

# RESEARCH ARTICLE

## Evolutionary Persistence of Functional Compensation by Duplicate Genes in *Arabidopsis*

Kousuke Hanada,\*†<sup>1</sup> Takashi Kuromori,\*<sup>1</sup> Fumiyoshi Myouga,\* Tetsuro Toyoda,† Wen-Hsiung Li,‡ and Kazuo Shinozaki\*

\*Gene Discovery Research Group, RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan; †Bioinformatics and Systems Engineering Division, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan; and ‡Department of Ecology and Evolution, University of Chicago

Knocking out a gene from a genome often causes no phenotypic effect. This phenomenon has been explained in part by the existence of duplicate genes. However, it was found that in mouse knockout data duplicate genes are as essential as singleton genes. Here, we study whether it is also true for the knockout data in *Arabidopsis*. From the knockout data in *Arabidopsis thaliana* obtained in our study and in the literature, we find that duplicate genes show a significantly lower proportion of knockout effects than singleton genes. Because the persistence of duplicate genes in evolution tends to be dependent on their phenotypic effect, we compared the ages of duplicate genes whose knockout mutants showed less severe phenotypic effects with those with more severe effects. Interestingly, the latter group of genes tends to be more anciently duplicated than the former group of genes. Moreover, using multiple-gene knockout data, we find that functional compensation by duplicate genes for a more severe phenotypic effect tends to be preserved by natural selection for a longer time than that for a less severe effect. Taken together, we conclude that duplicate genes contribute to genetic robustness mainly by preserving compensation for severe phenotypic effects in *A. thaliana*.

### Introduction

Knocking out a gene in an organism often causes no phenotypic effect. This phenomenon has been explained in part by the functional compensation of duplicate genes. If duplicate genes play a significant role in functional compensation, the proportion of duplicate genes with knockout phenotypic effects should be lower than the proportion for singleton genes. This is indeed the case in the yeast and in the nematode (Gu et al. 2003; Conant and Wagner 2004) but not in the mouse (Liang and Li 2007; Liao and Zhang 2007). Although recent reports pointed out that the mouse knockout data were biased in both duplication age and function (Su and Gu 2008; Makino et al. 2009), the relationship between phenotypic effect and gene duplication in the mouse appears to be weaker than those in the yeast and the nematode (Hannay et al. 2008). It was proposed that because mouse is more complex than nematode and yeast in terms of the number of different cell types, duplicate genes in mouse tend to undergo functional diversification instead of preserving functional compensation (Zhang 2003; Liao and Zhang 2007). To test this view, we examined the functional compensation by duplicate genes in *Arabidopsis thaliana*, using single-gene knockout data.

Another question is whether the functional compensation by duplicate genes persists for a long time. Because the loss of a redundant duplicate may cause no serious deleterious effect, a redundant copy may become lost quickly. However, many duplicate genes have been found to have preserved functional redundancy for a long period

of evolutionary time (Woollard 2005; Briggs et al. 2006; Tischler et al. 2006; DeLuna et al. 2008; Musso et al. 2008; Roux and Robinson-Rechavi 2008; Vavouri et al. 2008; Kafri et al. 2009). Also, theoretical studies suggest that functional compensation by duplicate genes can be preserved in evolution if it works for unexpected errors of gene function (e.g., null mutations; Nowak et al. 1997; Wagner 2000). Nevertheless, the evolutionary persistence of functional compensation has not been tested in detail in terms of natural selection. Therefore, we examined turnover rates and selection pressures on functional compensation for various phenotypic effects, using single-gene and multiple-gene knockout data.

### Materials and Methods

#### Phenotypic Effects of Single-Gene Knockout

We obtained 5041 insertional mutant lines from our mutant sources (Kuromori et al. 2006) and from the *Arabidopsis* Biological Resource Center (Columbus, OH). In each insertional mutant line, the tag was inserted into the coding region of a *TAIR7* gene. In many cases, a gene had multiple insertional mutant lines, but the phenotypic effect was the same for all lines. Thus, from the 5041 insertional mutant lines, we found 3871 genes that had an insertion(s) of *Ds* transposon or T-DNA. Among these 3871 genes with insertion, only 253 showed visible phenotypic changes. In addition, we found 1489 genes that had been described in the literature as inducing phenotypic change. In summary, among the 5360 genes included in this study, 1742 showed induced phenotypic changes, whereas 3618 did not (Table S1, Supplementary Material online).

