Functional Compensation of Primary and Secondary Metabolites by Duplicate Genes in *Arabidopsis thaliana*

Kousuke Hanada,*,†,[1,2,3] Yuji Sawada,†,[1,4] Takashi Kuromori,[1,4] Romy Klausnitzer,[1] Kazuki Saito,[1] Tetsuro Toyoda,[3] Kazuo Shinozaki,[1] Wen-Hsiung Li,*,[5,6] and Masami Yokota Hirai[1,4]

[1]Plant Science Center, RIKEN, Yokohama, Kanagawa, Japan

[2]Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Aichi, Japan

[3]Bioinformatics and Systems Engineering Division, RIKEN, Yokohama, Kanagawa, Japan

[4]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan

[5]Department of Ecology and Evolution, University of Chicago

[6]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

†These authors contributed equally to this work.

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

Associate editor: Takashi Gojobori

Abstract

It is well known that knocking out a gene in an organism often causes no phenotypic effect. One possible explanation is the existence of duplicate genes; that is, the effect of knocking out a gene is compensated by a duplicate copy. Another explanation is the existence of alternative pathways. In terms of metabolic products, the relative roles of the two mechanisms have been extensively studied in yeast but not in any multi-cellular organisms. Here, to address the functional compensation of metabolic products by duplicate genes, we quantified 35 metabolic products from 1,976 genes in knockout mutants of *Arabidopsis thaliana* by a high-throughput Liquid chromatography-Mass spectrometer (LC-MS) analysis. We found that knocking out either a singleton gene or a duplicate gene with distant paralogs in the genome tends to induce stronger metabolic effects than knocking out a duplicate gene with a close paralog in the genome, indicating that only duplicate genes with close paralogs play a significant role in functional compensation for metabolic products in *A. thaliana*. To extend the analysis, we examined metabolic products with either high or low connectivity in a metabolic network. We found that the compensatory role of duplicate genes is less important when the metabolite has a high connectivity, indicating that functional compensation by alternative pathways is common in the case of high connectivity. In conclusion, recently duplicated genes play an important role in the compensation of metabolic products only when the number of alternative pathways is small.

Key words: gene duplication, metabolic network, *Arabidopsis thaliana*, functional compensation, alternative pathway.

Introduction

Knocking out a gene in an organism often causes little phenotypic effect. One possible explanation is the existence of duplicate genes, where the effect of knocking out a gene is compensated by a duplicate copy. The role of functional compensation by duplicate genes has been examined in diverse organisms by comparing the proportion (P_E) of singleton genes whose single-gene knockout causes phenotypic effects to the P_E value for knocking out a duplicate gene (Wagner 2000; Gu et al. 2003; Conant and Wagner 2004; Liang and Li 2007; Liao and Zhang 2007; Hanada et al. 2009). If duplicate genes play a significant role in functional compensation, the P_E for duplicate genes should be lower than that of singleton genes. This is indeed the case in yeasts, worms, and plants (Gu et al. 2003; Conant and Wagner 2004; Hanada et al. 2009); in mice, no significant difference in P_E was found between singletons and duplicate genes in earlier studies (Liang and Li 2007; Liao and Zhang 2007) but was found when corrections for functional bias (Makino et al. 2009) and protein connectivity bias were made (Liang and Li 2009). The studies in mice indicated that the functional compensation by duplicate genes is affected by confounding factors. Therefore, it is of interest to study the role of network structure in the functional compensation by duplicate genes because the network structure of pathways may affect functional compensation.

