



Contribution of Functional Divergence Through Copy Number Variations to the Inter-Species and Intra-Species Diversity in Specialized Metabolites

Kazumasa Shirai and Kousuke Hanada*

Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Fukuoka, Japan

OPEN ACCESS

Edited by:

Takayuki Tohge,
Nara Institute of Science and
Technology (NAIST), Japan

Reviewed by:

Dirk Walther,
Max Planck Institute of Molecular
Plant Physiology, Germany
Qing Liu,
Commonwealth Scientific and
Industrial Research Organisation
(CSIRO), Australia

*Correspondence:

Kousuke Hanada
kohanada@bio.kyutech.ac.jp

Specialty section:

This article was submitted to
Plant Metabolism
and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 18 July 2019

Accepted: 08 November 2019

Published: 26 November 2019

Citation:

Shirai K and Hanada K (2019)
Contribution of Functional Divergence
Through Copy Number Variations to
the Inter-Species and Intra-Species
Diversity in Specialized Metabolites.
Front. Plant Sci. 10:1567.
doi: 10.3389/fpls.2019.01567

There is considerable diversity in the specialized metabolites within a single plant species (intra-species) and among different plant species (inter-species). The functional divergence associated with gene duplications largely contributes to the inter-species diversity in the specialized metabolites, whereas the intra-species diversity is due to gene dosage changes *via* gene duplications [i.e., copy number variants (CNVs)] at the intra-species level of evolution. This is because CNVs are thought to undergo associated with less functional divergence at the intra-species level of evolution. However, functional divergence caused by CNVs may induce specialized metabolite diversity at the intra-species and inter-species levels of evolution. We herein discuss the functional divergence of CNVs in metabolic quantitative trait genes (mQTGs). We focused on 5,654 previously identified mQTGs in 270 *Arabidopsis thaliana* accessions. The ratio of nonsynonymous to synonymous variations tends to be higher for mQTGs with CNVs than for mQTGs without CNVs within *A. thaliana* accessions, suggesting that CNVs are responsible for the functional divergence among mQTGs at the intra-species level of evolution. To evaluate the contribution of CNVs to inter-species diversity, we calculated the ratio of nonsynonymous to synonymous substitutions in the *Arabidopsis* lineage. The ratio tends to be higher for the mQTGs with CNVs than for the mQTGs without CNVs. Additionally, we determined that mQTGs with CNVs are subject to positive selection in the *Arabidopsis* lineage. Our data suggest that CNVs are closely related to functional divergence contributing to adaptations *via* the production of diverse specialized metabolites at the intra-species and inter-species levels of evolution.

Keywords: specialized metabolite, adaptation, *Arabidopsis*, copy number variant, gene duplication

INTRODUCTION

Plants produce various specialized metabolites, the diversity of which is closely related to adaptive evolution (Pichersky and Lewinsohn, 2011). Specialized metabolites vary among different species as well as within single species (Chan et al., 2010; Weigel, 2012; Carreno-Quintero et al., 2013; Alseekh et al., 2015; Matsuda et al., 2015). The diversity of the specialized metabolites resulted from gene duplications among various plant species. We previously revealed that copy number variants

(CNVs) derived from gene duplications are associated with specialized metabolites (Shirai et al., 2017).

Gene duplications contribute to the diversity in specialized metabolites because of two possible effects. The first effect is functional divergence. After gene duplication events, the copied genes tend to accumulate nonsynonymous mutations because of relaxed selection pressures (Scannell and Wolfe, 2008). Consequently, the copied genes induce functional divergence (Ohno, 1970), ultimately leading to the variability in the specialized metabolites among various plant species (Hanada et al., 2008; Kliebenstein and Osbourn, 2012; Panchy et al., 2016). The second effect involves gene dosage changes. Specifically, gene duplications increase gene dosage (Ohno, 1970). In particular, CNVs are believed to be the main cause of intra-species gene dosage changes (Zmienko et al., 2014). There is experimental evidence that the abundance of specialized metabolites within a single species is critically controlled by altered gene dosages due to CNVs (Kliebenstein, 2001). However, it remains unclear whether CNVs associated with specialized metabolites tend to induce functional divergence at the genomic scale.