#### Identification of Singletons, Duplicates, and Closest Paralog Genes

The amino acid sequences of *A. thaliana* (TAIR7) and mouse genes (NCBIM37.55) were obtained from TAIR ([www.arabidopsis.org](http://www.arabidopsis.org)) and Ensembl ([www.ensembl.org](http://www.ensembl.org)).

<sup>1</sup> These authors contributed equally to this work.

Key words: duplicate, *Arabidopsis thaliana*, phenotypic effect, functional compensation, selection pressure and genetic robustness.

E-mail: whli@uchicago.edu; kohanada@psc.riken.jp.

*Genome Biol. Evol.* Vol. 2009:409–414.

doi:10.1093/gbe/evp043

Advance Access publication October 29, 2009

**Table 1**  
**Numbers of Duplicates and Singletons with or without Knockout Effects on Phenotypes**

Phenotypic changes	Number of duplicates (D)	Number of singletons (S)	D/S ratio	<i>P</i> value (Yes vs. No)
Yes (either our study or literature)	1,168	98	12.0	$2.0 \times 10^{-11}$
Yes (our study)	166	19	8.7	$1.6 \times 10^{-5}$
Yes (literature)	1,002	79	12.7	$4.4 \times 10^{-8}$
No (our study)	2,903	122	23.8	

The longest amino acid sequence in a gene locus was used as the representative protein sequence. Similarity searches of all against all genes were conducted using BlastP (Altschul et al. 1997). A singleton protein was defined as a protein that did not match any other protein in the Blast search with the *E* value  $\leq 1$ . To find the closest paralog of a protein, we aligned the protein and all proteins obtained in the Blast search by ClustalW (Thompson et al. 1994) and estimated the amino acid similarity between sequences. The protein with the highest similarity to the one under study was defined as the closest paralog. Following the definition of Gu et al. (2003), a gene is defined as a duplicate if it showed  $\geq 30\%$  similarity and  $\geq 50\%$  alignable region at the protein level with the closest paralog. The synonymous divergence (Ks) and the nonsynonymous divergence (Ka) between the target gene and the closest paralog were estimated by the modified Nei–Gojombori method (Zhang et al. 1998).

#### Defining Gene Families and Tandem Duplicate Clusters

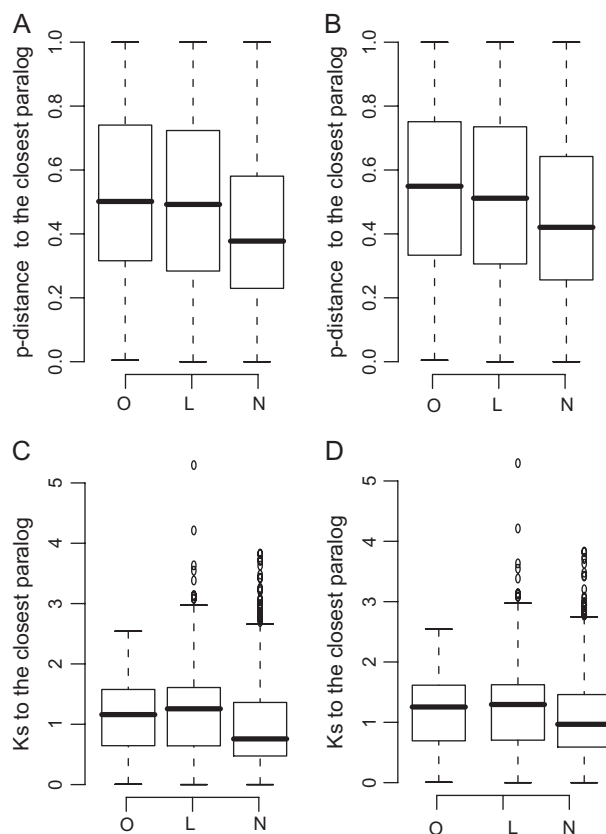
To define gene families in *A. thaliana*, all-against-all similarity searches of protein sequences were conducted using BlastP with an *E* value cutoff of  $10^{-5}$ . Based on the *E* values, we generated similarity clusters representing gene families with the Markov clustering program (van Dongen 2000). As reported earlier (Hanada et al. 2008), tandem duplicates were defined as genes in any gene pair, T1 and T2, that 1) belong to the same gene family, 2) are located within 100 kb of each other, and 3) are separated by at most ten nonhomologous (not in the same gene family as T1 and T2) genes.