In the budding yeast, the metabolic network has been used to study functional compensation by either duplicate genes or alternative pathways in knockout analysis (Papp et al. 2004; Blank et al. 2005; Kuepfer et al. 2005; Segre et al. 2005; Deutscher et al. 2006; Harrison et al. 2007; DeLuna et al. 2008; Wang and Zhang 2009). Both duplicate genes and alternative pathways contribute to functional compensation in the yeast metabolic network, but the relationship between duplicate genes and alternative pathways is poorly understood. Furthermore, knockout analysis has not been used to address functional compensation in the metabolic network of multi-cellular organisms. *Arabidopsis thaliana* is an excellent model organism for addressing these issues because many knockout mutants have
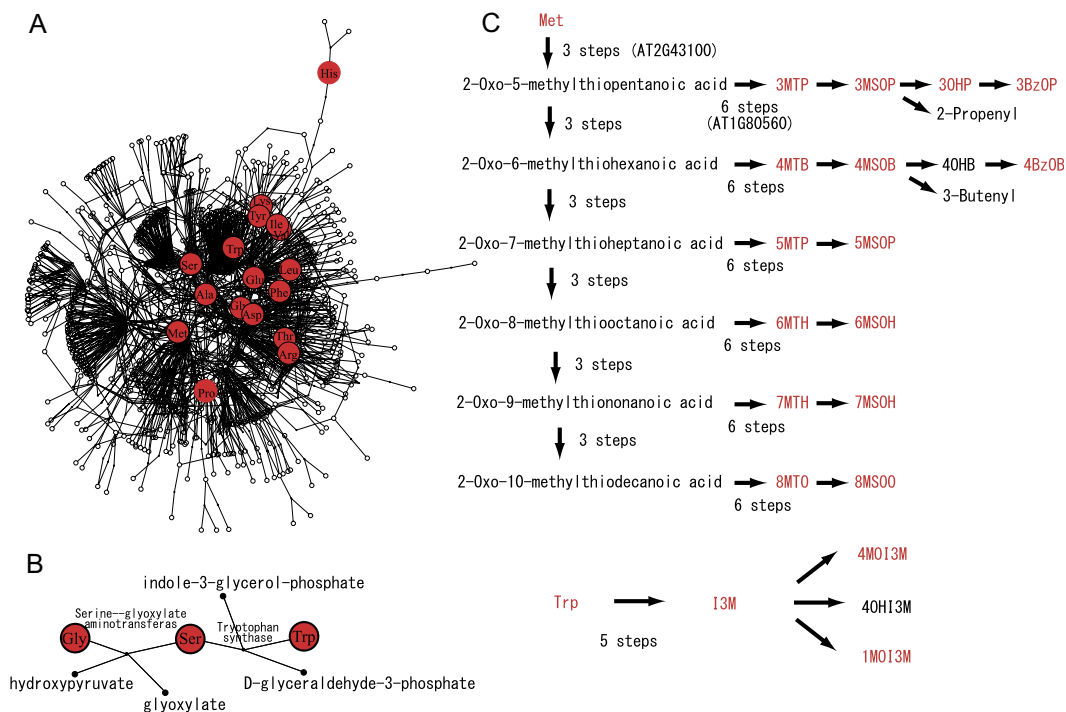
FIG. 1. Metabolic pathways in 35 metabolites in *Arabidopsis*. (*A*) Open circles and dots denote chemical compounds and enzyme reactions, respectively. The chemical compounds of 17 amino acids (red circles) were quantified in this study. All metabolic pathways synthesized by less than five enzyme reactions from the 17 amino acids are shown in this figure. All the pathways were collected from the TAIR Aracyc database (version 5.0). As small molecules, including water, O_2, ATP, NADP, NADPH, polyphosphates, CO_2, H^+, phosphates, ADP, UDP, Co-A, NAD, NADH, AMP, and Acetyl-CoA, are involved in many reactions or are used as carriers for transferring electrons, they were excluded from this map. (*B*) As an example of the structure of a metabolic pathway, two reactions as a minimum metabolic pathway from Gly to Trp are shown. (*C*) Metabolic pathway of 18 secondary metabolites in *Arabidopsis*. The chemical compounds shown in red were quantified in this study. Abbreviation of each metabolite is shown in supplementary table S1, Supplementary Material online. The metabolic pathways are modified from Hirai et al. (2007).

been generated and many metabolic networks have been identified (Mueller et al. 2003; Zhang et al. 2005). In this study, we focused on 35 metabolites, including 17 primary metabolites that are essential amino acids conserved in all organisms and 18 secondary metabolites that are produced specifically in *Brassicacease*, such as broccoli and cabbage (Fahey et al. 2001). Because the structures of the metabolic networks in the 17 primary and the 18 secondary metabolites tend to show high and low connectivity, respectively (supplementary table S1, Supplementary Material online, fig. 1A and C), primary metabolites seem to have higher functional compensation through alternative pathways than secondary metabolites. Therefore, it is of interest to examine the compensatory role of duplicate genes considering the two different network structures separately. In this study, we examined 35 types of metabolic production in the available knockout mutants of 1,976 genes in *A. thaliana* by a high-throughput Liquid chromatography-Mass spectrometer (LC-MS) analysis and the functional compensation of metabolites by either duplicate genes or alternative pathways.

