the plant journal

SEB Society for Experimental Biology

The Plant Journal (2013) **75,** 157–171 doi: 10.1111/tpj.12197

TECHNICAL ADVANCE/RESOURCE

GABI-DUPLO: a collection of double mutants to overcome genetic redundancy in *Arabidopsis thaliana*

Cordelia Bolle[1], Gunnar Huep[2], Nils Kleinbölting[2], Georg Haberer[3], Klaus Mayer[3], Dario Leister[1,*] and Bernd Weisshaar[2]

[1]*Lehrstuhl für Molekularbiologie der Pflanzen (Botanik), Department Biologie I, Ludwig-Maximilians-Universität München, Großhaderner Str. 2, D-82152, Planegg-Martinsried, Germany,*

[2]*Genome Research, Department of Biology, Bielefeld University, 33594 Bielefeld, Germany, and*

[3]*MIPS, Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany*

Received 11 March 2013; revised 28 March 2013; accepted 4 April 2013; published online 10 April 2013.
*For correspondence (e-mail leister@lmu.de).

SUMMARY

Owing to duplication events in its progenitor, more than 90% of the genes in the *Arabidopsis thaliana* genome are members of multigene families. A set of 2108 gene families, each consisting of precisely two unlinked paralogous genes, was identified in the nuclear genome of *A. thaliana* on the basis of sequence similarity. A systematic method for the creation of double knock-out lines for such gene pairs, designated as DUPLO lines, was established and 200 lines are now publicly available. Their initial phenotypic characterisation led to the identification of seven lines with defects that emerge only in the adult stage. A further six lines display seedling lethality and 23 lines were lethal before germination. Another 14 lines are known to show phenotypes under non-standard conditions or at the molecular level. Knock-out of gene pairs with very similar coding sequences or expression profiles is more likely to produce a mutant phenotype than inactivation of gene pairs with dissimilar profiles or sequences. High coding sequence similarity and highly similar expression profiles are only weakly correlated, implying that promoter and coding regions of these gene pairs display different degrees of diversification.

Keywords: *Arabidopsis thaliana*, double mutant, gene duplication, genetic redundancy, segmental duplication, technical advance.

INTRODUCTION

Only 9.7% of the approximately 27 000 protein-coding genes (TAIR 10; http://www.arabidopsis.org) in the nuclear genome of the model plant *Arabidopsis thaliana* code for unique proteins; all others have at least one additional homologue (Armisen *et al.*, 2008). Moreover, many of the latter class form larger gene families, with about 40% being accounted for by families with more than five members and some comprising more than 100 members (AGI, 2000; Cannon *et al.*, 2004; Armisen *et al.*, 2008). Genes can be amplified by several mechanisms (for recent reviews see: Van de Peer *et al.*, 2009; Paterson *et al.*, 2010; Rutter *et al.*, 2012). Single-gene duplications such as tandem duplications generate tightly linked copies, whereas duplications resulting from more widespread restructuring processes, such as polyploidization (duplication of the whole genome) or duplications of single chromosomes (aneuploidy) or large chromosomal segments, can lead to the rearrangement of large chromosomal regions and thereby to segmental duplications. Evidence for at least three major large-scale duplication events has been detected in the genome of *A. thaliana*, all of them predating the divergence of Arabidopsis species from *Brassica* species about 14.5–20.4 million years ago (Bowers *et al.*, 2003; Jiao *et al.*, 2011). The most recent duplication event encompassed 70–89% of the genes in the genome (Simillion *et al.*, 2002; Blanc *et al.*, 2003; Bowers *et al.*, 2003).

Because duplicated genes are redundant in function immediately after their formation, they are exposed to genetic fractionation processes (AGI, 2000; He and Zhang, 2005). One gene copy may be lost with no deleterious effects. Mutations in one or both gene copies may lead either to a change in gene dosage or to functional divergence

between the two. The latter may involve the gain of a novel function (neofunctionalization) or both gene copies can retain different subsets of the functional properties of the ancestral gene (subfunctionalization) (Lynch and Conery, 2000; Rutter *et al.*, 2012). Duplicated genes can therefore possess overlapping or redundant functions – especially if the duplication event is recent – and various instances of such genetic redundancy have been reported. Among these instances are several photosynthesis-related proteins, such as PsaD, PsaE, plastocyanin and PGRL1, which are encoded by segmentally duplicated genes and for which knock-out of both gene copies uncovers phenotypes not seen in either single mutant (Ihnatowicz *et al.*, 2004, 2007; DalCorso *et al.*, 2008; Pesaresi *et al.*, 2009). Moreover, it has been demonstrated that 22 core cell-cycle genes have evolved from segmental duplications (Vandepoele *et al.*, 2002). Duplications can also lead to the evolution of extended gene families that exhibit various degrees of diversification in sequence and expression pattern (Kolukisaoglu *et al.*, 2002; Meyers *et al.*, 2003; Cannon *et al.*, 2004; Leister, 2004; Remington *et al.*, 2004; Abel *et al.*, 2005; Jiang *et al.*, 2006; Maher *et al.*, 2006; Nakano *et al.*, 2006; Kong *et al.*, 2007; Wang *et al.*, 2009; Liu *et al.*, 2010; Charon *et al.* 2012).

Several collections of sequence-indexed T-DNA insertion mutants have already been constructed for *A. thaliana* (reviewed in Bolle *et al.*, 2011). In total, over 600 000 mutant lines have been generated and flanking sequence tags (FSTs) of over 325 000 T-DNA insertion lines have been mapped to the reference genome sequence (*A. thaliana* accession Columbia 0). More recently, a project was initiated with the intention of generating two independent homozygous T-DNA insertion lines for each nuclear Arabidopsis gene (O'Malley and Ecker, 2010). Such homozygous lines are a prerequisite for the direct assessment of gene function and will enable efficient genome-wide forward genetic screens. Gene silencing by small interfering RNAs or by artificial microRNAs is also an important tool in the study of gene function and can also be applied to the silencing of gene families (Ossowski *et al.*, 2008). However, whereas double mutations caused by T-DNA insertions are stably inherited, leakiness and silencing of RNA-based gene inactivation approaches have been occasionally observed (e.g. Gupta *et al.*, 2002; Weigel *et al.*, 2003). Therefore, lines generated by RNA-based approaches and the DUPLO lines represent complementary approaches to inactivate gene pairs, whereas RNA-based approaches alone are the method of choice to inactivate simultaneously three or more gene copies, including tandem copies.

Here we report on the generation of a collection of double knock-out mutants (DMs) specifically targeted to the members of highly homologous, but genetically unlinked, gene pairs, and derived from sequence-indexed T-DNA mutant lines with a Col-0 genetic background. This collection, termed 'GABI-DUPLO', is available at the Nottingham Arabidopsis Stock Centre (NASC) and will complement the existing sequence-indexed single-mutant collections by allowing one to bypass the effects of genetic redundancies arising from segmental duplications in systematic forward and reverse genetic screens. An initial phenotypic characterisation of 200 GABI-DUPLO lines is described here; we demonstrate that the probability that a DM line will display a phenotype increases with the degree of similarity in expression pattern or protein sequence between the two genes considered.

RESULTS

Selection of gene pairs for the GABI-DUPLO collection

To unmask the functions of genes that are represented by two copies in the *A. thaliana* genome through the analyses of DM phenotypes, we selected gene pairs that were highly likely to have redundant or overlapping functions. To this end, we employed two selection criteria, namely overall sequence similarity and number of homologous gene copies. In the first step, we selected protein-coding *A. thaliana* genes that have high sequence similarity at the amino acid sequence level ($\geq$60% similar amino acids and <20% gaps in a global Needleman–Wunsch alignment). This analysis resulted in a set of 26 982 pairwise combinations of closely related genes fulfilling this criterion (Figure 1). To identify true gene pairs, combinations that had at least one additional (third) homologue in the *A. thaliana* genome at the same similarity threshold were eliminated. This action resulted in a set of 2594 gene pairs. Paired genes located $\leq$7.5 Mbp apart on the same chromosome were not considered further as their proximity would have prevented the efficient generation of DMs, so that only 2108 gene pairs were ultimately selected for analysis (Figure 1).

When these 2108 gene families, each made up of exactly two genetically unlinked members, were analysed for their known or predicted molecular function, as well as the biological process they are involved in and the subcellular localization of their products (Figures 2 and S1), the set showed no obvious enrichment for any particular category (relative to the entire *A. thaliana* genome). However, genes coding for proteins whose molecular function is unknown (14 versus 20% within the whole genome), biological context unclear (5% versus 8%) or subcellular localization undetermined (9 versus 19%) are slightly underrepresented among the 2108 gene pairs, whereas the functional categories 'transcription factor activity' and 'DNA/RNA binding' are slightly overrepresented (6/9 versus 3/7%).

The GABI-Kat and SALK collections were analysed for the presence of T-DNA insertion alleles of both members of each of the 2108 gene pairs (Alonso *et al.*, 2003; Kleinboelting *et al.*, 2012). To obtain loss-of-function alleles of the genes of interest, only alleles with a (predicted)

insertion in the region between the start and stop codon (i.e. exon or intron regions and referred to here as 'CDS + intron' or 'CDSi' insertions) were considered. Such insertions were identified for 1294 of the gene pairs (Figure 1). However, because some CDSi insertions were either:

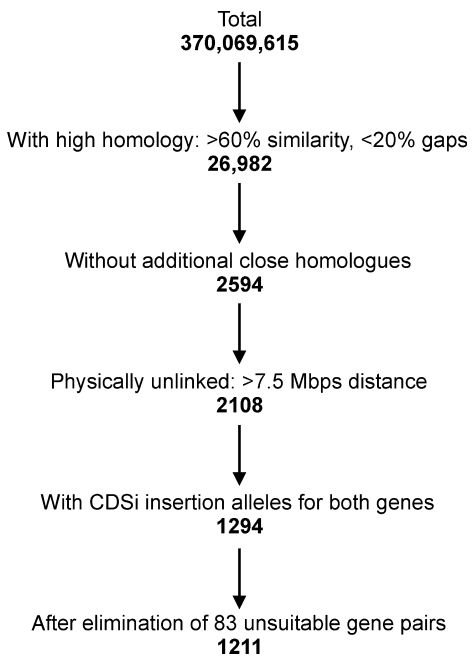

Figure 1. Flow chart that depicts the selection of lines for the DUPLO collection. The total number of all possible pairwise combinations for the 27 206 protein-coding genes in *A. thaliana* is 370 069 615. The stepwise selection of gene pairs is illustrated. Note that the number of gene pairs with CDSi alleles for both genes depends on the availability of insertion alleles and can change either when new alleles become available or when flanking sequence tag (FST)-based predictions are modified, for instance upon re-annotation of genes in the TAIR database. In the final selection step, gene pairs were excluded if the insertion was not found to be located within the CDS (after updating of TAIR to version 10), or the allele could not be confirmed, or if one or both parental lines were known to be lethal. For further details see main text and Experimental Procedures.

