

Review

Co-Regulation of Clustered and Neo-Functionalized Genes in Plant-Specialized Metabolism

Takayuki Tohge ^{1,*} and Alisdair R. Fernie ^{2,*}

¹ Graduate School of Biological Science, Nara Institute of Science and Technology (NAIST), Ikoma 630-0192, Japan

² Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany

* Correspondence: tohge@bs.naist.jp (T.T.); fernie@mpimp-golm.mpg.de (A.R.F.); Tel.: +81-(743)-72-5480 (T.T.); +49-(0)331-567-8211 (A.R.F.)

Received: 1 April 2020; Accepted: 4 May 2020; Published: 13 May 2020

Abstract: Current findings of neighboring genes involved in plant specialized metabolism provide the genomic signatures of metabolic evolution. Two such genomic features, namely, (i) metabolic gene cluster and (ii) neo-functionalization of tandem gene duplications, represent key factors corresponding to the creation of metabolic diversity of plant specialized metabolism. So far, several terpenoid and alkaloid biosynthetic genes have been characterized with gene clusters in some plants. On the other hand, some modification genes involved in flavonoid and glucosinolate biosynthesis were found to arise via gene neo-functionalization. Although the occurrence of both types of metabolic evolution are different, the neighboring genes are generally regulated by the same or related regulation factors. Therefore, the translation-based approaches associated with genomics, and transcriptomics are able to be employed for functional genomics focusing on plant secondary metabolism. Here, we present a survey of the current understanding of neighboring genes involved in plant secondary metabolism. Additionally, a genomic overview of neighboring genes of four model plants and transcriptional co-expression network neighboring genes to detect metabolic gene clusters in *Arabidopsis* is provided. Finally, the insights functional genomics have provided concerning the evolution and mechanistic regulation of both the formation and operation of metabolic neighboring clusters is discussed.

Keywords: plant specialized metabolism; gene cluster; tandem gene duplication; co-regulation network; neo-functionalization; metabolic evolution

1. Introduction

Plants produce a huge variety of specialized metabolites (secondary metabolites), which have been characterized as the defense metabolism of plants and the arsenals they develop come about via adaptation to and natural selection caused by ecological niches and environmental factors [1–3]. Generally, the chemical diversity of plant specialized metabolites has been greatly expanded, either by horizontal gene transfer from other species, or by functional diversification following tandem gene duplication during metabolic evolution [4–6]. Such gene duplication is often found as a key innovator of metabolic evolution in expanding the metabolic diversity of plant metabolism, including species-specific specialized metabolism. Species-specific neighboring genes, therefore, have been the focus of considerable research [7].

Metabolic gene clusters constructed by the genomic neighboring of biosynthetic genes were originally discovered in the operons of bacterial genomes [8], with operon-like gene clusters later also being found, for genes associated with primary and secondary metabolism, in plant genomes [6]. Metabolic gene clusters for terpenoid biosynthesis have been observed in several land plant species.

So far, terpenoid biosynthesis in rice [9–11], tomato [12,13], oat [14], *Arabidopsis thaliana* [4,15,16], Brassica plant species [17] and *Lotus japonica* [18] are reported to contain such metabolic gene clusters [7,19]. Additionally, biosynthetic genes of isoquinoline alkaloids in opium poppy, bezoxazinones in maize as well as cyanogenic glucosides in lotus, cassava and sorghum were found within metabolic gene clusters [19]. However, the core genes involved in land-plant conserved secondary metabolism, such as flavonoid and hydroxycinnamate pathways, were not yet found as a gene cluster in any plant species. Given that metabolic gene clusters have only been found in secondary metabolism which arose relatively recently after plant speciation, such gene clusters tend to be found in species-specific secondary metabolism.

