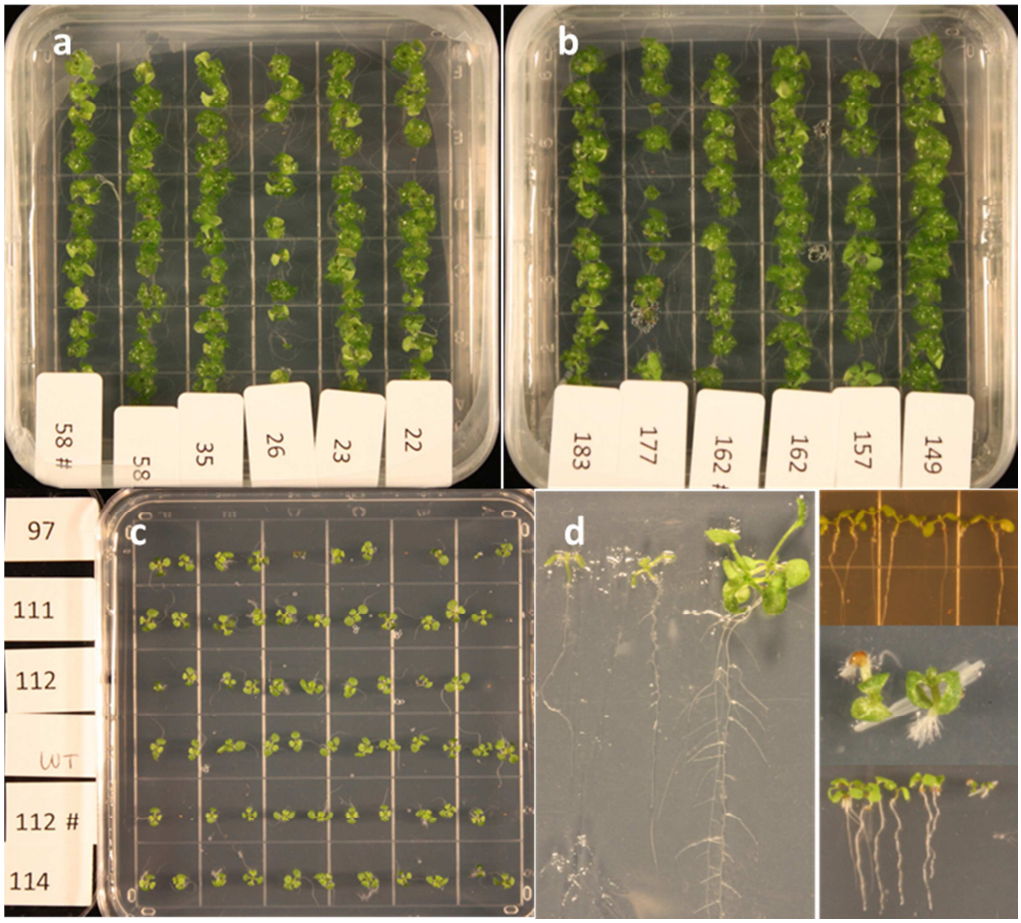

Figure 3.5 DR HM segregating phenotypes. a. Mutant 26 in ABA; b. Mutant 177 in ABA; c. Mutant 96 in cold growth; d. Mutants 137 (left) and 135 (right) in root screen. The numbers represent SALK lines, which can be found in Table S3.2. WT represents wild type Col 0.

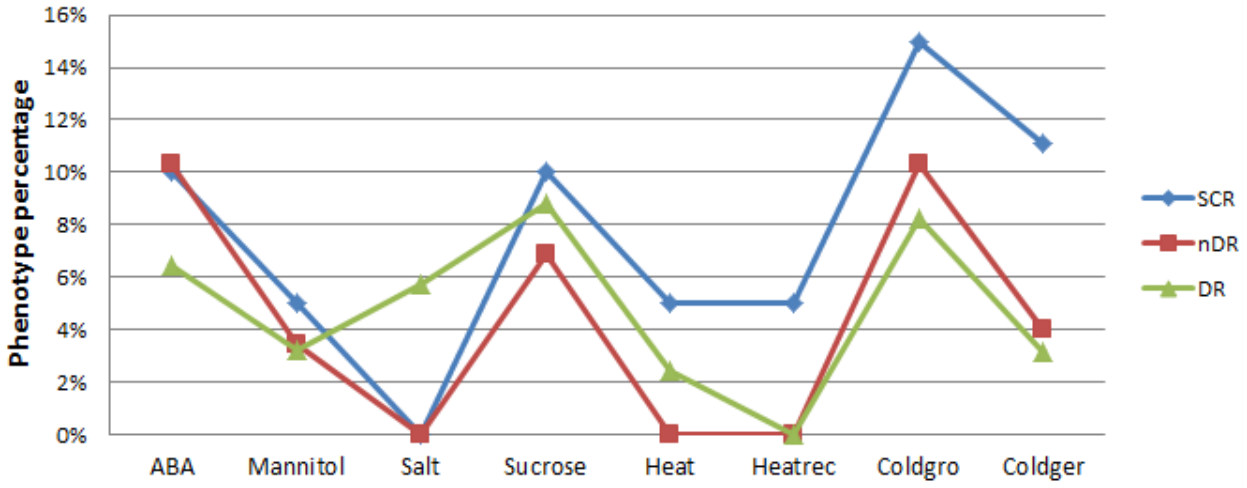

Figure 3.6 Stress phenotype percentage comparison for HM SALK lines. Heatrec stands for ‘heat recovery’. Colgro stands for ‘cold growth’. Coldger stands for ‘cold germination’.

CHAPTER 4

DUPLICATION AND OVEREXPRESSION OF DUPLICATION-RESISTANT GENES: A GLIMPSE INTO POTENTIAL OUTCOMES

Chengbo Zhou, Kenneth Feldmann, Kathryn Millward and Andrew H. Paterson. To be submitted to Physiologia Plantarum.

ABSTRACT

In contrast to the highly duplicated nature of plant genomes, 'duplication-resistant' (DR) genes have been recurrently restored to low dosage following independent genome duplications in various species. Their low copy status, representing a highly biased fate after gene duplication, implies that one copy of these genes is necessary for maintaining proper fitness but that multiple copies are less beneficial than single copies, or even deleterious. Complementing a screen for knockout (KO) phenotypes that focused on the necessity of DR genes, a phenotype search was conducted on genotypes with DR genes artificially duplicated or overexpressed, to examine the consequences of having multiple copies of these genes. Specifically, Arabidopsis addition (AD) lines each containing an extra copy of a target gene to mimic gene duplication, and overexpression (OE) lines with DR genes downstream of 35S promoters to amplify gene expression, were phenotyped in a similar manner as KO mutants. The observed AD phenotypes show DR genes to be more prone than other single copy genes when duplicated to cause phenotypes, all under stress conditions and most being mild. Yet, AD and OE phenotypes of DR genes are not always negative, as stress tolerant phenotypes occurred at an almost equal frequency as sensitive ones. Seedling phenotypes, absent in AD lines, were found in OE lines at a higher percentage than KO mutants. Genes with both OE and KO phenotypes, constituting approximately a third of the OE sample, are considered dosage sensitive. Thus, dosage sensitivity may be one major cause of duplication resistance, yet is not prevalent in organelle-targeting genes and those with a lethal KO phenotype. These observations increase knowledge of duplication resistance and its intricate nature. The plant resources created in this study warrant further investigation and will certainly benefit continuing effort to understand both duplication resistance and plant gene function generally.

INTRODUCTION

The fates of duplicated genes are an important part of gene and genome evolution [1]. To retain or purge a copy of each newly duplicated gene pair can have an enormous collective impact on genome biology and evolution, especially following whole genome duplication in which tens of thousands of genes are duplicated simultaneously. While several gene functional groups are preferentially preserved in duplicate [2-7] and comprise large multigene families, Most gene functional groups show post-duplication gene preservation/loss rates that are indistinguishable from the genome-wide average. Such 'neutral' loss of duplicated genes presumably involves inactivating mutations opposed by very weak selection [8], although active mechanisms of gene elimination have also been suggested [9].

A small group of 'duplication-resistant' (DR) genes [10] have been recurrently restored to low dosage following independent duplication events in taxa as divergent as plants, fishes, and yeast. The existence of DR genes is most renowned in angiosperm genomes, all of which experienced one or more whole genome duplications (WGDs) [11] and numerous small scale duplications (SSDs) [12]. Thanks to the richness of ancient duplication events, various groups of plant DR genes have been identified based mainly on their copy numbers among several plant species, and their features described [13-15]. While some common features of these different gene sets have been discerned (Chapter 3), a clear explanation of duplication resistance is not yet evident.

The defining characteristic of restoring low gene dosage, observed in divergent plant genomes with varying evolutionary rates, suggests that duplicated copies of DR genes disappear relatively rapidly. Sequence divergence due to fast evolution [16] and sequence erosion by accumulated mutations [17] are among causes of duplicate gene removal. Given that DR genes

are widely conserved, they are presumably slow-evolving [15], making the more likely scenario that duplicated copies were silenced and the non-functional sequences degenerated within a relatively short period. Therefore, it is reasonable to assume that their duplication may lead to some negative effect that was selected against.

It is not rare that gene duplication reduces fitness of an organism. Several severe human diseases are a direct consequence of gene duplication, for example, Charcot-Marie-Tooth (CMT), Parkinsons and Alzheimers diseases [18]. Such gene duplication-associated mutations, on the other hand, have been rarely reported in plants, perhaps because plant genomes are relatively tolerant to duplications, which occur frequently at various scales by natural causes. In Arabidopsis, where transgenic techniques are well developed, one can artificially 'duplicate' a chosen gene by transforming a copy into the genome. The resulting transgenic 'addition (AD) lines' (detailed in 'Materials and Methods'), allow visualization of direct outcomes from DR gene duplication, and thus may aid in the interpretation of duplication resistance.

For disorders resulting from gene duplication, the dosage increase is usually the cause [19-21]. The normal function of a protein complex, regulatory network or pathway may depend on a proper relative dosage of each member to others, and the duplication of one member would disturb the dosage balance [22]. Further, elevated protein amounts alone may cause problems. A high concentration of certain proteins may trigger the formation of aggregates [18, 23] or encourage non-specific binding to other proteins, preventing themselves or others from fulfilling their duties. Therefore, increasing gene expression level is considered an alternative strategy to evaluate duplication effects.

Overexpression (OE) lines, in which target genes are constitutively overexpressed by the CaMV 35S promoter (detailed in 'Materials and Methods'), were made to examine the

consequences of dosage increase for DR genes. The expression elevation in OE lines is expected to be more striking than that of gene duplication, thus may magnify mild AD phenotypes and render them visible. For genes that are not constitutively and ubiquitously expressed, ectopic expression may also occur, yet is likely not to mask the overexpression effect in most OE lines, as was implied by the similar phenotypes induced by 35S promoter and 35S enhancer [24-30] (which theoretically does not alter the endogenous expression pattern [31]).

As a powerful tool in gene functional study, OE lines will contribute to the interpretation of duplication resistance. Being an important variable in affecting evolutionary fate after gene duplication, gene function holds an irreplaceable piece toward solving the puzzle of persistent single copy gene dosage. Overexpression (OE) phenotypes, in many cases opposite the knockout (KO) phenotype of the same gene [25, 27, 32, 33], reflect gene function as KO phenotypes do. For example, an early flowering KO phenotype may correspond to a late flowering OE phenotype, reconfirming the protein's presumed role in flowering time control. Besides being a nice complement to KO phenotypes, OE lines would also shed light on functions of genes without KO phenotypes or lacking homozygous KO mutants [30].

In order to test the effect of duplication or expression increase for DR genes, and also to continue their functional investigation following our previous KO mutant study (Chapter 3), we made AD and OE lines for 11 and 64 DR candidates whose SALK lines were involved in the KO phenotype study, respectively. These lines, together with the AD and OE lines of some control genes, were screened for their root and stress-response phenotypes as were KO mutants (Chapter 3). 'Soil growth phenotypes', were also noted while growing AD or OE seedlings on soil under normal conditions. The phenotypic features of DR genes, compared to non-DR genes and KO mutants, were described.

MATERIALS AND METHODS

Target genes for AD and OE mutants

The initial Arabidopsis DR gene list contains strict singletons (SSs), which are single copy in five plant species (*Arabidopsis thaliana, Carica papaya, Populus trichocarpa, Vitis vinifera, and Oryza sativa*), and protein domain (PD) genes, which are single copy in *Arabidopsis thaliana* and *Oryza sativa* and have protein functional (Pfam) domains that are significantly enriched in singletons [34].

The initial DR candidates conserved in at least 15 of the 16 genomes (*Manihot esculenta, Ricinus communis, Populus trichocarpa, Medicago truncatula, Glycine max, Cucumis sativus, Prunus persica, Arabidopsis thaliana|Arabidopsis lyrata, Carica papaya, Citrus sinensis|Citrus clementina, Eucalyptus grandis, Vitis vinifera, Mimulus guttatus, Aquilegia coerulea, Sorghum bicolor|Setaria italica|Oryza sativa|Brachypodium distachyon, Zea mays,* where average counts were taken for two arabidopsis, four diploid grass, and two citrus, with the exception that the count cannot be zero as long as one lineage is nonzero) were considered the final DR candidates (Table S3.1). From 179 final DR candidates with their SALK lines ordered (Table S3.2), 11 and 64 genes were randomly selected for making addition (AD) and overexpression (OE) lines, respectively (Table M4.1; Table M4.2).

The initial DR candidates conserved in less than 15 of the 16 genomes were included in the control group and named ‘nDR’. The control group also contains single copy random (SCR) genes, which are Arabidopsis singletons but not all single copy in the other four species (*Carica papaya, Populus trichocarpa, Vitis vinifera, and Oryza sativa*). From the 36 nDR and 29 SCR genes with their SALK lines ordered, AD lines were created for 14 genes and OE lines were

made for four genes (Table M4.1; Table M4.2). As the number of control genes in AD and OE mutants was small, SCR and nDR genes were considered a single control group instead of two.

Creating AD and OE lines

Both OE and AD lines were transgenic. OE transgenic constructs expressed full length cDNA under the cauliflower mosaic virus 35S promoter. AD transgenic constructs contained the genomic region of a target gene including the 1kb upstream and 500bp downstream flanking sequences, to attempt to capture their endogenous promoters and possible regulatory sequence downstream, respectively. The maximum length of flanking sequences were included in AD constructs for genes located less than 1kb away from their upstream neighboring genes and/or less than 500nt from their downstream neighboring genes. If a target gene has a distance less than 1kb from its upstream gene or less than 500bp from its downstream genes, the non-coding region flanking the gene was included as much as possible.

Genomic DNA was extracted using the Promega Wizard genomic DNA purification kit; colonies containing the BAC and cDNA were ordered from ABRC. BAC DNA was extracted following David Mark's protocol in Molecular Cloning 3rd edition [35], while cDNA was extracted by the Qiagen QIAprep Spin Miniprep Kit.

Upon proper dilution (typically 1:1000), DNA was amplified by PCR, using Takara's PrimeSTAR HS DNA polymerase to avoid errors during amplification. The PCR product was cleaned (Qiagen Qiaquick PCR Purification kit), and the size checked on an agarose gel. Adenine (A) overhangs were added to the cleaned PCR product with correct size. PCR products with A overhangs were integrated into a Gateway entry vector (Invitrogen pCR8/GW/TOPO TA Cloning kit), which was then transformed into one shot competent E. coli cells.

After patching on selective LB plates, plasmid DNA was extracted from each patch and digested with proper restriction enzymes. DNA from the patch(s) with desired insertion was extracted (Qiagen QIAprep Spin Miniprep Kit) and sequenced. For entry clones without a mismatch, a LR clonase reaction was performed with the destination vector (pEarleyGate 100 for OE constructs, and pEarleyGate 301 for AD constructs). The product was transformed into E.coli. DNA was extracted from desired clones using the procedure described in the above paragraph.

The border sequence of the destination vector was sequenced. The destination vector with the correct border sequence was transferred into GV3103 Agrobacterium through electroporation. Colonies with successful transformation was identified by Colony PCR. The confirmed colony is ready for Arabidopsis transformation.

Arabidopsis transformation method is modified from US patent 6353155 (http://patents.com/us-6353155.html). For each transgenic construct, 50ml LB with OD600 >2.0 was diluted with 100ml of 6% sucrose solution to reach an OD600 between 0.8 and 1. Silwet-L77 (0.03%) was added to the diluted LB solution as a surfactant. About ¾ of the solution was poured into a large weigh boat, in which we dipped the inflorescence of each chosen Arabidopsis plant for ten seconds. Dipped plants were laid horizontally in a flat covered by a clear plastic dome, kept in dark at room temperature overnight, then moved to a growth chamber.

Seeds harvested from the dipped plants were germinated on plates containing 50 ug/ml kanamycin antibiotics. Resistant seedlings were transferred to soil and their seeds collected at maturation.

Phenotyping

ABA, mannitol, salt, sucrose, heat, heat recovery, cold growth and root screens were done on AD and OE lines. Seeds were sterilized by 50% clorox solution with Triton X-100 (two

drops per 100ml solution) for 8 minutes, and then rinsed three times with sterile water. Sterilized seeds were planted on 0.5xMS (Murashige and Skoog) medium with 5g/L sucrose in heat, heat recovery, cold growth and root screens. MS medium (5g/L sucrose, 5g/L 2-(N-morpholino)ethanesulfonic acid (MES)) supplemented with 1.5μM ABA and 375mM mannitol was applied in ABA and mannitol screens, respectively. For salt screen, MS medium (5g/L sucrose) was supplemented with 125mM NaCl, and for sucrose screen, MS medium contained 300mM sucrose.

Six lines or five lines with control (wildtype Col0) were planted per square plate (15 x 100 x 100mm), and each line occupied one column of the bottom 6x6 grid. In each line about 10-15 seeds were planted evenly on the surface of the medium, in a straight line parallel to one side of a plate.

After planting seeds, plates were put in a growth chamber with standard growth condition (22°C, 100 μEinsteins light intensity and 16:8 hr light:dark cycle) except for root, heat and heat recovery plates. Root plates were placed at 4°C for 3 days before being kept vertical under standard growth condition, and heat/heat recovery plates were put in a 34°C growth chamber. After 7 days, heat recovery plates were moved to the 22°C growth chamber and cold growth plates were moved to a 10°C growth chamber.

Plates were photographed after 4 weeks (for the first round of screens on AD lines), or every week for 2 to 4 weeks starting at the 14th day after planting. The process from seed planting to the last photographing was referred as one trial. The photographs were visually assessed, and one phenotype score was assigned to each mutant line per plate based on plant size or root length (in root screen). Plants smaller than the control were considered sensitive, while plants bigger than the control were thought as tolerant. In most cases, the score is an integer

ranging from 0 to 5, with 0 representing ‘no germination’, 1 for ‘very sensitive’, 2 for ‘somewhat sensitive’, 3 for ‘normal’, 4 for ‘somewhat tolerant’, and 5 for ‘very tolerant’. In root screens, scores represented the root length, with 1 being ‘short’, 2 being ‘slightly short’, 3 being normal, 4 being ‘slightly long’, and 5 being ‘long’. When only a few (<50%) individuals within a line showed phenotypes, the score ‘seg’ (for segregating) was assigned.

For each line, average of scores at different time points was calculated. The average, in combination of the frequency of score ‘3’, was used to generate a trial score, which would be 1 (average<2.5 and no ‘3’), 2 (average<2.5 and one ‘3’), 3 (2.5≤ average ≤3.5), 4 (average>3.5 and one ‘3’), and 5 (average>3.5 and no ‘3’). The trial scores were combined to create a final score, if two trials were done. Similar scores were interpreted as an increase in the level of confidence with which sensitivity or tolerance was inferred. For example, if a mutant was assigned score 2 in both trials, then the final score would be 1. In contrast, distinctive scores were interpreted as a decrease in the level of confidence with which sensitivity or tolerance was inferred. For example, if a mutant was 1 in one trial and 3 in the other, the final score would be 2. If there was only one score for a line, that score was ‘single read’, which was not considered valid result. In OE lines, the photographs of heat screen were missing and there were all ‘single reads’ in heat recovery screen. Therefore, no results were displayed for the two screens on OE lines.

RESULTS

AD phenotypes

No soil growth phenotype was found in the AD lines of DR or control genes. In the root screen, one DR AD line (A3_11) segregated for seeds failing to germinate. A3_11 also segregated for this phenotype in every other screen, suggesting that the non-germination phenotype is not specific to root or any stress but likely to be visible under normal conditions.

Seeds which did not germinate were hard to detect in soil when a majority of the population was normal, perhaps contributing to its lack of soil growth phenotype. In addition, A3_11 seedlings were larger in the heat recovery screen and smaller in sucrose and cold growth screens (Figure 4.1) compared to WT, presumably reflecting stress-triggered aspects of its phenotype.

In four of the seven stress screens, namely, salt, sucrose, heat and cold growth, photographs were available for only one trial, in which a single sample of seedlings for each line was evaluated once per week for up to four weeks. Consequently, the results were considered preliminary. The number of phenotypes in these four screens were generally higher than the other three (ABA, mannitol, and heat recovery) (Table 4.1), possibly due to false-positive phenotypes. Nevertheless, the DR group also has more AD lines showing stress phenotype(s) than non-DR group in screens with two trials (Table 4.1), indicating that duplication of DR genes is more likely to change phenotype under stress conditions than other conserved single copy genes.

In all three stress screens with two trials, the numbers of AD lines with phenotypes were higher in DR than control, where no phenotype was detected (Table 4.1). Three DR genes had a phenotype in their AD lines, two being striking. In addition to A3_11, A1_49 was slightly sensitive in ABA screen and A2_4 was tolerant in heat recovery screen (Figure 4.1). A1_49 corresponded to a knockout phenotype, being dwarf with curled leaves, whereas no phenotype was detected in KO3_11 (SALK_076441C) and KO2_4 (SALK_027781C). Two of the three DR genes (1_49 and 2_4) had a predicted function related to transcription or translation, a major functional category in DR KO mutants revealing striking stress phenotypes, with one involved in RNA modification and the other being a transcription regulator. The remaining gene (3_11),

whose AD phenotype (non-germination) might not be stress-triggered, was a nucleoside triphosphate hydrolase.

The greater number of DR AD lines than controls with phenotypes was also evident in screens with only one trial, except heat and cold (Table 4.1) where the DR phenotype number was similar to or less than the control. This observation is consistent with the conclusion from analysis of KO lines (Chapter 3) that DR genes are not highly involved in temperature stress response. Consistency with KO analysis was also found in salt screen, in which the phenotype number was highest in DR AD lines (Table 4.1), further supporting the special correlation between salt response and duplication resistance (Chapter 3).

Unlike DR KO phenotypes, a majority of which were sensitive (Table 3.4; Table3.5), the numbers of tolerant and sensitive DR AD phenotypes (non-segregating) were similar (Table 4.2). In the control group, there were even more tolerant AD phenotypes than sensitive AD phenotypes (Table 4.2). Collectively, the available data suggested that adding an extra copy might often confer stress tolerance for Arabidopsis single copy genes that are conserved in other species.

As was seen in A1_49, AD and KO phenotypes (ABA sensitivity vs. leaf abnormality) were not necessarily connected in a direct manner. Yet for two genes, AD phenotypes were similar to knockout or knockdown phenotypes. In the DR group, both AD and KO mutants of AT3G17670 were sensitive to sucrose, indicating a possible role of this gene in sugar metabolism. In the control group, the AD and a presumed knockdown mutant of AT1G50170 (SALK_086731C, which is viable while the null mutant of the gene should be lethal) were both slightly tolerant to cold stress, suggesting its involvement in cold response in addition to being essential for plant survival [36].