(i) unavailable; (ii) no longer annotated as CDSi insertions in version 10 of the TAIR database; (iii) could not be confirmed by polymerase chain reaction (PCR); or (iv) known to be lethal; we finally obtained mutant alleles for 1211 gene pairs that fulfilled all criteria for setting up a DM collection (Figure 1) (for a full list see: https://www.gabi-kat.de/double-mutant-lists/duplo-gene-pairs.html). Obviously, this list reflects the availability of insertion alleles at the time of writing and will be expanded as new insertion alleles become available.

Generation of homozygous DM lines

The generation of homozygous DMs usually requires at least 8 months. Parental single-mutant lines were first analysed by PCR and amplicon sequencing for the presence of the T-DNA insertion allele of the gene of interest. These 'confirmation amplicon' sequences were evaluated to deduce the position of the insertion, and only confirmed CDSi alleles were used for further steps in the workflow. If the putative CDSi allele could not be confirmed by sequencing, we tried to obtain a more suitable allele from the GABI-Kat or SALK collection. If this confirmation was not possible, the respective pair was removed from the workflow (see Figure 1). Lines that contained confirmed CDSi alleles were then genotyped using two different primer pairs. The 'genotyping amplicon' targets the wild-type allele, spans the entire insertion site in the mutant line and is generated only if the inserted T-DNA is not present. The 'confirmation amplicon' (see above) targets the insertion allele and is identical to the amplicon sequenced during confirmation (see Experimental Procedures and Figure S2).

In the next step, the parental lines that were homozygous or at least heterozygous for the insertion in each member of a given gene pair were crossed. F1 plants were allowed to self-pollinate if the presence of the two expected insertion alleles could be confirmed by PCR (Figure 3). About 54

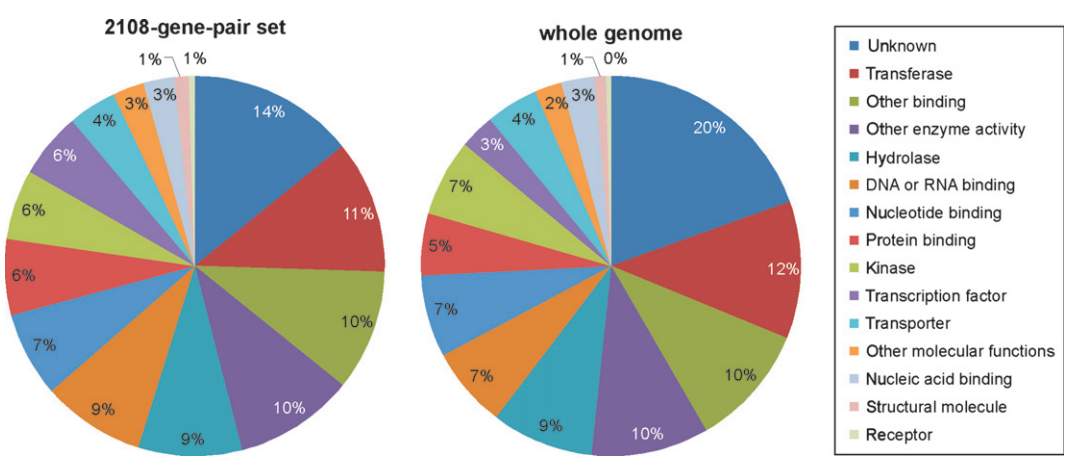

Figure 2. Distribution of the proteins encoded by the 2108 gene pairs according to their predicted molecular function.
The molecular functions of the proteins encoded by the 2108 gene pairs (left panel) and by the entire *A. thaliana* genome as control (right panel) were determined and displayed in a pie chart using the Gene Ontology (GO) annotation tool on the TAIR website (http://www.arabidopsis.org/tools/bulk/go/index.jsp).

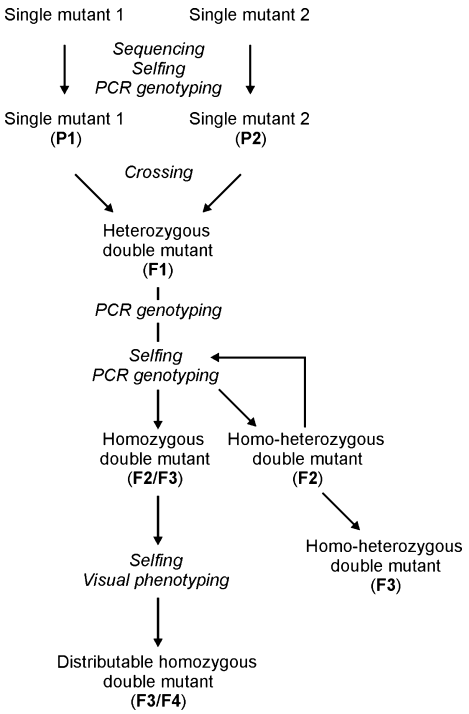

Figure 3. Workflow used for the systematic generation of DMs.
Parental lines (P1 and P2) were crossed to generate an F1 generation. After selfing, the F2 generation was genotyped to isolate double knock-out mutants (DMs). If no homozygous DM could be obtained, this step was repeated in the F3 generation. Seeds from homozygous DM lines were bulked in the next generation for distribution. Phenotyping was done in the F3 or F4 generation.

individual F2 plants were then propagated on soil and genotyped by PCR to identify plants that were homozygous for both insertions. Genotyping of the DMs with respect to the insertions in the two duplicated genes was performed by four PCRs, amplifying 'confirmation' and 'genotyping amplicons' for each of the two mutant alleles (Figure 3).

If no DUPLO line could be identified in the F2 generation, the offspring of an F2 line that was homozygous for one insertion and heterozygous for the other was allowed to self. The segregating F3 generation was propagated on plates of sucrose-containing media to allow heterotrophic growth, and genotyped by PCR (Figure 3). Propagation on sucrose-containing medium not only permitted the rescue of some mutants with defects that interfere with photoautotrophy or normal seedling development but also enabled us to identify non-germinating seeds, which are characteristic for embryo lethality. If segregation of the F3 seeds also failed to yield any homozygous DM, these pairs were classified as 'homozygous DM absent' and are likely to be lethal as DMs. In the F3 or F4 generation the genotype of homozygous DMs was rechecked and plants were evaluated phenotypically in the greenhouse.

So far, we have generated 200 DMs that were either homozygous for both mutations, or homozygous for one mutation and heterozygous for the other, and that are designated in the following as 'DUPLO lines' (Table S1). According to the Plant Genome Duplication Database (http://chibba.agtec.uga.edu/duplication/index/home) 145 (or 72%) of the 200 gene pairs are located within large segmental duplications (Tang *et al.*, 2008a,b).

DUPLO lines that display mutant phenotypes at the adult stage

During phenotypic evaluation at the F3 or F4 stage we looked for changes in growth and development at different time points. The first evaluation for size and pigmentation took place 14 days after germination (dag). Leaf coloration, rosette size, leaf form and early bolting were evaluated at 30 dag. Once the plants had flowered, their branching pattern, overall height and fertility were monitored. Seven DUPLO lines that deviated from WT plants with respect to chlorophyll content, size or developmental behaviour were identified (Table 1), and this class of mutants was referred to as DMs with an 'adult phenotype', because these plants survived to the adult stage and into the reproductive phase. All of these mutants displayed reduced or dwarfish growth, and three mutants had a reduced photosynthetic efficiency. These lines are described in more detail in the following. Unless otherwise noted, the parental single mutants behaved like WT.

The *DUPLO-195* line is a cross of the loss-of-function line of *TON1 recruiting motif 3* (*trm3*, *at1g18620*) and *trm4* (*at1g74160*). Some members of the TRM protein family have been shown to interact with TON1 and are involved in microtubule organisation at the cortex (Drevensek *et al.*, 2012). The DM and *trm4* show late flowering and a reduced height of the inflorescence, whereas *trm3* did not show any visually discernible phenotype (Figure 4a). In addition, DM leaves are smaller and occasionally rolled and display a reduced effective quantum efficiency of photosystem II (Φ_{II}) (Table 2 and Figure 4a).

The *DUPLO-932* line is dwarfed as it forms very small rosettes (about 30% of the WT diameter), and flowering is much delayed under long-day greenhouse conditions (Table 2 and Figure 4b). The molecular function of the two genes mutated (*At2g41770* and *At3g57420*), which are expressed ubiquitously in the plant, is not yet known, and the parental lines show no deviation from WT.

In the *DUPLO-998* line (*at5g18180 at3g03920*), two genes that code for the GAR1 subunit of the H/ACA ribonucleoprotein complex, which has been shown in mammalian systems to have a pseudouridine synthase activity and is important for rRNA folding and ribosome function (Watkins and Bohnsack, 2012), are defective. Although these two genes appear to be the only GAR1 orthologues in Arabidopsis, the DM plants are viable but display reduced growth and later flowering (Table 2 and Figure 4c).