Neo-functionalization, by contrast, which has been occurred following tandem gene duplication and results in differential substrate specificity of the independent gene copies, is also identified as a key factor of metabolic evolution [7]. Following secondary metabolite analysis in *A. thaliana* natural, several enzymatic decoration genes involved in the modification of core glucosinolate [20–22] and flavonoid [23] structures, were characterized by the identification of genomic features indicative of neo-functionalization of tandem duplicated genes. It is important to note that gene duplication is considerably more prominent in plant genomes [24], and that this is a likely cause of the fact that they are also metabolically considerably more diverse than other organisms [25,26]. The fact that such metabolic polymorphism is generally found among natural accessions, but not among related plant species, except occasionally in the case of close relatives such as *Arabidopsis lyrata*, these neo-functionalized genes are generally regarded to be recently evolved genes. As described above, although the core biosynthetic genes of flavonoid production have never been detected as a metabolic gene cluster, genes encoding the decorative reactions were. Such genes are responsible for a major proportion of the diversity of flavonoids. This fact urges caution to pay attention to different aspects of diversity, dependent on the chemical nature of the compounds in question. It furthermore underlines that the investigation of both core and decorative secondary reactions are needed, in order to fully understand the radiance of metabolic evolution.

In order to understand metabolic evolution, the analysis of neighboring genes in plant genomes would appear to be a sensible starting point. However, since on average, 65% of annotated genes in plant genomes have a duplicate copy [27], the detection of key genomic region and genes of species specific specialized metabolism is highly complex. This has been addressed computationally with the programs PhytoClust [28] and PlantiSmash [29] being highly usable tools that allowed web-based searches for gene clusters. However, unfortunately, their operation appears to have been discontinued. However, all is not lost, since genomic comparison via cross-species comparison can be employed for the detection of specific genes, including gene clusters, as well as neo-functionalized gene pairs. Additionally, the co-expression networks analysis based on the correlation co-efficient between expressed genes in different tissues and growth conditions, is another approach to identify functional gene clusters. Indeed, some neighboring genes in *Arabidopsis* are co-expressed [30] and, for example, the metabolic cluster genes involved in thalianol and marnerial biosynthesis in *Arabidopsis* are co-regulated [16]. We will return to the mechanism underlying this co-regulation in mechanisms by which clustered genes are co-expressed, below.

Here, we review the current understanding of neighboring metabolic genes, including both metabolic gene clusters and tandem neo-functionalized genes. To discuss the detection of neighboring genes located in the gene cluster and neo-functionalization in model plants, we also evaluated the duplication rate of metabolic genes (cytochrome P450, CYP; 2-oxoglutarate-dependent dioxygenase, 2ODD; terpene synthase, TPS; UDP-sugar dependent glycosyltransferase family 1, UGT1; polyketide synthase, PKS) in the genomes of *A. thaliana*, *Oryza sativa*, *Solanum lycopersicum* and *Lotus japonica*. Furthermore, we screened the transcriptional correlation of neighboring genes in transcriptomics datasets. We also present a case study of the approach of detecting metabolic gene clusters, via co-expression analyses focused on plant specialized metabolism.

2. Neighboring Genes of Plant Specialized Metabolism

2.1. Gene Clusters Found in Plant Specialized Metabolism

Biosynthetic gene clusters are a genomic region containing at least three different classes of enzymatic genes involved in the same biosynthetic pathway [31]. To date, most of the gene clusters of specialized metabolisms characterized in land plants belong to terpenoid biosynthesis (Figure 1A). In *A. thaliana*, the biosynthesis of two triterpenes, thalianol and marneral were found as metabolic gene clusters [4,15,16]. These clusters contain oxidosqualene cyclase(OSC)-like terpene synthases (thalianol synthase, THAS; marneral synthase, MRN1), P450s (thalian-diol hydroxylase, AtCYP708A2, THAH; thalian-diol desaturase, AtCYP705A5, THAD; AtCYP705A12), BAHD acyltransferase and oxidase (marneral oxidase, MRO) (Figure 1A). The gene cluster located in the genomic synteny of thalianol gene cluster in *A. thaliana* and *A. lyrata* was additionally found in the *Capsella rubella* genome. Interestingly, the structure of this gene cluster in *C. rubella* is diverse and rather, corresponds to the production of tirucallol [17]. This metabolic diversification is created by the functional diversification of both the terpene synthase and P450 in this conserved genomic region. The overexpression of THAS and MRN1 was reported to lead to a dwarf phenotype, whilst the knockout of these genes resulted in longer roots and delayed flowering [4,15]. Importantly, all thalianol biosynthetic genes in both *A. thaliana* and *A. lyrata* showed root-specific gene expression, but tirucallol biosynthetic genes in *C. rubella* displayed bud-specific gene expression pattern. These results suggest that the physiological functions of thalianols and tirucallols are different in plant species, although the origin and evolutionary occurrence of their biosynthesis prior to their genetic diversification are predicted to be the same.