OE phenotypes

There are only four control genes with OE lines, making it difficult to discern patterns differentiating these from DR genes. Therefore in this section, we focused on describing the phenotypes exhibited by DR OE lines with a brief mention of non-DR OE phenotypes.

While no soil growth phenotype was found in the four non-DR OE lines, ten of the 64 DR OE lines had soil growth phenotypes revealed by 25% or more of T1 seedlings (Table 4.3). Similar to KO mutants with soil growth phenotypes, most (seven) of the ten OE lines were 'small' (Table 4.3). Four of the small OE lines had no other mutant phenotypes, making 'small' the main category of OE soil growth phenotypes. The other equally large category is 'fertility', containing four OE lines with either reduced or no fertility. The remaining two OE lines had 'flowering' and 'leaf' phenotypes, respectively.

Five of the ten OE lines corresponded to genes with observed or published knockout or knockdown phenotypes (Table S4.1), among which one was similar to the OE phenotype. The knockdown mutant of AT3G17590 (BSH for 'bushy', a subunit of the SWI/SNF complex) is sterile [37], and its overexpression line (OE6) has reduced fertility. Among the other four genes with different KO and OE phenotypes, two proteins (NDHO and ATVPS11) are also components of protein complexes, and the other two are involved in the electron transfer chains of chloroplast and mitochondrion respectively (Table S4.1).

Despite the relatively high above ground phenotype percentage (PP), the root PP (3%) of DR OE lines and DR KO mutants (5%) were both low. The OE root phenotypes were mild and not associated with any OE soil growth phenotype. Unlike KO root mutants, half of which involved chloroplast-targeting genes (Chapter 3), neither of the two OE lines with root phenotypes corresponded to chloroplast-located proteins. OE3_36, with slightly long roots, had

slightly short roots in its corresponding KO mutant (SALK_090814C) In addition to root phenotypes, stress phenotypes were also found in KO3_36 (SALK_090814C) and OE3_36. The other OE root mutant, OE141, did not have any phenotype observed in its KO mutant. No root phenotypes were detected in non-DR OE lines.

All OE stress phenotypes were mild (scored '2' or '4') in DR and non-DR genes. The overall DR OE stress PP, defined as the percentage of OE lines showing a phenotype in at least one stress screen (45%; 29/64), was much higher than the overall DR KO stress PP (29%). A high stress PP was also found in the non-DR group, where three of the four lines had a stress phenotype. About 51% (19/37) of DR OE stress phenotypes were tolerant, whereas in DR KO mutants the tolerant phenotypes constituted of all observed stress phenotypes (Chapter 3). Similarly to KO mutants, DR OE stress PPs varied among the screens and the cold growth PP was the highest (Table 4.4).

Four DR OE lines had both stress and soil growth phenotypes (penetrance $\geq$ 25%). The remaining 25 DR genes with only OE stress phenotypes had a variety of observed KO phenotypes, including one in 'root', five in 'stress', four in 'soil growth (SG)', and two in 'potentially lethal (PL)' (Table S4.2). Among DR genes with stress-related OE and KO phenotypes, none had a predicted function directly pointing to stress signaling. Instead, three were unknown and two were related to transcription or translation regulation (Table S4.3). OE and KO phenotypes were similar for two DR genes (AT1G01920 and AT1G62250): OE79 and KO79 (SALK_060683C) were sensitive in the salt screen, and OE100 and KO100 (SALK_065936) were sensitive in the cold growth screen. In contrast, the other three DR genes showed OE and KO phenotypes in different stress screens, suggesting that they might participate

in multiple stresses' pathways or processes that were not directly related to stress but weakened plant in general once disturbed.

Removing those with OE soil growth phenotypes (penetrance>=25%), there were 29 DR genes with no observed OE phenotype. Fourteen of the 29 DR genes had observed KO phenotypes, half of which were PL and the other half stress related (Table 4.5; Table S4.4). Compared to genes with observed OE phenotypes, the proportion of KO PL phenotypes was higher in genes without observed OE phenotypes, while the proportion of KO soil growth phenotypes was much lower (Table 4.5). This observation also held true for published KO phenotype data (Table 4.5), indicating that the severity of KO phenotypes was not positively correlated with the likelihood of having OE phenotypes in DR genes.

AD phenotypes seemed to be more similar to OE phenotypes than KO phenotypes of the same genes. AD and OE phenotypes appeared in the same stress screen in four of the five genes (AT2G20980, AT1G56345, AT1G63980, AT1G50170) for which phenotypes were detected in both AD and OE lines, while only two genes showed KO and OE phenotypes in the same stress screen. Interestingly, OE phenotypes were not severe versions of AD phenotypes in the four genes. Both AD and OE phenotypes were slightly tolerant (score '4') in two genes, while the AD phenotype was slightly tolerant and OE phenotype was slightly sensitive (score '2') in the other two genes.

DISCUSSION

The persistent single copy status of some genes, herein called DR genes, has long been puzzling to evolutionary biologists. To unveil the basis of duplication resistance, which represents important mechanisms for gene and genome evolution, a thorough understanding of DR genes' function will be extremely helpful, if not absolutely necessary [13] (Chapter 3).

Among the various approaches for gene function analysis, we previously investigated knockout (KO) mutants, the phenotype screens of which constituted the initial phase of our research on DR genes. In the present stage, Arabidopsis AD and OE lines were made and phenotyped, to further advance functional investigation and test empirically whether DR genes confer a fitness reduction when present in more than one copy or expressed at a high level. The phenotypic features of these lines, as summarized and explained below, shed new light on this unique population of plant genes and their choice to stay single.

No soil growth phenotype was observed in DR AD lines, indicating that extra copies of these genes generally do not confer a drastic change. This finding was expected, because given that hundreds of DR genes exist, if the duplication of most DR genes resulted in visible morphological alternation, the newborn polyploid species would die quickly. Instead, the duplication effect is more likely to be mild, compatible with our observation, or take numerous generations to accumulate to a detectable level. The mild effect of an extra copy may be more visible on population level than on individual plant. For example, duplication of a DR gene may confer disadvantage in competition with other plants, causing continuous decrease in frequency of plants with the duplicated gene in a population. These possibilities could be tested by methods designed for revealing subtle phenotypes or by multigenerational study that has been used in knockout mutants [38].

More AD phenotypes were found in DR genes than in other single copy genes, consistent with the expectation that duplication of DR genes is more likely to affect phenotype than non-DR singletons. Interestingly, some DR AD lines showed mild stress tolerance. Most mild tolerant phenotypes came from stress screens with only one trial, thus needing further validation. However, the appearance of a striking tolerant phenotype (A2_4) in such a small sample implies

that the duplication of some DR genes may indeed confer stress tolerance. Much further testing is needed to determine if a mild degree of stress tolerance, superficially seeming to be a potential fitness advantage, translates into increased or decreased overall seed output. Functions mitigating stresses in unfavorable environments sometimes do more harm than good in optimal environments, where their consistent activation may compromise normal growth and reproduction. Indeed, increased expression of stress tolerance genes has been reported to cause reduced stature or fitness [39, 40]. Such adverse effects, though perhaps individually mild, may add on one another and be intensified in a genome duplication event.

Similar to the findings from our KO phenotype study, genes that may affect transcription or translation dominated the promising stress-response candidates identified in AD or OE lines. Among the two genes with striking AD stress phenotypes, one is involved in DNA or RNA metabolism and the other is annotated as a transcription regulator. Though no OE stress phenotype is striking, five genes with both OE and KO stress phenotypes have known annotations (RNA modification and SET domain protein) related to transcription or translation. Therefore, in this broad functional category of DR genes, some may serve as good targets for molecular breeding of crops with improved stress tolerance.

Seedling phenotypes with greater than or equal to 25% penetrance occurred in ten DR OE lines, making the phenotype percentage (15.6%) higher than that of DR HM KO lines (9.2%). Though ‘small’ remains a main feature for these OE lines, no pigment phenotype was detected, despite the fact that five of the ten proteins are located in chloroplast. In combination with previous results (Chapter 3), chloroplast-targeting DR genes tend not to cause pigment abnormality in either KO or OE lines. Instead, they affected a wide range of processes including root or leaf development, stress response and reproduction. On the other hand, ‘fertility’ was a

large subgroup of OE soil growth phenotypes, containing four OE lines with reduced fertility. Little or no fertility, found in 30% of 1262 activation-tagging lines revealing a phenotype [41], seems to be one of the major outcomes of Arabidopsis gene overexpression. Therefore, whether DR genes are more prone to fertility (and fitness) reduction than other genes upon overexpression remains unknown, requiring testing of more OE control genes.

Genes possessing both OE and KO phenotypes, compromising about one third of the DR gene sample with OE lines, show behavior that is consistent with dosage sensitivity. For this fraction of DR genes, dosage sensitivity may have contributed to their persistent single copy status, as gene removal and gene duplication are associated with dosage change. Three of the five dosage sensitive proteins (BSH, NDHO and ATVPS11) with OE soil growth phenotypes were components of protein complex (Table S4.1). Moreover, most (six of nine) DR genes encoding subunits of protein complexes are dosage sensitive (Table S4.5), supporting what would be predicted from the dosage balance hypothesis [42]. In contrast, only a small proportion (28%; 10/36) of chloroplast or mitochondrion targeting DR genes are dosage sensitive (Table S4.6), a finding not compatible with the notion that they are in dosage balance with their organelle genome-encoded partners [14, 23]. Although the test sample is small, the frequency of AD phenotypes seems to be lower in organelle-targeting DR genes (50%; 3/6) than other sampled DR genes (80%; 4/5), suggesting that neither dosage increase nor gene duplication in this DR subset is prone to cause immediate harm. Instead, the long term existence of a duplicated copy may be problematic. For example, mutations accumulated on the regulatory or coding sequence of either copy may interrupt the normal functioning of the pathway in which the gene participated.

The respective effects of knockout and overexpression are neither similar nor opposite in most cases. Genes with OE soil growth phenotypes have various KO phenotypes, including those detectable only on biochemical level, stress sensitivity and lethality (Table S4.1). Likewise, a wide range of KO phenotypes were found in genes with OE stress phenotypes. Different KO and OE phenotypes of the same gene occurred in non-DR genes as well, suggesting that such incongruity is a general phenomenon rather than a feature specific to DR genes. The chance of showing OE phenotypes differs among genes in distinct KO phenotype groups: all sampled DR genes with soil growth KO phenotypes have OE phenotypes, while only a minority of DR samples with potentially or confirmed lethal KO mutants cause detectable phenotypes when overexpressed. Therefore, mechanisms other than dosage sensitivity may be responsible for the duplication resistance of essential DR genes. For example, their functions, representing perhaps the most fundamental metabolic steps, might be unable to split into subfunctions and lack the potential to evolve novelty, rendering the duplicated copies useless.

A few genes had OE and KO phenotypes observed in the same screen, relating their functions to root growth, fertility, or a particular stress. The opposite root phenotypes in OE3_36 and KO3_36 suggested that root growth was promoted by the overexpression of AT1G56345 and retarded at its knockout. Being a pseudouridine synthase, AT1G56345 may positively regulate the translation of genes that induce root growth. Of the three genes where the KO and OE phenotypes were similar, one (AT_3G17590) was implicated in stoichiometry, the disturbance of which by knockout and overexpression may affect the same process thus exert similar phenotypic effect [18]. Given the intricacy of stress signaling network, however, perturbation on different functional pathways can lead to the same phenotype. The expression profile may be distinctive in the KO and OE lines of AT1G01920, a SET domain protein, yet

some of the transcriptional changes in both lines may lead to salt sensitivity. The correlation of AT1G62250's function with cold response, on the other hand, serves as a starting point to study this unknown gene. Collectively, these OE phenotypes helped to point out or narrow down a realm of possibilities that future functional studies could focus on.

In addition to providing insights on duplication resistance and gene function, the abundance of AD and OE phenotypes shows these lines to be practical tools in investigating the effects of duplication and overexpression of DR genes. Application of OE and especially AD lines in a larger gene sample is expected to reveal more phenotypes, which may help to validate the previous findings and lead to new discoveries. Further, analysis other than phenotype screens could be done using available AD or OE plants. Through evaluating the expression of key growth or stress signaling regulators in some promising lines, we might be able to connect DR genes with known pathways, thus providing more clarity on their functions. The possible consequence of having a mutation-bearing duplicated copy can be visualized via plants transformed with artificially mutated AD constructs. AD or OE constructs could also be transformed into other plant species, for extensive functional study of DR genes. Together, the plant and phenotype resources described here, along with the transgenic constructs, will contribute to future investigation in related research areas.

REFERENCES

1. Feldmann KA, Wierzbicki AM, Reiter RS, Coomber SA: T-DNA Insertion Mutagenesis in Arabidopsis - a Procedure for Unraveling Plant Development. Plant Molecular Biology 2 1991, 212:563-574.

2. Chapman BA, Bowers JE, Feltus FA, Paterson AH: Buffering crucial functions by paleologous duplicated genes may impart cyclicality to angiosperm genome duplication. Proc Natl Acad Sci U S A 2006, 103:2730-2735.

3. Seoighe C, Gehring C: Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. Trends in Genetics 2004, 20:461-464.

4. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y: Modeling gene and genome duplications in eukaryotes. Proceedings of the National Academy of Sciences of the United States of America 2005, 102:5454-5459.

5. Blanc G, Wolfe KH: Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 2004, 16:1679-1691.

6. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH: Unraveling Ancient Hexaploidy through Multiply-aligned Angiosperm Gene Maps. Genome Research 2008, 18:1944-1954.

7. Paterson AH, Chapman BA, Kissinger J, Bowers JE, Feltus FA, Estill J, Marler BS: Convergent retention or loss of gene/domain families following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces, and Tetraodon. Trends in Genetics 2006, 22:597-602.

8. Haldane JBS: The part played by recurrent mutation in evolution. The American Naturalist 1933, 67:5–19.

9. Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M: Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. PLoS Biol 2010, 8:e1000409.

10. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH: Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 2008, 18:1944-1954.

11. Wang Y, Wang X, Paterson AH: Genome and gene duplications and gene expression divergence: a view from plants. Ann N Y Acad Sci 2012, 1256:1-14.

12. Barker MS, Baute GJ, Liu S-L: Duplications and turnover in plant genomes. Springer; 2012.

13. Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J, dePamphilis CW: Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. BMC Evol Biol 2010, 10:61.

14. De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y: Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci U S A 2013, 110:2898-2903.

15. Armisen D, Lecharny A, Aubourg S: Unique genes in plants: specificities and conserved features throughout evolution. BMC Evol Biol 2008, 8:280.

16. Domazet-Loso T, Tautz D: An evolutionary analysis of orphan genes in Drosophila. Genome Res 2003, 13:2213-2219.

17. Lynch M: The Evolutionary Fate and Consequences of Duplicate Genes. Science 2000, 290:1151-1155.

18. Conrad B, Antonarakis SE: Gene duplication: a drive for phenotypic diversity and cause of human disease. Annu Rev Genomics Hum Genet 2007, 8:17-35.

19. Murakami T, Garcia CA, Reiter LT, Lupski JR: Charcot-Marie-Tooth disease and related inherited neuropathies. Medicine (Baltimore) 1996, 75:233-250.

20. Singleton A, Gwinn-Hardy K: Parkinson's disease and dementia with Lewy bodies: a difference in dose? Lancet 2004, 364:1105-1107.

21. Hardy J: Amyloid double trouble. Nat Genet 2006, 38:11-12.

22. Veitia RA: Gene dosage balance: deletions, duplications and dominance. Trends Genet 2005, 21:33-35.

23. Edger PP, Pires JC: Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res 2009, 17:699-717.

24. Gao D, Appiano M, Huibers RP, Chen X, Loonen AE, Visser RG, Wolters A-MA, Bai Y: Activation tagging of ATHB13 in Arabidopsis thaliana confers broad-spectrum disease resistance. Plant Mol Biol 2014, 86:641-653.

25. Bueso E, Ibañez C, Sayas E, Muñoz-Bertomeu J, Gonzalez-Guzmán M, Rodriguez PL, Serrano R: A forward genetic approach in Arabidopsis thaliana identifies a RING-type ubiquitin ligase as a novel determinant of seed longevity. Plant Science 2014, 215:110-116.

26. Kim S-G, Lee S, Kim Y-S, Yun D-J, Woo J-C, Park C-M: Activation tagging of an Arabidopsis SHI-RELATED SEQUENCE gene produces abnormal anther dehiscence and floral development. Plant Mol Biol 2010, 74:337-351.

27. Fang W, Wang Z, Cui R, Li J, Li Y: Maternal control of seed size by EOD3/CYP78A6 in Arabidopsis thaliana. Plant Journal 2012, 70:929-939.

28. Wang R, Liu X, Liang S, Ge Q, Li Y, Shao J, Qi Y, An L, Yu F: A subgroup of MATE transporter genes regulates hypocotyl cell elongation in Arabidopsis. J Exp Bot 2015, 66:6327-6343.

29. Gomez MD, Urbez C, Perez-Amador MA, Carbonell J: Characterization of constricted fruit (ctf) mutant uncovers a role for AtMYB117/LOF1 in ovule and fruit development in Arabidopsis thaliana. PLoS One 2011, 6:e18760.

30. Shao J, Liu X, Wang R, Zhang G, Yu F: The over-expression of an Arabidopsis B3 transcription factor, ABS2/NGAL1, leads to the loss of flower petals. PLoS One 2012, 7:e49861.

31. Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, et al: Activation tagging in Arabidopsis. Plant Physiology 2000, 122:1003-1013.

32. Qin LX, Li Y, Li DD, Xu WL, Zheng Y, Li XB: Arabidopsis drought-induced protein Di19-3 participates in plant response to drought and high salinity stresses. Plant Mol Biol 2014, 86:609-625.

33. Zhang JZ: Overexpression analysis of plant transcription factors. Curr Opin Plant Biol 2003, 6:430-440.

34. Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC: Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. Trends Genet 2006, 22:597-602.

35. Sambrook J, Fritsch EF, Maniatis T: Molecular cloning. Cold spring harbor laboratory press New York; 1989.

36. Saha K, Webb ME, Rigby SE, Leech HK, Warren MJ, Smith AG: Characterization of the evolutionarily conserved iron-sulfur cluster of sirohydrochlorin ferrochelatase from Arabidopsis thaliana. Biochem J 2012, 444:227-237.

37. Brzeski J, Podstolski W, Olczak K, Jerzmanowski A: Identification and analysis of the Arabidopsis thaliana BSH gene, a member of the SNF5 gene family. Nucleic Acids Res 1999, 27:2393-2399.

38. Gilliland LU, McKinney EC, Asmussen MA, Meagher RB: Detection of deleterious genotypes in multigenerational studies. I. Disruptions in individual Arabidopsis actin genes. Genetics 1998, 149:717-725.

39. Yoshida T, Mogami J, Yamaguchi-Shinozaki K: ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. Curr Opin Plant Biol 2014, 21:133-139.

40. Atkinson NJ, Urwin PE: The interaction of plant biotic and abiotic stresses: from genes to the field. J Exp Bot 2012, 63:3523-3543.

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

41. Birchler JA, Riddle NC, Auger DL, Veitia RA: Dosage balance in gene regulation: biological implications. Trends Genet 2005, 21:219-226.

Table 4.1 Number of AD lines with phenotype(s). Heatrec stands for ‘heat recovery’. Coldgro stands for ‘cold growth’.

	Screens with two trials					Screens with one trial					
	ABA	**Mannitol**	**Heatrec**	**Root**	**Overall[1]**	**Salt**	**Sucrose**	**Heat**	**Coldgro**	**Overall[2]**	**Overall[3]**
DR	2	1	2	1	3	4	3	1	2	6	7
Control	0	0	0	0	0	1	1	2	2	5	5

1. Number of AD lines with phenotype in at least one of the ABA, mannitol, heat recovery and root screens; the number remains the same without the root screen.

2. Number of AD lines with phenotype in at least one of the salt, sucrose, heat and cold growth screens.

3. Number of AD lines with phenotype in at least one of the eight screens; the number remains the same without root screen.

Table 4.2 Number of sensitive, tolerant, and segregating phenotypes. Sensitive phenotypes are with score 0, 1 or 2. Tolerant phenotypes are with score 4 or 5. Seg stands for ‘segregating’.

	Screens with two trials				**Screens with one trial**				
	Sensitive	**Tolerant**	**Seg**	**Total[1]**	**Sensitive**	**Tolerant**	**Seg**	**Total[2]**	**Total[3]**
DR	1	1	4	6	3	4	2	9	15
Control	0	0	0	0	1	5	0	6	6

1. The sum of 'sensitive', 'tolerant' and 'seg' in screens with two trials (excluding root screen).

2. The sum of 'sensitive', 'tolerant' and 'seg' in screens with one trial.

3. The sum of 'sensitive', 'tolerant' and 'seg' in all stress screens.

Table 4.3 Soil growth phenotypes of DR OE lines. Penetrance equals to the percentage of T1 plants showing a phenotype.

Mutant_ID	Gene	Penetrance	Category	Description
OE2	AT1G74880	25%	small	small
OE6	AT3G17590	73%	fertility	Reduced fertility
OE46	AT3G09580	50%	small	small
OE49	AT5G48440	42%	small	small
OE66	AT2G05170	67%	fertility	small and purple; no seeds
OE120	AT2G43400	75%	flowering	large, late flowering
OE122	AT2G46060	75%	small	very small
OE128	AT3G09210	>90%	fertility	small, sterile
OE161	AT4G30840	25%	fertility	small, sterile
OE162	AT4G31460	>50%	leaf	wavy and curled leaves

Table 4.4 Total number of DR OE lines and phenotype percentages in each stress screen. Total number of OE lines is the number of OE lines with a valid phenotype score. Coldgro stands for 'cold growth'.

	ABA	**Mannitol**	**Salt**	**Sucrose**	**Coldgro**
Total number of OE lines	61	61	61	62	63
Phenotype percentage	7%	0%	16%	15%	22%

Table 4.5 Proportions of genes with KO phenotypes in OE lines with and without phenotypes. Observed KO phenotypes are from our previous KO phenotype screen (Chapter 3). Published KO phenotypes are from Lloyd and Meinke's dataset [1] and other publications regarding to specific genes. ESN stands for essential. MRP stands for morphological (detailed in [1])

OE lines	**Observed KO phenotypes**				**Published KO phenotype[1]**	
	PL	**Soil growth**	**Root**	**Stress**	**ESN**	**MRP**
with phenotype	0.06	0.11	0.06	0.17	0.06	0.23
without phenotype	0.21	0.00	0.00	0.21	0.14	0.07

1. Lloyd J, Meinke D: A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. Plant Physiology 2012, 158:1115-1129.

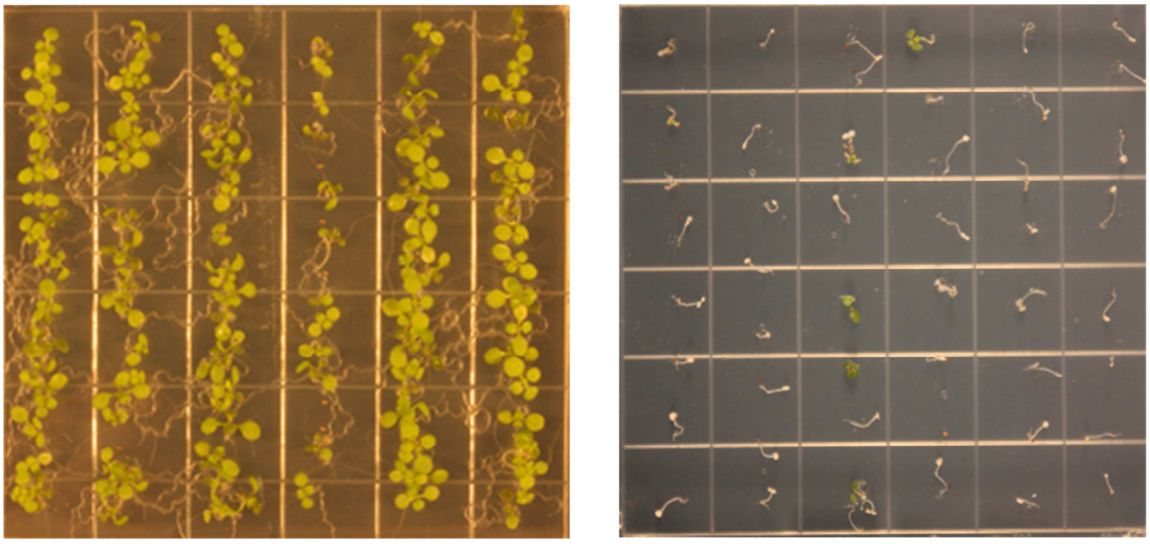

Figure 4.1 Phenotypes of AD lines. Left figure shows A3_11 in cold growth screen. Right figure shows A2_4 in heat recovery screen.