Materials and Methods

Construction of Metabolic Networks in 17 Primary Metabolites and 18 Secondary Metabolites
Based on Aracyc 5.0 (ftp://ftp.plantcyc.org/), we generated an *Arabidopsis* metabolic network of 17 primary metabo-

lites (essential amino acids) (fig. 1A and supplementary table S1, Supplementary Material online). The reactions between metabolites occur via enzymes (fig. 1B). Metabolic pathways synthesized by less than five enzyme reactions of the 17 essential amino acids are shown in figure 1A. Genes encoding enzymes were annotated by Aracyc 5.0. Compounds of water, O_2, ATP, NADP, NADPH, polyphosphates, CO_2, H^+, phosphates, ADP, UDP, Co-A, NAD, NADH, AMP, and Acetyl-CoA, were excluded from the network because they are used as carriers for transferring electrons in many reactions (Ma and Zeng 2003). We also generated an *Arabidopsis* metabolic network of 18 secondary metabolites (glucosinolates) (Hirai et al. 2007) (fig. 1C and supplementary table S1, Supplementary Material online).

Quantification of 35 Metabolic Products in Knockout Mutants
From the Ds transposon single-copy insertion lines we established previously (Kuromori et al. 2006), we used 2,234 knockout mutants in which the Ds transposon was homozygously inserted into the coding regions of 1,976 genes. To avoid including pseudogenes and erroneously predicted genes, we focused on the 1,976 genes that were highly expressed in available microarray data (Schmid et al. 2005; Kilian et al. 2007; Goda et al. 2008). For each gene mutant,

seeds were harvested from an F3 individual plant described in Kuromori et al. (2006). A total of 200 seeds of each independent mutant were homogenized using a mixer mill MM 200 (Retsch) in 80 μl of extraction buffer (40% acetonitrile in H_2O with 25 μM hydroxyphenyl-glucosinolate and 50 μM norleucine as internal standards). The extracts were diluted with 500 μl of LC–MS grade H_2O and centrifuged (1,000 × g) for 5 min. The supernatants were filtered through CAPTIVA 0.45 μm filter (Varian) and subjected to Ultra Performance Liquid Chromatography (UPLC) Waters-quadrupole mass spectrometry analysis. Thirty-five metabolites were separated on UPLC through a reverse phase column (50 × 2.1 mm, HSS T3 1.8 μm; Waters) and detected using ZQ mass spectrometers (Waters). MassLynx software version 4.0 (Waters) was used to control all instruments and calculate peak areas. We did not perform an analytical replicate because the analytical error of this analysis using UPLC-quadrupole MS is <10%. All our mutants were originated from five parental strains (Kuromori et al. 2006). To infer knockout effect, the absolute value of peak area between a mutant and the parental strain (wild type) was calculated and normalized by the quantile normalization in 35 metabolites using the Bioconductor (www.bioconductor.org) Affy package in the R software environment (www.r-project.org). The normalized value was defined to represent knockout effect in a metabolite. In 258 genes, two mutant lines independently had an insertion in the same gene. The mean normalized values for mutants targeting the same gene were used to represent metabolic effect (M_E) by knocking out the gene. The M_E of each gene in 35 metabolites are shown in supplementary table S2, Supplementary Material online.

Identification of Duplicate Genes and Singletons

The amino acid sequences of *A. thaliana* (TAIR7) genes were obtained from TAIR (www.arabidopsis.org). Similarity searches against all genes were conducted using BlastP (Altschul et al. 1997). To find the closest paralog of a gene, we aligned the gene and all genes obtained in the Blast search by ClustalW (Thompson et al. 1994) and estimated the amino acid similarity between sequences. The gene with the highest similarity to the gene under study was defined as the closest paralog. A gene was defined as "duplicated" if it showed a ≥30%, ≥50%, ≥70%, or ≥90% similarity and alignable region at the protein level with the closest paralog. A singleton protein was defined as a protein that did not match any other proteins in the Blast search with the E value ≥0.01. Out of the 1,976 genes, 123, 379, 602, 633, and 95 genes are singletons and duplicate genes with ≥30%, ≥50%, ≥70%, and ≥90% similarity and coverage, respectively. A total of 144 genes are not classified as singleton or duplicate genes.