Table 1 List of DUPLO lines displaying a mutant phenotype at the adult or seedling stage. The Pearson correlation coefficient (PCC) from an all-against all comparison of 170 microarray experiments based on the Affymetrix ATH1 platform is provided in the last column

Pair ID	Gene pair	Protein description	Phenotype	PCC
Adult phenotype				
195	*At1g18620; At1g74160*	TRM3; TRM4	Smaller plant size, later flowering	0.71
932	*At2g41770; At3g57420*	Expressed proteins	Dwarfed, later flowering	0.88
998	*At5g18180; At3g03920*	H/ACA ribonucleoprotein complex, subunit Gar1/Naf1 proteins	Smaller plant size, later flowering	0.61
2261	*At1g04300; At5g43560*	TRAF-like superfamily proteins	Bushy plant, shorter petioles, stunted	0.92
2268	*At3g11910; At5g06600*	UBP13; UBP12	Extremely small, dies often on soil; Ewan *et al.*, 2011 (conditional phenotype)	0.91
2589	*At1g66180; At5g37540*	Eukaryotic aspartyl protease family proteins	Smaller plant size	0.68
2647	*At1g51340; At3g08040*	MATE efflux family protein; FRD3, MAN1	Smaller plant size, less chlorophyll	0.61
Seedling lethal				
211	*At1g76090; At1g20330*	Sterol methyltransferase 3; CVP1	Seedling extremely small, not viable; Carland *et al.*, 2010 (smaller and bushy, short siliques, poor seed yield)	0.78
713	*At1g76490; At2g17370*	HMGR1; HMGR2	Seedling extremely small, not viable; Suzuki *et al.*, 2009 (male gametophytes lethal)	0.68
1569	*At5g47420; At4g17420*	Tryptophan RNA-binding attenuator protein-like proteins	Seedling extremely small, dies early	0.89
2098	*At4g24330; At5g49945*	Proteins of unknown function (DUF1682)	Seedling extremely small, not viable	0.91
2490	*At1g23310; At1g70580*	GGT1; GGT2	Seedling extremely small, dies early	0.76
2543	*At1g74910; At2g04650*	ADP-glucose pyrophosphorylase family proteins	Seedling extremely small, not viable	0.72

The two genes (*At1g04300* and *At5g43560*) defective in the *DUPLO-2261* line code for tumor necrosis factor (TNF) receptor-associated factor (TRAF)-like proteins. Mammalian TRAF proteins have important signalling functions, especially as interacting partners of the TNF (Lee and Choi, 2007). While *DUPLO-2261* seedlings are only slightly smaller than WT seedlings, adult DMs are strongly stunted and bushier, although their siliques are WT-like in size and the plants are fertile (Table 2 and Figure 4d). Additionally, the efficiency of photosynthesis is reduced. Both genes seem to be expressed in all parts of the plant (Arabidopsis eFP Browser; Winter *et al.*, 2007), with *At1g04300* being expressed at higher levels in the later stages of seed development.

The *DUPLO-2268* line is defective for the two ubiquitin-specific proteases UBP12 (At5g06600) and 13 (At3g11910), which show very high similarity in both sequence (95.2% at protein level) and expression pattern (Pearson correlation coefficient (PCC) = 0.91). The DM plants are extremely small compared with the parental single mutants and produce few seeds, if any (Table 2 and Figure 4e). Plants have to be pre-grown on medium to allow survival on soil. Also the parental line *ubp12* is not able to grow on soil, whereas *ubp13* is only slightly smaller than WT. *DUPLO-2268*/*ubp12 ubp13* resembles phenotypically the triple mutant *ubp15 ubp16 ubp17* (Liu *et al.*, 2008), which suggests additional redundancy within this gene family. Ewan *et al.* (2011) described *ubp12 ubp13* plants as being more susceptible to the virulent *Pseudomonas syringae* pv. tomato than WT, but displaying a WT-like growth behaviour. The loss-of-function allele for *UBP13* used by Ewan *et al.* (2011) differed from the one used in our study, a finding that might explain the discrepancies in the phenotypes observed.

The genes inactivated in *DUPLO-2589* encode eukaryote-specific aspartyl proteases (*At1g66180* and *At5g37540*). Genes for more than 50 A1-type aspartic proteases have been identified in Arabidopsis but the functions of their products are essentially unknown (Beers *et al.*, 2004). Some functions may be important for disease resistance and programmed cell death. The phenotype of the DM was not very pronounced, but the stature of the plants was slightly smaller than WT (Table 2 and Figure 4f).

The *DUPLO-2647* line corresponds to the DM *at1g51340 at3g08040*. *At3g08040* codes for the MATE efflux family protein MAN1 or FERRIC REDUCTASE DEFECTIVE3 (FRD3). FRD3 has previously been shown to control iron localization and to act as an efflux transporter of the efficient iron chelator citrate, and the *frd3-1* mutant has been characterised as

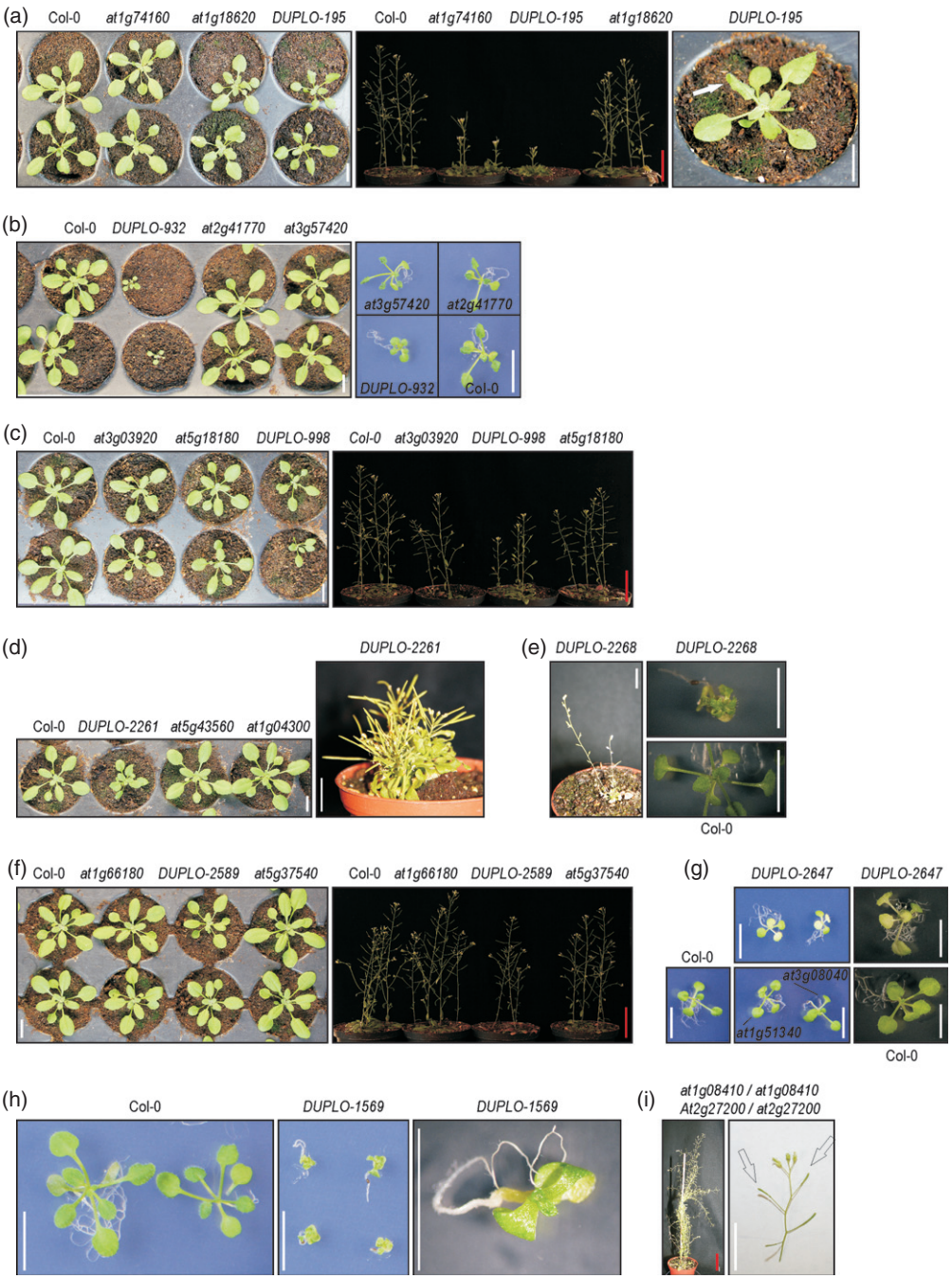

Figure 4. Phenotypes of selected DUPLO lines.
(a) *DUPLO-195* has a smaller rosette and shorter inflorescence compared with wild-type (WT) (Col-0). The phenotype is more severe than in the parental line *at1g74160*. Additionally, rolled leaves can be observed in the double knock-out mutants (DM) (see arrow).
(b) For *DUPLO-932* a smaller and more compact rosette can be observed.
(c) *DUPLO-998* is slightly smaller compared with the WT and the parental single-mutant lines.
(d) *DUPLO-2261* is only slightly smaller compared with WT at 30 dag (left panel), but the adult plant displays a stunted phenotype with fertile siliques (right panel).
(e) *DUPLO-2268* is dwarfed but capable of flowering and setting seeds if pre-grown on medium.
(f) *DUPLO-2589* is slightly smaller compared with WT and the parental single-mutant lines.
(g) A chlorotic phenotype can be observed for *DUPLO-2647*, which impairs autotrophic growth.
(h) *DUPLO-1569* displays strong growth retardation, which leads to seedling lethality.
(i) *DUPLO-810* plants, which are heterozygous for the *at2g27200* mutant allele, have many floral stems, but are infertile because the siliques do not develop (arrows).
Phenotypes were imaged on soil at the rosette stage (30 dag), at flowering (39 dag) or on media plates after 14 days. White bar = 1 cm, red bar = 5 cm.