CHAPTER 5

CONCLUSION

In this study, stress and root phenotype screens were conducted on three genotypes of DR genes, namely, knockout mutant (KO), addition (AD) and overexpression (OE) lines, in order to better understand DR gene functions and to investigate their duplication and dosage increase effects. The KO and AD lines of some randomly selected non-DR Arabidopsis single copy genes were also screened for phenotypes, as controls.

A variety of phenotypes, including seedling phenotypes under normal conditions, were found in KO mutants (SALK lines) of DR genes. Of the intended screens, including drought (ABA, mannitol), salt, sucrose, heat, heat recovery, cold growth, cold germination, and root screens, KO mutant phenotypes were detected in all but one (heat recovery). Plants grown on soil for genotyping and fresh seed production also exhibited various phenotypes. For SALK lines with T-DNA insertion but no HM plant detected, two thirds had phenotypes suggesting lethality. Accordingly, our observed KO phenotypes were divided into four main categories: stress, root, soil growth, and potentially lethal (PL).

The percentage of mutant lines showing phenotypes, referred to as phenotype percentage (PP), is used to compare the KO phenotypes of DR genes and controls. Unlike what public datasets suggested, our direct observations find DR genes to have higher percentages of potentially lethal and soil growth phenotypes than other singletons (‘OT’ in Chapter 1). Therefore, it is possible that DR genes are relatively more important to Arabidopsis than other singletons.

The stress PP of DR genes is lower than the primary control group (SCR), and similar to a supplemental control group (nDR), suggesting that DR genes are not highly involved in response to the tested stresses. Rather than reacting to environmental stimuli, DR genes may tend to have housekeeping functions, the disturbance of which might cause phenotypes more severe than those visible only under stress conditions. This hypothesis is supported by higher proportions of PL and soil growth phenotypes in DR genes than other singletons. Alternatively, DR genes' KO mutants might be more responsive to stresses causing DNA damage, such as UV light and γ-ray, which would be an intriguing screen to pursue in further study.

Compared to the control group, DR genes had lower proportions of ABA and cold sensitive phenotypes, suggesting less involvement in ABA-mediated stress signaling and lack of cold positive regulators. In contrast, salt phenotypes were only found in DR genes, and did not co-occur with mannitol phenotypes, indicating that DR genes tend to function in response to the non-osmotic component of salinity stress. IT is not yet clear if, or how, these phenotypic features may be related to duplication resistance.

DR genes that may affect transcription or translation of other genes, for example, transcription regulator or those involved in DNA or RNA metabolism, had a high knockout PP. Chloroplast –targeting DR genes, on the other hand, were slightly less likely to have phenotypes upon knockout than other DR genes. Chloroplast-targeting DR genes corresponded to various KO phenotypes, and occurred at a high frequency in phenotype categories such as root, sucrose and cold germination.

It is expected that some of our observed phenotypes were different from published loss-of-function phenotypes of the same genes, whose root or stress phenotypes may not have been previously screened for. Inconsistency between observed and published phenotypes for some

genes may come from non-target T-DNA insertions, or knockdown mutants, in which the more severe published phenotypes were absent. As the published phenotypes were usually confirmed by expression analysis, different alleles or complementary test, mutants used in our study were likely to be knockdown. Therefore, further validation of our observed phenotypes is necessary, though the observed phenotypes should remain true for the SALK lines used in this study.

AD lines, made by transforming the cloned genomic region of the target genes into Col0 Arabidopsis, were used to attempt to visualize gene duplication effects. Due to the small number of genes with AD lines and the similar sample sizes of DR and the control, phenotype number, defined as the number of mutants with phenotypes, was used to compare AD phenotypes of DR and control genes. DR genes have a higher phenotype number than control genes, indicating that DR genes are more likely to cause a visible change upon duplication than other singletons.

No soil growth phenotype was detected in AD lines, suggesting that the effect of gene duplication is normally weak. Striking stress phenotypes of AD lines, including one stress tolerance phenotype, were found in the DR group but not in the control group. The appearance of stress tolerance phenotypes in such a small sample of DR AD lines was intriguing from a practical standpoint and indicates that duplication of some DR genes may lead to better vegetative growth under stress conditions, but whether their duplication improves reproductive performance remains unknown.

OE lines, in which the transgenic constructs contain cDNAs under the control of the 35S promoter, were expected to reveal the consequences of large gene dosage increase. Soil growth phenotypes were found in DR OE lines at a higher frequency than DR KO lines, with the main features being small stature and reduced fertility. Half of genes with OE soil growth phenotypes had no KO phenotype, suggesting that OE lines would be useful in functional study of genes

without KO phenotypes. The other half of genes with OE soil growth phenotypes correspond to various KO phenotypes that are not limited to seedling alternations, reflecting different OE and KO effects of the same DR genes. Indeed, genes with OE stress phenotypes have a wide range of KO phenotypes as well.

About a third of DR genes with OE lines have both KO and OE phenotypes, indicating that some DR genes are dosage sensitive. Dosage sensitivity may have contributed to duplication resistance, as gene removal and duplication are accompanied by dosage change. Such dosage sensitivity at the phenotype level was found in genes that are predicted to be dosage sensitive. Most DR genes encoding components of protein complexes revealed phenotypes in their KO and OE lines, while the proportion of chloroplast-targeting DR genes with KO and OE phenotypes was slightly lower than one third.

Validation of associations between phenotypes and the intended gene duplication or overexpression is an important next step. AD and OE phenotypes may not always be caused by duplication and overexpression, respectively. For example, the transgene constructs may be located in other genes, revealing KO phenotypes that are unrelated to the target genes. Or, co-supression may have happened, reducing gene product. Though no evidence suggests that these situations occurred frequently, some of the observed phenotypes will inevitably be proved artifactual during the validation process.

Nevertheless, the abundant AD and OE phenotypes suggest that DR genes can have immediate, visible duplication and overexpression outcomes, which may be an important next step in understanding the basis of duplication resistance. In addition to detailed examination of AD/OE lines with phenotypes, such lines could be created for more genes and studied in various ways. In summary, our results mark one step toward better understanding of DR genes'

biological functions, and accordingly toward the basis of single-copy dosage regulation. The trends observed in this study provide several worthwhile future research directions, and show that more exciting discoveries can be made using the largely unexplored mutant collection.

ADDITIONAL FILES

Additional table S2.1-S2.3

Table S2.1 GO percentage in each subset of single copy genes and all genes. The numbers of genes with certain GO terms in each subset and in whole genome were obtained from TAIR.

GO groups and categories	GO percentage in each subset and whole genome (WG)*						
GO Biological Process	**DR**	**SD**	**LSS**	**OT**	**OT-F**	**OT-P**	**WG**
cell organization and biogenesis	18%	28%	2%	18%	24%	13%	15%
developmental processes	12%	21%	4%	17%	19%	13%	14%
DNA or RNA metabolism	9%	9%	0%	6%	9%	3%	3%
electron transport or energy pathways	4%	4%	0%	4%	5%	4%	3%
other biological processes	6%	11%	2%	11%	14%	9%	15%
other cellular processes	52%	70%	7%	50%	60%	38%	52%
other metabolic processes	46%	55%	5%	45%	57%	34%	50%
protein metabolism	13%	36%	1%	16%	20%	10%	19%
response to abiotic or biotic stimulus	10%	21%	2%	12%	16%	7%	15%
response to stress	13%	23%	3%	12%	15%	8%	17%
signal transduction	3%	9%	1%	3%	5%	3%	8%
transcription,DNA-dependent	7%	9%	1%	7%	9%	6%	11%
transport	9%	19%	2%	12%	16%	10%	14%
unknown biological processes	45%	30%	64%	45%	31%	56%	26%
GO Cellular Component	**DR**	**SD**	**LSS**	**OT**	**OT-F**	**OT-P**	**WG**
cell wall	0%	0%	0%	1%	1%	0%	3%
chloroplast	29%	19%	3%	25%	31%	18%	15%
cytosol	5%	17%	0%	8%	11%	4%	7%
ER	3%	6%	0%	3%	4%	2%	3%
extracellular	4%	4%	7%	3%	3%	5%	10%
Golgi apparatus	2%	6%	0%	2%	4%	2%	4%
mitochondria	14%	11%	40%	16%	14%	22%	12%
nucleus	33%	38%	14%	38%	40%	36%	35%
other cellular components	2%	4%	0%	3%	6%	2%	5%
other cytoplasmic components	27%	40%	2%	27%	38%	19%	27%

other intracellular components	25%	36%	1%	26%	38%	18%	21%
other membranes	14%	28%	3%	14%	20%	12%	18%
plasma membrane	4%	9%	2%	5%	7%	4%	12%
plastid	12%	13%	0%	10%	14%	7%	6%
ribosome	1%	2%	0%	2%	2%	2%	2%
unknown cellular components	6%	0%	15%	5%	3%	7%	6%
GO Molecular Function	**DR**	**SD**	**LSS**	**OT**	**OT-F**	**OT-P**	**WG**
DNA or RNA binding	8%	4%	1%	5%	9%	3%	14%
hydrolase activity	8%	17%	0%	7%	10%	4%	12%
kinase activity	1%	2%	0%	2%	2%	1%	5%
nucleic acid binding	3%	4%	0%	2%	2%	1%	5%
nucleotide binding	3%	9%	0%	4%	7%	2%	10%
other binding	8%	13%	1%	10%	18%	6%	26%
other enzyme activity	11%	19%	0%	11%	13%	5%	10%
other molecular functions	2%	2%	1%	2%	2%	2%	3%
protein binding	6%	11%	1%	9%	13%	7%	11%
receptor binding or activity	0%	0%	1%	0%	0%	1%	1%
structural molecule activity	1%	4%	0%	2%	2%	2%	2%
transcription factor activity	1%	0%	0%	1%	1%	1%	6%
transferase activity	7%	15%	0%	8%	12%	3%	12%
transporter activity	1%	6%	0%	3%	3%	2%	5%
unknown molecular functions	49%	28%	67%	48%	31%	62%	25%

*GO percentage=the number of genes with certain GO term/total number of genes

Table S2.2 **GO percentage fold of change comparison.** The GO percentages are in Table S2.1. Bold numbers indicate relatively large fold of change.

GO groups and categories	GO percentage fold of change compared to whole genome level*					
GO Biological Process	**DR**	**SD**	**LSS**	**OT**	**OT-F**	**OT-P**
cell organization and biogenesis	0.2	0.9	-0.9	0.2	0.6	-0.1
developmental processes	-0.1	0.5	-0.7	0.2	0.3	-0.1
DNA or RNA metabolism	**1.6**	**1.5**	-0.9	0.8	**1.7**	0.0
electron transport or energy pathways	0.5	0.6	-0.9	0.7	0.9	0.6
other biological processes	-0.6	-0.3	-0.9	-0.3	-0.1	-0.4
other cellular processes	0.0	0.3	-0.9	0.0	0.1	-0.3
other metabolic processes	-0.1	0.1	-0.9	-0.1	0.2	-0.3
protein metabolism	-0.3	0.9	-0.9	-0.1	0.1	-0.5
response to abiotic or biotic stimulus	-0.3	0.4	-0.9	-0.2	0.0	-0.5
response to stress	-0.2	0.4	-0.8	-0.2	-0.1	-0.5
signal transduction	-0.7	0.0	-0.8	-0.6	-0.4	-0.6
transcription,DNA-dependent	-0.4	-0.2	-0.9	-0.4	-0.2	-0.5
transport	-0.3	0.4	-0.9	-0.1	0.1	-0.3
unknown biological processes	0.7	0.1	**1.5**	0.7	0.2	**1.1**
GO Cellular Component	**DR**	**SD**	**LSS**	**OT**	**OT-F**	**OT-P**
cell wall	-0.9	-1.0	-0.9	-0.8	-0.6	-1.0
chloroplast	**1.0**	0.3	-0.8	0.7	**1.1**	0.2
cytosol	-0.3	**1.5**	-1.0	0.2	0.6	-0.5
ER	0.0	**1.2**	-0.9	0.1	0.5	-0.3
extracellular	-0.6	-0.6	-0.3	-0.7	-0.7	-0.5
Golgi apparatus	-0.4	0.6	-1.0	-0.4	-0.1	-0.6
mitochondria	0.2	-0.1	**2.5**	0.4	0.2	0.9
nucleus	0.0	0.1	-0.6	0.1	0.2	0.0
other cellular components	-0.6	-0.2	-0.9	-0.5	0.2	-0.6
other cytoplasmic components	0.0	0.5	-0.9	0.0	0.4	-0.3
other intracellular components	0.2	0.7	-0.9	0.2	0.8	-0.2
other membranes	-0.2	0.6	-0.9	-0.2	0.1	-0.3
plasma membrane	-0.7	-0.3	-0.9	-0.6	-0.5	-0.7
plastid	**1.2**	**1.3**	-0.9	0.9	**1.5**	0.3
ribosome	-0.5	0.2	-1.0	-0.2	0.1	0.0
unknown cellular components	-0.1	-1.0	**1.5**	-0.2	-0.5	0.1
GO Molecular Function	**DR**	**SD**	**LSS**	**OT**	**OT-F**	**OT-P**
DNA or RNA binding	-0.4	-0.7	-1.0	-0.6	-0.4	-0.8
hydrolase activity	-0.3	0.5	-1.0	-0.4	-0.1	-0.7
kinase activity	-0.9	-0.6	-1.0	-0.7	-0.5	-0.9

nucleic acid binding	-0.4	-0.2	-1.0	-0.6	-0.7	-0.8
nucleotide binding	-0.7	-0.1	-1.0	-0.5	-0.2	-0.8
other binding	-0.7	-0.5	-1.0	-0.6	-0.3	-0.8
other enzyme activity	0.1	0.9	-1.0	0.1	0.2	-0.5
other molecular functions	-0.4	-0.3	-0.8	-0.4	-0.4	-0.5
protein binding	-0.4	0.0	-0.9	-0.2	0.2	-0.4
receptor binding or activity	-0.7	-1.0	0.4	-0.4	-0.7	0.0
structural molecule activity	-0.5	**1.2**	-0.8	0.0	0.0	-0.2
transcription factor activity	-0.8	-1.0	-1.0	-0.8	-0.9	-0.8
transferase activity	-0.4	0.3	-1.0	-0.4	0.1	-0.7
transporter activity	-0.8	0.3	-1.0	-0.4	-0.3	-0.5
unknown molecular functions	0.9	0.1	**1.7**	0.9	0.2	**1.5**

*Fold of change = (GO percentage of subset-GO percentage of whole genome)/GO percentage of whole genome

Table S2.3 Number of sampled genes, genes with knockout phenotypes and phenotype percentage in the three datasets. PP stands for phenotype percentage.

		4000DS			Hanada			Lloyd and Meinke			
Subsets	**Total**	**Total[1]**	**Phenotype[2]**	**PP**	**Total[1]**	**Phenotype[2]**	**PP**	**Total[1]**	**Phenotype[2]**	**PP**	**Average PP**
LSS	1669	58	0	0.0%	60	0	0.0%	1669	6	0.4%	0.1%
OT-P	790	54	0	0.0%	69	5	7.2%	790	56	7.1%	4.8%
OT-F	1221	142	11	7.7%	172	24	14.0%	1221	229	18.8%	13.5%
OT	2011	196	11	5.6%	241	29	12.0%	2011	285	14.2%	10.6%
DR	409	46	2	4.3%	63	6	9.5%	409	49	12.0%	8.6%
SD	47	7	0	0.0%	8	2	25.0%	47	11	23.4%	16.1%

1. Total number of sampled genes (in each subset) in a dataset.

2. Number of mutants with phenotypes.

Additional table S3.1-S3.14

Table S3.1 DR, nDR and SCR candidates. DR and nDR genes listed here include those without SALK lines. SS stands for 'strict singleton'. PD stands for 'protein domain'.

Gene	Group	Subgroup	Gene	Group	Subgroup	Gene	Group	Subgroup
AT1G01760	DR	SS	AT3G52860	DR	SS	AT4G19350	nDR	SS
AT1G01920	DR	PD	AT3G52905	DR	SS	AT5G42760	nDR	SS
AT1G01930	DR	SS	AT3G56210	DR	SS	AT5G63290	nDR	PD
AT1G02870	DR	SS	AT3G56290	DR	SS	AT3G14910	nDR	SS
AT1G03760	DR	SS,PD	AT3G56510	DR	SS	AT1G76060	nDR	PD
AT1G04130	DR	PD	AT3G56570	DR	PD	AT3G19220	nDR	SS
AT1G04985	DR	SS	AT3G56820	DR	SS	AT4G01880	nDR	SS
AT1G05060	DR	SS	AT3G56840	DR	PD	AT1G75200	nDR	PD
AT1G06510	DR	SS	AT3G58470	DR	SS	AT3G09850	nDR	PD
AT1G07130	DR	SS	AT3G59490	DR	SS	AT3G02220	nDR	SS
AT1G07645	DR	SS	AT3G59650	DR	SS	AT2G43640	nDR	SS
AT1G07970	DR	SS	AT3G60850	DR	SS	AT5G58220	nDR	SS
AT1G08030	DR	SS	AT3G61080	DR	SS	AT2G29530	nDR	PD
AT1G08710	DR	SS	AT3G62140	DR	SS	AT5G48240	nDR	SS
AT1G09010	DR	SS	AT3G62370	DR	SS	AT2G36885	nDR	SS
AT1G10030	DR	SS	AT3G62810	DR	PD	AT1G01710	nDR	SS
AT1G10830	DR	SS	AT4G00560	DR	SS	AT2G01640	nDR	SS
AT1G12370	DR	SS	AT4G02110	DR	PD	AT4G16444	nDR	SS
AT1G12650	DR	SS	AT4G03200	DR	SS	AT1G26180	nDR	SS
AT1G12790	DR	SS	AT4G06676	DR	SS	AT5G51545	nDR	SS
AT1G13990	DR	SS	AT4G10090	DR	SS	AT5G05560	nDR	SS
AT1G14345	DR	SS	AT4G11980	DR	SS	AT5G27390	nDR	SS
AT1G14620	DR	SS	AT4G13330	DR	SS	AT1G68080	nDR	SS
AT1G15980	DR	SS	AT4G13670	DR	SS	AT2G01590	nDR	SS
AT1G16970	DR	PD	AT4G13950	DR	PD	AT3G18760	nDR	SS
AT1G17680	DR	PD	AT4G15180	DR	PD	AT5G39940	nDR	SS
AT1G18730	DR	SS	AT4G15520	DR	PD	AT1G27752	nDR	SS
AT1G19140	DR	SS	AT4G17370	DR	SS	AT4G33140	nDR	SS
AT1G21350	DR	SS	AT4G17540	DR	SS	AT3G63390	nDR	SS
AT1G21370	DR	SS	AT4G17760	DR	SS	AT3G60810	nDR	SS
AT1G21840	DR	SS	AT4G18460	DR	SS	AT5G14910	nDR	SS
AT1G22700	DR	PD	AT4G18470	DR	SS	AT2G41760	nDR	SS
AT1G26660	DR	PD	AT4G19070	DR	SS	AT1G48200	nDR	SS
AT1G26840	DR	SS	AT4G19400	DR	SS	AT3G15150	nDR	SS
AT1G27530	DR	SS	AT4G20060	DR	SS	AT5G49800	nDR	SS
AT1G28560	DR	SS	AT4G20350	DR	SS	AT4G38490	nDR	SS

AT1G31780	DR	SS	AT4G21720	DR	SS	AT4G31150	nDR	SS
AT1G31860	DR	SS	AT4G21770	DR	SS,PD	AT1G33400	nDR	PD
AT1G32370	DR	SS	AT4G23660	DR	PD	AT4G21620	nDR	PD
AT1G33810	DR	SS	AT4G26370	DR	SS	AT3G55080	nDR	PD
AT1G34770	DR	SS	AT4G27750	DR	SS	AT2G44580	nDR	SS
AT1G36310	DR	SS	AT4G28020	DR	SS	AT1G07020	nDR	SS
AT1G42990	DR	SS	AT4G28660	DR	SS	AT3G09085	nDR	SS
AT1G45150	DR	SS	AT4G29520	DR	SS	AT2G25720	nDR	SS
AT1G48270	DR	SS	AT4G29890	DR	SS	AT3G24080	nDR	SS
AT1G48360	DR	SS	AT4G30840	DR	SS	AT4G29560	nDR	SS
AT1G50910	DR	SS	AT4G31460	DR	SS	AT4G13690	nDR	SS
AT1G52530	DR	SS	AT4G31770	DR	SS	AT3G27520	nDR	SS
AT1G53120	DR	SS,PD	AT4G34140	DR	PD	AT3G20490	nDR	SS
AT1G53645	DR	SS	AT4G34700	DR	PD	AT2G44820	nDR	SS
AT1G54990	DR	SS	AT4G35760	DR	SS	AT4G09680	nDR	SS
AT1G55280	DR	SS	AT4G35987	DR	SS	AT5G37480	nDR	SS
AT1G56345	DR	PD	AT4G37020	DR	SS	AT2G25625	nDR	SS
AT1G60460	DR	SS	AT4G37210	DR	PD	AT5G16610	nDR	SS
AT1G60600	DR	PD	AT4G37510	DR	SS	AT5G12920	nDR	SS
AT1G61690	DR	PD	AT4G38020	DR	PD	AT2G05320	nDR	SS
AT1G62250	DR	SS	AT4G38090	DR	SS	AT5G24314	nDR	SS
AT1G64050	DR	SS	AT4G39680	DR	PD	AT4G16410	nDR	SS
AT1G65020	DR	SS	AT5G01160	DR	SS	AT5G37590	nDR	PD
AT1G65230	DR	SS	AT5G01300	DR	SS	AT4G18400	nDR	SS
AT1G65900	DR	SS	AT5G03770	DR	SS	AT1G53200	nDR	SS
AT1G66080	DR	SS	AT5G06830	DR	SS	AT4G00790	nDR	SS
AT1G66330	DR	SS	AT5G07380	DR	SS	AT1G48580	nDR	SS
AT1G67180	DR	PD	AT5G10320	DR	SS	AT4G15030	nDR	SS
AT1G70570	DR	SS	AT5G10460	DR	SS	AT5G22820	nDR	SS
AT1G71340	DR	SS	AT5G10620	DR	SS	AT5G52290	nDR	SS
AT1G73350	DR	SS	AT5G11030	DR	SS	AT5G17070	nDR	SS
AT1G74640	DR	SS	AT5G11450	DR	SS	AT5G37580	nDR	SS
AT1G74880	DR	SS	AT5G11640	DR	SS	AT1G71760	nDR	SS
AT1G76050	DR	PD	AT5G11980	DR	SS	AT1G43245	nDR	SS
AT1G76250	DR	SS	AT5G13070	DR	SS	AT1G51080	nDR	SS
AT1G76450	DR	SS	AT5G13240	DR	SS	AT2G24970	nDR	SS
AT1G77030	DR	SS	AT5G14520	DR	PD	AT2G24830	nDR	PD
AT1G77230	DR	PD	AT5G15390	DR	PD	AT5G63135	nDR	SS
AT1G77320	DR	PD	AT5G15750	DR	PD	AT2G17972	nDR	SS
AT1G77350	DR	SS	AT5G15802	DR	SS	AT4G22550	nDR	SS
AT1G77550	DR	SS	AT5G17240	DR	PD	AT4G29660	nDR	SS
AT1G78650	DR	SS	AT5G17670	DR	SS	AT2G01100	nDR	SS
AT1G80410	DR	PD	AT5G19130	DR	SS	AT4G17610	nDR	PD