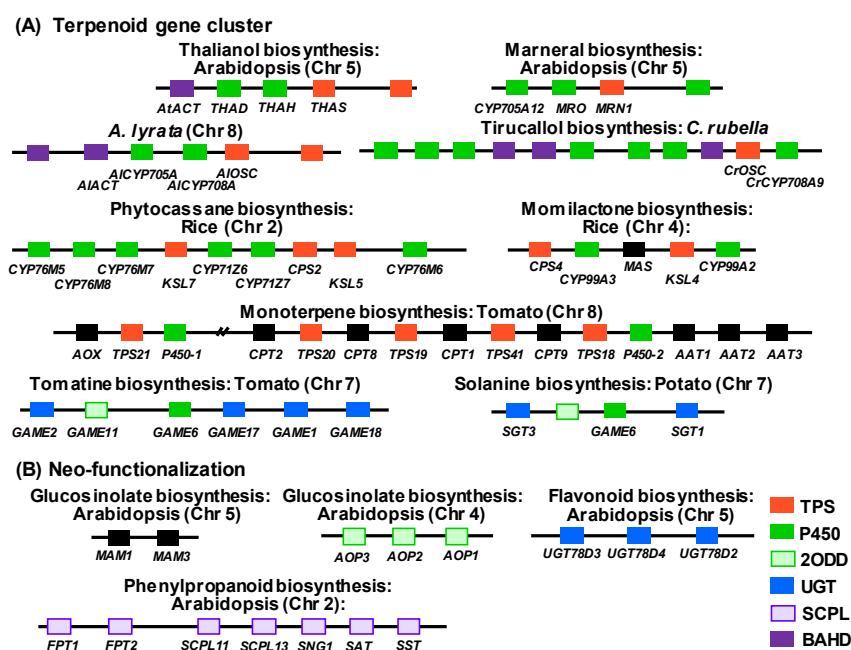


Figure 1. Neighboring genes of plant secondary metabolism. (A) Reported terpenoid gene cluster in plants. (B) Genomic neo-functionalization in plant secondary metabolism. Abbreviations: ACT, acyltransferase; THAH, thalianol hydroxylase; THAS, thalianol synthase; MRO, putative marneral oxidase; MRN, marneral synthase; baruol synthase, BARS1; OSC, oxidosqualene synthase; CPS, labdadienyl/copalyl synthases; KSL, kaurene synthase-like gene; MAS, momilactone A synthase; SAD, saponin-deficient; AMY, oxidosqualene cyclase; GAME, glycoalkaloid metabolism; SGT, sterol alkaloid glycosyltransferase; AOX, aldehyde oxidase; TPS, terpene synthases; CPT, *cis*-prenyl transferases; AAT, alcohol acyl transferases; TPS, terpene synthase; SCPL, serine carboxypeptidase-like; P450, CYP; 2ODD, 2-oxoglutarate-dependent dioxygenase; UGT, UDP-sugar dependent glycosyltransferase family.