Table 2 Physiological characterisation of mutant lines with phenotypes at the adult stage. At least six plants of DM lines, parental lines and WT were grown side-by-side

	Φ_{II}	Length of longest leaf (in % of WT)	Flowering time (dag)	Length of inflorescence (in % of WT at 39 dag)	Length of inflorescence (% of mature WT plants)
WT (Col-0)	0.72 ± 0.02	100.00 ± 19.74	28.0 ± 1.4	100.00 ± 10.1	100.0 ± 6.5
DUPLO-195	0.68 ± 0.03*	68.87 ± 8.96*	>31*	17.5 ± 12.5*	47.3 ± 17.6*
at1g18620	0.70 ± 0.02	97.49 ± 5.93	27.1 ± 0.5	103.1 ± 15.6	101.2 ± 7.5
at1g74160	0.72 ± 0.04	80.75 ± 6.08	>31*	17.1 ± 15.2*	49.9 ± 14.2*
DUPLO-932	0.71 ± 0.01	29.90 ± 10.4*	>31*	13.7 ± 4.6*	42.4 ± 5.9*
at3g57420	0.72 ± 0.02	99.04 ± 14.75	27.7 ± 2.2	121.4 ± 12.5	99.3 ± 4.8
at2g41770	0.74 ± 0.02	92.75 ± 11.69	27.2 ± 1.1	99.3 ± 15.8	91.3 ± 8.1
DUPLO-998	0.73 ± 0.01	73.94 ± 12.18*	>31*	85.6 ± 9.9*	98.2 ± 9.3
at5g18180	0.74 ± 0.02	83.77 ± 2.66	28.1 ± 0.4	104.1 ± 4.9	98.7 ± 6.3
at3g03920	0.70 ± 0.03	77.38 ± 23.11	26.6 ± 1.2	104.1 ± 14.8	86.0 ± 11.2
DUPLO-2261	0.69 ± 0.01*	75.29 ± 17.06*	27.2 ± 0.3	24.0 ± 9.4*	30.7 ± 4.0*
at5g43560	0.72 ± 0.01	97.79 ± 7.62	25.9 ± 1.3*	106.5 ± 18.1	89.9 ± 6.1
at1g04300	0.71 ± 0.03	100.32 ± 4.95	25.4 ± 1.4*	127.1 ± 16.8	93.1 ± 12.5
DUPLO-2268	nd	nd	nd	nd	nd
at5g06600	nd	nd	nd	nd	nd
at3g11910	0.72 ± 0.02	71.43 ± 17.15*	28.8 ± 2.6	77.3 ± 11.9*	84.4 ± 9.9*
DUPLO-2589	0.71 ± 0.01	100.53 ± 4.96	26.8 ± 1.3	99.0 ± 10.2	85.7 ± 12.1*
at5g37540	0.72 ± 0.03	105.92 ± 6.90	26.3 ± 2.1	95.1 ± 13.8	92.2 ± 16.2
at1g66180	0.74 ± 0.02	83.88 ± 7.63	>31*	80.7 ± 10.4*	94.4 ± 14.7
DUPLO-2647	nd	nd	nd	nd	nd
at1g51340	0.73 ± 0.01	98.03 ± 6.14	>31*	102.6 ± 12.6	100.2 ± 6.1
at3g08040	0.64 ± 0.06*	31.02 ± 7.79*	>31*	5.3 ± 10.0*	27.2 ± 18.1*

nd, not determined (because homozygous plants were not able to germinate directly on soil).
The length of the longest leaf (blade plus petiole) and Φ_{II} (as a marker for photosynthetic activity) were determined at 30 dag. Flowering time was measured in dag upon appearance of the bolt. The length of the inflorescence was measured after 39 dag and upon maturation.
Statistical significant differences from WT ($P < 0.05$) are indicated (*).

chlorotic and partially sterile (Green and Rogers, 2004; Roschzttardtz *et al.*, 2011). Our *frd3* mutant was also much smaller than WT, but showed normal fertility, although also the photosynthetic efficiency was reduced (Table 2). The T-DNA insertion in the DUPLO allele was close to the 3′-end of the *FRD3* gene, whereas it was more in the middle of the gene in the *frd3-1* allele. Both parental lines were flowering later than WT. *DUPLO-2647* was small and yellowish when grown on medium (Figure 4g), but some plants survived and set seeds after being transferred onto soil. This finding suggests that loss of the second homologous gene (*At1g51340*) has a additive effect on the mutant phenotype caused by the loss of *At3g08040*.

DUPLO lines that display lethality at the seedling stage

The second class of phenotypes observed in DUPLO lines is characterised by their failure to reach the reproductive phase, either because the seedling dies soon after germination or it is extremely small and stops growing prior to the adult phase ('seedling lethal' phenotype in Table 1). Six lines fall into this category and are described in the following paragraphs.

DUPLO-211 combines mutations in the *CVP1 (At1g20330)* and *SMT3* (*At1g76090*) genes, which code for sterol 24-carbon methyltransferases that catalyse the synthesis of structural sterols. The *cvp1 smt3* DM has been described before as being smaller and bushy, with short siliques and poor seed yield (Carland *et al.*, 2010). The more dramatic phenotype observed here might be associated with the fact that we used different mutant alleles.

In *DUPLO-713*, two genes (*HMGR1/At1g76490* and *HMGR2/At2g17370*) for 3-hydroxy-3-methylglutaryl CoA reductase (HMGR), which catalyses the first committed step in the mevalonate (MVA) pathway for isoprenoid biosynthesis in the cytoplasm, are inactivated. The *hmgr1* mutant was smaller than WT, but not as drastically dwarfed as reported previously (Suzuki *et al.*, 2004). DM plants germinated and grew into very small and inviable seedlings, whereas the DM described by Suzuki *et al.* (2009) showed male gametophyte lethality.

In the *DUPLO-1569* line, a gene pair coding for two tryptophan RNA-binding attenuator protein-like proteins (At5g47420 and At4g17420), predicted to be localized in the chloroplast and containing at least three transmembrane domains, are inactivated. DM plants are very small and die early (Figure 4h). Inactivation of *At4g17420* alone in the heterozygous *at5g47420* background already causes a drastic reduction in plant size, and these very small plants do not flower.

In the *DUPLO-2098* line, genes for two proteins of unknown function (At4g24330 and At5g49945) are disrupted. This line produces extremely small and inviable seedlings. The WT proteins are predicted to possess two transmembrane domains and to participate in the secretory pathway.

DUPLO-2490 (*at1g23310 at1g70580*) seedlings also die early when grown heterotrophically. At1g23310 and At1g70580 are both predicted to be localized in the peroxisome, and a glyoxylate aminotransferase activity has been assigned to each seedling.

In *DUPLO-2543*, genes for two ADP-glucose pyrophosphorylase family proteins (At1g74910 and At2g04650) are disrupted. DM seedlings are extremely tiny and die after a few weeks even when grown on sucrose-containing media.

DUPLO lines that display lethality prior to germination

The third and largest group of lines are those that are unable to generate doubly mutant homozygotes. Even resegregation of at least 100 seeds on sucrose-containing medium led only to lines that remained heterozygous for one insertion and homozygous for the other. One must therefore assume that the DM dies at a very early stage, and this 'absence of homozygotes' suggests that the DM is either gametophytic or embryo lethal. For some of these DM lines (*DUPLO-151*, *-238*, *-1241*, *-2212* and *-2380*) lethality had been reported in the literature but using different alleles (Table 3). When the parental single mutants of DM lines that were incapable of producing homozygous DM embryos were re-investigated, it was found that in four cases (*DUPLO-110*: *at2g42910*; *DUPLO-2394*: *at2g24500*; *DUPLO-2666*: *at5g48950*; *DUPLO-2753*: *at4g08430*) already parental single mutants might be lethal.

About half of these gene pairs with essential functions seems to be important for metabolic processes. The proteins they encode are predicted to be phosphoribosyltransferases, UDP-galactose transporters, glycosyl hydrolases, nucleoside triphosphate hydrolases, sphingolipid

Table 3 List of DUPLO lines that are lethal before germination

Pair ID	Gene pair	Protein description	Phenotype	PCC
110	*At2g42910; At1g10700*	Phosphoribosyltransferase family protein; PRS3	Parent *At2g42910* may be lethal	0.69
151	*At2g02810; At1g14360*	UTR1; UTR3	Reyes *et al.*, 2010 (gametophytic lethal)	0.89
191	*At5g15870; At1g18310*	Glycosyl hydrolase family 81 proteins	–	0.37
213	*At1g76300; At1g20580*	SMD3; snRNP family protein	–	0.82
222	*At1g76850; At1g21170*	Exocyst complex components SEC5	–	0.74
238	*At4g10040; At1g22840*	Cytochrome C-2; Cytochrome C-1	Welchen *et al.*, 2012 (arrest of embryo development)	0.74
601	*At1g21190; At1g76860*	snRNP family proteins	–	0.84
810	*At1g08410; At2g27200*	P-loop containing nucleoside triphosphate hydrolases superfamily proteins	–	0.70
1022	*At1g16180; At3g06170*	Serine-domain containing serine and sphingolipid biosynthesis proteins	–	0.85
1241	*At5g40650; At3g27380*	Succinate dehydrogenase 2-2; Succinate dehydrogenase 2-1	León *et al.*, 2007 (essential for gametophyte development)	0.92
1361	*At2g38700; At3g54250*	MVD1; GHMP kinase family protein	–	0.86
1545	*At3g22440; At4g14900*	FRIGIDA-like proteins	–	0.87
1641	*At5g55120; At4g26850*	VTC5; VTC2	Dowdle *et al.*, 2007 (seedling lethal)	0.66
2212	*At5g09740; At5g64610*	HAM2; HAM1	Latrasse *et al.*, 2008 (lethal)	0.86
2253	*At1g53850; At3g14290*	PAE1; PAE2	–	0.95
2257	*At5g66140; At3g51260*	PAD2; PAD1	–	0.94
2341	*At1g17890; At1g73250*	GER2; GER1	–	0.85
2349	*At1g74040; At1g18500*	IPMS2; IPMS1	–	0.85
2361	*At5g60550; At3g45240*	GRIK2; GRIK1	–	0.84
2380	*At1g65660; At4g37120*	SMP1; SMP2	Clay and Nelson, 2005 (lethal)	0.83
2394	*At2g24500; At4g31420*	FZF; Zinc-finger protein 622	Parent *at2g24500* may be lethal	0.82
2666	*At1g48320; At5g48950*	Thioesterase superfamily proteins	Parent *at5g48950* may be lethal	0.59
2753	*At5g45570; At4g08430*	Ulp1 protease family proteins	Parent *at4g08430* may be lethal	0.49

PCC, Pearson correlation coefficient.

For all these lines, no homozygous double knock-out mutants (DM) plant could be obtained. Additional information, including the used T-DNA insertion alleles are provided in Table S1.

biosynthesis proteins, succinate dehydrogenases, mevalonate diphosphate decarboxylases, guanylyltransferases, histone acetyltransferases, NAD(P)-binding Rossmann-fold superfamily proteins, 2-isopropylmalate synthases and thioesterases (Table 3). Other gene products are constituents of multiprotein complexes such as cytochrome c, snRNP core proteins, exocyst complex components, small nuclear ribonucleoprotein family proteins, subunits of proteasomes and pre-mRNA splicing interacting factor. Only a few proteins, such as FRIGIDA-like proteins, kinases, Ulp1 protease family proteins and zinc-finger proteins, seem to be involved in signal transduction.