Results and Discussion

Validation of Metabolic Effects in the Present Data

To determine whether the present data captured the metabolic production by knocking out genes, we searched for enzymatic genes related to any production of the 35 metabolites based on the AracyC database (ftp://ftp.plantcyc.org/). Our data set had 261 enzymatic genes. A total of 15 enzymatic genes are linked to the production of each of the 17 primary metabolites (essential amino acids) via one-step enzyme reaction (supplementary table S3, Supplementary Material online). To examine the production change of each metabolite by knocking out one of these enzymatic genes, we compared the M_E of each primary metabolite targeted by 15 genes with the M_E of the primary metabolites by knocking out other 246 enzymatic genes. The M_E by knocking out 15 genes related to primary metabolites was significantly higher than the M_E by knocking out other enzymatic genes ($P < 0.05$ by the Wilcoxon test, supplementary fig. S1A, Supplementary Material online). For 18 secondary metabolites (glucosinolates derived from either methionine or tryptophan) (Hirai et al. 2007), our data set had only two known enzymatic genes related to the production of the secondary metabolites (fig. 1C). We compared the M_E of each secondary metabolite targeted by the two genes with the M_E of secondary metabolites by knocking out other 261 enzymatic genes and found that the M_E in the two genes was significantly larger than the M_E in the other 261 enzymatic genes ($P < 0.05$ by the Wilcoxon test, supplementary fig. S1B, Supplementary Material online). Although 15 (primary metabolites) and 2 (secondary metabolites) genes may not be enough to represent the correct trend of metabolic effect by knocking out enzymatic genes, our metabolic profile encompassed the metabolic effect expected by known enzymatic reactions.

Sequence Divergences in Genes Whose Knockout Mutants Cause Changes in Metabolic Production

Because not only enzymatic genes but also other genes are related to metabolic products, we examined the metabolic products by individually knocking out as many genes as possible. Using our available knockout mutants of 1,976 genes, we compared the compensatory role by singleton genes with that by duplicate genes in 35 metabolic products. If duplicate genes contribute to functional compensation in metabolites, the overall metabolic effect of knocking out duplicate genes should be lower than that by knocking out singleton genes. However, duplicate genes have different similarities and alignable lengths (coverages) with the closest paralog in the genome. Based on the similarity and coverage at the protein level (see Materials and Methods), we generated four types of duplicate genes: with ≥30%, ≥50%, ≥70%, or ≥90% similarity and coverage to the closest paralog, and we compared the summed M_E in 35 metabolites between singletons and each group of duplicate genes (fig. 2). Although there was no significant difference between singleton and duplicate genes with $a \geq$ 30%, 50%, or 70% protein sequence similarity and coverage to the closest paralog ($P > 0.05$ by the Wilcoxon test), the duplicate genes with $a \geq 90$% protein sequence similarity and coverage with the closet paralog showed a significantly lower metabolic effect than singleton genes ($P < 0.05$ by the Wilcoxon test). These results indicate that in
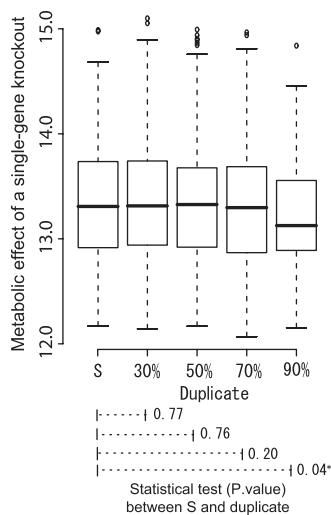
FIG. 2. Metabolic effect of a single-gene knockout in singletons and duplicate genes. The metabolic effect (M_E) of each metabolite in a gene is the normalized absolute value of peak area between a mutant and the parental strain. As the overall metabolic effect by knocking out a gene, the log-scale of summed M_E in 35 metabolites is represented in the Y axis. Genes are classified into singletons (S) and duplicate genes (D). Duplicate genes are classified into ≥30%, ≥50%, ≥70%, and ≥90% duplicate gene groups based on the sequence similarity and coverage at the protein level with the closest paralog (see Materials and Methods). The distributions of summed M_E values are shown as box plots with the solid horizontal line indicating the median value. The box represents the inter quartile range (25–75%), and the dotted line indicates the first to the 99th percentile. The summed M_E is compared between singletons and duplicate genes. The group of ≥90% duplicate genes has a significantly lower M_E than the group of singleton genes ($P <$ 0.05), but other groups of duplicate genes do not have different M_E than the group of singleton genes ($P >$ 0.05).