Diterpenoid gene clusters in rice species have been found to be the key genomic region corresponding to the productivity of diverse antifungal-phytoalexins. Biosynthesis of three triterpene-type phytoalexins; phytocassanes, momilactones and oryzalexins were found to be produced by these clusters [9–11]. They commonly contain *ent*-CDP (chimera diterpene) synthase type diterpene synthases (CPS), kaurene synthase-like (KSL), and P450s (CYPs) (Figure 1A). Generally, these rice diterpenoid phytoalexins are induced in response to fungal and bacterial infections [10], however, the momilactones are highly accumulated in grain husks. The physiological functions of phytocassanes, momilactones and oryzalexins have been suggested to be different. Importantly, the gibberellin biosynthetic gene cluster, comprising the biosynthetic genes OsCPS1, kaurene synthase (OsKS1), CYP genes (*ent*-kaurene oxidase, OsKO; *ent*-kaurenoic acid oxidase, OsKAO), has a highly similar structure to that of the rice phytoalexin gene cluster. A tomato monoterpene gene cluster containing five terpene synthase genes (SITPS), *cis*-prenyl transferase (CPTs) and P450s, was similarly identified by the detection of orthologue gene clusters via comparative genomics between *S. lycopersicum*, *S. pennellii*, *S. habrochaites*, *S. pimpinellifolium* and *S. tuberosum* [13] (Figure 1A). Similarly, a steroidal alkaloid gene cluster also accounting for triterpene derived secondary metabolites, that was conserved in both the tomato and potato genomes, was identified [12] (Figure 1A). Furthermore, a second steroidal alkaloid gene cluster was recently observed in a multiomics study of tomato domestication and improvement [32]. All three steroidal alkaloid gene clusters contain P450, 2ODDs and UGTs.

2.2. Neo-functionalization Following Tandem Gene Duplication

The current framework of genetics-based strategies coupled with metabolomic and transcriptomic approaches have been largely performed in functional genomics approaches, aimed at the identification of key genes and genomic regions involved in species- or accession-specific secondary metabolism [7]. Within these studies, considerable neo-functionalization has been observed in the genomic regions originated by tandem duplications in Arabidopsis. In the metabolite profiling of *Arabidopsis thaliana* accessions, the intra-species metabolic polymorphism of glucosinolates was identified with the *AOP2/3* (Aliphatic glucosinolate 2-oxo acid-dependent dioxygenase) and *MAM1/3* (Methylthioalkylmalate synthase 1 and 3) tandem gene duplication region [20–22]. Furthermore, the metabolomics analysis of floral secondary metabolites among Arabidopsis natural accessions revealed that flavonol-phenylacyltransferase (AtFPT2) was found as a neo-functionalization of the serine carboxypeptidase-like (SCPL) tandem duplicated genomic region [23]. In this genomic region containing a total of seven SCPL genes, four acyltransferase genes, including AtFPT2, have been experimentally confirmed. These FPT genes have slightly different substrate specificities with regard to their phenylacyl acceptors, but not their phenylacyl donors. Interestingly, a comparative genomics analysis revealed that the tandem gene duplication of FPT genes were different between close Brassica relatives with the exception of *A. lyrata*. These genes are predicted to be the result of neo-functionalization following tandem gene duplication and to have relatively recently evolved, because gene duplication occurred among natural accessions and is not conserved in any but the most closely related Brassica species. Since the decorative reaction catalyzed by AtFPT2 could render much higher tolerance against UV-B irradiance stress, the gene deletion of AtFPT2 is thought to be selected against by natural light stress. This example demonstrates how duplication and neo-functionalization can effectively expand the metabolic diversity of secondary metabolism. However, it only tells part of the story, since protection against light stress in other species has arisen by convergent evolution. Indeed, we recently screened the KNApSack database for phenylacylated flavonoids and found that rare Arabidopsis was in no means unique in harboring such compounds [33]. Indeed, one of those phenylacylated flavonols was previously identified to confer UV-B protection to spruce [34]. Similarly, wide screens of the metabolic repertoire of rice and the wild barley quinke revealed that the 5-*O*-glycosylation, 7-*O*-glycosylation and 8-*C*-pentosylation of flavones conferred UV tolerance to these species, respectively [35,36]. However, whilst an interesting aside, these examples indicate the limitations of the approaches we are championing here, rather than highlighting their utility, so we will not dwell on them further. Suffice to say, metabolic

profiling of association mapping panels and in particular the checking of decorative modifications [37] may ultimately prove to be a highly effective pre-selection of genomic regions in which to search for either metabolic gene clusters or neo-functionalized genes.