AT1G80420	DR	SS,PD	AT5G20600	DR	SS	AT5G62440	nDR	PD
AT2G02590	DR	SS	AT5G20935	DR	SS	AT3G08610	nDR	SS
AT2G04270	DR	SS	AT5G22130	DR	SS	AT4G27390	nDR	SS
AT2G04560	DR	SS	AT5G23290	DR	PD	AT1G57540	nDR	SS
AT2G05170	DR	SS	AT5G23395	DR	SS	AT4G32915	nDR	SS
AT2G13840	DR	SS	AT5G23520	DR	SS	AT5G16020	nDR	SS
AT2G15890	DR	SS	AT5G25480	DR	SS	AT4G10330	nDR	SS
AT2G17900	DR	PD	AT5G25500	DR	SS	AT3G17170	nDR	SS
AT2G18850	DR	PD	AT5G37055	DR	SS	AT5G67490	nDR	SS
AT2G18950	DR	PD	AT5G37290	DR	SS	AT1G12244	nDR	SS
AT2G19270	DR	SS	AT5G38900	DR	SS	AT1G21500	nDR	SS
AT2G19640	DR	PD	AT5G39410	DR	SS	AT2G48070	nDR	SS
AT2G19870	DR	PD	AT5G40660	DR	SS	AT2G41120	nDR	SS
AT2G20790	DR	SS	AT5G41150	DR	SS	AT3G52230	nDR	SS
AT2G20940	DR	SS	AT5G41190	DR	SS	AT5G26880	nDR	PD
AT2G20980	DR	SS	AT5G41880	DR	SS	AT5G27400	nDR	SS
AT2G21970	DR	SS	AT5G42370	DR	SS	AT3G46560	nDR	PD
AT2G22370	DR	SS	AT5G43750	DR	SS	AT4G10100	nDR	SS
AT2G22650	DR	SS	AT5G45310	DR	SS	AT5G18540	nDR	SS
AT2G25605	DR	SS	AT5G46850	DR	SS	AT1G61570	nDR	PD
AT2G26540	DR	SS	AT5G47570	DR	SS	AT5G59140	nDR	SS
AT2G26680	DR	SS	AT5G48440	DR	PD	AT5G50810	nDR	PD
AT2G30410	DR	SS	AT5G48470	DR	SS	AT2G25570	nDR	SS
AT2G31040	DR	SS	AT5G49550	DR	SS	AT1G67785	nDR	SS
AT2G31890	DR	SS	AT5G50320	DR	PD	AT1G74340	nDR	SS
AT2G31955	DR	PD	AT5G50375	DR	SS	AT3G14430	nDR	SS
AT2G32900	DR	SS	AT5G51020	DR	SS	AT2G01755	nDR	SS
AT2G33255	DR	SS	AT5G51040	DR	SS	AT3G59840	nDR	SS
AT2G34090	DR	SS	AT5G51130	DR	SS	AT3G04640	nDR	PD
AT2G35360	DR	SS	AT5G51220	DR	SS	AT5G19970	nDR	SS
AT2G36895	DR	SS	AT5G52190	DR	SS	AT5G03560	nDR	SS
AT2G37560	DR	SS	AT5G52880	DR	SS	AT5G09830	nDR	SS
AT2G39090	DR	PD	AT5G54855	DR	SS	AT5G03460	nDR	SS
AT2G39910	DR	SS	AT5G55500	DR	SS	AT4G35980	nDR	SS
AT2G40316	DR	SS	AT5G57950	DR	SS	AT5G50930	nDR	SS
AT2G40430	DR	SS	AT5G59440	DR	SS	AT5G61220	nDR	PD
AT2G40570	DR	SS	AT5G59460	DR	SS	AT5G22875	nDR	SS
AT2G41530	DR	SS	AT5G60410	DR	PD	AT3G45050	nDR	SS
AT2G41950	DR	SS	AT5G61330	DR	SS	AT4G22600	nDR	SS
AT2G42780	DR	SS	AT5G61850	DR	SS	AT4G27380	nDR	SS
AT2G43360	DR	PD	AT5G62140	DR	SS	AT4G04614	nDR	SS
AT2G43400	DR	SS	AT5G62760	DR	SS	AT2G38570	nDR	SS
AT2G44520	DR	PD	AT5G63200	DR	PD	AT2G34585	nDR	SS

AT2G44870	DR	SS	AT5G63460	DR	PD	AT3G52070	nDR	SS
AT2G45990	DR	SS	AT5G64250	DR	SS	AT3G20470	nDR	PD
AT2G46060	DR	SS	AT5G64680	DR	SS	AT4G14020	nDR	PD
AT2G46200	DR	SS	AT5G65660	DR	SS	AT2G39725	nDR	PD
AT2G47760	DR	SS	AT5G66090	DR	SS	AT5G57860	nDR	SS
AT3G02820	DR	SS	AT5G13050	nDR	SS	AT5G67290	nDR	PD
AT3G03100	DR	SS	AT4G37830	nDR	SS	AT2G41260	nDR	PD
AT3G04560	DR	SS	AT5G49510	nDR	PD	AT3G14480	nDR	PD
AT3G04600	DR	SS	AT1G72440	nDR	SS	AT1G19490	nDR	SS
AT3G04950	DR	SS	AT1G63980	nDR	PD	AT3G23450	nDR	PD
AT3G05210	DR	SS	AT2G02500	nDR	SS	AT3G55790	nDR	PD
AT3G05625	DR	SS	AT5G52110	nDR	SS	AT5G66840	nDR	PD
AT3G05760	DR	SS	AT1G63680	nDR	SS	AT2G33855	nDR	SS
AT3G09180	DR	SS	AT3G27110	nDR	SS	AT3G03341	nDR	SS
AT3G09210	DR	SS	AT1G73820	nDR	SS	AT3G09860	nDR	SS
AT3G09430	DR	SS	AT3G04260	nDR	PD	AT3G56870	nDR	SS
AT3G09580	DR	SS	AT1G64510	nDR	SS	AT1G70200	nDR	SS
AT3G10370	DR	PD	AT1G44920	nDR	SS	AT3G25165	nDR	PD
AT3G10572	DR	SS	AT5G15680	nDR	SS	AT4G39160	nDR	SS
AT3G11620	DR	SS	AT2G44360	nDR	SS	AT3G15280	nDR	SS
AT3G12040	DR	SS	AT1G30480	nDR	PD	AT4G11510	nDR	PD
AT3G12210	DR	SS	AT1G71790	nDR	SS	AT2G14660	nDR	SS
AT3G12260	DR	PD	AT3G17930	nDR	SS	AT4G17590	nDR	SS
AT3G13226	DR	SS	AT2G33180	nDR	SS	AT4G06534	nDR	SS
AT3G13940	DR	SS	AT5G63000	nDR	SS	AT1G75550	nDR	PD
AT3G14110	DR	PD	AT3G26580	nDR	SS	AT1G58250	SCR	SD
AT3G14900	DR	SS	AT4G03150	nDR	SS	AT2G01170	SCR	SD
AT3G15110	DR	SS	AT1G08220	nDR	SS	AT3G24630	SCR	SD
AT3G15180	DR	SS	AT2G36740	nDR	SS	AT3G57990	SCR	SD
AT3G16270	DR	SS	AT2G45520	nDR	SS	AT4G12070	SCR	SD
AT3G16760	DR	PD	AT5G54080	nDR	SS	AT5G06120	SCR	SD
AT3G16990	DR	SS	AT5G19050	nDR	SS	AT5G40940	SCR	OT-P
AT3G17590	DR	SS	AT2G31440	nDR	SS	AT5G24090	SCR	SD
AT3G17670	DR	SS,PD	AT5G53080	nDR	PD	AT2G23090	SCR	SD
AT3G18730	DR	PD	AT1G73740	nDR	SS	AT1G78780	SCR	SD
AT3G20070	DR	SS	AT4G02405	nDR	SS	AT1G08370	SCR	SD
AT3G20480	DR	SS	AT3G46220	nDR	SS	AT3G13130	SCR	OT-P
AT3G21820	DR	PD	AT2G20495	nDR	SS	AT2G03667	SCR	OT-F
AT3G22990	DR	SS	AT4G02725	nDR	SS	AT4G24265	SCR	OT-P
AT3G24315	DR	SS	AT2G36145	nDR	SS	AT3G51580	SCR	OT-F
AT3G24560	DR	SS	AT3G01160	nDR	SS	AT3G21360	SCR	OT-F
AT3G25120	DR	SS	AT3G54230	nDR	PD	AT2G13440	SCR	OT-F
AT3G25470	DR	SS	AT3G21350	nDR	SS	AT4G39690	SCR	OT-F

AT3G26085	DR	SS	AT4G31790	nDR	SS	AT4G18593	SCR	SD
AT3G26710	DR	SS	AT5G08060	nDR	SS	AT1G08280	SCR	SD
AT3G28760	DR	SS	AT5G02130	nDR	SS	AT2G34050	SCR	OT-F
AT3G32930	DR	SS	AT5G02710	nDR	SS	AT1G50170	SCR	OT-F
AT3G46200	DR	SS	AT5G24670	nDR	SS	AT4G39450	SCR	OT-F
AT3G47850	DR	SS	AT5G51170	nDR	SS	AT2G27900	SCR	OT-F
AT3G48120	DR	SS	AT4G35910	nDR	SS	AT2G44970	SCR	SD
AT3G48500	DR	SS	AT5G06410	nDR	SS	AT2G19940	SCR	OT-F
AT3G49890	DR	SS	AT3G57910	nDR	PD	AT5G62390	SCR	SD
AT3G51010	DR	SS	AT3G27050	nDR	PD			
AT3G51820	DR	PD	AT2G44760	nDR	SS			

Table S3.2 SALK lines genotype results. Green cells in column 'Gene' represent genes with two SALK lines. HM stands for 'homozygous'. HT stands for 'heterozygous'. WT stands for 'wildtype'. The numbers in column 'HM', 'HZ', and 'WT' are the numbers of plants with corresponding genotypes.

Mutant_ID	Gene	Gene group	SALK	Insertion position	HM	HT	WT	Genotype group
1	AT5G15750	DR	SALK_093583C	5utr	12	0	0	HM
100	AT1G62250	DR	SALK_065936	intron	3	4	4	HM
101	AT1G65020	DR	SALK_010039	exon	5	6	5	HM
102	AT1G65230	DR	SALK_130615	exon	10	0	0	HM
103	AT1G65900	DR	SALK_122807	intron	2	2	4	HM
104	AT1G66080	DR	SALK_058334	exon	12	0	0	HM
105	AT1G70570	DR	SALK_078468C	intron	1	0	8	HM
106	AT1G73350	DR	SALK_020247	intron	0	4	17	PL
107	AT1G76250	DR	SALK_019804	intron	7	4	8	HM
108	AT1G77230	DR	SALK_131710C	exon	4	0	0	HM
109	AT1G77350	DR	SALK_124373	5utr	0	0	13	WT
11	AT3G25470	DR	SALK_043556C	exon	2	3	2	HM
111	AT2G04270	DR	SALK_093546C	exon	2	4	3	HM
112	AT2G19270	DR	SALK_043882	intron	0	0	11	WT
114	AT2G20940	DR	SALK_082302	exon	0	0	7	WT
115	AT2G22370	DR	SALK_027178C	intron	10	4	12	HM
116	AT2G22650	DR	SALK_147486	exon	6	0	0	HM
117	AT2G34090	DR	SALK_122423C	exin	10	0	0	HM
118	AT2G36895	DR	SALK_071847	exon	0	0	16	WT
119	AT2G42780	DR	SALK_145132	exon	10	0	7	HM
120	AT2G43400	DR	SALK_007870	exon	3	5	15	HM
121	AT2G45990	DR	SALK_141449C	exon	19	0	0	HM
122	AT2G46060	DR	SALK_135304C	exon	12	12	0	HM
123	AT3G03100	DR	SALK_132527	intron	0	0	7	WT
124	AT3G04560	DR	SALK_004031C	exon	6	0	0	HM
125	AT3G05625	DR	SALK_081322	exon	0	0	7	WT
127	AT3G09180	DR	SALK_012449	exon	1	3	10	HM
128	AT3G09210	DR	SALK_039502	exon	4	1	16	HM
13	AT4G00560	DR	SALK_138740C	exon	10	1	1	HM
130	AT3G10572	DR	SALK_132193	exon	0	7	14	PL
131	AT3G13940	DR	SALK_054381	exon	2	12	0	HM
132	AT3G14110	DR	SALK_002383C	intron	4	3	0	HM
133	AT3G15110	DR	SALK_151651C	exon	6	0	0	HM
134	AT3G15180	DR	SALK_010908	exon	0	1	23	PL

135	AT3G20070	DR	SALK_148785	exon	4	5	9	HM
136	AT3G21820	DR	SALK_026154C	exon	2	3	5	HM
137	AT3G24315	DR	SALK_040675C	exon	6	0	0	HM
138	AT3G24560	DR	SALK_099074	exon	0	0	18	WT
139	AT3G26085	DR	SALK_030049C	exon	6	0	0	HM
140	AT3G28760	DR	SALK_083209C	exon	6	0	0	HM
141	AT3G48120	DR	SALK_136585	exon	11	1	0	HM
142	AT3G49890	DR	SALK_121199C	intron	4	0	7	HM
143	AT3G51010	DR	SALK_066359	intron	1	5	13	HM
144	AT3G51820	DR	SALK_112733C	exon	0	11	13	PL
145	AT3G56210	DR	SALK_022623	exon	0	12	11	PL
146	AT3G56510	DR	SALK_101389	intron	0	0	7	WT
147	AT3G56820	DR	SALK_016215	exon	6	2	1	HM
148	AT3G59490	DR	SALK_008961	intron	9	0	1	HM
149	AT3G60850	DR	SALK_052869	exon	0	5	5	PL
15	AT2G33255	DR	SALK_145197C	exon	2	3	0	HM
150	AT3G62140	DR	SALK_027223	exon	0	0	7	WT
151	AT3G62370	DR	SALK_046903C	intron	6	0	0	HM
152	AT4G10090	DR	SALK_100099	intron	0	0	7	WT
153	AT4G15520	DR	SALK_027418C	intron	10	0	0	HM
154	AT4G17370	DR	SALK_010608C	exon	6	0	0	HM
155	AT4G17540	DR	SALK_043122	exon	6	4	13	HM
157	AT4G18470	DR	SALK_018281	exon	3	0	5	HM
158	AT4G26370	DR	SALK_013094C	intron	9	0	1	HM
159	AT4G27750	DR	SALK_014032	exon	0	3	1	PL
160	AT4G29890	DR	SALK_026428	exon	5	1	5	HM
161	AT4G30840	DR	SALK_095344C	exon	10	1	0	HM
162	AT4G31460	DR	SALK_062358	exon	0	12	0	OT
163	AT4G34700	DR	SALK_030356	exon	9	0	2	HM
164	AT4G37020	DR	SALK_069628	exon	16	1	0	HM
165	AT4G38020	DR	SALK_055128	exon	0	0	19	WT
166	AT5G03770	DR	SALK_035981C	intron	6	0	0	HM
167	AT5G62140	DR	SALK_008702	exon	1	6	11	HM
169	AT5G42370	DR	SALK_019576	exon	0	0	14	WT
170	AT5G55500	DR	SALK_042226C	exon	6	0	0	HM
171	AT5G52190	DR	SALK_048298	exon	9	0	0	HM
172	AT5G20935	DR	SALK_050034	intron	1	6	0	HM
173	AT5G10320	DR	SALK_057370C	exon	6	0	0	HM
174	AT5G15802	DR	SALK_078544C	intron	4	1	1	HM
175	AT5G54855	DR	SALK_081219C	intron	6	0	0	HM
176	AT5G61330	DR	SALK_088403	intron	0	0	19	WT
177	AT5G11030	DR	SALK_089074	exon	1	2	0	HM
178	AT5G17240	DR	SALK_097673C	exon	6	0	0	HM

179	AT5G51130	DR	SALK_100446C	intron	1	7	1	HM
18	AT2G39090	DR	SALK_109171C	intron	21	11	0	HM
180	AT5G59460	DR	SALK_104801	intron	0	0	7	WT
182	AT5G22130	DR	SALK_116293	intron	0	0	11	WT
183	AT5G11980	DR	SALK_122096	exon	0	0	12	WT
184	AT4G38090	DR	SALK_029673C	exon	12	0	0	HM
19	AT3G13226	DR	SALK_110892C	exon	6	0	0	HM
1_15	AT4G13330	DR	SALK_122684C	promoter	12	0	0	HM
1_49	AT4G21770	DR	SALK_039518C	5utr	6	1	0	HM
PD7	AT4G21770	DR	SALK_149232C	intron	10	1	0	HM
1_6	AT2G40316	DR	SALK_014092C	intron	3	0	0	HM
2	AT1G74880	DR	SALK_097351C	exon	12	0	0	HM
23	AT1G16970	DR	SALK_123114C	exon	12	0	0	HM
24	AT5G49550	DR	SALK_118923C	intron	6	0	0	HM
26	AT5G20600	DR	SALK_073773C	exon	6	0	0	HM
28	AT2G44870	DR	SALK_082864C	intron	6	0	0	HM
2_1	AT1G76450	DR	SALK_086261C	5utr	0	0	6	WT
2_2	AT1G80420	DR	SALK_056275C	exon	0	0	10	WT
2_4	AT5G13240	DR	SALK_027781C	3utr	10	0	0	HM
33	AT5G41190	DR	SALK_021098C	exon	0	0	19	WT
34	AT1G61690	DR	SALK_133410C	exon	1	6	7	HM
35	AT2G37560	DR	SALK_027788C	exon	0	8	10	PL
36	AT5G14520	DR	SALK_026359C	exon	0	0	7	WT
37	AT2G40430	DR	SALK_012561C	exon	0	3	8	PL
38	AT2G31955	DR	SALK_037143C	exon	6	0	0	HM
39	AT4G13670	DR	SALK_096411C	exon	4	1	1	HM
3_1	AT2G20980	DR	SALK_053315C	3utr	12	0	0	HM
3_11	AT5G62760	DR	SALK_076441C	intron	3	1	6	HM
3_2	AT2G31890	DR	SALK_035413C	5utr	9	0	0	HM
SS1	AT2G31890	DR	SALK_088986	exon	0	4	20	PL
3_23	AT3G10370	DR	SALK_080169C	intron	0	7	5	PL
3_35	AT3G17670	DR	SALK_021028C	exon	6	2	0	HM
3_36	AT1G56345	DR	SALK_090814C	exon	4	5	0	HM
3_4	AT3G16270	DR	SALK_131068C	promoter	11	0	0	HM
3_8	AT3G46200	DR	SALK_025038C	intron	13	12	0	HM
3_9	AT3G61080	DR	SALK_059076C	exon	0	0	12	WT
40	AT5G15390	DR	SALK_005531C	5utr	6	0	0	HM
42	AT2G41530	DR	SALK_002548C	intron	2	3	0	HM
43	AT5G23395	DR	SALK_044358C	intron	16	2	0	HM
44	AT5G23290	DR	SALK_057848C	5utr	5	5	0	HM
45	AT5G19130	DR	SALK_143842	exon	0	10	5	PL
46	AT3G09580	DR	SALK_058610C	exon	12	9	0	HM
47	AT1G18730	DR	SALK_056498C	exon	6	0	0	HM

49	AT5G48440	DR	SALK_063996C	exon	6	0	0	HM
5	AT5G63460	DR	SALK_081993C	intron	6	0	0	HM
50	AT2G26540	DR	SALK_065522C	exon	0	8	4	PL
51	AT1G04130	DR	SALK_073054C	exon	6	0	0	HM
52	AT2G02590	DR	SALK_034951C	exon	12	0	0	HM
55	AT1G77550	DR	SALK_108909C	exon	12	8	0	HM
56	AT5G48470	DR	SALK_069893C	intron	0	11	12	PL
57	AT2G25605	DR	SALK_138104C	exon	6	0	0	HM
58	AT5G11450	DR	SALK_024527C	intron	6	0	0	HM
6	AT3G17590	DR	SALK_073635C	exon	6	0	0	HM
61	AT1G12650	DR	SALK_116493	exon	0	0	7	WT
62	AT4G11980	DR	SALK_065249C	intron	0	0	7	WT
63	AT2G47760	DR	SALK_040296c	intron	6	0	0	HM
64	AT3G26710	DR	SALK_055566C	exon	0	6	13	PL
65	AT2G31040	DR	SALK_057229C	exon	4	0	0	HM
66	AT2G05170	DR	SALK_124074	exon	0	8	15	PL
67	AT2G17900	DR	SALK_127952C	intron	6	0	0	HM
69	AT1G26840	DR	SALK_021894	exon	0	0	7	WT
7	AT1G06510	DR	SALK_074725C	exon	6	0	0	HM
70	AT5G37290	DR	SALK_091235	intron	4	12	10	HM
71	AT1G53120	DR	SALK_099429C	exon	7	0	0	HM
74	AT1G52530	DR	SALK_036820	5utr	6	0	0	HM
75	AT2G30410	DR	SALK_131921	exon	0	15	7	PL
76	AT4G03200	DR	SALK_079828C	intron	10	0	0	HM
77	AT3G25120	DR	SALK_094748	exon	3	6	2	HM
78	AT4G31770	DR	SALK_122077C	intron	12	0	0	HM
79	AT1G01920	DR	SALK_060683C	intron	9	2	10	HM
8	AT3G05210	DR	SALK_077000C	exon	17	0	0	HM
80	AT1G02870	DR	SALK_018438	exon	0	0	18	WT
81	AT1G03760	DR	SALK_038314	intron	0	3	18	PL
82	AT1G05060	DR	SALK_034347	intron	1	3	8	HM
83	AT1G07645	DR	SALK_102268	intron	2	7	23	HM
84	AT1G08710	DR	SALK_137276	exon	16	0	0	HM
86	AT1G10830	DR	SALK_068653	exon	16	0	0	HM
87	AT1G12370	DR	SALK_000335	exon	0	0	7	WT
88	AT1G12790	DR	SALK_127447C	exon	3	1	8	HM
89	AT1G13990	DR	SALK_041341	exon	0	0	11	WT
9	AT3G56570	DR	SALK_131900C	intron	6	0	0	HM
90	AT1G15980	DR	SALK_137420	exon	1	3	14	HM
91	AT1G17680	DR	SALK_122606	exon	0	0	7	WT
92	AT1G26660	DR	SALK_109676	exon	0	4	15	PL
93	AT1G27530	DR	SALK_040508	intron	6	0	0	HM
94	AT1G31860	DR	SALK_027157	intron	0	7	16	PL