Arabidopsis only duplicate genes with a highly similar paralog in the genome play a significant role in functional compensation of metabolites. Therefore, duplicate genes categorized by ≥90% were used in the following analyses.

Compensation of Duplicate Genes Associated with Multiple Metabolic Changes

In functional compensations of phenotypic changes in *Arabidopsis*, a more severe phenotypic effect tends to be better compensated by gene duplication than a less severe effect (Hanada et al. 2009). To address whether the same phenomenon is observed in metabolic changes or not, we examined the compensational role of duplicate genes in metabolic changes. The metabolic severity was measured by the number of metabolic changes. To quantify the number of metabolic changes, we defined the threshold of significant metabolic changes at the top and bottom 5% peak areas for each metabolite. Among the 1,976 genes, 790 knocked out genes did not cause any changes in metabolic production. The remaining 1,186 genes were classified into 734 and 452 genes that caused changes in a small (1–2) and a large (≥3) number of metabolic productions, respectively. Genes in each group were classified into singleton and duplicate genes. Table 1 shows that genes

Table 1. Numbers of Duplicate Genes and Singletons with Knockout Effects on 35 Metabolic Productions.

Number of Metabolic Changes	Number of Singletons (S)	Number of Duplicates (D)	S/D Ratio	P value (0 vs. 1–2 or ≥3)
0	47	49	0.96	
1–2	50	33	1.52	0.040
≥3	26	13	2.00	0.027

whose knockout mutants caused either small or large number of metabolic changes have a significantly higher ratio of singleton to duplicate genes than the ratio for genes whose knockout mutants did not cause any change in the 35 metabolites under study ($P <$ 0.05 by the chi-square test). The ratio of singletons to duplication genes increases as the number of metabolic changes increases (table 1), indicating that functions associated with changes of multiple metabolic products tend to be compensated by gene duplication.

Functional Compensation by Duplicate Genes in Metabolic Networks versus Number of Alternative Pathways

As shown in fig. 1, the synthesis of a primary metabolite has many alternative pathways, whereas the synthesis of a secondary metabolite has few. The compensatory role of duplicate genes for metabolites may depend on the number of alternative pathways. If alternative pathways contribute to the compensatory role in metabolic production, the proportion of duplicate genes causing changes in metabolites with few alternative pathways should be higher than the proportion of genes causing changes in a metabolite with many alternative pathways.

We compared the functional compensation by duplicate genes among different metabolites. Among the 1,976 genes under study, 1,240 and 1,152 knocked out genes did not cause any changes in primary and secondary metabolic production, respectively. The remaining genes that caused changes in primary or secondary metabolites were classified into two groups: those inducing a small number (1–2) of metabolic changes and those inducing a large number (≥3) of metabolic changes. Genes in each group were classified into singletons and duplicate genes. To examine the functional compensation by either duplicate genes or alternative pathways among metabolites, we examined the ratio of singletons to duplicate genes in each group (tables 2 and 3). The ratio was not significantly different between genes causing no metabolite change and genes causing a small or a large number of changes in primary metabolic production (table 2). In contrast, the ratio for genes causing a large number of changes in secondary metabolic production was significantly larger than that for the genes that caused no secondary metabolite changes (table 3, $P <$ 0.05 by the chi-square test). These results indicate that duplicate genes have a higher probability of functional compensation for changes in secondary metabolic production, and duplicate genes have a lower

Table 2. Numbers of Duplicate Genes and Singletons with Knockout Effects on Primary Metabolic Productions.