The decorative enzymes of secondary metabolism, such as UGT1, glycoside hydrolase family 1-type gene (BGLU), BEAT (Benzylalcohol-*O*-acetyltransferase)/AHCT (anthocyanin-*O*-hydroxycinnamoyl transferase)/HCBT (anthranilate-*N*-hydroxycinnamoyl/benzoyltransferase)/DAT (deacetylvinidoline-4-*O*-acetyltransferase) (BAHD) genes and SCPL genes, are often found next to highly similar “tandem genes”. These decorative enzymatic genes are additionally generally vital in the creation of the chemical diversity inherent in plant secondary metabolism. For example, in *Arabidopsis* flavonoid biosynthesis, anthiocyanin-3-*O*-glycoside-2''-*O*-phenylacyltransferase [38] and flavonol-3-*O*-glycoside-2''-*O*-phenylacyltransferase [23] were identified in tandem with a BAHD and a SCPL gene, respectively. Such functional convergence of enzymatic properties sometimes renders the identification of such features difficult. For example, the *FTP1* and *FTP2* genes described above were not discussed in the initial definition of the SCPL cluster [39,40] (Figure 1B). Additionally, in some cases of neo-functionalized genes which occur by convergent evolution, the protein sequence does not show a higher similarity between proteins which have exactly the same function (e.g. flavonoid glactocyltransferase of grapevine, kiwi and Vinga) [41,42]. Given the difficulty in such cases, we suggest the adoption of an integrative approach, taking into account protein sequence similarity, as well as proximately alongside transcriptional co-regulation.

2.3. Co-expression Networks of Neighboring Genes for the Discovery of Metabolic Cluster Genes and Neo-functionalized Genes

Co-expression network analysis, which is assessed via the analysis of large-scale transcriptomics data, has greatly aided the elucidation of gene annotation and functional genomics in a broad range of plant species [7,43–45]. Co-expression network analysis is well-developed in plant science, however, such multi-gene coefficient based approaches can still be further refined by data optimization strategies, including the use of target-defined sub-datasets [43,46] and targeted gene network analysis [47,48]. Neighboring gene sets found in genome-wide gene annotation have been tested as an approach for the prediction of operon-like gene clusters in the *Arabidopsis* genome [30,49]. Gene ontology (GO), or KEGG-based target co-expression analysis, revealed that some biosynthetic/catabolic genes in pathways, such as phospholipid degradation and porphyrin and chlorophyll metabolism clusters, are highly correlated gene groups [49]. Such co-expression network analyses are able to identify metabolic gene clusters constructed by genes which are transcriptionally co-regulated in certain tissues and/or stress conditions. In order to carry out a co-expression network approach for neighboring genes, the genomic regions containing putative metabolic gene clusters and tandem gene duplication must first be assigned. For example, a genomic survey of putative specialized metabolic gene clusters in four major model plants (*A. thaliana*, *Oryza sativa*, *Solanum lycopersicum* and *Lotus japonica*) is evaluated.

Table 1. Genomic survey for prediction of the metabolic gene clusters and neo-functionalization.

Plant Species	Annotated Metabolic Gene Cluster	Tandem Gene Duplication of Single Gene Family				
		P450	2ODD	TPS	PKS	UGT
<i>A. thaliana</i>	39	28	16	6	5	15
<i>O. sativa</i>	34	57	15	10	10	37
<i>S. lycopersicum</i>	50	30	24	8	9	30
<i>L. japonicus</i>	18	29	7	5	4	7

The respective genomic regions of putative gene clusters and neo-functionalized gene pairs were defined for five gene families, namely P450, 2ODD, TPS and PKS, and UGT1, which are the cardinal gene families of plant-specialized metabolism (Table 1). In our genomic survey, 109 genomic regions in *A. thaliana*, 163 genomic regions in *O. sativa*, 151 genomic regions in *S. lycopersicum* and 70 genomic

regions in *L. japonica*, were found as either gene clusters and/or tandem gene duplications containing more than three genes located in the same genomic region. Amongst these genomic region, single tandem gene duplications of single gene families, indicative of potential neo-functionalized genes, were found in 70 regions in *A. thaliana*; 129 regions in *O. sativa*; 101 regions in *S. lycopersicum* and 52 regions in *L. japonica*. The results of our genomic survey include known metabolic gene clusters, for example, thalianol and marneral biosynthetic genes in Arabidopsis [17,49], steroidal glycoalkaloids in tomato [12], triterpene-type phytoalexin biosynthetic genes [10], cyanogenic glucoside biosynthesis in *L. japonica* [50], as well as novel putative metabolic gene clusters (39 regions in *A. thaliana*; 34 regions in *O. sativa*; 50 regions in *S. lycopersicum* and 18 regions in *L. japonica* (Table 1)).