## Results and Discussion

### Functional Compensation by Duplicate Genes

We examined the phenotypic changes in each of the 3871 knockout mutants and found phenotypic changes in only 253 genes. Furthermore, from the literature we found 1489 genes whose knockout mutants showed phenotypic changes. We then computed the ratio of duplicate and singleton genes with and without phenotypic changes (table 1). Genes with phenotypic changes have a significantly lower proportion of duplicate genes and a higher proportion of singleton genes than genes without phenotypic changes ( $P = 2.0 \times 10^{-11}$ ,  $\chi^2$  test), indicating that the functional compensation is significantly more common in duplicate genes than in singleton genes in the *Arabidopsis* genome.

Because the age (sequence divergence) of duplicate genes may influence the chance of functional compensa-



**Fig. 1.**—Sequence divergence in genes with and without phenotypic effects in single-gene knockout. Sequence divergences are in terms of the p-distance (proportion of amino acid differences) and the synonymous distance (Ks) between a knockout gene and the closest paralog. The three data sets (O, L, and N) used are 1) with phenotypic changes observed in our study (O), 2) with phenotypic changes reported in the literatures (L), and 3) no phenotypic changes observed in our study (N). **A.** The p-distance in the entire data. **B.** The p-distance in the data without tandem duplicates. **C.** The Ks in the entire data. **D.** The Ks in the data without tandem duplicates. The distributions of p-distance and Ks values are shown as box plots with the thick solid horizontal line indicating the median value, the box representing the interquartile range (25–75%), and the dotted lines indicating the first to the 99th percentile. The p-distance and Ks are compared between genes with and without phenotypic effects. Genes with phenotypic effects have a significantly higher p-distance and Ks to the closest paralog than genes without any phenotypic effects ( $P < 0.01$  in p-distance,  $P < 0.01$  in Ks). See Table S2 (Supplementary Material online) for *P* values.

tion, we computed the p-distance (proportion of amino acid differences) and the synonymous distance (Ks) between a knockout gene and its closest paralog. Because tandem duplicates may have a higher chance of gene conversion than nontandem duplicates (Gao and Innan 2004), we also did the same analyses using the data set without tandem duplicates. In both data sets, genes without phenotypic changes have, on average, significantly lower p-distance and Ks to the closest paralog than genes with phenotypic change ( $P < 0.01$  in p-distance and  $P < 0.01$  in Ks; fig. 1 and Table S2, Supplementary Material online). These results indicate a higher probability of functional compensation for recently duplicated genes than for anciently duplicated genes. Note that the conclusions drawn from our own data are essentially the same as those drawn from the data collected from the literature (table 1 and fig. 1).