We herein discuss the functional divergence of CNVs associated with specialized metabolites. For this discussion, we performed additional analyses involving our previously published data. On the basis of the analyses, we propose that CNVs induce functional divergence that generates various specialized metabolites during the evolution of *A. thaliana*.

Functional Divergence of CNVs at the Intra-Species Level of Evolution

It is believed that CNVs mainly cause quantitative changes rather than qualitative changes (Zmienko et al., 2014), likely because of an insufficient amount of time for CNVs to accumulate nonsynonymous mutations leading to the diversity in specialized metabolites. However, several studies have identified a few nonsynonymous mutations responsible for the functional divergence of genes related to specialized metabolites (Chye et al., 2000; Yu et al., 2015; Bunsupa et al., 2016). These reports suggest CNVs may induce functional divergence.

To examine the functional divergence of duplicated genes, the selection pressure based on the ratio between the nonsynonymous mutation/substitution rate and the synonymous mutation/substitution rate is useful (Hanada et al., 2009). High and low selection pressures are associated with functional divergence and constraint, respectively. Therefore, we estimated the $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratio, which is the ratio between the number of nonsynonymous mutations relative to the number of nonsynonymous sites (dN_{SNP}) and the number of synonymous mutations relative to the number of synonymous sites (dS_{SNP}) (Nei and Gojoborit, 1986; Hanada et al., 2009), for 27,130 annotated protein-coding genes in 270 *A. thaliana* accessions.

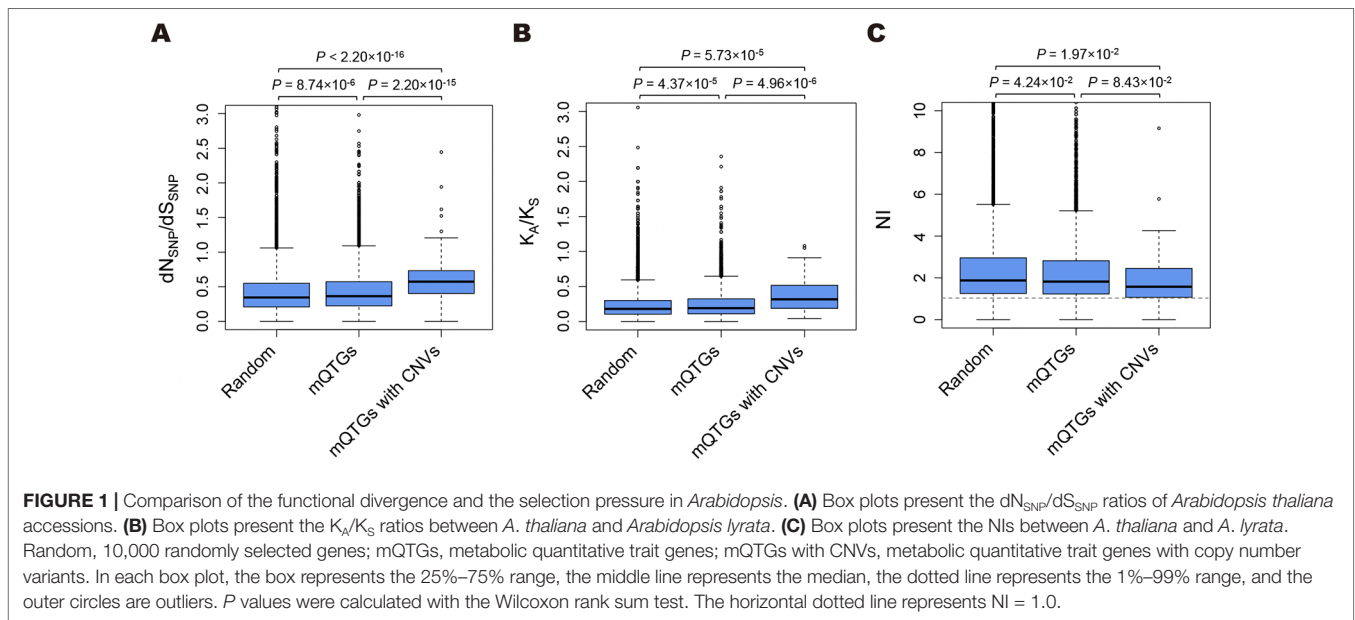
The single nucleotide polymorphism (SNP) data examined in this study were compiled from 270 *A. thaliana* accessions analyzed