Having the candidates of metabolic gene clusters and neo-functionalized genes in hand, a co-expression approach was conducted. A total of 507 Arabidopsis genes which are putatively annotated as metabolic gene cluster, were used for the co-expression network analysis by ATTED-II [51,52]. Figure 2A provides a global overview of the co-expression of clustered genes in Arabidopsis. Within these co-regulated networks, three networks could be detected as neighboring gene co-expression networks in Arabidopsis. One of the co-expressed neighboring genes that we revealed were terpenoid gene clusters (Figure 2B) of thalianol and marneral biosynthesis, which are known specialized metabolic gene clusters within the Arabidopsis genome. The other co-expressed network contains brassinosteroids inactivator 1 (BIA1) gene involved in brassinosteroid homeostasis (Figure 2C) [53]. Furthermore, a putative gene cluster containing both P450s and TPSs, as well as the gene encoding baruol synthase (BARS1), was detected as a co-expressed metabolic gene cluster-like genomic region (Figure 2D and 2E). Whilst this putative gene cluster will clearly need to be validated experimentally, the fact that we re-found already characterized Arabidopsis gene clusters is highly reassuring. As previously reported [30,49], such a co-expression network approach combined with the genomic survey of neighboring genes is highly useful to identify metabolic gene clusters. It will be interesting in future studies to assess whether it is equally useful at identifying neo-functionalized genes. The ever-increasing wealth of annotated plant genomes, both in quantity and more recently in quality, alongside a similar torrent of transcriptomic datasets, will likely greatly expand the capacity of this approach. Additionally, a cross-species comparative genomics between close relatives or accessions will likely be highly useful to understand the evolution of metabolic neighboring clusters, as recently reported in the cross species comparison of terpenoid biosynthesis in three Brassicaceae species [17].

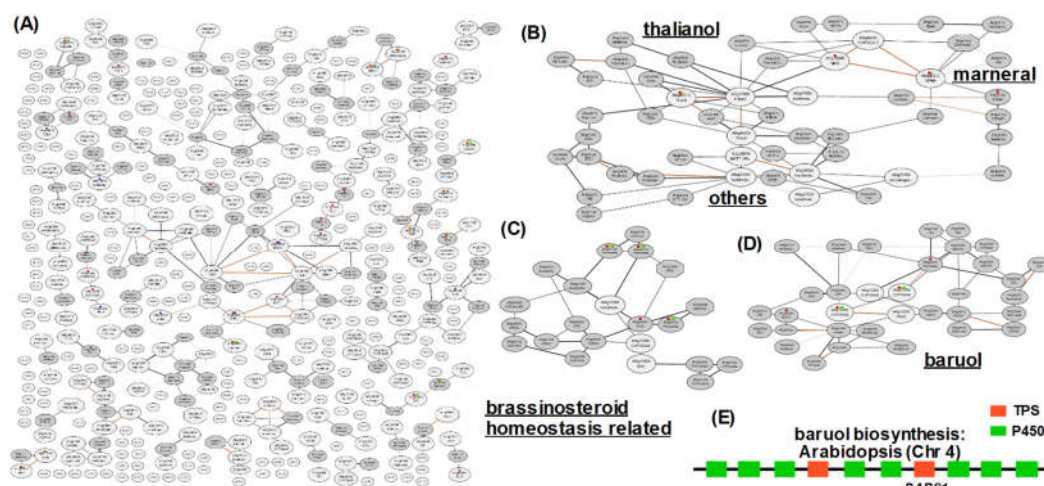


Figure 2. Co-expression analysis of neighboring genes in Arabidopsis genome. (A) a global overview of the co-expression of clustered genes in Arabidopsis, (B) a co-expression network of terpenoid metabolic gene clusters, (C) a co-expression network of baruol synthase, (D) a co-expression network of putative brassinosteroids homeostasis related genes, (E) putative gene cluster of baruol biosynthesis. Abbreviations: terpene synthases; P450, CYP.