Number of Metabolic Changes	Number of Singletons (S)	Number of Duplicates (D)	S/D Ratio	P value (0 vs. 1–2 or ≥3)
0	81	66	1.23	
1–2	34	20	1.70	0.25
≥3	8	9	0.90	0.51

FIG. 3. Relationship between connectivity and the ratio of duplicate and singleton genes with knockout effects on metabolic productions. Each dot represents a metabolite. The log-scaled number of metabolic pathways is represented in X axis. To quantify the number of metabolic changes, we defined the threshold of significant metabolic changes at the top and bottom 5% peak areas for each metabolite. By the threshold, genes whose knockout mutant induced metabolic changes are identified in 35 metabolites. Identified genes are classified into singleton and duplicate genes categorized by ≥90%. The ratio of duplicate to singleton genes is represented in Y axis. The connectivity is positively correlated with the ratio of duplicate to singleton genes (correlation coefficient = 0.4, P = 0.01).
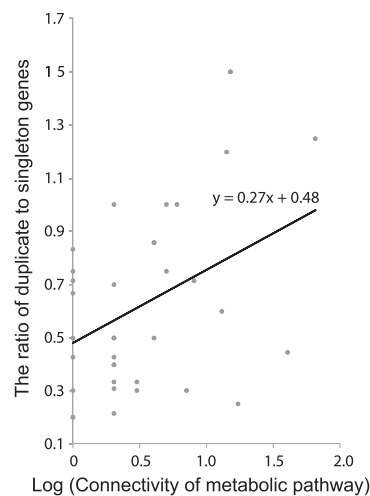
probability of functional compensation for changes in primary metabolic production. It seems that functional compensation by gene duplication is not important for changes of primary metabolites because of the existence of many alternative pathways. These results indicate that alternative pathways play an important role in functional compensation for primary metabolic production in *A. thaliana*.

Although we found that overall metabolic changes tend to be compensated by only a duplicate gene with a close paralog in the genome (fig. 1 and table 1), we now recognized that the metabolites compensated by gene duplication are mainly secondary metabolites. Therefore, functional compensation may be observed by more distant paralog once we focused on only secondary metabolites. We then performed the same analysis using duplicate genes categorized by ≥30%, ≥50%, and ≥70% in secondary metabolites (supplementary table S4, Supplementary Material online). Although the ratio of singletons to duplication genes increases as the number of secondary metabolic changes increases, genes whose knockout mutants caused either a small or a large number of metabolic changes do not have any significantly different ratios compared with the ratio for genes whose knockout mutants did not cause any change in the 35 metabolites. These results indicate that in *Arabidopsis*, only duplicate genes with a highly similar paralog in the genome play a significant role in functional compensation of secondary metabolites.

Without the classification of primary and secondary metabolites, we furthermore examined the compensational role of duplicate genes in metabolic changes. In each metabolite, we identified genes whose knockout mutant induced metabolic changes. Identified genes were classified into singleton and duplicate genes. Assuming that connectivity (the number of metabolic pathways) is an indicator of alternative pathways, we examined the relationship between connectivity and the ratio of duplicate to singleton genes in 35 metabolites. Then, the connectivity is positively correlated with the proportion of duplicate genes (correlation coefficient = 0.4, P = 0.01, fig. 3). The results strongly

suggest that duplicate genes tend to have functional compensation in the case of low connectivity and duplicate genes tend not to have functional compensation in the case of high connectivity. Thus, by the analysis without any classification of metabolites, we found that the compensatory role of duplicate genes is more important when the metabolite has a low connectivity.

Concluding Remarks

We found that both duplicate genes and alternative pathways contribute to the genetic robustness of metabolic effects in *A. thaliana*. In particular, gene duplication is important for the compensation of functions associated with multiple metabolic products but is less important for the compensation of metabolites with many alternative pathways.