in several studies (<http://1001genomes.org>, 1001 Genomes; Mouille et al., 2006; Cao et al., 2011; Gan et al., 2011; Schmitz et al., 2013; Shirai et al., 2017). A total of 7,624,270 SNPs were included. For each accession, nonsynonymous and synonymous variations were annotated according to the TAIR10 database with the SnpEff program (<https://www.arabidopsis.org>; Cingolani et al., 2012). Of the 7,624,270 SNPs, 1,330,920 were located in 27,130 annotated protein-coding genes in the reference *A. thaliana* genome. For each of the 270 accessions, the 1,330,920 SNPs were classified as 733,796 nonsynonymous and 597,124 synonymous mutations in the 27,130 coding sequences. There was an average of 27 nonsynonymous and 22 synonymous mutations in each of the 27,130 coding sequences. Because the number of nonsynonymous and synonymous sites in codons varied, we calculated the number of synonymous and nonsynonymous sites in all 27,130 coding sequences with scripts that we developed following the Nei-Gojobori method (Nei and Gojoborit, 1986; Hanada et al., 2009). The 27,130 genes were classified as metabolic quantitative trait genes (mQTGs) and mQTGs with CNVs. We previously predicted 5,654 mQTGs for 1,335 specialized metabolites in *A. thaliana* (Shirai et al., 2017). In that study, mQTGs were detected by combining a genome-wide association study (GWAS) and a metabolite-transcriptome correlation analysis (MTCA). This method enabled the prediction of mQTGs with a lower false positive rate than that of the general GWAS method. Genes with CNVs were previously detected by comparing genomic read counts among *A. thaliana* accessions (Gan et al., 2011). Of the 27,130 genes, 929 were predicted as genes with CNVs ($P < 0.05$).

To assess whether the functional divergence of CNVs is associated with the diversity in specialized metabolites, we compared the $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratios among mQTGs, mQTGs with CNVs, and randomly selected genes (**Figure 1A** and **Supplementary Table S1**). The $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratios were significantly higher for the mQTGs and mQTGs with CNVs than for the randomly selected genes (Wilcoxon rank sum test: $P < 0.001$; **Figure 1A**). Moreover, the ratios of mQTGs with CNVs were also significantly higher than the ratios of mQTGs (Wilcoxon rank sum test: $P < 0.001$; **Figure 1A**). These results imply that nonsynonymous variations tend to accumulate in mQTGs more frequently than in genes not associated with specialized metabolites. Specifically, mutations that alter the amino acid sequence accumulated in mQTGs with CNVs at a higher rate than in mQTGs without CNVs. These findings suggest that the diversity in specialized metabolites due to CNVs is the result of the functional divergence of mQTGs in addition to gene dosage changes at the genomic scale. Additionally, mQTGs with CNVs tended to be associated with a larger number of specialized metabolites than mQTGs without CNVs (Wilcoxon rank sum test: $P = 1.92 \times 10^{-3}$; **Supplementary Figure S1**), implying that the functional divergence derived from CNVs enhances the divergence of specialized metabolites.

It was unclear whether CNVs induce functional divergence for mQTGs only or for other genes as well. Therefore, we compared the $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratios of randomly selected genes and non-mQTGs with CNVs. The $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratios were significantly higher for the non-mQTGs with CNVs than for the randomly selected genes (Wilcoxon rank sum test: $P < 2.2 \times$

Abbreviations: CNV, copy number variant; mQTG, metabolic quantitative trait gene; NI, neutrality index; PacBio, Pacific Bioscience; SNP, single nucleotide polymorphism.



10^{-16} ; **Supplementary Figure S2**). Thus, CNVs are generally responsible for the functional divergence of genes at the intra-species level of evolution.