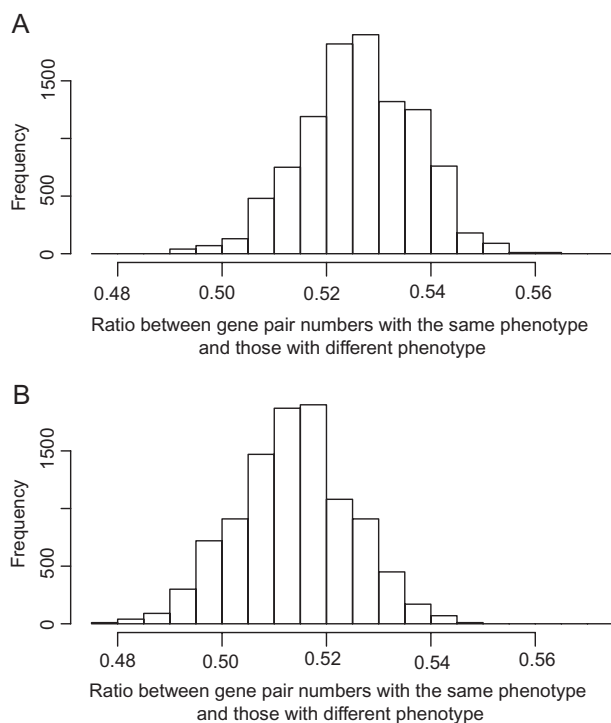


FIG. 2.—Relationship between gene family and phenotypic effect. Based on *Arabidopsis* gene families generated by the Markov Clustering algorithm, 10,000 pairs of genes are randomly chosen within a gene family (A) and between gene families (B). The ratio between the number of gene pairs with the same kind of phenotypes (seed, reproductive, vegetative, or conditional) and the number of gene pairs with different kind of phenotypes is significantly higher between genes in the same gene family than between genes in different gene families ( $P < 1 \times 10^{-15}$ ).

Therefore, the pooled data were used in the following analyses.

#### Persistence of a Phenotype in a Gene Family

Although functional compensation by duplicate genes tends to decrease with the age of duplicate genes, some proteins tend to keep functional redundancy for a long period (Kafri et al. 2008, 2009). It is possible that the rate of decrease may be different for different phenotypic effects. Note, however, that if a gene originally had multiple phenotypic effects, the rate of decrease for functional compensation to a particular phenotype may not be correctly inferred by functional compensation for the phenotype due to the deletion of an extant gene. To address whether a gene has been persistently related to a particular kind of phenotype in evolution or not, we examined the relationship between phenotypic effects and families of genes with phenotypic changes.

We inferred *Arabidopsis* gene families by the Markov Clustering algorithm (<http://micans.org/mcl/>). On the basis of the work of Meinke et al. (2003), we classified phenotypic changes into seed, reproductive, vegetative, and conditional phenotypes (Table S1, Supplementary Material online). Briefly, seed, reproductive, and vegetative phenotypes show abnormally visible changes with developmental stage, whereas the conditional phenotype shows abnormally visible changes only in response to either a biotic

or an abiotic treatment. Among the 1742 genes whose knockout mutants showed phenotypic changes, 279, 219, 882, and 362 were classified as genes whose knockout mutants showed changes as the seed, reproductive, vegetative, and conditional phenotypes, respectively. From the 1742 genes, we randomly chose 10,000 pairs of genes either within a gene family or between gene families, and examined the ratio between the number of gene pairs with the same kind of phenotypes (seed, reproductive, vegetative, or conditional) and the number of gene pairs with different kind of phenotypes. We compared the ratios within and between gene families (fig. 2) and found that genes in the same gene family tend to have higher ratios than genes from different families ( $P < 1 \times 10^{-15}$ , the *t*-test), indicating that a gene family tends to be persistently maintaining its phenotype.

#### Differential Persistence in Functional Compensation among Phenotypes

To examine the evolutionary persistence of functional compensation for different phenotypes, we compared the ages of duplicate genes among the four phenotypes in either the entire data set or the data set excluding tandem duplicates (fig. 3). In both data sets, genes whose knockout mutants showed changes in seed, reproductive, vegetative, conditional, and no-phenotypes are ranked in that order of oldest to youngest duplication dates. That is, it is likely that seed, reproductive, vegetative, conditional, and no-phenotypes are ranked in the order of longest to the shortest persistence of functional compensation.

If natural selection affects the persistence of functional compensation, the turnover rate is likely to be slower in genes with functional compensation to more severe phenotypes than genes with functional compensation to less severe phenotypes. We assume that the different phenotypes can be ranked in the order of highest to lowest significance for plant survival. The relationship between phenotype and significance can be explained as follows. Among the four phenotypes studied, an effect on seed phenotype might be most severe because a mutant with an altered seed phenotype may have embryonic defects and may not be able to germinate. Thus, functional compensation by duplicate genes for this phenotype may persist longer than the three other phenotypes. In the case of mutants with changes in reproductive phenotype, these changes largely affect the next generation and seem to reduce survivability. Therefore, the reproductive phenotype is generally severe but less severe for survival than the seed phenotype. For the vegetative phenotype, some mutations in this phenotype do not affect survival, so that this phenotype is likely to be less severe than that of either seed or reproductive phenotype. Changes in conditional phenotype appear only under certain treatments, so the significance of this phenotype is likely to be minimal. No-change phenotype is undoubtedly the least severe for plant survival. We therefore propose the order from the highest to the lowest significance in phenotypes as seed, reproductive, vegetative, and conditional and no-change.