95	AT1G36310	DR	SALK_135308C	exon	7	5	5	HM
96	AT1G42990	DR	SALK_050203C	exon	1	9	14	HM
97	AT1G48270	DR	SALK_027808	intron	16	1	11	HM
98	AT1G60460	DR	SALK_121547	intron	0	0	7	WT
99	AT1G60600	DR	SALK_021962C	intron	1	12	6	HM
113	AT2G19640	DR	SALK_024470C	exon	2	0	0	HM
206	AT1G45150	DR	SALK_013411C	exon	4	2	0	HM
220	AT1G77320	DR	SALK_201730C	exon	2	0	0	HM
232	AT2G32900	DR	SALK_017317C	exon	9	0	0	HM
242	AT3G04950	DR	SALK_146533C	intron	11	0	0	HM
246	AT3G12210	DR	SALK_019370C	intron	6	0	0	HM
248	AT3G14900	DR	SALK_123989	exon	0	6	2	PL
250	AT3G16990	DR	SALK_062985C	exon	7	0	0	HM
252	AT3G20480	DR	SALK_100275C	exon	7	0	0	HM
264	AT4G02110	DR	SALK_001578C	exon	4	0	0	HM
272	AT4G20350	DR	SALK_105865C	intron	9	0	0	HM
277	AT4G29520	DR	SALK_022300C	exon	11	0	0	HM
293	AT5G25480	DR	SALK_136635C	intron	11	0	0	HM
10	AT2G36885	nDR	SALK_019139C	exon	6	0	0	HM
12	AT3G46220	nDR	SALK_000583C	exon	1	9	0	HM
14	AT5G02710	nDR	SALK_148214C	exon	6	0	0	HM
16	AT5G67290	nDR	SALK_149999C	exon	6	0	0	HM
17	AT3G04260	nDR	SALK_110045C	exon	0	1	12	PL
1_10	AT3G27050	nDR	SALK_044969C	exon	4	8	0	HM
1_18	AT5G50930	nDR	SALK_119435C	exon	12	0	0	HM
1_19	AT5G59140	nDR	SALK_111402C	promoter	5	6	0	HM
1_3	AT1G68080	nDR	SALK_044417C	intron	9	0	0	HM
20	AT2G41120	nDR	SALK_113946C	exon	12	0	0	HM
21	AT2G33180	nDR	SALK_113601C	intron	0	0	0	OT
22	AT3G26580	nDR	SALK_118163C	exon	12	0	0	HM
25	AT2G41760	nDR	SALK_075466C	exon	12	7	0	HM
27	AT1G75200	nDR	SALK_076701C	exon	15	0	2	HM
29	AT5G48240	nDR	SALK_098728C	intron	0	20	9	PL
2_3	AT3G20490	nDR	SALK_023330C	exon	7	0	0	HM
3	AT4G02405	nDR	SALK_092870C	exon	6	0	0	HM
30	AT1G61570	nDR	SALK_104396C	exon	9	3	6	HM
31	AT3G17170	nDR	SALK_102663C	exon	6	0	0	HM
32	AT1G44920	nDR	SALK_137362C	intron	0	0	7	WT
3_12	AT3G09085	nDR	SALK_146703C	3utr	2	3	6	HM
3_3	AT1G27752	nDR	SALK_065549C	exon	9	2	22	HM
3_33	AT1G63980	nDR	SALK_041197C	exon	0	0	3	WT
3_34	AT4G17610	nDR	SALK_138076C	intron	3	4	5	HM
3_5	AT1G43245	nDR	SALK_135311C	5utr	9	0	0	HM

3_6	AT1G63680	nDR	SALK_126518C	exon	0	5	1	PL
3_7	AT2G44820	nDR	SALK_089034C	promoter	11	1	0	HM
SS2	AT2G44820	nDR	SALK_017223	3utr	2	4	5	HM
4	AT3G27110	nDR	SALK_082409C	exon	16	0	1	HM
41	AT5G53080	nDR	SALK_006120C	3utr	0	0	7	WT
48	AT1G19490	nDR	SALK_053908C	exon	12	0	0	HM
53	AT1G01710	nDR	SALK_067535C	5utr	5	0	0	HM
54	AT1G51080	nDR	SALK_133472C	exon	5	9	4	HM
60	AT5G37590	nDR	SALK_063987C	exon	2	1	4	HM
68	AT1G08220	nDR	SALK_044812C	intron	12	0	2	HM
72	AT5G19050	nDR	SALK_104522C	intron	6	0	0	HM
73	AT1G70200	nDR	SALK_092951C	exon	1	0	0	HM
1_26	AT1G58250	SCR	SALK_021431C	5utr	11	0	0	HM
1_28	AT2G01170	SCR	SALK_107641C	exon	17	0	11	HM
1_30	AT3G24630	SCR	SALK_137126C	5utr	10	0	0	HM
SCR3	AT3G24630	SCR	SALK_027123	promoter	9	9	2	HM
1_31	AT3G57990	SCR	SALK_001614C	5utr	12	0	0	HM
1_34	AT4G12070	SCR	SALK_104199C	5utr	0	5	21	PL
1_37	AT5G06120	SCR	SALK_100541C	exon	12	0	0	HM
1_47	AT3G52260	SCR	SALK_024791C	exon	12	0	0	HM
2_5	AT5G40940	SCR	SALK_063501C	exon	11	0	0	HM
2_6	AT5G24090	SCR	SALK_095362C	3utr exon	9	2	0	HM
3_14	AT1G78780	SCR	SALK_045403C	promoter	9	0	0	HM
3_15	AT1G08370	SCR	SALK_064670C	5utr	0	8	0	OT
3_16	AT3G13130	SCR	SALK_076776C	exon	11	0	0	HM
3_17	AT2G03667	SCR	SALK_145050C	intron	10	0	0	HM
3_18	AT4G24265	SCR	SALK_092556C	exon	10	1	0	HM
3_19	AT3G51580	SCR	SALK_053063C	promoter	0	0	12	WT
SCR4	AT3G51580	SCR	SALK_027835	intron	0	6	14	PL
3_20	AT3G21360	SCR	SALK_092692C	5utr	11	0	0	HM
3_21	AT2G13440	SCR	SALK_100713C	exon	8	0	0	HM
3_22	AT4G39690	SCR	SALK_087650C	exon	9	3	0	HM
3_24	AT4G18593	SCR	SALK_025575C	promoter	5	0	6	HM
3_25	AT1G08280	SCR	SALK_083443C	exon	12	0	0	HM
3_26	AT2G34050	SCR	SALK_040220C	exon	0	0	12	WT
3_27	AT1G50170	SCR	SALK_086731C	promoter	1	0	0	HM
SCR5	AT1G50170	SCR	SALK_001710	intron	0	0	11	WT
3_28	AT4G39450	SCR	SALK_026025C	exon	12	0	0	HM
3_29	AT2G27900	SCR	SALK_006327C	promoter	11	0	0	HM
SCR6	AT2G27900	SCR	SALK_026357	3utr exon	5	8	11	HM
3_30	AT2G44970	SCR	SALK_067058C	exon	0	0	0	OT
3_31	AT2G19940	SCR	SALK_138081C	intron	0	8	16	PL
3_32	AT5G62390	SCR	SALK_065883C	intron	4	0	8	HM

3_37	AT5G08540	SCR	SALK_100294C	exon	9	0	0	HM

Table S3.3 Observed and published phenotypes for PL SALK lines. In column 'Published phenotypes', ESN stands for 'essential', MRP stands for 'morphological', NOP stands for 'no obvious phenotype'. More detailed information regarding to these phenotype categories (ESN, MRP) can be found in Lloyd and Meinke, 2012 (cited in Chapter 3 main text).

mutant_ID	Gene	Group	Observed phenotype			Published phenotype
			On soil	Heterozygous siliques	Stress and root screen	
106	AT1G73350	DR	turned albino at 4 leaf stage			
130	AT3G10572	DR	tiny	small seeds		ESN
134	AT3G15180	DR	reduced fertility			
144	AT3G51820	DR				MRP
145	AT3G56210	DR	small			
149	AT3G60850	DR				
159	AT4G27750	DR	1/4 small and dying			MRP
35	AT2G37560	DR	tiny, small, yg	empty spaces		ESN
37	AT2G40430	DR				
3_23	AT3G10370	DR	lethal to small	small seeds	low germination rate (ABA, mannitol)	ESN
45	AT5G19130	DR			seg for tiny seedlings in sucrose	
50	AT2G26540	DR				NOP
56	AT5G48470	DR	small, lethal		seg for abnormal seedling (ABA, mannitol)	ESN
64	AT3G26710	DR	yellow green, lethal			MRP
66	AT2G05170	DR		small seeds	low germination rate (heat); cold sensitive	
75	AT2G30410	DR	small, embryo mut 2x			ESN
81	AT1G03760	DR				
92	AT1G26660	DR	small, yellow green			
94	AT1G31860	DR	lethal			ESN
SS1	AT2G31890	DR		albino seeds	seg for small plants or non-germiantion (sucrose)	
248	AT3G14900	DR				ESN
17	AT3G04260	nDR	small			MRP
29	AT5G48240	nDR	lethal	small seeds; empty spaces		
3_6	AT1G63680	nDR	seedling lethal; 1/4 small	albino seeds		ESN
1_34	AT4G12070	SCR	lethals, 1/4 tiny, no SAM			
3_31	AT2G19940	SCR				
SCR4	AT3G51580	SCR	seedling lethal			

Table S3.4 Soil growth phenotypes for HM SALK lines. In column 'Observed phenotype', 'YG' stands for 'yellow green', 'VLF' stands for 'very late flowering', 'RF' stands for 'reduced fertility'. Descriptions in column 'Published phenotype' came from references listed after this table or Lloyd and Meinke, 2012 (cited in Chapter 3 main text).

Mutant_ID	Gene	Observed phenotype			Published phenotype		
		Mutant	Phenotype	Category	Mutant	Phenotype	Group
107	AT1G76250	SALK_019804	YG, lethal	lethal			
11	AT3G25470	SALK_043556C	small, YG , early flowering	flowering			
115	AT2G22370	SALK_027178C	small curled leaves and YG; #11 VLF and no seed, #5 large leaves and RF; #8 small; #12 wt	multiple	SALK_027178[1]	cause a syndrome of related phenotypes affecting flowering time, inflorescence structure, and flower morphology.	MRP
122	AT2G46060	SALK_135304C	small lethals, retest	lethal			
127	AT3G09180	SALK_012449	small, dark green	pigment			
128	AT3G09210	SALK_039502	embryo lethals, retest	lethal			
13	AT4G00560	SALK_138740C	small, YG	pigment			
131	AT3G13940	SALK_054381	abnormal seeds	seed			
135	AT3G20070	SALK_148785	embryo lethals; #s10,11,12 fewer siliques, slightly smaller	seg	a non-SALK T-DNA insertion line (Syngenta)[2]	Embryo defective; Preglobular; Enlarged endosperm nuclei	ESN
160	AT4G29890	SALK_026428	small , YG	pigment			
177	AT5G11030	SALK_089074	seg small and abnormal	seg	SALK_063183[3]	defective in lateral root formation; fail to respond to exogenous IAA;	MRP
26	AT5G20600	SALK_073773C	small, silvery leaves	leaf			
34	AT1G61690	SALK_133410C	small, lethal	lethal			
40	AT5G15390	SALK_005531C	fewer leaf trichomes	leaf			
51	AT1G04130	SALK_073054C	YG, dying	pigment			
97	AT1G48270	SALK_027808	seg 32:8 for tiny seedlings	seg	CS540[4]	Abolishes Seed Dormancy; Enhances the Expression of Germition-Associated Genes; Accelerates Flowering and the Expression of Genes that Control Flowering.	MRP
99	AT1G60600	SALK_021962C	YG, lethal	lethal	Feldmann T-DNA line, no. 2755[5]	Seedling lethal without exogenous sucrose; Abnormal chloroplast development	ESN
PD7	AT4G21770	SALK_149232C	many small leaves	leaf			

1. Kim YJ, Zheng B, Yu Y, Won SY, Mo B, Chen X: The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. Embo Journal 2011, 30:814-822.

2. Tzafrir I, McElver JA, Liu Cm CM, Yang LJ, Wu JQ, Martinez A, Patton DA, Meinke DW: Diversity of TITAN functions in Arabidopsis seed development. Plant Physiology 2002, 128:38-51.

3. DiDonato RJ, Arbuckle E, Buker S, Sheets J, Tobar J, Totong R, Grisafi P, Fink GR, Celenza JL: Arabidopsis ALF4 encodes a nuclear-localized protein required for lateral root formation. Plant Journal 2004, 37:340-353.

4. Chen JG, Pandey S, Huang JR, Alonso JM, Ecker JR, Assmann SM, Jones AM: GCR1 can act independently of heterotrimeric G-protein in response to brassinosteroids and gibberellins in Arabidopsis seed germination. Plant Physiology 2004, 135:907-915.

5. Shimada H, Ohno R, Shibata M, Ikegami I, Onai K, Ohto MA, Takamiya K: Inactivation and deficiency of core proteins of photosystems I and II caused by genetical phylloquinone and plastoquinone deficiency but retained lamellar structure in a T-DNA mutant of Arabidopsis. Plant Journal 2005, 41:627-637.

Table S3.5 Phenotype scores for HM SALK lines in stress and root screens.

Mutant_ID	Gene	Group	ABA	Mannitol	Salt	Sucrose	Heat	Heatrec	Coldgro	Coldger	Root
1	AT5G15750	DR	3	3	3	3	3	3	sr	SR	1
2	AT1G74880	DR	3	3	3	3	3	3	3	NA	NA
5	AT5G63460	DR	3	3	3	3	3	3	3	3	sr
6	AT3G17590	DR	3	3	3	3	3	3	3	3	sr
7	AT1G06510	DR	3	3	3	3	sr	sr	3	3	sr
8	AT3G05210	DR	3	3	3	3	3	3	3	3	sr
9	AT3G56570	DR	3	3	3	3	sr	sr	3	3	sr
11	AT3G25470	DR	3	3	3	3	3	3	3	SR	sr
13	AT4G00560	DR	3	3	3	3	3	3	3	NA	NA
15	AT2G33255	DR	3	3	2	3	sr	sr	3	3	sr
18	AT2G39090	DR	3	3	2	3	3	3	3	3	sr
19	AT3G13226	DR	3	3	3	3	sr	sr	4	3	sr
23	AT1G16970	DR	3	3	3	3	3	sr	3	NA	3
24	AT5G49550	DR	3	3	3	3	sr	sr	3	SR	sr
26	AT5G20600	DR	seg	3	3	2	3	3	3	3	3
28	AT2G44870	DR	3	3	3	3	3	3	3	3	NA
34	AT1G61690	DR	1	sr	1	1	sr	NA	3	NA	NA
38	AT2G31955	DR	3	3	3	3	1	sr	3	3	sr
39	AT4G13670	DR	3	3	3	3	3	sr	3	YG	sr
40	AT5G15390	DR	3	3	3	3	sr	sr	3	3	sr
42	AT2G41530	DR	3	3	3	3	sr	sr	3	3	NA
43	AT5G23395	DR	3	3	3	3	3	3	3	3	NA
44	AT5G23290	DR	3	3	3	3	3	3	3	SR	sr
46	AT3G09580	DR	2	3	3	3	3	3	3	3	3
47	AT1G18730	DR	3	3	3	3	sr	sr	3	3	5
49	AT5G48440	DR	3	3	3	3	3	3	3	3	3
51	AT1G04130	DR	3	3	3	3	sr	sr	3	SR	1
52	AT2G02590	DR	3	3	3	3	sr	sr	3	3	sr
55	AT1G77550	DR	3	3	3	2	sr	3	3	3	NA
57	AT2G25605	DR	3	3	3	4	3	3	3	SR	NA
63	AT2G47760	DR	2	2	3	2	3	sr	4	3	sr
65	AT2G31040	DR	3	3	3	3	3	3	sr	YG	sr
67	AT2G17900	DR	3	2	3	3	sr	sr	3	3	NA
70	AT5G37290	DR	3	3	3	3	3	sr	3	SR	sr
71	AT1G53120	DR	3	3	3	3	3	3	3	NA	3
74	AT1G52530	DR	3	3	3	3	sr	sr	3	3	sr
76	AT4G03200	DR	4	3	3	3	sr	sr	3	SR	NA
77	AT3G25120	DR	4	3	3	3	3	3	3	NA	3
79	AT1G01920	DR	3	3	2	3	3	3	3	3	sr
82	AT1G05060	DR	3	3	3	3	sr	sr	3	3	sr

83	AT1G07645	DR	4	3	2	3	3	3	3	NA	3
84	AT1G08710	DR	3	3	3	3	3	3	3	3	sr
86	AT1G10830	DR	3	3	3	3	3	3	3	3	3
88	AT1G12790	DR	3	3	3	3	3	3	3	3	sr
90	AT1G15980	DR	3	3	3	3	sr	sr	3	3	sr
93	AT1G27530	DR	3	3	3	3	3	3	3	NA	3
95	AT1G36310	DR	3	3	3	2	sr	sr	3	3	3
96	AT1G42990	DR	3	3	3	3	3	3	3	NA	3
97	AT1G48270	DR	3	3	3	3	3	3	seg	NA	3
99	AT1G60600	DR	3	3	3	3	3	3	3	SR	3
100	AT1G62250	DR	3	3	3	3	3	3	2	NA	3
101	AT1G65020	DR	3	3	3	3	3	3	3	3	sr
102	AT1G65230	DR	3	3	3	2	3	3	3	3	sr
103	AT1G65900	DR	3	3	3	3	3	3	3	3	sr
104	AT1G66080	DR	3	3	NA	sr	2	3	NA	NA	NA
105	AT1G70570	DR	3	3	3	3	3	3	3	NA	2
107	AT1G76250	DR	NA	NA	NA	3	sr	sr	3	NA	3
108	AT1G77230	DR	3	3	3	3	sr	sr	3	3	sr
111	AT2G04270	DR	3	3	3	3	3	3	3	SR	3
113	AT2G19640	DR	3	3	3	3	5	3	NA	NA	NA
115	AT2G22370	DR	3	3	3	3	3	3	3	3	sr
116	AT2G22650	DR	3	3	3	3	3	3	3	3	sr
117	AT2G34090	DR	3	3	3	3	3	3	3	3	sr
119	AT2G42780	DR	4	3	3	3	3	3	3	NA	3
120	AT2G43400	DR	3	3	3	3	3	3	3	SR	sr
121	AT2G45990	DR	3	3	3	4	3	3	3	SR	sr
122	AT2G46060	DR	3	3	3	3	3	3	3	3	sr
124	AT3G04560	DR	3	3	3	3	3	3	3	3	sr
127	AT3G09180	DR	3	3	3	3	sr	sr	3	3	NA
128	AT3G09210	DR	3	3	3	3	3	3	3	3	3
131	AT3G13940	DR	3	3	3	3	3	3	3	NA	NA
132	AT3G14110	DR	3	3	3	3	3	3	NA	3	3
133	AT3G15110	DR	3	3	3	3	3	3	3	3	3
135	AT3G20070	DR	3	3	3	seg	3	3	3	SR	seg
136	AT3G21820	DR	3	3	3	3	sr	sr	3	3	3
137	AT3G24315	DR	3	3	3	3	3	3	3	3	seg
139	AT3G26085	DR	3	3	3	3	NA	NA	3	3	1
140	AT3G28760	DR	3	3	3	3	sr	NA	3	3	sr
141	AT3G48120	DR	3	3	3	3	sr	NA	3	3	sr
142	AT3G49890	DR	3	3	3	3	3	3	3	3	sr
143	AT3G51010	DR	3	3	3	3	3	3	3	NA	NA
147	AT3G56820	DR	3	4	3	3	sr	sr	3	NA	sr
148	AT3G59490	DR	3	3	3	3	sr	sr	3	3	sr

151	AT3G62370	DR	3	3	3	3	sr	sr	3	3	sr
153	AT4G15520	DR	3	3	3	3	sr	sr	3	3	sr
154	AT4G17370	DR	3	3	3	3	sr	sr	3	3	sr
155	AT4G17540	DR	3	2	3	3	3	3	3	3	3
157	AT4G18470	DR	3	3	3	3	3	3	3	3	3
158	AT4G26370	DR	3	3	3	3	3	3	3	SR	sr
160	AT4G29890	DR	3	3	3	3	sr	sr	4	3	sr
161	AT4G30840	DR	3	3	3	3	3	3	3	SR	3
163	AT4G34700	DR	3	3	3	3	sr	sr	sr	3	sr
164	AT4G37020	DR	3	3		3	sr	sr	3	3	sr
166	AT5G03770	DR	3	3	3	3	sr	sr	2	3	sr
167	AT5G62140	DR	3	3	3	4	sr	sr	5	CT	sr
170	AT5G55500	DR	3	3	3	3	NA	sr	3	SR	sr
171	AT5G52190	DR	3	3	2	3	sr	sr	3	3	sr
172	AT5G20935	DR	3	3	3	3	sr	sr	3	3	sr
173	AT5G10320	DR	3	3	3	3	sr	sr	3	3	sr
174	AT5G15802	DR	3	3	3	3	sr	sr	3	3	sr
175	AT5G54855	DR	3	3	3	3	sr	sr	3	SR	sr
177	AT5G11030	DR	seg	3	3	3	3	3	3	SR	3
178	AT5G17240	DR	3	3	3	3	sr	sr	3	3	sr
179	AT5G51130	DR	3	3	3	3	3	3	3	NA	3
184	AT4G38090	DR	3	3	3	3	sr	sr	3	SR	sr
206	AT1G45150	DR	2	3	3	3	3	sr	3	NA	3
220	AT1G77320	DR	3	3	3	3	3	sr	3	NA	3
242	AT3G04950	DR	3	3	sr	3	3	sr	3	NA	3
246	AT3G12210	DR	3	3	sr	3	3	sr	1	NA	3
250	AT3G16990	DR	3	3	3	3	3	sr	3	NA	3
252	AT3G20480	DR	3	3	3	3	3	sr	3	NA	3
264	AT4G02110	DR	3	3	sr	3	3	sr	3	NA	3
272	AT4G20350	DR	3	3	sr	3	3	sr	2	NA	3
277	AT4G29520	DR	3	3	3	3	3	sr	3	NA	3
293	AT5G25480	DR	3	3	NA	3	3	sr	3	NA	3
1_15	AT4G13330	DR	3	3	3	3	sr	sr	sr	3	3
1_6	AT2G40316	DR	3	3	3	3	3	sr	3	3	sr
2_4	AT5G13240	DR	3	3	3	3	sr	sr	3	3	sr
3_1	AT2G20980	DR	3	3	3	3	3	3	sr	SR	sr
3_11	AT5G62760	DR	3	3	3	3	3	3	3	3	sr
3_35	AT3G17670	DR	3	3	3	1	3	3	2	3	sr
3_36	AT1G56345	DR	3	3	3	1	3	3	2	3	2
3_4	AT3G16270	DR	3	3	2	3	sr	sr	3	3	sr
3_8	AT3G46200	DR	3	3	3	3	3	3	3	SR	sr
3	AT4G02405	nDR	3	3	3	3	sr	sr	3	3	3
4	AT3G27110	nDR	3	3	3	3	sr	sr	3	3	sr