Functional Divergence of CNVs at the Inter-Species Level of Evolution

It was recently reported that CNVs are associated with various phenotypic differences within a plant species (Lye and Purugganan, 2019). By contrast, the contribution of CNVs to inter-species diversity remains relatively unknown in plants.

The $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratio indicates the intra-species level of evolution. Therefore, to characterize the functional divergence of CNVs at the inter-species level of evolution, we examined the K_A/K_S ratio, which is the ratio between the number of nonsynonymous substitutions relative to the number of nonsynonymous sites (K_A) and the number of synonymous substitutions relative to the number of synonymous sites (K_S). The K_A/K_S ratio was estimated for 20,498 orthologs between *A. thaliana* and *Arabidopsis lyrata*. These orthologs were detected based on the reciprocal best hit (E-value < 1.0×10^{-3} and coverage > 90%) of a BLASTP (version 2.8.1) analysis of *A. thaliana* and *A. lyrata* (<https://www.arabidopsis.org>, TAIR10; <http://genome.jgi.doe.gov>, Phytozome v12: Alyrata_384_v2.1; Rawat et al., 2015; Boratyn et al., 2013). The coding sequences were aligned according to the amino acid sequences aligned by MAFFT (version 7.407) (Katoh and Standley, 2013). To evaluate the functional divergence between *A. thaliana* and *A. lyrata*, the nonsynonymous and synonymous substitutions in the 20,498 orthologs were counted. The K_A/K_S ratio was calculated according to Yang and Nielsen's method in the “yn00” program of PAML (version 4.8a) (Yang and Nielsen, 2000; Yang, 2007).

We compared the K_A/K_S ratios of mQTGs, mQTGs with CNVs, and randomly selected genes (**Figure 1B** and **Supplementary Table S1**). The mQTGs were found to have significantly higher K_A/K_S ratios than the randomly selected genes (Wilcoxon rank

sum test: $P < 0.001$; **Figure 1B**), indicating that functional divergence was more commonly detected for mQTGs than for the other genes. Additionally, the K_A/K_S ratios were higher for mQTGs with CNVs than for mQTGs and randomly selected genes (Wilcoxon rank sum test: $P < 0.001$; **Figure 1B**), suggesting that CNVs enhanced the functional divergence of mQTGs between *A. thaliana* and *A. lyrata*.

Selection Pressure for CNVs in a Species Lineage

A strong positive selection decreases the nucleotide diversity around target sites throughout the genome (i.e., selective sweep). The mQTGs with CNVs are more frequently affected by a selective sweep than the other genes in *A. thaliana* accessions (Shirai et al., 2017). This suggests that CNVs contribute to local adaptations at the intra-species level of evolution. The results of the present study suggest that CNVs contribute to the functional divergence of mQTGs at the inter-species and intra-species levels. However, it remains unclear whether positive or relaxed selection pressure controls mQTGs with CNVs at the inter-species level of evolution.

In earlier investigations, determining the selection pressure generally involved comparisons between variations at the inter-species and intra-species levels of evolution (McDonald and Kreitman, 1991; Rand and Kann, 1996; Smith and Eyre-Walker, 2002; Stoletzki and Eyre-Walker, 2011). These studies compared the number of nonsynonymous mutations (P_n), the number of synonymous mutations (P_s), the number of nonsynonymous substitutions (D_n), and the number of synonymous substitutions (D_s). The neutrality index (NI; i.e., $P_{n/s}/D_{n/s}$) is one of the parameters for comparing the variations and inferring the selection pressure (Rand and Kann, 1996). The NI quantifies the direction and extent of the difference from neutrality in which $P_{n/s}$ equals $D_{n/s}$. That is, an NI of 1 means the intra-species and inter-species functional divergences are the same. Additionally,

NI < 1 and NI > 1 reflect greater inter-species and intra-species functional divergences, respectively. Moreover, NI < 1 and NI > 1 represent the effects of positive and negative selection, respectively. We calculated the NI based on the variations of mQTGs with CNVs within *A. thaliana* accessions (intra-species) and between *A. thaliana* and *A. lyrata* (inter-species) among 20,214 genes. The P_n and P_s were estimated according to the SNPs of the 270 accessions (dN_{SNP}/dS_{SNP} calculation). The D_n and D_s were estimated based on the substitutions of the orthologs between *A. thaliana* and *A. lyrata* (K_A/K_S calculation).