Under this ranking of significance, we tested differences in either p-distance or Ks distance between different pairs of phenotypes by the Wilcoxon test (fig. 3 and Table S2,

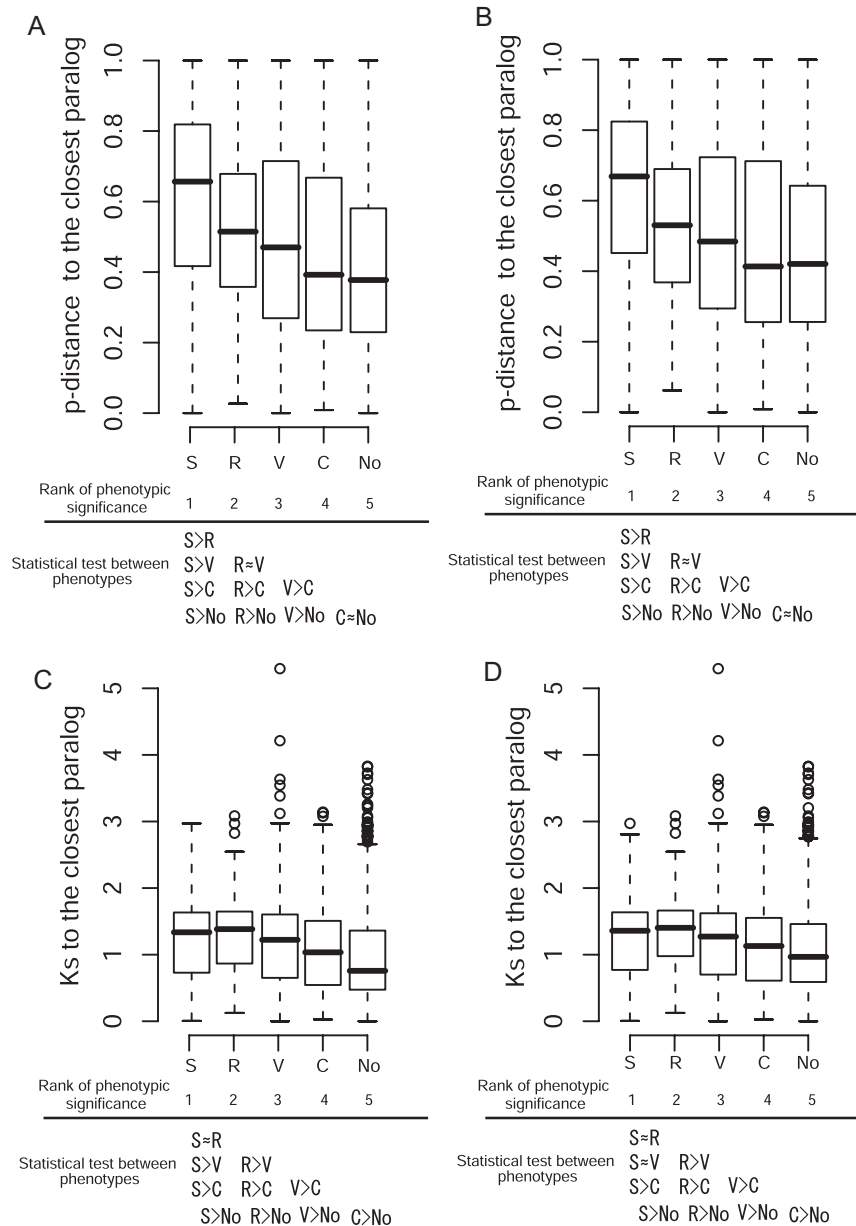


FIG. 3.—Relationship between sequence divergence and phenotypic effect. *A*. Relationship between the protein distance (p-distance) between a knockout gene and its closest paralog and phenotypic effect when the entire data were used. *B*. Relationship between the protein distance (p-distance) between a knockout gene and its closest paralog and phenotypic effect when the data without tandem duplicates were used. *C*. Relationship between the synonymous distance (Ks) between a knockout gene and its closest paralog and phenotypic effect when the entire data were used. *D*. Relationship between the synonymous distance (Ks) between a knockout gene and its closest paralog and phenotypic effect when the data without tandem duplicates were used. Phenotypic changes are classified into seed (S), reproductive (R), vegetative (V), conditional (C), or no (No) phenotypes. The distributions of p-distance and Ks values are shown as box plots with the thick solid horizontal line indicating the median value, the box representing the interquartile range (25–75%), and the dotted lines indicating the first to the 99th percentile. It appears that the order of phenotypic effect from the highest to the lowest significance is seed > reproduction > vegetative > conditional > no effect; see text for the ranking. Below each figure, significant differences are shown for each pair of phenotypes by the Wilcoxon test. “ $X > Y$ ” means that  $X$  is significantly greater than  $Y$  at  $P = 0.05$  and “ $X \approx Y$ ” means that  $X$  is not different from  $Y$  at the 5% level of significance.

Supplementary Material online). We found that genes whose knockout mutants showed more severe phenotypes tend to be more anciently duplicated, whereas those whose knockout mutants showed less severe phenotypes are more recently duplicated. Thus, this result supports our expectation that functional compensation to more severe phenotypes tends to persist for a longer time, whereas functional compensation to less severe phenotypes tends to disappear faster. Thus, this relationship between duplica-