In functional compensations of phenotypic changes in *Arabidopsis*, a more severe phenotypic effect tends to be better compensated by gene duplication than a less severe effect (Hanada et al. 2009). In this study, functions associated with changes of multiple metabolic products tend to be compensated by gene duplication (table 1). Therefore, the persistence of functional compensation for a more significant effect is supported in metabolic changes as well. However, duplicate genes play a more important role for the compensation of secondary metabolites, but a less important role for the compensation of primary metabolites (table 2), although primary metabolites are likely to be more essential than secondary metabolites for plant

Table 3. Numbers of Duplicate Genes and Singletons with Knockout Effects on Secondary Metabolic Productions.

Number of Metabolic Changes	Number of Singletons (S)	Number of Duplicates (D)	S/D Ratio	P value (0 vs. 1–2 or ≥3)
0	69	64	1.08	
1–2	35	25	1.40	0.317
≥3	19	6	3.17	0.016

survival. This inconsistency can be explained by the existence of many alternative pathways in primary metabolites. Another interesting finding is that only duplicate genes with a highly similar paralog showed functional compensation in metabolites. One explanation for this finding is that functional compensation of metabolic changes may be quickly lost in evolution. However, functional compensation by gene duplication is mainly in secondary metabolites. Because secondary metabolites, glucosinolates, are produced in only species close to *Arabidopsis*, most genes related to the production of secondary metabolites seem to have appeared recently in evolution. Indeed, it has been reported that many genes related to secondary metabolites were expanded recently by gene duplication (Hanada et al. 2008). Taken together, it is likely that functionally overlapping genes associated with the production of secondary metabolites were expanded recently by gene duplication in the *Arabidopsis* lineage.

Supplementary Material

Supplementary tables S1–S4 and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe. oxfordjournals.org/).

Acknowledgments

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.

Blank LM, Kuepfer L, Sauer U. 2005. Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol.* 6:R49.

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, Vetsigian K, Shoresh N, Hegreness M, Colon-Gonzalez M, Chao S, Kishony R. 2008. Exposing the fitness contribution of duplicated genes. *Nat Genet.* 40:676–681.

Deutscher D, Meilijson I, Kupiec M, Ruppin E. 2006. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat Genet.* 38:993–998.

Fahey JW, Zalcmann AT, Talalay P. 2001. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56:5–51.

Goda H, Sasaki E, Akiyama K, et al. (28 co-authors). 2008. The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J.* 55:526–542.

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

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

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.

Harrison R, Papp B, Pal C, Oliver SG, Delneri D. 2007. Plasticity of genetic interactions in metabolic networks of yeast. *Proc Natl Acad Sci U S A.* 104:2307–2312.

Hirai MY, Sugiyama K, Sawada Y, et al. (14 co-authors). 2007. Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc Natl Acad Sci U S A.* 104:6478–6483.

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

Kuepfer L, Sauer U, Blank LM. 2005. Metabolic functions of duplicate genes in Saccharomyces cerevisiae. *Genome Res.* 15:1421–1430.

Kuromori T, Wada T, Kamiya A, et al. (12 co-authors). 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.

Liang H, Li WH. 2009. Functional compensation by duplicated genes in mouse. *Trends Genet.* 25:441–442.

Liao BY, Zhang J. 2007. Mouse duplicate genes are as essential as singletons. *Trends Genet.* 23:378–381.

Ma H, Zeng AP. 2003. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 19:270–277.

Makino T, Hokamp K, McLysaght A. 2009. The complex relationship of gene duplication and essentiality. *Trends Genet.* 25:152–155.

Mueller LA, Zhang P, Rhee SY. 2003. AraCyc: a biochemical pathway database for Arabidopsis. *Plant Physiol.* 132:453–460.

Papp B, Pal C, Hurst LD. 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429:661–664.

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of Arabidopsis thaliana development. *Nat Genet.* 37:501–506.

Segre D, Deluna A, Church GM, Kishony R. 2005. Modular epistasis in yeast metabolism. *Nat Genet.* 37:77–83.

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.

Wagner A. 2000. Robustness against mutations in genetic networks of yeast. *Nat Genet.* 24:355–361.

Wang Z, Zhang J. 2009. Abundant indispensable redundancies in cellular metabolic networks. *Genome Biol Evol.* 2009:23–33.

Zhang P, Foerster H, Tissier CP, Mueller L, Paley S, Karp PD, Rhee SY. 2005. MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiol.* 138:27–37.