In *DUPLO-810*, genes *At2g27200* and *At1g08410* for two members of the P-loop-containing nucleoside triphosphate hydrolase superfamily are disrupted. Each of the parental single mutants was partially sterile (Figure 4i).

DUPLO lines that show phenotypes under non-standard conditions or at the molecular level

DUPLO lines that do not exhibit a visually discernible phenotype under greenhouse conditions might still display a phenotype when exposed to special conditions (such as stress or special lighting conditions) or at the molecular level (for instance altered transcript or metabolite profiles). The prevalence of such lines in the DUPLO collection was assessed on the basis of literature searches. These lines revealed that, for 14 DMs with counterparts in the DUPLO collection, phenotypes had been identified which would not have been detectable in our initial phenotypic screen of greenhouse-grown plants (Table 4). This finding indicates that the detailed analysis of DMs can lead to the discovery of additional phenotypes. For example, the DM *pif4 pif5* (*DUPLO-2409*) shows a hypersensitive phenotype at lower fluence rates of far-red light (Lorrain *et al.*, 2009), whereas the DM *far1 fhy3* (*DUPLO-1196*) has an elongated hypocotyl under far-red light, even longer than those hypocotyls seen in each of the single-mutant parents (Wang and Deng, 2002). The counterpart of the *DUPLO-1202* line showed no visible phenotype under our growth conditions but, in a more detailed analysis, Rossini *et al.* (2006) have demonstrated that the DM shows a slight reduction in the chlorophyll content in mature leaves and a less

Table 4 List of DUPLO lines displaying a phenotype under non-standard conditions or at the molecular level

Pair ID	*Gene pair*	Protein description	Phenotype	References	PCC
148	*At2g02950; At1g14280*	PKS1; PKS2	Involved in phyA-mediated VLFR	Lariguet *et al.* (2003)	0.74
380	*At3g19280; At1g49710*	FucTA; FucTB	Variation of the complex-type *N*-glycans	Strasser *et al.* (2004)	0.68
1196	*At4g15090; At3g22170*	FAR1; FHY3	Elongated hypocotyl under far-red light	Wang and Deng (2002)	0.85
1202	*At4g14690; At3g22840*	ELIP2; ELIP1	Reduced chlorophyll content, less accumulation of zeaxanthin in high light,; altered germination under abiotic stress	Rossini *et al.* (2006), Rizza *et al.* (2011)	0.79
1210	*At2g48110; At3g23590*	REF4; RFR1	Enhanced expression of phenylpropanoid biosynthetic genes, increased accumulation of downstream products	Bonawitz *et al.* (2012)	0.75
2376	*At1g51590; At3g21160*	MNS1; MNS2	Aberrant Man(8)GlcNAc(2) accumulation	Liebminger *et al.* (2009)	0.83
2409	*At2g43010; At3g59060*	PIF4; PIL6	Hypersensitive to lower fluence rates of far-red light	Lorrain *et al.* (2009)	0.81
2412	*At4g31920; At2g25180*	ARR10; ARR12	Inhibition of root elongation, green callus formation from root explants	Yokoyama *et al.* (2007)	0.81
2429	*At5g45110; At4g19660*	NPR3; NPR4	Elevated PR-1 expression, enhanced resistance to virulent bacterial and oomycete pathogens	Zhang *et al.* (2006)	0.80
2440	*At1g50890; At4g27060*	SP2L; CN, SPR2, TOR1	Enhanced right-handed helical growth	Yao *et al.* (2008)	0.79
2474	*At5g47230; At4g17490*	ERF5; ERF6	Increased susceptibility to *B. cinerea*, JA-induced gene expression reduced	Moffat *et al.* (2012)	0.77
2488	*At2g31650; At1g05830*	ATX1; ATX2	Involved in *FLC* activation and histone methylation	Yun *et al.* (2012)	0.76
2653	*At1g27320; At2g01830*	AHK3; CRE1	Reduced responsiveness to cytokinin	Franco-Zorrilla *et al.* (2005), Riefler *et al.* (2006)	0.60
2668	*At1g77130; At3g18660*	GUX3; GUX1	Shorter GlcA and MeGlcA side chains on xylan in the cell wall	Lee *et al.* (2012)	0.59

PCC, Pearson correlation coefficient.

accumulation of zeaxanthin in high light conditions. For the DM corresponding to *DUPLO-380* structural variations in complex-type *N*-glycans have been detected using MALDI-ToF mass spectrometry (Strasser *et al.*, 2004).

The relationship between mutant phenotype frequency and gene pair characteristics

Taking all phenotypes together, we were able to attribute some phenotypic and/or developmental deviation from the WT to 50 of the 200 DUPLO lines investigated. For all DUPLO gene pairs, a multidimensional Arabidopsis mRNA expression dataset comprising 1765 expression arrays was used to rank these pairs according to the degree of similarity between their expression profiles by global correlation. The expression dataset covers a large range of biological processes, including hormone and light treatments, mutant analysis, biotic and abiotic stresses, various tissues and developmental stages (Haberer *et al.*, 2006).

To investigate whether a correlation exists between similarity of expression pattern and the frequency of DM mutant phenotypes observed, the frequency of DM phenotypes was plotted against the PCC values for the corresponding gene pairs (Figure 5). This comparison clearly showed that gene pairs with very similar expression profiles, i.e. a high PCC value, are more likely to result in DM phenotypes than gene pairs with lower PCC values. Similarly, gene pairs with high sequence similarity also displayed a trend towards a high mutant phenotype frequency (Figure 5). When we correlated expression similarity and coding sequence similarity of the gene pairs, we obtained PCC values of 0.40 (for the 50 DUPLO lines with phenotypes) and 0.24 (for all 200 DUPLO lines), indicating that high expression similarity is not necessarily associated with high coding sequence similarity and vice versa. In consequence, our data suggest that duplicated genes that display high coding sequence similarity or expression profiles more frequently result in DM phenotypes.

Public availability of DUPLO lines

The DUPLO lines are available through NASC, together with the corresponding parental single-mutant alleles. For DUPLO lines that are not viable or do not produce sufficient seed (see above and Table S1), the segregating F2 or F3 population will be distributed. Information on the parental alleles, primers used for genotyping and the initial phenotypic evaluation will be available at http://www.gabi-kat.de/duplo.html (see Figure S3).

DISCUSSION

Gene knock-outs are an important tool in the study of gene functions in the model plant *A. thaliana* (reviewed in Bolle *et al.*, 2011), but when additional gene copies with overlapping or even redundant functions exist their utility is limited. About one-sixth of the genes in the Arabidopsis genome are present in multiple copies which may have overlapping or even fully redundant functions. In such cases, single-gene knock-outs are likely to be uninformative. Because we selected pairs of clearly paralogous genes that have no other close relatives in the genome, the function(s) they mediate should be fully inactivated in the DUPLO lines, as interference due to functional overlap with a third gene is quite unlikely. Practical applications of DUPLO lines include the phenotypic characterisation of individual DMs in reverse genetics approaches, but also extend to systematic forward genetics screens such as those already established for conventional (single) mutant collections. Because most of the unwanted secondary mutations have been segregated out in the F2 and F3 generations, phenotypic effects unrelated to the inactivated gene pairs should be rare.

Comparison with published data

When the phenotypes observed in this study were compared with published results, some differences were

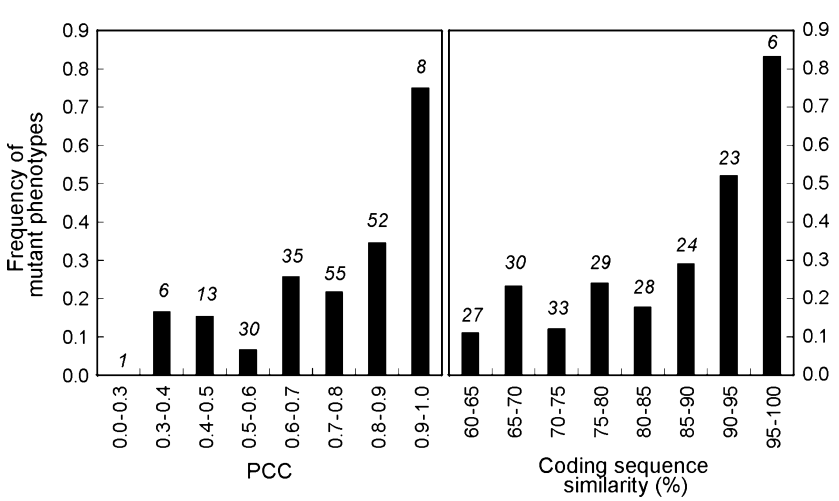

Figure 5. Effects of coding sequence or expression pattern similarity of gene pairs on double knock-out mutants (DM) phenotype frequency.
In the left (right) panel, the frequency of DM phenotypes among the 200 DUPLO lines is plotted against the Pearson correlation coefficient (PCC) value (sequence similarity) for the members of each gene pair. The number of lines assigned to the particular intervals of PCC values and coding sequence similarities is given above each column.

observed. In several cases we observed less severe mutant phenotypes than those described previously. Thus, the lines *DUPLO-2332* and *-2572* and their parental single mutants were viable, despite the fact that embryo lethality had previously been reported for other alleles of one parental single mutant of each gene pair (*DUPLO-2332*: SCO3, Albrecht *et al.*, 2010; *DUPLO-2572:* BUB3.1, Lermontova *et al.*, 2008). Similarly, for one of the two genes mutated in *DUPLO-2693* (*At1g76620*), which codes for a protein of unknown function, a mutant phenotype with altered seed pigmentation was observed (Bryant *et al.*, 2011; *pde339*). In our hands, *DUPLO-2693* was WT-like.