We found that mQTGs and mQTGs with CNVs tend to have a lower NI than the randomly selected genes (Wilcoxon rank sum test: $P < 0.05$; **Figure 1C** and **Supplementary Table S1**). These results indicate that mQTGs and mQTGs with CNVs enhanced the inter-species functional divergence over the intra-species functional divergence. To address whether mQTGs with CNVs are associated with positive selection due to functional divergence, we examined the proportion of mQTGs with CNVs in positively selected genes and in other genes. We defined positively selected genes as a gene with NI < 1 and a significant difference between $P_{n/s}$ and $D_{n/s}$ (false discovery rate < 0.05 according to the chi-squared test; **Supplementary Table S1**). The proportion of mQTGs with CNVs (0.37% = 4/1,076) was significantly higher for positively selected genes than for the other genes (0.22% = 45/19,138) (chi-squared test: $P = 2.42 \times 10^{-49}$; **Supplementary Table S2**). These results imply that CNVs tend to be contained in the mQTGs related to the adaptive evolution of *A. thaliana*.

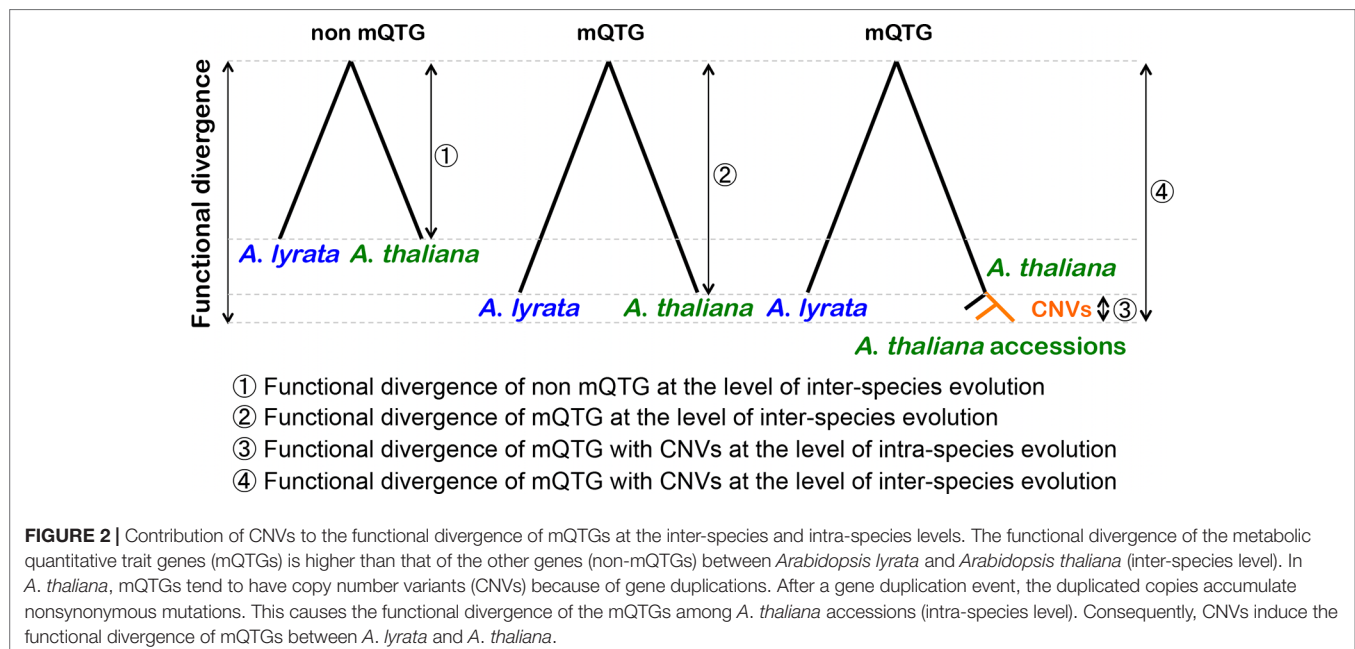
The NI reportedly leads to the incorrect determination of natural selection when there is an insufficient number of substitutions and mutations (Stoletzki and Eyre-Walker, 2011). Therefore, we validated the inferred selection pressure based on the direction of selection (DoS) (Stoletzki and Eyre-Walker, 2011). The DoS was defined as $D_n/(D_n + D_s) - P_n/(P_n + P_s)$. Additionally, DoS > 0 and DoS < 0 represent the effect of positive and negative