10	AT2G36885	nDR	3	3	3	3	3	sr	3	3	sr
12	AT3G46220	nDR	3	3	3	3	sr	sr	3	3	sr
14	AT5G02710	nDR	3	3	3	3	sr	sr	3	3	sr
16	AT5G67290	nDR	3	3	3	3	sr	sr	3	3	sr
20	AT2G41120	nDR	4	3	3	3	3	3	3	NA	3
22	AT3G26580	nDR	3	3	3	3	3	3	3	NA	3
25	AT2G41760	nDR	3	2	3	3	sr	sr	2	3	sr
27	AT1G75200	nDR	3	3	3	3	sr	sr	3	SR	3
30	AT1G61570	nDR	3	3	3	3	3	3	3	NA	NA
31	AT3G17170	nDR	3	3	3	3	sr	sr	sr	3	NA
48	AT1G19490	nDR	3	3	3	3	sr	sr	3	3	3
53	AT1G01710	nDR	3	3	3	3	sr	sr	sr	CS	sr
54	AT1G51080	nDR	2	3	3	2	sr	sr	3	3	sr
60	AT5G37590	nDR	3	3	3	3	3	3	3	SR	sr
68	AT1G08220	nDR	seg	3	3	3	3	3	3	3	NA
72	AT5G19050	nDR	3	3	3	3	sr	sr	3	3	sr
73	AT1G70200	nDR	3	3	3	3	sr	sr	3	3	sr
1_10	AT3G27050	nDR	3	3	3	3	sr	sr	3	3	3
1_18	AT5G50930	nDR	3	3	3	3	sr	sr	3	3	2
1_19	AT5G59140	nDR	3	3	3	3	seg	sr	sr	SR	3
1_3	AT1G68080	nDR	3	3	3	3	3	3	3	3	sr
2_3	AT3G20490	nDR	3	3	3	3	3	3	2	3	sr
3_12	AT3G09085	nDR	3	3	3	1	sr	sr	sr	SR	sr
3_3	AT1G27752	nDR	4	3	3	3	3	3	3	NA	3
3_34	AT4G17610	nDR	3	3	3	seg	3	sr	2	3	3
3_5	AT1G43245	nDR	3	3	3	3	sr	sr	3	SR	sr
3_7	AT2G44820	nDR	3	3	3	3	3	3	3	3	sr
1_26	AT1G58250	SCR	3	3	3	3	3	sr	3	SR	3
1_28	AT2G01170	SCR	3	3	3	3	3	3	3	3	3
1_30	AT3G24630	SCR	3	3	3	2	sr	sr	3	SR	3
1_31	AT3G57990	SCR	3	3	3	3	sr	sr	3	SR	3
1_37	AT5G06120	SCR	3	3	3	3	sr	sr	3	3	3
1_47	AT3G52260	SCR	3	3	3	3	sr	sr	3	3	sr
2_5	AT5G40940	SCR	3	3	3	3	sr	sr	3	SR	3
2_6	AT5G24090	SCR	3	3	3	3	sr	sr	3	3	3
3_14	AT1G78780	SCR	3	3	3	3	sr	sr	sr	CS	3
3_16	AT3G13130	SCR	3	3	3	3	sr	sr	2	SR	3
3_17	AT2G03667	SCR	3	3	3	3	sr	sr	3	NA	sr
3_18	AT4G24265	SCR	3	2	3	3	sr	sr	3	NA	sr
3_20	AT3G21360	SCR	3	3	3	3	3	3	3	SR	3
3_21	AT2G13440	SCR	1	3	3	3	0	0	3	CS	3
3_22	AT4G39690	SCR	2	3	3	3	3	3	4	3	2
3_24	AT4G18593	SCR	3	3	3	3	sr	sr	3	3	sr

3_25	AT1G08280	SCR	3	3	3	3	sr	sr	3	3	3
3_29	AT2G27900	SCR	3	3	3	3	sr	sr	3	3	3
3_32	AT5G62390	SCR	3	3	3	3	3	3	2	3	3
3_37	AT5G08540	SCR	3	3	3	2	sr	sr	3	SR	sr

Table S3.6 HM SALK lines with striking stress phenotype. Protein descriptions are from TAIR. In column 'Stress or root phenotype', numbers in the parentheses are phenotype scores.

Mutant_ID	Gene	Group	Protein description	Stress or root phenotype
34	AT1G61690	DR	phosphoinositide binding	ABA(1), salt(1), sucrose(1)
38	AT2G31955	DR	cofactor of nitrate reductase and xanthine dehydrogenase 2	heat(1)
39	AT4G13670	DR	plastid transcriptionally active 5	coldger(yg)
65	AT2G31040	DR	ATCGL160, an integral thylakoid protein that facilitates assembly of the membranous part of the chloroplast ATPase.	coldger(yg)
113	AT2G19640	DR	ASH1-RELATED 2, ASHR2, SDG39, SET DOMAIN PROTEIN 39	heat(5)
167	AT5G62140	DR	unknown	sucrose(4), coldgro(5), coldger(CT)
246	AT3G12210	DR	DNA binding; involved in DNA repair; Helix-hairpin-helix DNA-binding motif	coldgro(1)
3_35	AT3G17670	DR	tetratricopeptide repeat (TPR)-containing protein	sucrose(1), coldgro(2)
3_36	AT1G56345	DR	Pseudouridine synthase family protein; RNA modification	sucrose(1), coldgro(2), root(2)
3_12	AT3G09085	nDR	Protein of unknown function (DUF962)	sucrose(1)
3_14	AT1G78780	SCR	pathogenesis-related family protein	coldger(cs)
3_21	AT2G13440	SCR	glucose-inhibited division family A protein	ABA(1), heat(0), heatrec(0), coldger(cs)

Table S3.7 HM SALK lines with mild stress phenotype. Protein descriptions are from TAIR.

In column 'Stress or root phenotype', numbers in the parentheses are phenotype scores.

Mutant_ID	Gene	Group	Protein description	Stress or root phenotype
15	AT2G33255	DR	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein	salt(2)
18	AT2G39090	DR	ANAPHASE-PROMOTING COMPLEX 7; contains TRP domain	salt(2)
19	AT3G13226	DR	regulatory protein RecX family protein	coldgro(4)
46	AT3G09580	DR	FAD/NAD(P)-binding oxidoreductase family protein	ABA(2)
55	AT1G77550	DR	tubulin-tyrosine ligases;tubulin-tyrosine ligases	sucrose(2)
57	AT2G25605	DR		sucrose(4)
63	AT2G47760	DR	asparagine-linked glycosylation 3	ABA(2), mannitol(2), sucrose(2), coldgro(4)
67	AT2G17900	DR	SET domain group 37	mannitol(2)
76	AT4G03200	DR	catalytics; Protein of unknown function DUF255, Thioredoxin fold, Six-hairpin glycosidase-like, Thioredoxin-like fold;	ABA(4)
77	AT3G25120	DR	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	ABA(4)
79	AT1G01920	DR	SET domain protein	salt(2)
83	AT1G07645	DR	dessication-induced 1VOC superfamily protein	ABA(4), salt(2)
95	AT1G36310	DR	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	sucrose(2)
100	AT1G62250	DR		coldgro(2)
102	AT1G65230	DR		sucrose(2)
104	AT1G66080	DR		heat(2)
119	AT2G42780	DR	RNA polymerase II transcription factor SIII, subunit A; regulation of transcription	ABA(4)
121	AT2G45990	DR		sucrose(4)
147	AT3G56820	DR		mannitol(4)
155	AT4G17540	DR		mannitol(2)
166	AT5G03770	DR	KDO transferase A	coldgro(2)
171	AT5G52190	DR	Sugar isomerase (SIS) family protein	salt(2)
206	AT1G45150	DR		ABA(2)
272	AT4G20350	DR	oxidoreductases; contains oxoglutarate/iron-dependent oxygenase	coldgro(2)
3_4	AT3G16270	DR	ENTH/VHS family protein; intracellular protein transport	salt(2)
1_30	AT3G24630	SCR	TON1 Recruiting motif (TRM) protein; TRM34	sucrose(2)
3_16	AT3G13130	SCR		coldgro(2)
3_18	AT4G24265	SCR		mannitol(2)

3_22	AT4G39690	SCR	Mitochondrial inner membrane protein Mitofilin	ABA(2), coldgro(4), root(2)
3_32	AT5G62390	SCR	BCL-2-associated athanogene 7	coldgro(2)
3_37	AT5G08540	SCR		sucrose(2)
20	AT2G41120	nDR		ABA(4)
25	AT2G41760	nDR		mannitol(2), coldgro(2)
54	AT1G51080	nDR		ABA(2), sucrose(2)
2_3	AT3G20490	nDR		coldgro(2)
3_3	AT1G27752	nDR	Ubiquitin system component Cue protein	ABA(4)
3_34	AT4G17610	nDR	tRNA/rRNA methyltransferase (SpoU) family protein	sucrose(seg), coldgro(2)

Table S3.8 P values for phenotype percentage comparison. Fisher exact test was applied for comparison ‘SCR vs. DR’ and ‘nDR vs. DR’. Z test was applied for ‘(SCR+nDR) vs. DR’.

Comparison	ABA	Mannitol	Salt	Sucrose	Heat	Heatrec	Coldgro	Coldger	Overall	Striking
SCR vs. DR	0.6302	0.5319	0.5933	1.0000	0.4564	0.1418	0.3952	0.1790	0.3085	0.6494
nDR vs. DR	0.4373	1.0000	0.3470	1.0000	1.0000	1.0000	0.7155	1.0000	1.0000	1.0000
(SCR+nDR) vs. DR	0.1995	0.3909	0.0434	0.4467	0.4379	0.0575	0.2056	0.1541	0.3087	0.4140

Table S3.9 HM SALK lines that are considered knockdown or correspond to an obsolete gene. Phenotype data in column 'Published phenotypes' are from references listed after this table or Lloyd and Meinke, 2012 (cited in Chapter 3 main text).

Mutant_ID	Gene	Group	Published phenotype	ABA	Mannitol	Salt	Sucrose	Heat	Heatrec	Coldgro	Coldger	Root
58	AT5G11450	DR	MRP[1]	3	3	3	4	3	3	3	3	3
78	AT4G31770	DR	ESN	4	3	3	3	sr	sr	sr	3	NA
1_49	AT4G21770	DR	NULL	3	3	3	3	2	3	3	3	sr
3_2	AT2G31890	DR	NULL	3	3	3	3	3	3	3	3	sr
3_27	AT1G50170	SCR	ESN	3	3	3	3	0	3	4	3	3
3_28	AT4G39450	SCR	NULL	3	3	3	3	sr	sr	3	3	3

1. Roose JL, Frankel LK, Bricker TM: Developmental defects in mutants of the PsbP domain protein 5 in Arabidopsis thaliana. *PLoS One* 2011, 6:e28624.

Table S3.10 HM SALK lines without observed phenotypes and with published phenotypes.

Mutant_ID	Gene	Our mutant	Mutant with published phenotypes	Published phenotype description
6	AT3G17590	SALK_073635C	RNAi line	Plants with >=80% of the WT mRNA didn't have any phenotype; Plants with <=15% mRNA were bushy.[1]
44	AT5G23290	SALK_057848C	CT955299	smaller than WT; sensitive to 100mM NaCl; slightly sensitive to 12mM LiCl and 300mM Mannitol[2]
86	AT1G10830	SALK_068653C	SALK_136385, CS859876, SALK_057053, and SALK_057915C	defect in greening after transfering from dark to light[3]
88	AT1G12790	SALK_127447C	SALK_127447	shorter siliques; a portion of flowers have less pollens; [4]
111	AT2G04270	SALK_093546C	SALK_093546	albino on agar plates; albino seedings on soil [5]
120	AT2G43400	SALK_007870	SALK_007870	shorter siliques and lower number of seeds[6]
132	AT3G14110	SALK_002383C	flu; T-DNA insertion mutant, not SALK	growth inhibition after transferring from dark to light[7]
157	AT4G18470	SALK_018281	EMS mutant	much smaller than WT[8]
163	AT4G34700	SALK_030356	SALK_097732	late flowering at 40 days; looks smaller (not very striking but visible) than WT at 8 week; shorter root (not quite striking)[9]
220	AT1G77320	SALK_201730C	Versailles collection of T-DNA lines	very short silique and very few seeds (1.5 per silique); seeds slightly larger than WT[10]

1. Brzeski J, Podstolski W, Olczak K, Jerzmanowski A: Identification and analysis of the Arabidopsis thaliana BSH gene, a member of the SNF5 gene family. Nucleic Acids Res 1999, 27:2393-2399.

2. Rodriguez-Milla MA, Salinas J: Prefoldins 3 and 5 play an essential role in Arabidopsis tolerance to salt stress. Mol Plant 2009, 2:526-534.

3. Chen Y, Li F, Wurtzel ET: Isolation and characterization of the Z-ISO gene encoding a missing component of carotenoid biosynthesis in plants. Plant Physiology 2010, 153:66-79.

4. Wijeratne AJ, Chen CB, Zhang W, Timofejeva L, Ma H: The Arabidopsis thaliana PARTING DANCERS gene encoding a novel protein is required for normal meiotic homologous recombination. Molecular Biology of the Cell 2006, 17:1331-1343.

5. Mudd EA, Sullivan S, Gisby MF, Mironov A, Kwon CS, Chung WI, Day A: A 125 kDa RNase E/G-like protein is present in plastids and is essential for chloroplast development and autotrophic growth in Arabidopsis. J Exp Bot 2008, 59:2597-2610.

6. Ishizaki K, Larson TR, Schauer N, Fernie AR, Graham IA, Leaver CJ: The critical role of Arabidopsis electron-transfer flavoprotein: Ubiquinone oxidoreductase during dark-induced starvation. Plant Cell 2005, 17:2587-2600.

7. Kim C, Meskauskiene R, Zhang SR, Lee KP, Ashok ML, Blajecka K, Herrfurth C, Feussner I, Apel K: Chloroplasts of Arabidopsis Are the Source and a Primary Target of a Plant-Specific Programmed Cell Death Signaling Pathway. Plant Cell 2012, 24:3026-3039.

8. Mosher RA, Durrant WE, Wang D, Song JQ, Dong XN: A comprehensive structure-function analysis of Arabidopsis SNI1 defines essential regions and transcriptional repressor activity. Plant Cell 2006, 18:1750-1765.

9. Han L, Qin G, Kang D, Chen Z, Gu H, Qu LJ: A nuclear-encoded mitochondrial gene AtCIB22 is essential for plant development in Arabidopsis. J Genet Genomics 2010, 37:667-683.

10. Grelon M, Gendrot G, Vezon D, Pelletier G: The Arabidopsis MEI1 gene encodes a protein with five BRCT domains that is involved in meiosis-specific DNA repair events independent of SPO11-induced DSBs. Plant Journal 2003, 35:465-475.

Table S3.11 HM SALK lines without observed phenotype. Protein descriptions are from TAIR.

Mutant_ID	Gene	Group	Protein description
2	AT1G74880	DR	subunit NDH-O of NAD(P)H:plastoquinone dehydrogenase complex
5	AT5G63460	DR	SAP domain-containing protein
7	AT1G06510	DR	
8	AT3G05210	DR	a homolog of human ERCC1 protein
9	AT3G56570	DR	SET domain-containing protein; Rubisco methyltransferase family protein
23	AT1G16970	DR	KU70 homolog
24	AT5G49550	DR	BLOS2; Putative homolog of mammalian BLOC-1 Subunit 2
28	AT2G44870	DR	
42	AT2G41530	DR	S-formylglutathione hydrolase
43	AT5G23395	DR	Mia40, a component of the mitochondrial intermembrane space assembly machinery
49	AT5G48440	DR	FAD-dependent oxidoreductase family protein
52	AT2G02590	DR	
70	AT5G37290	DR	ARM repeat superfamily protein
71	AT1G53120	DR	RNA-binding S4 domain-containing protein
74	AT1G52530	DR	Hus1-like protein
82	AT1G05060	DR	
84	AT1G08710	DR	F-box family protein
90	AT1G15980	DR	a novel subunit of the chloroplast NAD(P)H dehydrogenase complex
93	AT1G27530	DR	Ubiquitin-conjugating enzyme/RWD-like, Ubiquitin-fold modifier-conjugating enzyme 1
96	AT1G42990	DR	ATBZIP60; contains a bZIP DNA binding domain and a putative transmembrane domain
101	AT1G65020	DR	
103	AT1G65900	DR	
108	AT1G77230	DR	Tetratricopeptide repeat (TPR)-like superfamily protein
116	AT2G22650	DR	FAD-dependent oxidoreductase family protein
117	AT2G34090	DR	MATERNAL EFFECT EMBRYO ARREST 18, MEE18
124	AT3G04560	DR	
133	AT3G15110	DR	
136	AT3G21820	DR	ATXR2, HISTONE-LYSINE N-METHYLTRANSFERASE ATXR2, SDG36, SET DOMAIN PROTEIN 36
137	AT3G24315	DR	Sec20 family protein
140	AT3G28760	DR	3-dehydroquinate synthase, prokaryotic-type
141	AT3G48120	DR	
142	AT3G49890	DR	
143	AT3G51010	DR	
148	AT3G59490	DR	
151	AT3G62370	DR	heme binding
153	AT4G15520	DR	tRNA/rRNA methyltransferase (SpoU) family protein

154	AT4G17370	DR	Oxidoreductase family protein
158	AT4G26370	DR	antitermination NusB domain-containing protein
161	AT4G30840	DR	Transducin/WD40 repeat-like superfamily protein
164	AT4G37020	DR	
170	AT5G55500	DR	beta-1,2-xylosyltransferase
172	AT5G20935	DR	
173	AT5G10320	DR	
174	AT5G15802	DR	
175	AT5G54855	DR	Pollen Ole e 1 allergen and extensin family protein
178	AT5G17240	DR	SET domain group 40
179	AT5G51130	DR	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
184	AT4G38090	DR	Ribosomal protein S5 domain 2-like superfamily protein
242	AT3G04950	DR	SEC-C motif
250	AT3G16990	DR	Haem oxygenase-like, multi-helical
252	AT3G20480	DR	tetraacyldisaccharide 4'-kinase family protein
264	AT4G02110	DR	transcription coactivators; contains BRCT domain
277	AT4G29520	DR	Saposin B
293	AT5G25480	DR	DNA methyltransferase-2; ATDNMT2
1_15	AT4G13330	DR	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
1_6	AT2G40316	DR	
2_4	AT5G13240	DR	transcription regulators; Maf1 regulator, RNA polymerase III transcriptional repressor, MAF1
3_1	AT2G20980	DR	MCM10, minichromosome maintenance 10
3_11	AT5G62760	DR	P-loop containing nucleoside triphosphate hydrolases superfamily protein
3_8	AT3G46200	DR	nudix hydrolase homolog 9
3	AT4G02405	nDR	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
4	AT3G27110	nDR	Peptidase family M48 family protein
10	AT2G36885	nDR	
12	AT3G46220	nDR	
14	AT5G02710	nDR	
16	AT5G67290	nDR	FAD-dependent oxidoreductase family protein
22	AT3G26580	nDR	Tetratricopeptide repeat (TPR)-like superfamily protein
27	AT1G75200	nDR	flavodoxin family protein / radical SAM domain-containing protein
30	AT1G61570	nDR	translocase of the inner mitochondrial membrane 13
31	AT3G17170	nDR	Translation elongation factor EF1B/ribosomal protein S6 family protein
48	AT1G19490	nDR	Basic-leucine zipper (bZIP) transcription factor family protein
60	AT5G37590	nDR	Tetratricopeptide repeat (TPR)-like superfamily protein
68	AT1G08220	nDR	ATPase assembly factor ATP10, mitochondria
72	AT5G19050	nDR	alpha/beta-Hydrolases superfamily protein
73	AT1G70200	nDR	RNA-binding (RRM/RBD/RNP motifs) family protein
1_10	AT3G27050	nDR	
1_18	AT5G50930	nDR	Histone superfamily protein

1_19	AT5G59140	nDR	BTB/POZ domain-containing protein
1_3	AT1G68080	nDR	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
3_5	AT1G43245	nDR	SET domain-containing protein
3_7	AT2G44820	nDR	
1_26	AT1G58250	SCR	Golgi-body localisation protein domain; RNA pol II promoter Fmp27 protein domain
1_28	AT2G01170	SCR	bidirectional amino acid transporter 1
1_31	AT3G57990	SCR	
1_37	AT5G06120	SCR	ARM repeat superfamily protein
1_47	AT3G52260	SCR	Pseudouridine synthase family protein
2_5	AT5G40940	SCR	putative fasciclin-like arabinogalactan protein 20
2_6	AT5G24090	SCR	chitinase A
3_17	AT2G03667	SCR	Asparagine synthase family protein
3_20	AT3G21360	SCR	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
3_24	AT4G18593	SCR	dual specificity protein phosphatase-related
3_25	AT1G08280	SCR	Glycosyltransferase family 29 (sialyltransferase) family protein
3_29	AT2G27900	SCR	Vacuolar protein sorting-associated protein 54

Table S3.12 DR HM SALK lines with observed phenotypes. In column 'Chloroplast' and 'Mitochondrion', 'Y' means the corresponding protein is located in chloroplast or mitochondrion, while 'N' means the protein is not located in chloroplast or mitochondrion.

Mutant_ID	Gene	Phenotype group	Chloroplast	Mitochondrion
1	AT5G15750	root	N	N
11	AT3G25470	soil growth	N	N
13	AT4G00560	soil growth	N	N
15	AT2G33255	stress mild	Y	N
18	AT2G39090	stress mild	N	N
19	AT3G13226	stress mild	N	N
26	AT5G20600	soil growth/stress mild	N	N
34	AT1G61690	stress striking	N	N
38	AT2G31955	stress striking	N	Y
39	AT4G13670	stress striking	Y	N
40	AT5G15390	soil growth	Y	N
46	AT3G09580	stress mild	Y	N
47	AT1G18730	root	Y	N
51	AT1G04130	soil growth/root	N	N
55	AT1G77550	stress mild	Y	N
57	AT2G25605	stress mild	N	N
63	AT2G47760	stress mild	N	Y
65	AT2G31040	stress striking	Y	N
67	AT2G17900	stress mild	N	N
76	AT4G03200	stress mild	Y	N
77	AT3G25120	stress mild	Y	Y
79	AT1G01920	stress mild	N	N
83	AT1G07645	stress mild	N	N
95	AT1G36310	stress mild	Y	N
99	AT1G60600	PL	Y	N
100	AT1G62250	stress mild	N	Y
102	AT1G65230	stress mild	Y	N
104	AT1G66080	stress mild	N	N
105	AT1G70570	root	Y	N
113	AT2G19640	stress striking	N	N
115	AT2G22370	soil growth	N	N
119	AT2G42780	stress mild	N	N
121	AT2G45990	stress mild	Y	N
127	AT3G09180	soil growth	N	N
131	AT3G13940	soil growth	N	N

135	AT3G20070	PL	N	N
139	AT3G26085	root	Y	N
147	AT3G56820	stress mild	N	Y
155	AT4G17540	stress mild	N	N
160	AT4G29890	soil growth	Y	N
166	AT5G03770	stress mild	N	Y
167	AT5G62140	stress striking	Y	N
171	AT5G52190	stress mild	N	N
177	AT5G11030	soil growth	N	N
206	AT1G45150	stress mild	N	N
246	AT3G12210	stress striking	N	N
272	AT4G20350	stress mild	N	N
3_35	AT3G17670	stress striking	N	N
3_36	AT1G56345	stress striking/root	N	Y
3_4	AT3G16270	stress mild	N	N
PD7	AT4G21770	soil growth	Y	N
106	AT1G73350	PL	N	N
130	AT3G10572	PL	N	Y
134	AT3G15180	PL	N	N
145	AT3G56210	PL	Y	N
159	AT4G27750	PL	N	N
35	AT2G37560	PL	N	N
3_23	AT3G10370	PL	N	Y
56	AT5G48470	PL	Y	N
64	AT3G26710	PL	Y	N
66	AT2G05170	PL	N	N
75	AT2G30410	PL	N	N
92	AT1G26660	PL	N	N
94	AT1G31860	PL	Y	N
SS1	AT2G31890	PL	Y	N
45	AT5G19130	PL	N	N

Table S3.13 Number and percentage of chloroplast-targeting (CT) genes in DR genes with stress phenotypes in each screen.