Conversely, in several cases we observed more severe mutant phenotypes than reported before. Thus the phenotype of *DUPLO-2268* (very small und barely viable on soil) was more severe than the one described by Ewan *et al.* (2011), who found no obvious growth defect. This difference might be related to the fact that different alleles were used for *UBP12* and *UBP13*, although the DUPLO insertions were closer to the 3′ end of the coding regions than the insertions in the allele used by Ewan *et al.* (2011). Also for *DUPLO-211* (*cvp1 smt3*) we observed a more pronounced phenotype: here seedlings were lethal, whereas the DM described by Carland *et al.* (2010) (which carried a different *smt3* and *cvp1* allele; the *cvp1* allele was derived from EMS mutation screen in contrast to the DUPLO lines which carries a T-DNA insertion close to the 3′-end of the gene) was characterised by discontinuities in the vein pattern of the cotyledon, defective root growth, loss of apical dominance, sterility, and homeotic floral transformations. The *vtc2 vtc5* line, analysed by Dowdle *et al.* (2007), is disrupted at a step in the ascorbate biosynthetic pathway and shows growth arrest immediately upon germination and the cotyledons subsequently bleach. Its counterpart, *DUPLO-1641*, generated from two different alleles, was lethal before germination in our hands, and even supplementation with L-galactose failed to rescue seedlings. In this case the insertion in our allele used for *vtc5* was closer to the ATG codon compared with the ones used by Dowdle *et al.* (2007), although localized within an intron. The allele used for *vtc2* by Dowdle *et al.* (2007) had a single base exchange at the predicted 3′ splice site of intron 5, which produces a truncated mRNA at a reduced level, whereas the DUPLO allele contains a T-DNA insertion between the sixth and seventh exon. The alleles used to generate *DUPLO-713* also differ from those examined by Suzuki *et al.* (2009): for *hmg1* both insertions are within the first exon, but for *hmg2* the allele used by Suzuki *et al.* (2009), was in the third exon, whereas the DUPLO allele had an insertion in the first exon). This finding could explain the difference between the phenotypes found – gametophytic lethality in the earlier report and seedling lethality in this study.

In summary, we found eight instances in which the DM phenotypes observed in this study differed from those described before. The most plausible explanation for such differences is that they result from the use of different mutant alleles as insertions within the intron or close to the 3′-part of the coding region might not lead to complete knock-out lines. This situation in turn implies that, for DMs also, combinations of at least two different single-mutant alleles have to be analysed before one can unambiguously assign functions to gene pairs. In cases of discrepancies like those reported above, a third independent DM line needs to be analysed, together with all available single and DMs, under the same growth conditions.

Frequency of mutant phenotypes

Our initial phenotyping analysis showed that 13 (or 6.5%) of the 200 DUPLO lines display a visually discernible phenotype (adult phenotype and seedling lethal) under greenhouse conditions. This is higher than the frequency of visible phenotypes observed among 4000 lines with transposon insertions in their coding regions; there 140 lines (or 3.5%) with visually discernible phenotypes were found (Kuromori *et al.*, 2006). Although the number of DUPLO lines is still too small to allow robust statistics, this difference (by a factor of almost two) might indicate that the DUPLO collection is enriched for lines with phenotypic variations (adult phenotypes and seedling lethal) when compared with single-gene mutant collections. Moreover, we observed 23 lines that were lethal before germination and 14 lines for which phenotypes that become manifest under non-standard conditions or at the molecular level had been reported in the literature. Thus, 50 DUPLO lines (or 25%) displayed a mutant phenotype. Because the distribution of gene functions in the set of 2108 gene pairs (from which the 200 DUPLO lines were selected) was similar to that for the whole genome (see Figure 2), we speculate that these 200 gene pairs were enriched for housekeeping functions which, when mutated, result in obvious phenotypes visible in our initial screen (see below). Within the set of 50 lines with phenotypic deviations from WT, lines that contained mutated gene pairs with either highly similar expression pattern or sequence were clearly overrepresented (see Figure 5). Because we found no strong correlation between high coding sequence similarity and highly similar expression profiles, it appears that promoter and coding regions of gene pairs can display different degrees of diversification. Assuming that the 50 gene pairs with clear mutant phenotypes are enriched for housekeeping functions and possess either similar promoter regions (resulting in similar expression profiles) or similar coding sequences, we speculate that housekeeping genes are under a special type of post-duplication selection. This type of selection conserves promoter or coding regions, or sometimes both (see PCC of 0.40 for sequence similarity and expression similarity in case of the 50 gene pairs resulting in phenotypes), and might serve to stabilise

theexpression level of gene products with housekeeping functions.

EXPERIMENTAL PROCEDURES

Selection of candidate genes

To identify paralogous genes coding for very similar proteins, we conducted an all-against-all BlastP comparison of all *A. thaliana* genes (TAIR version 6) and selected for gene pairs with an e-value $\leq 10^{-10}$. Preselected pairs were subjected to an optimal global alignment and paralogues with <20% gaps and a minimal alignment similarity $\geq$60% were retained. These relations were projected onto a similarity graph, on which genes were represented by nodes and retained gene pairs were connected by edges. Paralogous gene clusters were computed as connected components. Gene pairs that were <7.5 Mbp apart in the genome were not considered further. This step removed tandemly repeated and genetically closely linked genes, for which it would be difficult to obtain homozygous DMs by genetic crosses. The availability either in the GABI-Kat or SALK collection of alleles with an insertion within the coding region was a prerequisite for the generation of DM lines in the Col-0 background. During the course of the project the annotation was updated to version 10 of TAIR, and the locations of insertions in lines assigned to DUPLO pairs were re-evaluated. Insertions outside of the coding region or the 5′-UTR were swapped for better alleles, if available.

To evaluate the over- and under-representation of genes in the collection, the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 was used (Huang da *et al.*, 2009a, b). The graphs for functional categorisation were generated using the TAIR annotation tool (http://www.arabidopsis.org/tools/bulk/go/index.jsp).

Expression similarities were calculated as PCCs from an all-against-all comparison of 170 microarray experiments (1765 individual expression arrays) based on the Affymetrix ATH1 platform. The experiments comprise the AtGenExpress and the NASC release three datasets (http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl). Probe sets were realigned to the TAIR version 6 annotation and analysed as described previously (Haberer *et al.*, 2006).

Growth conditions

All lines were grown on standard soil in the greenhouse. Plants were fertilised with a liquid nitrogen–phosphate–potassium fertiliser as recommended by the manufacturer. For growth on agar medium, seeds were surface-sterilised and sown on agar plates containing 1× MS (Sigma-Aldrich, http://www.sigmaaldrich.com/germany.html), 3% sucrose, 0.01% *myo*-inositol, 0.05% MES, vitamins (biotin, nicotinic acid, pyridoxine, thiamine). For selection of T-DNA-containing T2 plants of GABI-Kat lines, 5.25 μg ml^{-1} sulfadiazine was added to the plates. Segregation of the F2 or F3 generation was in part performed on 1× MS medium including 1% sucrose.

Selection of primers

For each individual T-DNA insertion allele, an allele-specific primer was designed using scripts based upon the program PRIMER3 (Rozen and Skaletsky, 2000), taking into account the predicted T-DNA insertion position, the orientation of the T-DNA insertion, a desired amplicon size of 400–760 bp, and a T_m of about 60.5°C. The allele-specific primers were used together with T-DNA border primers, which are specific either for the GABI-Kat or SALK T-DNA. We refer to the resulting fragments as 'confirmation amplicons'. The T-DNA insertion positions were deduced from the sequences of the confirmation amplicons (Kleinboelting *et al.*, 2012). Amplicons designed to show presence/absence of the wild-type alleles ('genotyping amplicons') are based on primer pairs that span the deduced insertion sites. In this case, one of the allele-specific primers for the confirmation amplicon was routinely used together with a primer based on genomic sequences located at the opposite side of the insertion and designed with PRIMER3 (Figure S2). The size range of the genotyping amplicons was 540–1260 bp. If synthesis of the initially designed amplicon failed, primers were redesigned. All primers shown in this work and on the website were experimentally verified.

Genotyping

Genomic DNA of the parental lines was prepared from ground leaf material with a modified CTAB-DNA protocol (Dellaporta *et al.*, 1983; Jobes *et al.*, 1995). Genomic DNA of the following generations was extracted from 5 to 10 mg of leaf material from a 3- to 5-week-old plant. After addition of 400 μl of extraction buffer (0.2 M Tris/HCl pH 7.5; 0.25 M NaCl; 0.025 M EDTA; 0.5% SDS) and a stainless steel ball (5 mm diameter) the leaf material was homogenised by shaking the tubes for 3 min at 30 Hz in a Retsch homogeniser (Retsch Mixer Mill 300; Qiagen, http://www.qiagen.com). The suspension was centrifuged at 13 000 ***g*** and the supernatant was transferred into a new tube. An equal volume of isopropanol was added, mixed, and the mixture was centrifuged after a 2-min incubation at room temperature. The supernatant was transferred to a new tube and 2.5 volumes of 70% ethanol were added. After mixing and centrifugation the supernatant was removed and the pellet was dried for 30 min. Finally, the DNA was dissolved in 100 μl of H_2O by shaking at room temperature. DNA was stored at −20°C. For GABI-Kat lines the template DNA was tested for the presence of a T-DNA insertion with the SUL^r ORF-specific primers Sul2 and Sul4 (see Figure S2). For SALK lines, the SALK T-DNA-specific primers G171 and G172 were used (see Figure S2).

PCR was performed using either the primers for the confirmation or the genotyping amplicon. PCR was carried out with 1–2 μl of genomic DNA in a final reaction volume of 50 μl containing 17 mM $MgCl_2$. Col-0 DNA was used for control of the genotyping amplicon PCR (positive result expected) and confirmation amplicon PCR (negative result expected). For the confirmation of FST-based insertion-site predictions, genomic DNA isolated from a pool of up to 25 seedlings was used as template in the confirmation amplicon PCRs. For GABI-Kat lines, the border PCR primers for the confirmation amplicons were 8474 for left border (LB) FSTs and 3144 for right border (RB) FSTs. For SALK lines the LB primer was LBb3. If a confirmation PCR failed, a second reaction was performed using a different gene-specific primer. Confirmation amplicons were purified with the ExoSAP-IT kit (USB, USA) and sequenced with a gene-specific primer and a border primer. The border sequencing primers were 8409 for LB FSTs and 3144 for RB FSTs in the case of GABI-Kat lines and R210 in the case of SALK lines (LB). The gene-specific primers were those used for the PCRs. Sequencing results were compared to the genomic sequence of *A. thaliana* (TAIR version 10 genome sequence and annotation dataset) using the BLAST algorithm (Altschul *et al.*, 1990). When the insertion-site prediction for at least one of the two sequences was consistent with the FST-based prediction, the respective allele was regarded as confirmed in the line. If no confirmation amplicon could be obtained, a suitable replacement was chosen, when further lines

were available. After confirmation of the insertions and the establishment of the genotyping amplicon using Col-0 DNA as described above, genomic DNA of 12–18 individual T2 plants was prepared and the plants were genotyped in order to identify predominantly homozygous parental lines.