selection, respectively. We defined positively selected genes as genes with DoS > 0 and a significant difference between $P_{n/s}$ and $D_{n/s}$ (false discovery rate < 0.05 according to the chi-squared test; **Supplementary Table S1**). We examined the proportion of mQTGs with CNVs in positively selected genes and in other genes. Similar to the results of our analyses of NI, the proportion of mQTGs with CNVs (0.37% = 4/1,090) was significantly higher for positively selected genes than for the other genes (0.23% = 46/19,440) (chi-squared test: $P = 6.12 \times 10^{-46}$; **Supplementary Table S3**). Thus, the DoS analysis supported the NI results.

Conclusion and Perspectives

The current study examined the relationship between CNVs and the functional divergence of mQTGs at the inter-species and intra-species levels of evolution (**Figure 2**). Gene duplications induce nonsynonymous mutations *via* relaxed selection pressures. The CNVs derived from gene duplications seem to have accelerated nonsynonymous mutations. Thus, the mQTGs with CNVs have a high functional divergence at the intra-species level of evolution. Additionally, this intra-species functional divergence increases the inter-species functional divergence of the mQTGs. In fact, the functional divergence of mQTGs with CNVs tends to be high between *A. thaliana* and *A. lyrata*. Therefore, CNVs contribute to the functional divergence related to the diversity in specialized metabolites at the inter-species and intra-species levels. Consequently, CNVs tend to contribute to adaptations at the inter-species and intra-species levels. We propose that CNV is an important adaptive mechanism for generating diverse specialized metabolites in plants.

Our analyses are based on SNP calling with short-read sequencing. When SNPs are predicted in genes with CNVs based on the short reads, the SNPs are detected in the representative sequence of copied genes. The SNP detection over- or



under-estimates the number of SNPs depending on the number of copied genes. In this study, we focused on only the rate of nonsynonymous or synonymous mutations. It is unlikely that the miscalling of SNPs between nonsynonymous and synonymous mutations is biased. Therefore, we believe that the effect of miscalling is limited for our analyses.

In the past 10 years, short-read sequencing has mainly been applied in investigations at the genome scale. Unfortunately, detecting structural variants is difficult based on short-read sequencing (van Dijk et al., 2018). Therefore, there have been relatively few studies on the CNVs in plants. However, third-generation sequencing platforms, such as Pacific Bioscience (PacBio), that can generate long reads (> 5 kb) have recently been applied for plant genomic research (Zhang et al., 2016; Fukushima et al., 2017; Lan et al., 2017; Baek et al., 2018; Edger et al., 2019). The long-read sequencing data may enable the accurate detection of structural variants (Jiao and Schneeberger, 2017; van Dijk et al., 2018). For example, structural variants were recently detected by PacBio in a tropical maize inbred line (Yang et al., 2019). If this experimental approach becomes more affordable, CNVs in plants may be more easily detected. Therefore, in the near future, it will be possible to verify conclusions in other plant species.

DATA AVAILABILITY STATEMENT

The datasets for this study are available in the 1001 Genomes (<http://1001genomes.org>), TAIR10 (<https://www.arabidopsis.org>), and Phytozome v12 (<http://genome.jgi.doe.gov>) databases.