	ABA	Mannitol	Salt	Sucrose	Heat	Heatrec	Coldgro	Coldger
Number of DR genes with phenotypes	8	4	7	11	3	0	10	3
Number of CT DR genes with phenotypes	3	0	1	5	0	0	2	3
Percentage of CT DR genes	38%	0%	14%	45%	0%	NA	20%	100%

Table S3.14 Phenotype comparison among three published datasets. The three published datasets, '4000DS', 'Hanada' and 'Lloyd and Meinke' were introduced in Chapter 2. Column 'Gene' includes genes involved in both 4000DS and Hanada datasets. 'NP' stands for 'no phenotype'. 'NA' stands for 'not available'. Other letters in 'Phenotype Group' are abbreviations of phenotype categories: the ones in '4000DS' and 'Hanada' can be found in Hanada et al., 2009 (cited in Chapter 3 main text) , and those in 'Lloyd and Meinke' can be found in Lloyd and Meinke, 2012 (cited in Chapter 3 main text).

Gene	Phenotype Group		
	4000DS	Hanada	Lloyd and Meinke
AT2G48070	NP	S	V
AT1G12790	R	R	NA
AT1G21840	NP	C	H
AT2G47760	NP	NP	B
AT3G14900	NP	S	S
AT3G48500	NP	NP	L
AT4G27750	S	V	V
AT5G52290	NP	NP	R
AT1G20050	NP	S	S
AT1G79040	NP	NP	B
AT4G38240	NP	NP	H
AT1G76060	NP	S	S
AT2G33250	NP	V	NA
AT4G00800	NP	NP	G
AT1G21600	V	V	L
AT1G32080	V	V	NA
AT5G03455	S	V	H
AT5G48390	R	R	NA
AT4G10180	V	S	V
AT5G66120	NP	S	NA
AT1G21760	NP	NP	P
AT1G63970	V	V	L
AT1G71440	NP	S	S
AT2G33800	NP	S	S
AT2G34470	NP	C	H
AT2G36230	NP	S	G

AT2G48120	V	S	L
AT3G09090	NP	R	R
AT3G22590	NP	NP	T
AT3G55250	NP	S	V
AT4G00310	NP	NP	G
AT4G02980	NP	S	S
AT4G21320	NP	C	P
AT4G32260	V	V	L
AT5G14800	NP	NP	S
AT5G45380	NP	C	H
AT5G50320	NP	V	V
AT1G27760	NP	NP	V
AT3G01780	NP	R	G
AT3G54690	NP	NP	G
AT5G05680	V	V	S
AT1G67630	S	R	S
AT1G08190	NP	NP	L
AT1G65260	R	V	NA
AT1G78620	NP	S	NA
AT1G79790	NP	S	NA
AT1G79850	NP	V	V
AT2G26460	NP	NP	V
AT3G05680	NP	S	S
AT3G63410	NP	V	L
AT4G30720	NP	NP	V
AT1G58250	NP	V	V
AT1G64790	NP	V	S
AT3G48110	NP	S	S
AT1G67140	NP	NP	V
AT1G08260	NP	S	S
AT1G15710	V	V	NA
AT1G16630	NP	S	NA
AT1G21280	V	V	NA
AT1G21640	R	V	NA
AT1G22700	V	V	L
AT1G22940	V	V	L
AT1G24490	NP	V	V
AT1G43710	NP	NP	G
AT1G55370	NP	NP	B
AT1G55870	NP	V	S
AT1G56200	NP	NP	S

AT1G64520	NP	V	V
AT1G67370	R	R	R
AT1G80070	NP	S	S
AT1G80410	NP	NP	G
AT2G20860	R	R	NA
AT2G37690	NP	S	NA
AT2G45330	NP	NP	S
AT3G07060	NP	S	S
AT3G14060	NP	V	NA
AT3G15830	S	R	NA
AT3G27420	R	R	NA
AT3G29320	NP	NP	V
AT3G57860	NP	NP	R
AT3G58140	NP	S	NA
AT3G63420	NP	C	V
AT3G63490	V	S	S
AT4G00020	NP	NP	G
AT4G01690	V	V	NA
AT4G09980	NP	S	S
AT4G29400	V	V	NA
AT5G01630	NP	NP	H
AT5G08400	V	V	NA
AT5G14320	NP	S	S
AT5G14660	NP	V	L
AT5G19660	NP	C	H
AT5G44040	R	V	NA
AT5G57030	NP	NP	B
AT1G01860	NP	C	P
AT1G02910	R	V	V
AT1G04020	NP	C	V
AT1G10840	NP	NP	L
AT1G11070	V	V	NA
AT1G15690	NP	V	V
AT1G21680	V	V	NA
AT1G42550	NP	NP	C
AT1G63990	R	R	R
AT1G74140	V	V	NA
AT2G37970	NP	NP	P
AT2G38670	NP	S	S
AT3G01020	NP	NP	V
AT3G13170	NP	R	R

AT3G27530	NP	NP	V
AT3G55830	NP	V	L
AT3G58650	R	V	NA
AT3G61730	NP	NP	V
AT4G03280	NP	V	B
AT4G18480	V	S	L
AT4G35040	NP	NP	H
AT5G19820	NP	S	S
AT5G27010	V	V	NA
AT5G62500	NP	V	V
AT5G66570	NP	V	V
AT1G02205	R	R	V
AT1G05340	NP	V	NA
AT1G08130	NP	NP	S
AT1G18500	NP	S	V
AT1G28380	V	V	V
AT1G71230	NP	V	C
AT2G16650	NP	V	NA
AT2G25850	NP	NP	L
AT2G26550	NP	V	V
AT3G01120	NP	S	B
AT3G01440	NP	NP	B
AT3G01510	NP	NP	B
AT3G03860	V	V	NA
AT3G04790	NP	S	S
AT3G12080	NP	S	S
AT3G27750	NP	S	S
AT3G59220	NP	C	H
AT4G00600	V	V	NA
AT4G00620	NP	S	S
AT4G39640	V	V	V
AT5G12210	NP	NP	V
AT5G19530	V	V	V
AT1G01030	NP	NP	R
AT1G01060	NP	NP	T
AT1G01120	NP	V	V
AT1G01280	NP	R	R
AT1G01360	R	R	NA
AT1G01420	V	V	NA
AT1G01480	NP	NP	V
AT1G01570	V	V	NA

AT1G01580	V	V	NA
AT1G01950	NP	NP	V
AT1G02580	S	S	S
AT1G02730	V	V	V
AT1G03660	V	V	NA
AT1G03790	NP	NP	P
AT1G04110	NP	V	C
AT1G04220	NP	NP	V
AT1G04250	NP	V	V
AT1G04400	V	V	T
AT1G04470	V	V	NA
AT1G04540	NP	R	NA
AT1G04820	NP	V	V
AT1G04870	NP	NP	T
AT1G05550	R	R	NA
AT1G05600	NP	NP	S
AT1G05630	NP	V	C
AT1G05700	V	V	NA
AT1G05760	NP	C	I
AT1G05990	NP	NP	C
AT1G06040	NP	V	V
AT1G06490	NP	NP	V
AT1G07180	V	V	NA
AT1G07630	V	V	V
AT1G07890	V	V	V
AT1G08070	NP	S	NA
AT1G08810	NP	C	C
AT1G09090	NP	NP	V
AT1G09100	NP	NP	H
AT1G09270	NP	NP	I
AT1G09560	R	R	NA
AT1G09970	NP	NP	V
AT1G10270	NP	S	S
AT1G10370	NP	NP	P
AT1G10470	NP	V	P
AT1G10910	NP	V	S
AT1G10920	NP	C	I
AT1G11000	NP	NP	P
AT1G11210	R	V	NA
AT1G11220	R	R	NA
AT1G11370	V	V	NA

AT1G12110	NP	V	B
AT1G12240	NP	NP	V
AT1G12980	NP	S	V
AT1G14750	NP	R	R
AT1G14870	NP	NP	H
AT1G16060	NP	NP	V
AT1G16150	NP	C	V
AT1G17110	NP	NP	V
AT1G17140	NP	NP	V
AT1G17840	V	V	V
AT1G18100	NP	NP	H
AT1G18410	V	V	NA
AT1G18890	NP	NP	P
AT1G19250	NP	C	I
AT1G20090	NP	V	C
AT1G20110	NP	V	V
AT1G20960	NP	S	S
AT1G20990	R	V	NA
AT1G21270	NP	C	L
AT1G21690	NP	NP	S
AT1G21700	R	V	V
AT1G21970	NP	S	S
AT1G22090	NP	NP	S
AT1G22260	NP	NP	R
AT1G22400	NP	NP	I
AT1G22740	V	NP	NA
AT1G23090	NP	NP	V
AT1G23310	NP	C	V
AT1G23400	NP	NP	S
AT1G23420	NP	R	R
AT1G24260	R	R	NA
AT1G24460	V	V	NA
AT1G25490	NP	V	V
AT1G26790	V	V	NA
AT1G27080	NP	NP	R
AT1G27320	NP	NP	H
AT1G27450	NP	R	R
AT1G28300	NP	S	S
AT1G29260	NP	C	V
AT1G29600	S	R	NA
AT1G31470	NP	R	G

AT1G31880	NP	V	V
AT1G32060	NP	V	NA
AT1G32450	V	V	B
AT1G32640	V	V	NA
AT1G33240	NP	NP	C
AT1G35720	NP	C	P
AT1G48920	NP	V	V
AT1G49430	V	V	V
AT1G50040	NP	V	NA
AT1G51190	NP	V	V
AT1G51500	NP	V	V
AT1G54060	NP	NP	V
AT1G55020	NP	V	V
AT1G55380	NP	S	NA
AT1G56650	NP	C	H
AT1G58360	NP	C	H
AT1G61720	NP	R	R
AT1G61940	V	V	NA
AT1G62340	NP	V	L
AT1G62830	NP	V	T
AT1G62940	R	R	R
AT1G63440	NP	C	H
AT1G63880	NP	C	I
AT1G63900	NP	NP	I
AT1G64060	NP	V	V
AT1G64070	NP	C	I
AT1G64390	V	V	NA
AT1G64670	V	V	V
AT1G64760	R	R	NA
AT1G65360	NP	NP	G
AT1G65770	NP	NP	H
AT1G67730	NP	NP	S
AT1G68450	V	V	L
AT1G68560	R	R	V
AT1G69180	R	R	R
AT1G69440	R	V	T
AT1G69770	NP	NP	B
AT1G70750	V	V	NA
AT1G70910	NP	NP	V
AT1G74720	R	R	V
AT1G75030	V	V	NA

AT1G75100	NP	V	P
AT1G77860	NP	R	R
AT1G78000	NP	C	B
AT1G78580	NP	S	S
AT1G79840	NP	V	C
AT1G80080	NP	V	C
AT1G80100	NP	NP	C
AT1G80760	NP	NP	H
AT2G01050	R	R	NA
AT2G01140	NP	V	V
AT2G01190	NP	V	L
AT2G01390	R	R	S
AT2G01420	NP	S	S
AT2G01570	NP	NP	V
AT2G01800	R	R	NA
AT2G01830	NP	V	V
AT2G01940	NP	V	V
AT2G01950	NP	V	T
AT2G02000	R	V	NA
AT2G02150	NP	NP	S
AT2G02220	NP	V	V
AT2G02300	V	V	NA
AT2G03500	V	V	NA
AT2G17090	NP	NP	S
AT2G20180	NP	V	P
AT2G20750	NP	NP	P
AT2G21660	NP	NP	H
AT2G26490	NP	R	NA
AT2G26710	NP	NP	V
AT2G27050	NP	C	H
AT2G33730	S	R	NA
AT2G34650	NP	S	S
AT2G35350	NP	NP	R
AT2G36000	NP	S	S
AT2G38110	R	R	C
AT2G38740	V	V	NA
AT2G38770	NP	NP	S
AT2G41560	NP	NP	V
AT2G41850	NP	NP	P
AT2G44190	NP	S	S
AT2G46970	NP	V	P

AT2G47000	NP	V	B
AT2G47240	NP	NP	V
AT2G47750	NP	V	G
AT2G47940	NP	NP	S
AT3G01080	NP	NP	I
AT3G01140	NP	NP	C
AT3G01810	S	R	NA
AT3G02000	NP	R	R
AT3G02270	V	V	NA
AT3G03530	NP	NP	P
AT3G04260	V	V	V
AT3G06430	NP	S	S
AT3G06860	NP	V	L
AT3G07130	NP	C	R
AT3G07670	V	V	NA
AT3G08660	NP	V	NA
AT3G08970	NP	NP	G
AT3G09340	V	V	NA
AT3G10030	V	V	NA
AT3G10570	NP	NP	C
AT3G10800	NP	NP	P
AT3G13650	R	R	NA
AT3G14230	NP	S	L
AT3G14370	NP	V	P
AT3G14440	NP	C	H
AT3G16830	V	V	NA
AT3G18110	NP	NP	S
AT3G18390	NP	S	S
AT3G19210	NP	C	P
AT3G21560	V	V	B
AT3G21630	NP	C	I
AT3G22880	R	R	R
AT3G22960	V	V	NA
AT3G24220	NP	C	B
AT3G25250	NP	C	C
AT3G29635	R	R	NA
AT3G44260	NP	NP	H
AT3G44310	NP	C	H
AT3G45130	NP	NP	B
AT3G46550	NP	C	V
AT3G47500	NP	NP	P

AT3G47620	NP	S	P
AT3G48560	NP	S	NA
AT3G51570	V	V	NA
AT3G52280	NP	V	V
AT3G54460	R	R	NA
AT3G54920	NP	NP	I
AT3G57870	NP	S	G
AT3G59060	NP	C	P
AT3G60330	NP	NP	C
AT3G60460	NP	R	G
AT3G61510	NP	NP	V
AT3G61560	V	V	NA
AT3G62090	NP	NP	V
AT4G00120	R	R	NA
AT4G00220	NP	S	S
AT4G00650	NP	V	T
AT4G00730	NP	V	V
AT4G01020	R	NP	NA
AT4G01060	NP	V	C
AT4G01070	V	V	NA
AT4G01190	NP	NP	H
AT4G01370	NP	C	V
AT4G02060	NP	S	G
AT4G03570	V	V	NA
AT4G05200	R	R	NA
AT4G08150	NP	R	V
AT4G16950	NP	NP	I
AT4G20740	NP	S	S
AT4G20900	NP	R	R
AT4G21060	V	V	NA
AT4G23250	NP	NP	S
AT4G24190	NP	V	G
AT4G24860	NP	S	NA
AT4G26080	NP	V	H
AT4G26690	NP	V	C
AT4G27060	V	V	V
AT4G27600	V	V	L
AT4G28980	NP	NP	L
AT4G32551	R	R	R
AT4G33360	NP	NP	H
AT4G33650	NP	V	C

AT4G33790	NP	V	V
AT4G34131	V	V	NA
AT4G34940	NP	NP	G
AT4G37200	NP	V	L
AT4G38160	NP	NP	L
AT4G39620	NP	S	S
AT5G01540	NP	NP	H
AT5G01560	NP	NP	H
AT5G01820	NP	NP	P
AT5G01840	NP	NP	H
AT5G01920	NP	NP	C
AT5G04660	R	R	NA
AT5G04770	NP	NP	V
AT5G05850	NP	S	NA
AT5G06850	V	V	NA
AT5G07480	R	V	NA
AT5G07990	NP	R	R
AT5G08610	NP	V	V
AT5G08640	NP	NP	V
AT5G10120	V	V	NA
AT5G10330	NP	S	S
AT5G11060	V	V	NA
AT5G11320	R	R	NA
AT5G11890	NP	S	S
AT5G13320	NP	C	I
AT5G13480	NP	V	S
AT5G15450	NP	V	L
AT5G16750	NP	S	G
AT5G19520	NP	NP	B
AT5G20730	V	V	V
AT5G23630	R	R	V
AT5G23940	V	V	S
AT5G28030	NP	NP	T
AT5G39980	NP	S	S
AT5G43160	R	V	NA
AT5G43470	NP	C	I
AT5G43940	NP	V	V
AT5G45250	NP	V	I
AT5G45830	NP	S	V
AT5G46110	NP	V	B
AT5G46260	R	R	NA

AT5G47910	NP	C	V
AT5G48910	NP	V	V
AT5G49360	R	V	R
AT5G51100	V	V	V
AT5G53200	R	R	C
AT5G55740	NP	V	B
AT5G57160	NP	C	H
AT5G57180	NP	V	V
AT5G58090	R	R	NA
AT5G59920	NP	C	P
AT5G62920	NP	NP	P
AT5G64330	NP	V	P
AT5G64750	NP	C	H
AT5G66190	NP	V	V

Additional table M4.1-M4.2

Table M4.1 List of Addition (AD) lines.

Mutant_ID	Gene	Group
A1_49	AT4G21770	DR
A1_6	AT2G40316	DR
A2_1	AT1G76450	DR
A2_4	AT5G13240	DR
A3_1	AT2G20980	DR
A3_11	AT5G62760	DR
A3_2	AT2G31890	DR
A3_35	AT3G17670	DR
A3_36	AT1G56345	DR
A3_8	AT3G46200	DR
A3_9	AT3G61080	DR
A1_28	AT2G01170	SCR
A3_14	AT1G78780	SCR
A3_15	AT1G08370	SCR
A3_25	AT1G08280	SCR
A3_27	AT1G50170	SCR
A3_16	AT3G13130	SCR
A1_47	AT3G52260	SCR
A1_18	AT5G50930	nDR
A1_3	AT1G68080	nDR
A2_3	AT3G20490	nDR
A3_33	AT1G63980	nDR
A3_5	AT1G43245	nDR
A3_6	AT1G63680	nDR
A3_34	AT4G17610	nDR

Table M4.2 List of Overexpression (OE) lines.

Mutant_ID	Gene	Group
OE1	AT5G15750	DR
OE2	AT1G74880	DR
OE5	AT5G63460	DR
OE6	AT3G17590	DR
OE8	AT3G05210	DR
OE11	AT3G25470	DR
OE18	AT2G39090	DR
OE28	AT2G44870	DR
OE35	AT2G37560	DR
OE39	AT4G13670	DR
OE44	AT5G23290	DR
OE46	AT3G09580	DR
OE49	AT5G48440	DR
OE56	AT5G48470	DR
OE57	AT2G25605	DR
OE58	AT5G11450	DR
OE65	AT2G31040	DR
OE66	AT2G05170	DR
OE71	AT1G53120	DR
OE79	AT1G01920	DR
OE81	AT1G03760	DR
OE83	AT1G07645	DR
OE84	AT1G08710	DR
OE86	AT1G10830	DR
OE88	AT1G12790	DR
OE89	AT1G13990	DR
OE93	AT1G27530	DR
OE94	AT1G31860	DR
OE96	AT1G42990	DR
OE97	AT1G48270	DR
OE99	AT1G60600	DR
OE100	AT1G62250	DR
OE101	AT1G65020	DR
OE102	AT1G65230	DR
OE103	AT1G65900	DR
OE106	AT1G73350	DR
OE107	AT1G76250	DR
OE114	AT2G20940	DR
OE115	AT2G22370	DR
OE116	AT2G22650	DR

OE117	AT2G34090	DR
OE119	AT2G42780	DR
OE120	AT2G43400	DR
OE121	AT2G45990	DR
OE122	AT2G46060	DR
OE124	AT3G04560	DR
OE128	AT3G09210	DR
OE130	AT3G10572	DR
OE132	AT3G14110	DR
OE133	AT3G15110	DR
OE134	AT3G15180	DR
OE135	AT3G20070	DR
OE137	AT3G24315	DR
OE141	AT3G48120	DR
OE161	AT4G30840	DR
OE162	AT4G31460	DR
OE177	AT5G11030	DR
OE1_6	AT2G40316	DR
OE2_1	AT1G76450	DR
OE3_1	AT2G20980	DR
OE3_2	AT2G31890	DR
OE3_36	AT1G56345	DR
OE3_8	AT3G46200	DR
OE3_9	AT3G61080	DR
OE1_3	AT1G68080	nDR
OE2_3	AT3G20490	nDR
OE3_33	AT1G63980	nDR
OE3_27	AT1G50170	SCR

Additional table S4.1-S4.6

Table S4.1 Genes with OE soil growth phenotype and a KO phenotype.

Mutant_ID	Gene	Protein description	OE phenotypes	KO phenotype
OE2	AT1G74880	subunit NDH-O of NAD(P)H:plastoquinone dehydrogenase complex (Ndh complex) present in the thylakoid membrane of chloroplasts. This subunit is thought to be required for Ndh complex assembly.	small	no postillumination fluorescence transient; low sensitivity to antimycin A (AA); lack of shifted AG emissio[1]
OE6	AT3G17590	Encodes the Arabidopsis homologue of yeast SNF5 and represents a conserved subunit of plant SWI/SNF complexes	Reduced fertility	sexual sterility in addition to a characteristic "bushy" phenotype[2]
OE46	AT3G09580	FAD/NAD(P)-binding oxidoreductase family protein	small	Slightly sensitive in ABA screen
OE66	AT2G05170	Homologous to yeast VPS11. Forms a complex with VCL1 and AtVPS33. Involved in vacuolar biogenesis. The mRNA is cell-to-cell mobile.	small and purple; no seeds	poentially lethal (small seeds in heterozygous silique)
OE120	AT2G43400	ELECTRON-TRANSFER FLAVOPROTEIN:UBIQUINONE OXIDOREDUCTASE, ETFQO; Encodes a unique electron-transfer flavoprotein:ubiquinone oxidoreductase that is localized to the mitochondrion. Mutants are more sensitive to sugar starvation when plants are kept in the dark for long periods.	large, late flowering	shorter siliques and less seeds; Early senescence in the dark[3]

1. Courteille A, Vesa S, Sanz-Barrio R, Cazale AC, Becuwe-Linka N, Farran I, Havaux M, Rey P, Rumeau D: Thioredoxin m4 controls photosynthetic alternative electron pathways in Arabidopsis. Plant Physiology 2013, 161:508-520

2. Brzeski J, Podstolski W, Olczak K, Jerzmanowski A: Identification and analysis of the Arabidopsis thaliana BSH gene, a member of the SNF5 gene family. Nucleic Acids Res 1999, 27:2393-2399.

3. Ishizaki K, Larson TR, Schauer N, Fernie AR, Graham IA, Leaver CJ: The critical role of Arabidopsis electron-transfer flavoprotein:ubiquinone oxidoreductase during dark-induced starvation. Plant Cell 2005, 17:2587-2600.

Table S4.2 Genes with observed OE stress phenotypes and their corresponding KO phenotypes. Numbers in parentheses are phenotype scores. SG stands for 'soil growth'. PL stands for 'potentially lethal'. Phenotype categories in KO published can be found in Lloyd and Meinke, 2012 (cited in Table 4.5).