2.4. Mechanisms by which Clustered Genes are Co-expressed

In the case of the cross-species comparative genomics approach of triterpene biosynthetic genes in Brassica plant species, the structure of cluster genes conserved in genomic synteny are slightly different between *Arabidopsis* species and *Capsella*, for thalianol and tirucallol biosynthesis. Key genes of each pathway were, however, highly co-expressed in a tissue species manner in the respective species [17]. Taking into account the fact that these specialized metabolites are produced in certain tissues and under certain conditions of stress begs the question as to whether there is a mechanistic advantage to their being clustered. Intriguingly, transcription factors regulating plant specialized metabolic gene clusters are still largely unknown. Additionally, the complexity of the regulatory network of combination with miRNA-based post-translational regulation of plant specialized metabolism [54] reported in terpenoid biosynthesis [55] and flavonoid biosynthesis [56], possibly provide the unclear framework of co-expressed network. Indeed despite the fact that some regulators of metabolic gene clusters in bacteria are located in the target gene cluster [57], no such regulators have been reported from surveys of the neighboring genes in known plant clusters. As such, the co-expression network of neighboring genes approach might provide candidate regulators controlling metabolic gene clusters. That said, results from the Osbourn laboratory described that the phylogenetically conserved histone variant, H2A.Z, is essential for the normal expression of the abovementioned thalianol gene cluster of *Arabidopsis* [16,58]. Indeed, in their study, Nützmann and Osbourn revealed that the levels of thalianol hydroxylase were altered in five of six mutants defective in histone modifications and chromatin remodeling [16,59]. Thus, they indicate that the SWR1 chromatin remodeling complex is required for the incorporation of H2A.Z into the nucleosomes (Figure 3). Indeed, H2A.Z deposition has previously been demonstrated to activate the DAL gene cluster of yeast [60]. Further experiments in both the Arabinol and marenol gene clusters revealed that H2A.Z occupancy into nucleosomes within the gene clusters leads to a localized opening of the chromatin structure, and thereby facilitates cluster expression [16]. This mechanism is illustrated in Figure 3. It is, however, also important to note that chromatin regulation has been reported to affect the synthesis of a number of compounds, including phenylpropanoids, glucosinolates and gibberellins [16,61,62], which are not encoded by cluster genes. Therefore, the use of chromatin marks as a screening strategy for clusters should be approached with caution.

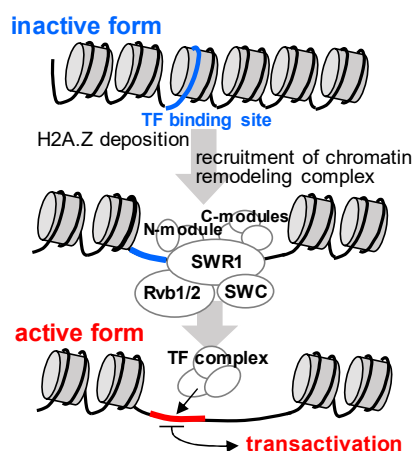


Figure 3. Overview of chromatin remodeling following H2A.Z deposition. N-module and C-module indicate histone and H2A.Z bindings.

3. Concluding Remarks and Future Prospects

Neighboring genes involved in plant specialized metabolism, such as metabolic gene cluster and neo-functionalized genes, are often key in metabolic evolution and metabolic diversification. The distance between biosynthetic genes involved in plant specialized metabolism seems to correlate to the age of the biosynthesis after its occurrence. Therefore, species-specific biosynthesis, namely

relatively recent biosynthetic innovations, show much clearer genetic signatures of metabolic evolution. As we report, the subjection of neighboring genes involved in plant specialized metabolism to genomic surveys and co-expression network analysis represents a simple way to find such structural features within plant genomes. Moreover, comparative genomics approaches can be integrated in order to allow translational genomics research. Furthermore, coupling these approaches with recent insights into histone modification and chromatin modelling offers a further route into identifying putative gene clusters. The ever-increasing number of plant genomes alongside massive increases in the amount and availability of transcriptomics data. This fact highly suggests that the approaches we describe here will become increasingly useful in future studies concerned with understanding both the evolution and metabolic regulation of pathways of plant-specialized metabolism.