REFERENCES

- Alseikh, S., Tohge, T., Wendenberg, R., Scossa, F., Omranian, N., Li, J., et al. (2015). Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell* 27, 485–512. doi: 10.1105/tpc.114.132266
- Baek, S., Choi, K., Kim, G. B., Yu, H. J., Cho, A., Jang, H., et al. (2018). Draft genome sequence of wild *Prunus yedoensis* reveals massive inter-specific hybridization between sympatric flowering cherries. *Genome Biol.* 19, 1–17. doi: 10.1186/s13059-018-1497-y
- Bunsuma, S., Hanada, K., Maruyama, A., Aoyagi, K., Komatsu, K., Ueno, H., et al. (2016). Molecular evolution and functional characterization of a bifunctional decarboxylase involved in lycopodium alkaloid biosynthesis. *Plant Physiol.* 171, 2432–2444. doi: 10.1104/pp.16.00639
- Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., et al. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* 43, 956–963. doi: 10.1038/ng.911
- Carreno-Quintero, N., Bouwmeester, H. J., and Keurentjes, J. J. B. (2013). Genetic analysis of metabolome-phenotype interactions: From model to crop species. *Trends Genet.* 29, 41–50. doi: 10.1016/j.tig.2012.09.006
- Chan, E. K. F., Rowe, H. C., and Kliebenstein, D. J. (2010). Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* 185, 991–1007. doi: 10.1534/genetics.109.108522
- Chye, M. L., Li, H. Y., and Yung, M. H. (2000). Single amino acid substitutions at the acyl-CoA-binding domain interrupt [4C]palmitoyl-CoA binding of ACBP2, an *Arabidopsis* acyl-CoA-binding protein with ankyrin repeats. *Plant Mol. Biol.* 44, 711–721. doi: 10.1023/A:1026524108095
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila*

AUTHOR CONTRIBUTIONS

KS analyzed the data and wrote the manuscript. KS and KH designed the data analysis method and revised and approved the manuscript.

FUNDING

This work was supported by Grants-in-Aid for Scientific Research (25710017, 15H02433, 17H03727, 18KK0176, 18H02420, and 19H05348; to KH) as well as research grants from the Takeda Science Foundation (to KH), the Sumitomo Foundation (to KH), Kurume Research Park (to KH), and the Asahi Glass Foundation (to KH).

ACKNOWLEDGMENTS

We thank the National Institute of Genetics of the Research Organization of Information and Systems for providing excellent supercomputer services. We also thank Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01567/full#supplementary-material>

- melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 6, 80–92. doi: 10.4161/fly.19695
- Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., et al. (2019). Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51, 541–547. doi: 10.1038/s41588-019-0356-4
- Fukushima, K., Fang, X., Alvarez-Ponce, D., Cai, H., Carretero-Paulet, L., Chen, C., et al. (2017). Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. *Nat. Ecol. Evol.* 1, 1–9. doi: 10.1038/s41559-016-0059
- Gan, X., Stegle, O., Behr, J., Steffen, J. G., Drewe, P., Hildebrand, K. L., et al. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477, 419–423. doi: 10.1038/nature10414
- Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K., and Shiu, S.-H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148, 993–1003. doi: 10.1104/pp.108.122457
- Hanada, K., Kuromori, T., Myouga, F., Toyoda, T., and Shinozaki, K. (2009). Increased expression and protein divergence in duplicate genes is associated with morphological diversification. *PLoS Genet.* 5, 1–7. doi: 10.1371/journal.pgen.1000781
- Jiao, W. B., and Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. *Curr. Opin. Plant Biol.* 36, 64–70. doi: 10.1016/j.pbi.2017.02.002
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kliebenstein, D. J., and Osbourn, A. (2012). Making new molecules - evolution of pathways for novel metabolites in plants. *Curr. Opin. Plant Biol.* 15, 415–423. doi: 10.1016/j.pbi.2012.05.005
- Kliebenstein, D. J. (2001). Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control