Of the parental lines, four to eight plants were analysed with the insertion- and gene-specific primers prior to crossing, and predominantly homozygous parents were selected for crossing. Of the F1 generation 2–4 reciprocally crossed plants were tested for both insertions. For the F2 generation 54 plants were tested with the confirmation and genotyping amplicons of both insertions. DMs were confirmed in the F3 or F4 generation. Segregation on plates made with sucrose-containing and sulfadiazine-containing medium (for GABI-Kat lines) allowed for calculation of the segregation rate of germinated versus non-germinated seeds or non-viable seedlings. For statistical analysis about 100 seeds were used.

Phenotypic analysis

Phenotypic analysis was performed in the F3 or F4 generation. Three to six plants per DM lines were grown in parallel to the WT. The plants were inspected visually after 1, 3 and 5 weeks. Rosette size, leaf coloration, flowering time, overall height, branching pattern and floral development were noted. For confirmation of the observed phenotype at least six DM plants were grown side-by-side with the parental lines and WT. Three-week-old seedlings were photographed to measure the size of the rosette indicated by the length of the longest leaf (blade plus petiole) and the effective quantum yield of photosystem II (Φ_{II}) was measured as a marker for photosynthetic activity (Leister *et al.*, 1999). Flowering time was measured in days after germination (dag) upon appearance of the bolt. The overall height of the flowering stem was measured after 39 dag and upon maturation. Additional abnormalities were documented by photographs.

In-silico analyses

To compare the expression in different tissues and developmental stages for the individual gene pairs, the eFP Browser was used (Winter *et al.*, 2007). Prediction of subcellular localization and cluster analysis of related plant protein sequences were performed using the Aramemnon database (Schwacke *et al.*, 2003).

ACKNOWLEDGEMENTS

We thank Mario Rosso and Yong Li for their effort to start this project and the selection of the alleles, the BMBF for funding (grant GABI-DUPLO; Förderkennzeichen: 0315055A) and Paul Hardy for critical reading of the manuscript. We acknowledge the technical assistance of Elvira Glatt, Renate Harder, Christine Gonzales-Serrano, Simona Horn, Michael Mulatsch, Susanne Olbrich, Helene Schellenberg, Laura Schröder, Nina Schmidt, Andrea Voigt and Marja-Leena Wilke.

SUPPORTING INFORMATION

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

Figure S1. Functional distribution of proteins covered by the DUPLO collection.

Figure S2. Overview of the primers used to analyse T-DNA insertion lines.

Figure S3. Screenshots of the DUPLOdb.

Table S1. List of the 200 DUPLO lines generated and distributed to NASC so far (as of March 2013).

REFERENCES

Abel, S., Savchenko, T. and Levy, M. (2005) Genome-wide comparative analysis of the IQD gene families in *Arabidopsis thaliana* and *Oryza sativa*. *BMC Evol. Biol.* **5**, 72.

AGI. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Albrecht, V., Simková, K., Carrie, C., Delannoy, E., Giraud, E., Whelan, J., Small, I.D., Apel, K., Badger, M.R. and Pogson, B.J. (2010) The cytoskeleton and the peroxisomal-targeted snowy cotyledon3 protein are required for chloroplast development in Arabidopsis. *Plant Cell*, **22**, 3423–3438.

Alonso, J.M., Stepanova, A.N., Leisse, T.J. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653–657.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

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

Beers, E.P., Jones, A.M. and Dickerman, A.W. (2004) The S8 serine, C1A cysteine and A1 aspartic protease families in Arabidopsis. *Phytochemistry*, **65**, 43–58.

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

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

Bonawitz, N.D., Soltau, W.L., Blatchley, M.R., Powers, B.L., Hurlock, A.K., Seals, L.A., Weng, J.K., Stout, J. and Chapple, C. (2012) REF4 and RFR1, subunits of the transcriptional coregulatory complex mediator, are required for phenylpropanoid homeostasis in Arabidopsis. *J. Biol. Chem.* **287**, 5434–5445.

Bowers, J.E., Chapman, B.A., Rong, J. and Paterson, A.H. (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, **422**, 433–438.

Bryant, N., Lloyd, J., Sweeney, C., Myouga, F. and Meinke, D. (2011) Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. *Plant Physiol.* **155**, 1678–1689.

Cannon, S.B., Mitra, A., Baumgarten, A., Young, N.D. and May, G. (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* **4**, 10.

Carland, F., Fujioka, S. and Nelson, T. (2010) The sterol methyltransferases SMT1, SMT2, and SMT3 influence Arabidopsis development through nonbrassinosteroid products. *Plant Physiol.* **153**, 741–756.

Charon, C., Bruggeman, Q., Thareau, V. and Henry, Y. (2012) Gene duplication within the Green Lineage: the case of TEL genes. *J. Exp. Bot.* **63**, 5061–5077.

Clay, N.K. and Nelson, T. (2005) The recessive epigenetic swellmap mutation affects the expression of two step II splicing factors required for the transcription of the cell proliferation gene STRUWWELPETER and for the timing of cell cycle arrest in the Arabidopsis leaf. *Plant Cell*, **17**, 1994–2008.

DalCorso, G., Pesaresi, P., Masiero, S., Aseeva, E., Schunemann, D., Finazzi, G., Joliot, P., Barbato, R. and Leister, D. (2008) A complex containing PGRL1 and PGR5 is involved in the switch between linear and cyclic electron flow in Arabidopsis. *Cell*, **132**, 273–285.

Dellaporta, S., Wood, J. and Hicks, J.B. (1983) A plant DNA minipreparation. *Plant Mol. Biol. Rep.* **1**, 19–21.

Dowdle, J., Ishikawa, T., Gatzek, S., Rolinski, S. and Smirnoff, N. (2007) Two genes in *Arabidopsis thaliana* encoding GDP-L-galactose phosphorylase are required for ascorbate biosynthesis and seedling viability. *Plant J.* **52**, 673–689.

Drevensek, S., Goussot, M., Duroc, Y. *et al.* (2012) The Arabidopsis TRM1-TON1 interaction reveals a recruitment network common to plant cortical microtubule arrays and eukaryotic centrosomes. *Plant Cell*, **24**, 178–191.

Ewan, R., Pangestuti, R., Thornber, S., Craig, A., Carr, C., O'Donnell, L., Zhang, C. and Sadanandom, A. (2011) Deubiquitinating enzymes AtUBP12 and AtUBP13 and their tobacco homologue NtUBP12 are negative regulators of plant immunity. *New Phytol.* **191**, 92–106.

Franco-Zorrilla, J.M., Martín, A.C., Leyva, A. and Paz-Ares, J. (2005) Interaction between phosphate-starvation, sugar, and cytokinin signaling in Arabidopsis and the roles of cytokinin receptors CRE1/AHK4 and AHK3. *Plant Physiol.* **138**, 847–857.

Green, L.S. and Rogers, E.E. (2004) FRD3 controls iron localization in Arabidopsis. *Plant Physiol.* **136**, 2523–2531.

Gupta, R., He, Z. and Luan, S. (2002) Functional relationship of cytochrome c_6 and plastocyanin in *Arabidopsis*. *Nature*, **417**, 567–571.

Haberer, G., Mader, M.T., Kosarev, P., Spannagl, M., Yang, L. and Mayer, K.F. (2006) Large-scale cis-element detection by analysis of correlated expression and sequence conservation between Arabidopsis and *Brassica oleracea*. *Plant Physiol.* **142**, 1589–1602.

He, X.L. and Zhang, J.Z. (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics*, **169**, 1157–1164.

Huang da, W., Sherman, B.T., Zheng, X., Yang, J., Imamichi, T., Stephens, R. and Lempicki, R.A. (2009a) Extracting biological meaning from large gene lists with DAVID. *Curr. Protoc. Bioinformatics*, Chapter 13, Unit 13.11.

Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.

Ihnatowicz, A., Pesaresi, P., Varotto, C., Richly, E., Schneider, A., Jahns, P., Salamini, F. and Leister, D. (2004) Mutants for photosystem I subunit D of *Arabidopsis thaliana*: effects on photosynthesis, photosystem I stability and expression of nuclear genes for chloroplast functions. *Plant J.* **37**, 839–852.

Ihnatowicz, A., Pesaresi, P. and Leister, D. (2007) The E subunit of photosystem I is not essential for linear electron flow and photoautotrophic growth in *Arabidopsis thaliana*. *Planta*, **226**, 889–895.

Jiang, D., Yin, C., Yu, A., Zhou, X., Liang, W., Yuan, Z., Xu, Y., Yu, Q., Wen, T. and Zhang, D. (2006) Duplication and expression analysis of multicopy miRNA gene family members in Arabidopsis and rice. *Cell Res.* **16**, 507–518.

Jiao, Y., Wickett, N.J., Ayyampalayam, S. *et al.* (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473**, 97–100.

Jobes, D.V., Hurley, D.L. and Thien, L.B. (1995) Plant DNA isolation: a method to efficiently remove polyphenolics, polysaccharides, and RNA. *Taxon*, **44**, 379–386.

Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P. and Weisshaar, B. (2012) GABI-Kat SimpleSearch: new features of the *Arabidopsis thaliana* T-DNA mutant database. *Nucleic Acids Res.* **40**, D1211–D1215.

Kolukisaoglu, H.U., Bovet, L., Klein, M., Eggmann, T., Geisler, M., Wanke, D., Martinoia, E. and Schulz, B. (2002) Family business: the multidrug-resistance related protein (MRP) ABC transporter genes in *Arabidopsis thaliana*. *Planta*, **216**, 107–119.

Kong, H., Landherr, L.L., Frohlich, M.W., Leebens-Mack, J., Ma, H. and dePamphilis, C.W. (2007) Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *Plant J.* **50**, 873–885.

Kuromori, T., Wada, T., Kamiya, A. *et al.* (2006) A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of Arabidopsis. *Plant J.* **47**, 640–651.