tion age and phenotypic significance supports the hypothesis that natural selection affects the persistence of functional compensation.

#### Different Selection Pressures on Functional Compensation among Phenotypes

We then tried to evaluate selection pressures using duplicate genes with direct evidence of functional

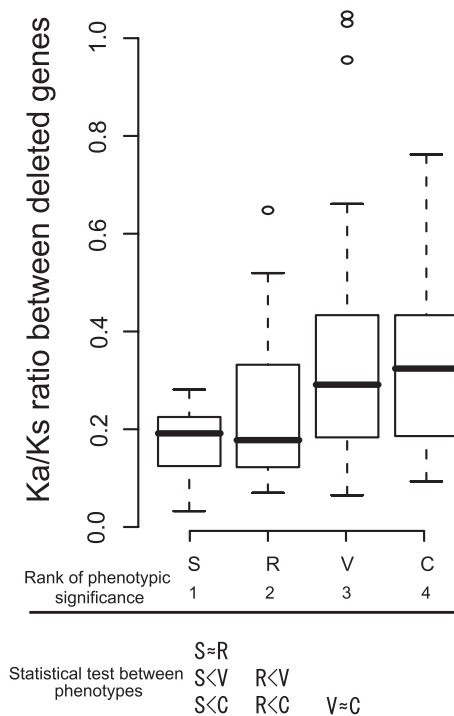


FIG. 4.—Selection pressures (Ka/Ks ratio) on paralogous genes whose multiple-knockout mutants show phenotypic changes, but single-knockout mutants do not selection pressures on paralogous genes whose multiple-knockout mutants show phenotypic changes in seed (S), vegetative (V), reproductive (R), or conditional (C) phenotypes are inferred by the ratio of the nonsynonymous substitution rate (Ka) to the synonymous substitution rate (Ks). The distributions of Ka/Ks ratios are shown as box plots with the thick solid horizontal line indicating the median value, the box representing the interquartile range (25–75%), and the dotted lines indicating the first to the 99th percentile. Significant differences are shown for each pair of phenotypes by the Wilcoxon test. “ $X > Y$ ” means that  $X$  is significantly greater than  $Y$  at  $P = 0.05$  and “ $X \approx Y$ ” means that  $X$  is not different from  $Y$  at the 5% level of significance.

compensation. “Direct evidence” means the observation of phenotypic change only upon knocking out multiple duplicate genes; knocking out a single gene did not induce phenotypic change (Table S1, Supplementary Material online). A total of 163 nontandem duplicate genes that were related to 75 kinds of phenotypic change were classified into 26, 89, 32, and 16 genes whose multiple-knockout mutants showed phenotypic changes in the seed, vegetative, reproductive, and conditional phenotypes, respectively.