- glucosinolate biosynthesis in *Arabidopsis*. *Plant Cell Online* 13, 681–693. doi: 10.1105/tpc.13.3.681
- Lan, T., Renner, T., Ibarra-Laclette, E., Farr, K. M., Chang, T.-H., Cervantes-Pérez, S. A., et al. (2017). Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. *Proc. Natl. Acad. Sci.* 114, E4435–E4441. doi: 10.1073/pnas.1702072114
- Lye, Z. N., and Purugganan, M. D. (2019). Copy number variation in domestication. *Trends Plant Sci.* 24, 352–365. doi: 10.1016/j.tplants.2019.01.003
- Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J. I., Ebana, K., et al. (2015). Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J.* 81, 13–23. doi: 10.1111/tpj.12681
- McDonald, J. H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351, 652–654. doi: 10.1038/351652a0
- Mouille, G., Witucka-wall, H., Bruyant, M., Loudet, O., Pelletier, S., Pauly, M., et al. (2006). Quantitative trait loci analysis of primary cell wall composition in *Arabidopsis* 1. *Plant Physiol.* 141, 1035–1044. doi: 10.1104/pp.106.079384.aestivum
- Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426. doi: 10.1093/oxfordjournals.molbev.a040410
- Ohno, S. (1970). *Evolution by Gene Duplication*.
- Panchy, N., Lehti-Shiu, M. D., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523
- Pichersky, E., and Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* 62, 549–566. doi: 10.1146/annurev-arplant-042110-103814
- Rand, D., and Kann, L. (1996). Excess amino acid polymorphism in mitochondrial among genes from *Drosophila*, Mice, and Humans. *Mol. Biol. Evol.* 13, 735–748. doi: 10.1093/oxfordjournals.molbev.a025634
- Rawat, V., Abdelsamad, A., Pietzenuk, B., Seymour, D. K., Koenig, D., Weigel, D., et al. (2015). Improving the annotation of *Arabidopsis* Lyrata using RNA-seq data. *PLoS One* 10, 1–12. doi: 10.1371/journal.pone.0137391
- Scannell, D. R., and Wolfe, K. H. (2008). A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. *Genome Res.* 18, 137–147. doi: 10.1101/gr.6341207
- Schmitz, R. J., Schultz, M. D., Urich, M., Nery, J. R., Pelizzola, M., Libiger, O., et al. (2013). Patterns of population epigenomic diversity. *Nature* 495, 193–198. doi: 10.1038/nature11968
- Shirai, K., Matsuda, F., Nakabayashi, R., Okamoto, M., Tanaka, M., Fujimoto, A., et al. (2017). A highly specific genome-wide association study integrated with transcriptome data reveals the contribution of copy number variations to specialized metabolites in *Arabidopsis thaliana* accessions. *Mol. Biol. Evol.* 34, 3111–3122. doi: 10.1093/molbev/msx234
- Smith, N. G. C., and Eyre-Walker, A. (2002). Adaptive protein evolution in *Drosophila*. *Nature* 415, 1022–1024. doi: 10.1038/4151022a
- Stoletzki, N., and Eyre-Walker, A. (2011). Estimation of the neutrality index. *Mol. Biol. Evol.* 28, 63–70. doi: 10.1093/molbev/msq249
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The third revolution in sequencing technology. *Trends Genet.* 34, 666–681. doi: 10.1016/j.tig.2018.05.008
- Weigel, D. (2012). Natural variation in *Arabidopsis*: from molecular genetics to ecological genomics. *Plant Physiol.* 158, 2–22. doi: 10.1104/pp.111.189845
- Yang, Z., and Nielsen, R. (2000). Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Mol. Biol. Evol.* 17, 32–43. doi: 10.1093/oxfordjournals.molbev.a026236
- Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., et al. (2019). Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat. Genet.* 51, 1052–1059. doi: 10.1038/s41588-019-0427-6
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yu, Q., Jalaludin, A., Han, H., Chen, M., Sammons, R. D., and Powles, S. B. (2015). Evolution of a double amino acid substitution in the 5-enolpyruvylshikimate-3-phosphate synthase in *eleusine indica* conferring high-level glyphosate resistance. *Plant Physiol.* 167, 1440–1447. doi: 10.1104/pp.15.00146
- Zhang, J., Chen, L. L., Sun, S., Kudrna, D., Copetti, D., Li, W., et al. (2016). Data descriptor: building two indica rice reference genomes with PacBio long-read and illumina paired-end sequencing data. *Sci. Data* 3, 1–8. doi: 10.1038/sdata.2016.76
- Zmienko, A., Samelak, A., Kozłowski, P., and Figlerowicz, M. (2014). Copy number polymorphism in plant genomes. *Theor. Appl. Genet.* 127, 1–18. doi: 10.1007/s00122-013-2177-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Shirai and Hanada. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.