Lariguet, P., Boccalandro, H.E., Alonso, J.M., Ecker, J.R., Chory, J., Casal, J.J. and Fankhauser, C. (2003) A growth regulatory loop that provides homeostasis to phytochrome a signaling. *Plant Cell*, **15**, 2966–2978.

Latrasse, D., Benhamed, M., Henry, Y., Domenichini, S., Kim, W., Zhou, D.X. and Delarue, M. (2008) The MYST histone acetyltransferases are essential for gametophyte development in Arabidopsis. *BMC Plant Biol.* **8**, 121.

Lee, S.Y. and Choi, Y. (2007) TRAF1 and its biological functions. *Adv. Exp. Med. Biol.* **597**, 25–31.

Lee, C., Teng, Q., Zhong, R. and Ye, Z.H. (2012) Arabidopsis GUX proteins are glucuronyltransferases responsible for the addition of glucuronic acid side chains onto xylan. *Plant Cell Physiol.* **53**, 1204–1216.

Leister, D. (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. *Trends Genet.* **20**, 116–122.

Leister, D., Varotto, C., Pesaresi, P., Niwergall, A. and Salamini, F. (1999) Large-scale evaluation of plant growth in *Arabidopsis thaliana* by non-invasive image analysis. *Plant Physiol. Biochem.* **37**, 671–678.

León, G., Holuigue, L. and Jordana, X. (2007) Mitochondrial complex II Is essential for gametophyte development in Arabidopsis. *Plant Physiol.* **143**, 1534–1546.

Lermontova, I., Fuchs, J. and Schubert, I. (2008) The Arabidopsis checkpoint protein Bub3.1 is essential for gametophyte development. *Front Biosci.* **13**, 5202–5211.

Liebminger, E., Huttner, S., Vavra, U. *et al.* (2009) Class I alpha-mannosidases are required for N-glycan processing and root development in *Arabidopsis thaliana*. *Plant Cell*, **21**, 3850–3867.

Liu, Y., Wang, F., Zhang, H., He, H., Ma, L. and Deng, X.W. (2008) Functional characterization of the Arabidopsis ubiquitin-specific protease gene family reveals specific role and redundancy of individual members in development. *Plant J.* **55**, 844–856.

Liu, Q., Zhang, C., Yang, Y. and Hu, X. (2010) Genome-wide and molecular evolution analyses of the phospholipase D gene family in poplar and grape. *BMC Plant Biol.* **10**, 117.

Lorrain, S., Trevisan, M., Pradervand, S. and Fankhauser, C. (2009) Phytochrome interacting factors 4 and 5 redundantly limit seedling de-etiolation in continuous far-red light. *Plant J.* **60**, 449–461.

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

Maher, C., Stein, L. and Ware, D. (2006) Evolution of Arabidopsis microRNA families through duplication events. *Genome Res.* **16**, 510–519.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H. and Michelmore, R.W. (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell*, **15**, 809–834.

Moffat, C.S., Ingle, R.A., Wathugala, D.L., Saunders, N.J., Knight, H. and Knight, M.R. (2012) ERF5 and ERF6 play redundant roles as positive regulators of JA/Et-mediated defense against *Botrytis cinerea* in Arabidopsis. *PLoS ONE*, **7**, e35995.

Nakano, T., Suzuki, K., Fujimura, T. and Shinshi, H. (2006) Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant Physiol.* **140**, 411–432.

O'Malley, R.C. and Ecker, J.R. (2010) Linking genotype to phenotype using the Arabidopsis unimutant collection. *Plant J.* **61**, 928–940.

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

Paterson, A.H., Freeling, M., Tang, H. and Wang, X. (2010) Insights from the comparison of plant genome sequences. *Annu. Rev. Plant Biol.* **61**, 349–372.

Pesaresi, P., Scharfenberg, M., Weigel, M. *et al.* (2009) Mutants, overexpressors, and interactors of Arabidopsis plastocyanin isoforms: revised roles of plastocyanin in photosynthetic electron flow and thylakoid redox state. *Mol. Plant*, **2**, 236–248.

Remington, D.L., Vision, T.J., Guilfoyle, T.J. and Reed, J.W. (2004) Contrasting modes of diversification in the Aux/IAA and ARF gene families. *Plant Physiol.* **135**, 1738–1752.

Reyes, F., Leon, G., Donoso, M., Brandizzi, F., Weber, A.P. and Orellana, A. (2010) The nucleotide sugar transporters AtUTr1 and AtUTr3 are required for the incorporation of UDP-glucose into the endoplasmic reticulum, are essential for pollen development and are needed for embryo sac progress in *Arabidopsis thaliana*. *Plant J.* **61**, 423–435.

Riefler, M., Novak, O., Strnad, M. and Schmülling, T. (2006) Arabidopsis cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism. *Plant Cell*, **18**, 40–54.

Rizza, A., Boccaccini, A., Lopez-Vidriero, I., Costantino, P. and Vittorioso, P. (2011) Inactivation of the ELIP1 and ELIP2 genes affects Arabidopsis seed germination. *New Phytol.* **190**, 896–905.

Roschzttardtz, H., Seguela-Arnaud, M., Briat, J.F., Vert, G. and Curie, C. (2011) The FRD3 citrate effluxer promotes iron nutrition between symplastically disconnected tissues throughout Arabidopsis development. *Plant Cell*, **23**, 2725–2737.

Rossini, S., Casazza, A.P., Engelmann, E.C., Havaux, M., Jennings, R.C. and Soave, C. (2006) Suppression of both ELIP1 and ELIP2 in Arabidopsis does not affect tolerance to photoinhibition and photooxidative stress. *Plant Physiol.* **141**, 1264–1273.

Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386.

Rutter, M.T., Cross, K.V. and Van Woert, P.A. (2012) Birth, death and subfunctionalization in the Arabidopsis genome. *Trends Plant Sci.* **17**, 204–212.

Schwacke, R., Schneider, A., van der Graaff, E., Fischer, K., Catoni, E., Desimone, M., Frommer, W.B., Flügge, U.I. and Kunze, R. (2003) ARAMEM-

NON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiol.* **131**, 16–26.

Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M. and Van de Peer, Y. (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **99**, 13627–13632.

Strasser, R., Altmann, F., Mach, L., Glossl, J. and Steinkellner, H. (2004) Generation of *Arabidopsis thaliana* plants with complex N-glycans lacking beta1,2-linked xylose and core alpha1,3-linked fucose. *FEBS Lett.* **561**, 132–136.

Suzuki, M., Kamide, Y., Nagata, N. *et al.* (2004) Loss of function of 3-hydroxy-3-methylglutaryl coenzyme A reductase 1 (HMG1) in Arabidopsis leads to dwarfing, early senescence and male sterility, and reduced sterol levels. *Plant J.* **37**, 750–761.

Suzuki, M., Nakagawa, S., Kamide, Y., Kobayashi, K., Ohyama, K., Hashinokuchi, H., Kiuchi, R., Saito, K., Muranaka, T. and Nagata, N. (2009) Complete blockage of the mevalonate pathway results in male gametophyte lethality. *J. Exp. Bot.* **60**, 2055–2064.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M. and Paterson, A.H. (2008a) Synteny and collinearity in plant genomes. *Science*, **320**, 486–488.

Tang, H., Wang, X., Bowers, J.E., Ming, R., Alam, M. and Paterson, A.H. (2008b) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* **18**, 1944–1954.

Van de Peer, Y., Fawcett, J.A., Proost, S., Sterck, L. and Vandepeole, K. (2009) The flowering world: a tale of duplications. *Trends Plant Sci.* **14**, 680–688.

Vandepoele, K., Simillion, C. and Van de Peer, Y. (2002) Detecting the undetectable: uncovering duplicated segments in Arabidopsis by comparison with rice. *Trends Genet.* **18**, 606–608.

Wang, H. and Deng, X.W. (2002) Arabidopsis FHY3 defines a key phytochrome A signaling component directly interacting with its homologous partner FAR1. *EMBO J.* **21**, 1339–1349.

Wang, J., Long, Y., Wu, B., Liu, J., Jiang, C., Shi, L., Zhao, J., King, G.J. and Meng, J. (2009) The evolution of *Brassica napus* FLOWERING LOCUS T paralogues in the context of inverted chromosomal duplication blocks. *BMC Evol. Biol.* **9**, 271.

Watkins, N.J. and Bohnsack, M.T. (2012) The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip. Rev. RNA*, **3**, 397–414.

Weigel, M., Varotto, C., Pesaresi, P., Finazzi, G., Rappaport, F., Salamini, F. and Leister, D. (2003) Plastocyanin is indispensable for photosynthetic electron flow in *Arabidopsis thaliana*. *J. Biol. Chem.* **278**, 31286–31289.

Welchen, E., Hildebrandt, T.M., Lewejohann, D., Gonzalez, D.H. and Braun, H.P. (2012) Lack of cytochrome c in Arabidopsis decreases stability of Complex IV and modifies redox metabolism without affecting Complexes I and III. *Biochim. Biophys. Acta*, **1817**, 990–1001.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J. (2007) An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS ONE*, **2**, e718.

Yao, M., Wakamatsu, Y., Itoh, T.J., Shoji, T. and Hashimoto, T. (2008) Arabidopsis SPIRAL2 promotes uninterrupted microtubule growth by suppressing the pause state of microtubule dynamics. *J. Cell Sci.* **121**, 2372–2381.

Yokoyama, A., Yamashino, T., Amano, Y., Tajima, Y., Imamura, A., Sakakibara, H. and Mizuno, T. (2007) Type-B ARR transcription factors, ARR10 and ARR12, are implicated in cytokinin-mediated regulation of protoxylem differentiation in roots of *Arabidopsis thaliana*. *Plant Cell Physiol.* **48**, 84–96.

Yun, J.Y., Tamada, Y., Kang, Y.E. and Amasino, R.M. (2012) Arabidopsis trithorax-related3/SET domain GROUP2 is required for the winter-annual habit of *Arabidopsis thaliana*. *Plant Cell Physiol.* **53**, 834–846.

Zhang, Y., Cheng, Y.T., Qu, N., Zhao, Q., Bi, D. and Li, X. (2006) Negative regulation of defense responses in Arabidopsis by two NPR1 paralogs. *Plant J.* **48**, 647–656.