Using the above genes, we examined whether the pressure of natural selection on functional compensation by duplicate genes depends on the phenotype involved. Because selection pressure on genes is commonly estimated by the ratio of the nonsynonymous distance (Ka) to the synonymous distance (Ks), we compared the Ka/Ks ratios of the genes whose knockout mutants showed phenotypic changes in deletion of multiple duplicate genes. Differences in Ka/Ks were tested between phenotypes by the Wilcoxon test (fig. 4 and Table S2, Supplementary Material online). Although the difference in Ka/Ks was not significant between some phenotype pairs, the Ka/Ks ratio increased with decreasing phenotypic significance as a whole, indicating

different selection pressures. Indeed, duplicate genes with functional compensation for more severe phenotypes tend to be under strong purifying selection (a low Ka/Ks ratio), and those with functional compensation for less severe phenotypes are subject to relaxed purifying selection (a high Ka/Ks ratio). Because recently (low Ks and p-distance values) and anciently (high Ks and p-distance values) duplicated genes tend to have high and low Ka/Ks ratios, respectively (Lynch and Conery 2000), the differences in Ks and p-distance may mislead the relationship between selection pressures and phenotypes in the case where genes with functional compensation for more severe phenotypes have high Ks and p-distance values. However, genes with functional compensation for more severe phenotypes tend to have lower Ks and P-distance values (Fig. S1, Supplementary Material online). That is, genes with functional compensation for more severe phenotypes have lower Ks values and lower Ka/Ks ratios in comparison with genes with functional compensation for less severe phenotypes (fig. 4 and Fig. S1, Supplementary Material online). Therefore, duplication age unlikely affects our conclusion. These results indicate that there is a stronger tendency for selection pressure to preserve functional compensation to severe phenotype by duplicate genes. Thus, duplicate genes contribute to genetic robustness mainly by preserving compensation of severe phenotypic effects in the *A. thaliana* genome.

## Conclusion

From single-gene knockout data in *A. thaliana* obtained in our study and from the literature, we found that duplicate genes play a significant role in functional compensation. This conclusion is similar to that in nematode and yeast but different from that in mouse (Gu et al. 2003; Conant and Wagner 2004; Liang and Li 2007; Liao and Zhang 2007). Because mouse is a more complex organism than nematode and yeast, it was proposed that mouse duplicate genes may tend to undergo functional divergence instead of preserving functional compensation (Zhang 2003; Liao and Zhang 2007). However, this view is not supported by our observation in *A. thaliana*. A possible reason for the difference between mouse and *A. thaliana* duplicate genes is differences in duplication ages because genes with functional compensation tend to be more recently duplicated (Fig. 1). To examine the duplication age between mouse and *A. thaliana*, we calculated the p-distance and the Ks between each duplicate gene and its closest paralog in mouse and *A. thaliana* (Fig. S2, Supplementary Material online). Both p-distance and Ks are, on average, significantly lower in *A. thaliana* than in mouse ( $P < 1 \times 10^{-167}$  in p-distance,  $P < 1 \times 10^{-120}$  in Ks) by the Wilcoxon test, indicating that *A. thaliana* genome has a higher proportion of recently duplicated genes than the mouse genome.

Another interesting finding in this study is that the evolutionary persistence of functional compensation by duplicate genes depends on the phenotype involved. That is, functional compensation by duplicate genes tends to persist for a longer time for a more severe phenotype than a less severe phenotype. We therefore propose that natural

selection retains functional compensation to increase genetic robustness, especially by preserving the functional compensation to more significant phenotype. Evidence of functional compensation by duplicate genes was provided by the observation of phenotypic change only upon knocking out multiple duplicate genes. Moreover, we found that the duplicate genes with functional compensation of more severe phenotypes tend to be under strong purifying selection (a low  $K_a/K_s$  ratio), and those with functional compensation of less severe phenotypes are subject to relaxed selection (a high  $K_a/K_s$  ratio). In conclusion, one reason why many duplicated genes are retained in the *Arabidopsis* genome is because some or many of them contribute to genetic robustness and are retained by natural selection.