Mutant_ID	Gene	OE phenotype	KO Observed	KO Published
OE1	AT5G15750	sucrose(4)	root(1)	
OE11	AT3G25470	salt(2), sucrose(4)	SG(flowering)	
OE35	AT2G37560	salt(4)	PL	ESN
OE44	AT5G23290	salt(4)		MRP
OE57	AT2G25605	ABA(4)	sucrose(4)	
OE58	AT5G11450	salt(2),coldgro(2)		MRP[1]
OE71	AT1G53120	salt(2)		
OE79	AT1G01920	salt(2),coldgro(4)	salt(2)	
OE81	AT1G03760	salt(2),coldgro(4)		
OE86	AT1G10830	coldgro(4)		MRP[2]
OE88	AT1G12790	coldgro(4)		MRP[3]
OE89	AT1G13990	ABA(2), sucrose(2)		
OE93	AT1G27530	salt(2), sucrose(2)		
OE96	AT1G42990	coldgro(2)		CND[4]
OE99	AT1G60600	sucrose(2)	SG(lethal)	ESN
OE100	AT1G62250	coldgro(2)	coldgro(2)	
OE101	AT1G65020	salt(4),coldgro(2)		
OE103	AT1G65900	coldgro(4)		
OE106	AT1G73350	coldgro(4)	PL	
OE107	AT1G76250	sucrose(4)		
OE115	AT2G22370	sucrose(2)	SG(multiple)	MRP[5]
OE121	AT2G45990	ABA(2)	sucrose(4)	
OE177	AT5G11030	sucrose(4),coldgro(2)	SG(seg)	MRP[6]
OE3_1	AT2G20980	coldgro(2)		
OE3_36	AT1G56345	salt(4),root(4)	sucrose(1),coldgro(2),root(2)	

1. Roose JL, Frankel LK, Bricker TM: Developmental defects in mutants of the PsbP domain protein 5 in Arabidopsis thaliana. PLoS One 2011, 6:e28624.

2. Chen Y, Li F, Wurtzel ET: Isolation and characterization of the Z-ISO gene encoding a missing component of carotenoid biosynthesis in plants. Plant Physiology 2010, 153:66-79.

3. Wijeratne AJ, Chen CB, Zhang W, Timofejeva L, Ma H: The Arabidopsis thaliana PARTING DANCERS gene encoding a novel protein is required for normal meiotic homologous recombination. Molecular Biology of the Cell 2006, 17:1331-1343.

4. Humbert S, Zhong S, Deng Y, Howell SH, Rothstein SJ: Alteration of the bZIP60/IRE1 pathway affects plant response to ER stress in Arabidopsis thaliana. PLoS One 2012, 7:e39023.

5. Zheng Z, Guan H, Leal F, Grey PH, Oppenheimer DG: Mediator subunit18 controls flowering time and floral organ identity in Arabidopsis. PLoS One 2013, 8:e53924.

6. Celenza JL, Grisafi PL, Fink GR: A Pathway for Lateral Root-Formation in Arabidopsis-Thaliana. Genes & Development 1995, 9:2131-2142.

Table S4.3 Genes with both OE and KO stress phenotypes. Protein descriptions are from TAIR. Numbers in parentheses are phenotype scores.

Mutant_ID	Gene	Protein Description	OE phenotype	KO Observed
OE57	AT2G25605	unknown protein	ABA(4)	sucrose(4)
OE79	AT1G01920	SET domain protein	salt(2),coldgro(4)	salt(2)
OE100	AT1G62250	unknown protein	coldgro(2)	coldgro(2)
OE121	AT2G45990	unknown protein	ABA(2)	sucrose(4)
OE3_36	AT1G56345	Pseudouridine synthase family protein	salt(4),root(4)	sucrose(1),coldgro(2),root(2)

Table S4.4 DR OE lines without phenotypes and their corresponding KO phenotypes. Numbers in parentheses are phenotype scores. PL stands for 'potentially lethal'. YG stands for 'yellow green'. Phenotype categories in KO published can be found in Lloyd and Meinke, 2012 (cited in Chapter 3 main text).

Mutant_ID	Gene	KO Observed	KO Published
OE5	AT5G63460		
OE8	AT3G05210		CND[1]
OE18	AT2G39090	salt(2)	
OE28	AT2G44870		
OE39	AT4G13670	coldger(YG)	CND[2]
OE56	AT5G48470	PL	
OE65	AT2G31040	coldger(YG)	MRP[3]
OE83	AT1G07645	aba(4), salt(2)	
OE84	AT1G08710		
OE94	AT1G31860	PL	ESN
OE97	AT1G48270	SG(seg)	CND[4]
OE102	AT1G65230	sucrose(2)	
OE114	AT2G20940		
OE116	AT2G22650		
OE117	AT2G34090		
OE119	AT2G42780	aba(4)	
OE124	AT3G04560		
OE130	AT3G10572	PL	ESN[5]
OE132	AT3G14110		ESN
OE133	AT3G15110		
OE134	AT3G15180	PL	
OE135	AT3G20070	PL	ESN
OE137	AT3G24315		
OE141	AT3G48120		
OE1_6	AT2G40316		
OE2_1	AT1G76450		
OE3_2	AT2G31890	PL	
OE3_8	AT3G46200		
OE3_9	AT3G61080		

1. Dubest S, Gallego ME, White CI: Roles of the AtErcc1 protein in recombination. Plant Journal 2004, 39:334-342.

2. Zhong L, Zhou W, Wang H, Ding S, Lu Q, Wen X, Peng L, Zhang L, Lu C: Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in Arabidopsis under heat stress. Plant Cell 2013, 25:2925-2943.

3. Ruhle T, Razeghi JA, Vamvaka E, Viola S, Gandini C, Kleine T, Schunemann D, Barbato R, Jahns P, Leister D: The Arabidopsis protein CONSERVED ONLY IN THE GREEN LINEAGE160 promotes the assembly of the membranous part of the chloroplast ATP synthase. Plant Physiology 2014, 165:207-226.

4. Chen JG, Pandey S, Huang JR, Alonso JM, Ecker JR, Assmann SM, Jones AM: GCR1 can act independently of heterotrimeric G-protein in response to brassinosteroids and gibberellins in Arabidopsis seed germination. Plant Physiology 2004, 135:907-915.

5. Goto S, Mano S, Nakamori C, Nishimura M: Arabidopsis ABERRANT PEROXISOME MORPHOLOGY9 Is a Peroxin That Recruits the PEX1-PEX6 Complex to Peroxisomes. Plant Cell 2011, 23:1573-1587.

Table S4.5 OE and KO phenotypes for genes encoding components of protein complexes. SG stands for 'soil growth'. PL stands for 'potentially lethal'. Phenotype categories in column 'KO published' can be found in Lloyd and Meinke, 2012 (cited in Table 4.5). Phenotypes in 'KO published' come from publications cited in Table 4.1, Table 4.2, Table 4.4 and Table 4.5.

Mutant_ID	Gene	OE phenotype categories	KO observed	KO published
OE2	AT1G74880	SG, stress		CLB
OE6	AT3G17590	SG		MRP
OE18	AT2G39090		stress	NULL
OE35	AT2G37560	stress	PL	ESN
OE44	AT5G23290	stress		MRP
OE66	AT2G05170	SG	PL	NULL
OE115	AT2G22370	stress	SG	MRP
OE119	AT2G42780		stress	NULL
OE3_1	AT2G20980	stress		NULL

Table S4.6 OE and KO phenotypes for genes encoding chloroplast located proteins. SG stands for 'soil growth'. PL stands for 'potentially lethal'. Phenotype categories in column 'KO published' can be found in Lloyd and Meinke, 2012 (cited in Table 4.5). Phenotypes in 'KO published' come from publications cited in Table 4.1, Table 4.2, Table 4.4 and Table 4.5.

Mutant_ID	Gene	OE phenotype	KO observed	KO Published
OE2	AT1G74880	SG, stress		CLB
OE5	AT5G63460			
OE28	AT2G44870			
OE39	AT4G13670		stress	CND
OE46	AT3G09580	SG	stress	
OE49	AT5G48440	SG		
OE56	AT5G48470		PL	
OE58	AT5G11450	stress		MRP
OE65	AT2G31040		stress	MRP
OE71	AT1G53120	stress		
OE86	AT1G10830	stress		MRP
OE88	AT1G12790	stress		MRP
OE89	AT1G13990	stress		
OE93	AT1G27530	stress		
OE94	AT1G31860		PL	ESN
OE99	AT1G60600	stress	PL	ESN
OE102	AT1G65230		stress	
OE103	AT1G65900	stress		
OE116	AT2G22650			
OE121	AT2G45990	stress	stress	
OE124	AT3G04560			
OE128	AT3G09210	SG		
OE132	AT3G14110			ESN
OE133	AT3G15110			
OE161	AT4G30840	SG, stress		
OE1_6	AT2G40316			
OE2_1	AT1G76450			
OE3_2	AT2G31890		PL	
OE3_9	AT3G61080			
OE3_27	AT1G50170	stress	stress (knockdown)	ESN

Additional figure S2.1-S2.3

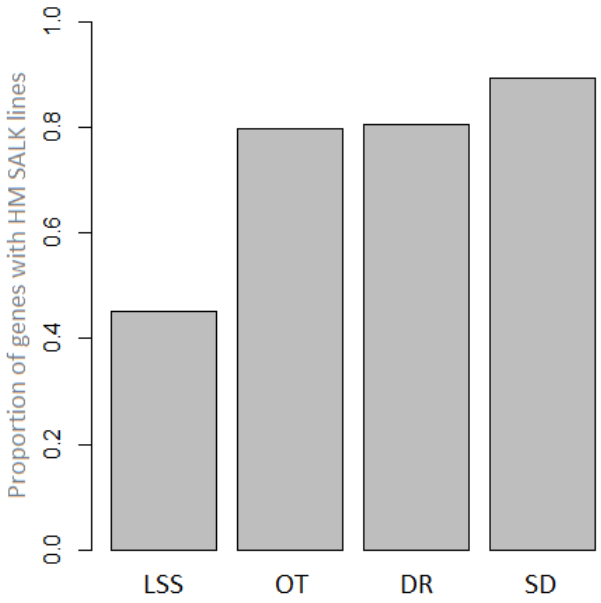

Figure S2.1 Proportion of genes with homozygous SALK lines in the four subsets.

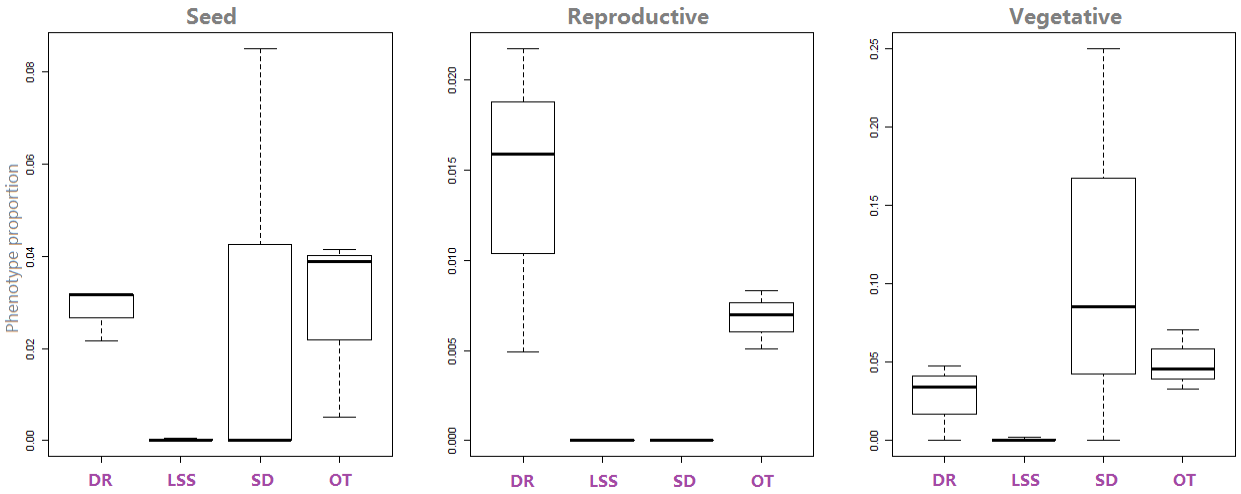

Figure S2.2 Phenotype proportion distribution among the three datasets in the three phenotypic categories.

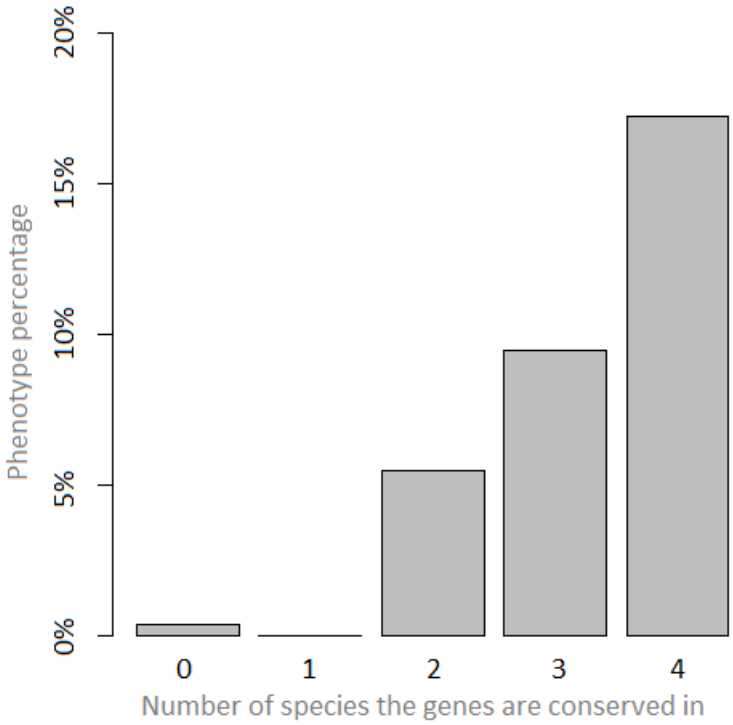

Figure S2.3 Phenotype percentages for single-copy genes with different conservation levels.

Additional figure M3.1-M3.6

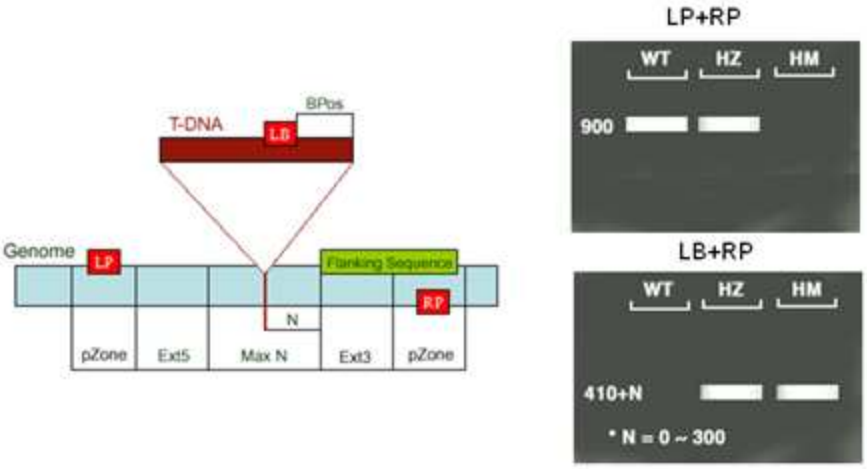

Figure M3.1 Expected gel image in SALK line genotyping (modified from http://signal.salk.edu/tdnaprimers.2.html).

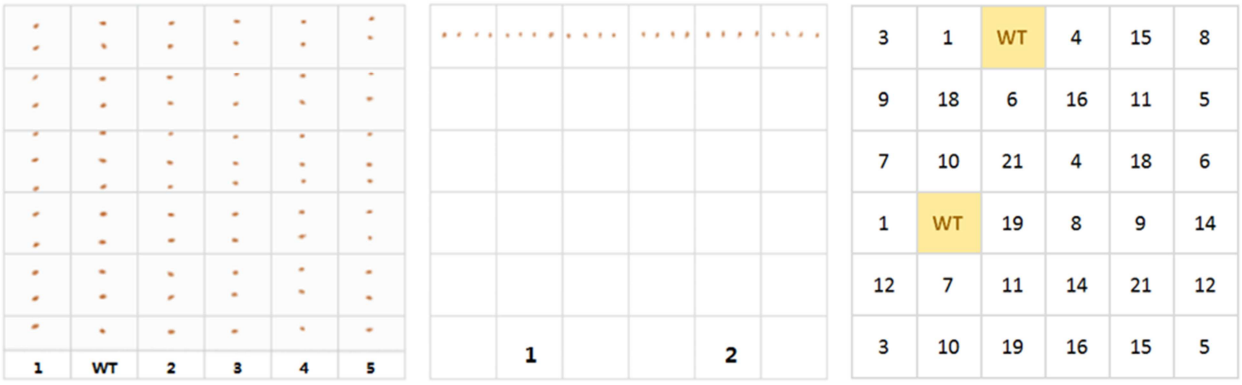

Figure M3.2 Plate arrangement for ABA, mannitol, sucrose, salt, cold growth, heat/heat recovery (left), root (middle) and cold germination screens (right). The numbers are mutant_IDs, each representing a SALK line. WT is the control, representing wild type Col0.

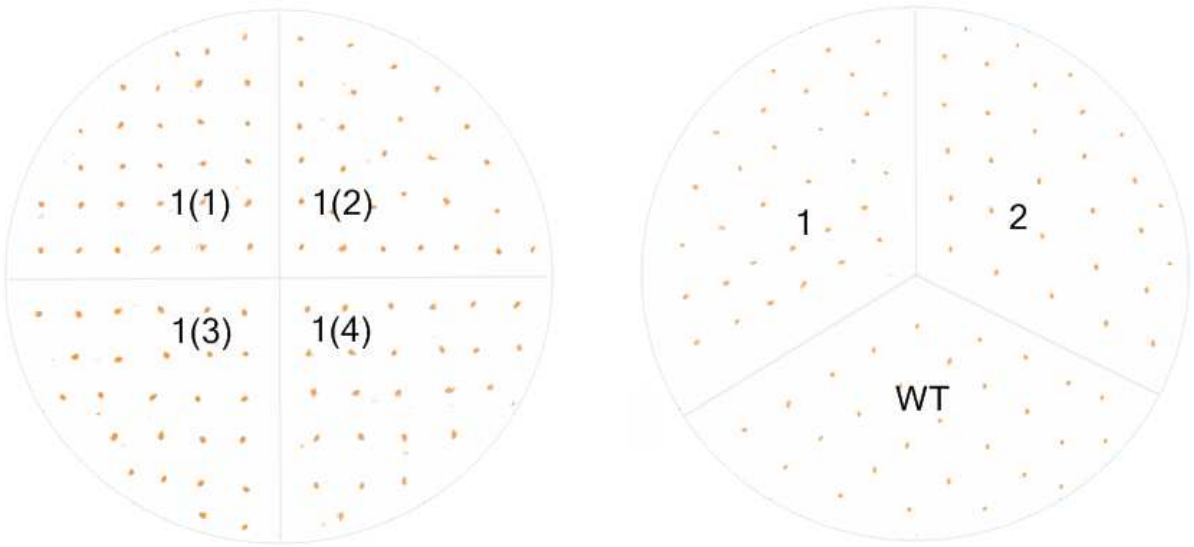

Figure M3.3 Plate arrangement for Kanamycin screens (left) and early heat, heat recovery and cold growth plates (right). The numbers in parentheses represent different individuals in the same SALK line.

Figure M3.4 Phenotype scores (blue numbers) based on plant size (A) and a typical 'seg' line (B).

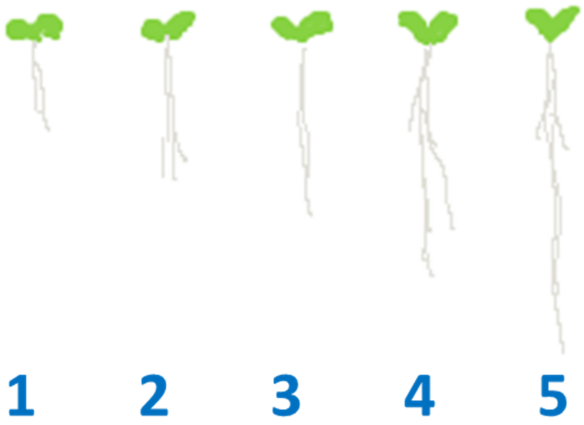

Figure M3.5 Root image and phenotype scores (blue numbers).

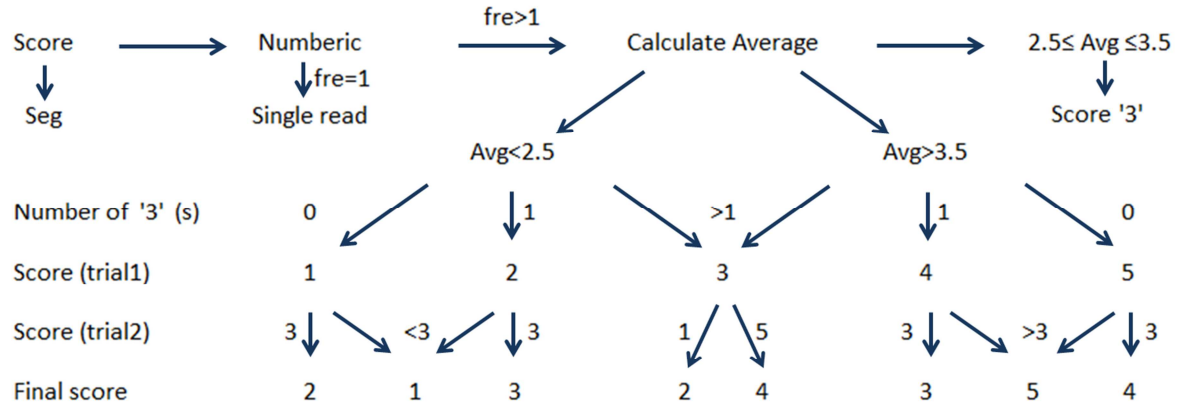

Figure M3.6 Phenotype scoring. ‘fre’ stands for ‘frequency’.

Additional figure S3.1-S3.2

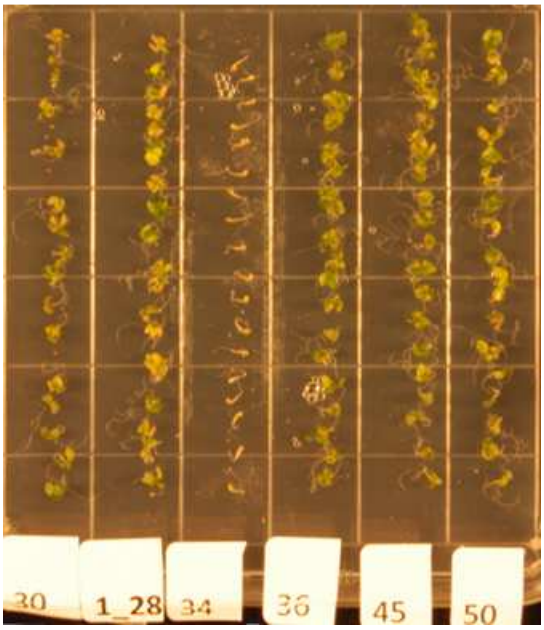

Figure S3.1. Phenotype of mutant 34 in mannitol screen.

Figure S3.2 Phenotype in cold germination screen for mutant 39. The two replicates of mutant 39 were circled.