### Supplementary Material

Supplementary materials Table S1, Table S2, Figure S1 and Figure S2 are available at *Genome Biology and Evolution* online ([http://www.oxfordjournals.org/our\\_journals/gbe/](http://www.oxfordjournals.org/our_journals/gbe/)).

### Acknowledgments

We thank Kenneth H Wolfe, Takashi Makino, and Nishiyama Tomoaki for discussions. We also thank TAIR (<http://www.arabidopsis.org/>) for giving us gene sequence data. This work was supported by the RIKEN Plant Science Center and by National Institutes of Health grant (GM30998 to W.H.L.).

### Literature Cited

- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Briggs GC, Osmont KS, Shindo C, Sibout R, Hardtke CS. 2006. Unequal genetic redundancies in *Arabidopsis*—a neglected phenomenon? *Trends Plant Sci.* 11:492–498.
- Conant GC, Wagner A. 2004. Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proc Biol Sci.* 271:89–96.
- DeLuna A, et al. 2008. Exposing the fitness contribution of duplicated genes. *Nat Genet.* 40:676–681.
- Gao LZ, Innan H. 2004. Very low gene duplication rate in the yeast genome. *Science.* 306:1367–1370.
- Gu Z, et al. 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature.* 421:63–66.
- Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH. 2008. Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148:993–1003.
- Hannay K, Marcotte EM, Vogel C. 2008. Buffering by gene duplicates: an analysis of molecular correlates and evolutionary conservation. *BMC Genomics.* 9:609.
- Kafri R, Dahan O, Levy J, Pilpel Y. 2008. Preferential protection of protein interaction network hubs in yeast: evolved functionality of genetic redundancy. *Proc Natl Acad Sci USA.* 105:1243–1248.
- Kafri R, Springer M, Pilpel Y. 2009. Genetic redundancy: new tricks for old genes. *Cell.* 136:389–392.
- Kuromori T, et al. 2006. A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of *Arabidopsis*. *Plant J.* 47:640–651.
- Liang H, Li WH. 2007. Gene essentiality, gene duplicability and protein connectivity in human and mouse. *Trends Genet.* 23:375–378.
- Liao BY, Zhang J. 2007. Mouse duplicate genes are as essential as singletons. *Trends Genet.* 23:378–381.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science.* 290:1151–1155.
- Makino T, Hokamp K, McLysaght A. 2009. The complex relationship of gene duplication and essentiality. *Trends Genet.* 25:152–155.
- Meinke DW, et al. 2003. A sequence-based map of *Arabidopsis* genes with mutant phenotypes. *Plant Physiol.* 131:409–418.
- Musso G, et al. 2008. The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. *Genome Res.* 18:1092–1099.
- Nowak MA, Boerlijst MC, Cooke J, Smith JM. 1997. Evolution of genetic redundancy. *Nature.* 388:167–171.
- Roux J, Robinson-Rechavi M. 2008. Developmental constraints on vertebrate genome evolution. *PLoS Genet.* 4:e1000311.
- Su Z, Gu X. 2008. Predicting the proportion of essential genes in mouse duplicates based on biased mouse knockout genes. *J Mol Evol.* 67:705–709.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- Tischler J, Lehner B, Chen N, Fraser AG. 2006. Combinatorial RNA interference in *Caenorhabditis elegans* reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. *Genome Biol.* 7:R69.
- van Dongen S. 2000. Graph clustering by flow simulation [PhD thesis]. [Utrecht (The Netherlands)]: University of Utrecht.
- Vavouri T, Semple JI, Lehner B. 2008. Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends Genet.* 24:485–488.
- Wagner A. 2000. The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. *Genetics.* 154:1389–1401.
- Woollard A. 2005. Gene duplications and genetic redundancy in *C. elegans*. *WormBook*, ed. The *C. elegans* Research Community, WormBook. p. 1–6.
- Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18:292–298.
- Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA.* 95:3708–3713.

Marta Wayne, Associate Editor

Accepted October 21, 2009