Author Contributions: Conceptualization, T.T. and A.R.F.; supervision, T.T. and A.R.F.; writing—original draft preparation, T.T. and A.R.F. All authors have read and agreed to the published version of the manuscript.

Funding: This article was funded by Japan Society for the Promotion of Science (JSPS) Scientific Research B (19H03249).

Acknowledgments: T.T. gratefully acknowledge partial support by NAIST and JSPS KAKENHI Grant-in-Aid for Scientific Research B (19H03249).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tohge, T.; Watanabe, M.; Hoefgen, R.; Fernie, A.R. Shikimate and Phenylalanine Biosynthesis in the Green Lineage. *Front. Plant Sci.* **2013**, *4*, 62, doi:10.3389/fpls.2013.00062.
2. Tohge, T.; Watanabe, M.; Hoefgen, R.; Fernie, A.R. The evolution of phenylpropanoid metabolism in the green lineage. *Crit. Rev. Biochem. Mol. Boil.* **2013**, *48*, 123–152, doi:10.3109/10409238.2012.758083.
3. Saito, K.; Yonekura-Sakakibara, K.; Nakabayashi, R.; Higashi, Y.; Yamazaki, M.; Tohge, T.; Fernie, A.R. The flavonoid biosynthetic pathway in Arabidopsis: Structural and genetic diversity. *Plant Physiol. Biochem.* **2013**, *72*, 21–34, doi:10.1016/j.plaphy.2013.02.001.
4. Field, B.; Osbourn, A.E. Metabolic Diversification—Independent Assembly of Operon-Like Gene Clusters in Different Plants. *Sci.* **2008**, *320*, 543–547, doi:10.1126/science.1154990.
5. Scossa, F.; Brotman, Y.; Lima, F.D.A.E.; Willmitzer, L.; Nikoloski, Z.; Tohge, T.; Fernie, A.R. Genomics-based strategies for the use of natural variation in the improvement of crop metabolism. *Plant Sci.* **2016**, *242*, 47–64, doi:10.1016/j.plantsci.2015.05.021.
6. Nützmann, H.-W.; Scazzocchio, C.; Osbourn, A. Metabolic Gene Clusters in Eukaryotes. *Annu. Rev. Genet.* **2018**, *52*, 159–183, doi:10.1146/annurev-genet-120417-031237.
7. Fernie, A.R.; Tohge, T. The Genetics of Plant Metabolism. *Annu. Rev. Genet.* **2017**, *51*, 287–310, doi:10.1146/annurev-genet-120116-024640.
8. Ames, B.N.; Hartman, P.E. The Histidine Operon. *Cold Spring Harb. Symp. Quant. Boil.* **1963**, *28*, 349–356, doi:10.1101/sqb.1963.028.01.049.
9. Shimura, K.; Okada, A.; Okada, K.; Jikumaru, Y.; Ko, K.-W.; Toyomasu, T.; Sassa, T.; Hasegawa, M.; Kodama, O.; Shibuya, N.; et al. Identification of a Biosynthetic Gene Cluster in Rice for Momilactones. *J. Boil. Chem.* **2007**, *282*, 34013–34018, doi:10.1074/jbc.m703344200.
10. Swaminathan, S.; Morrone, D.; Wang, Q.; Fulton, D.B.; Peters, R.J. CYP76M7 is an ent-cassadiene C11 α -hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* **2009**, *21*, 3315–3325, doi:10.1105/tpc.108.063677.
11. Miyamoto, K.; Fujita, M.; Shenton, M.R.; Akashi, S.; Sugawara, C.; Sakai, A.; Horie, K.; Hasegawa, M.; Kawaide, H.; Mitsuhashi, W.; et al. Evolutionary trajectory of phytoalexin biosynthetic gene clusters in rice. *Plant J.* **2016**, *87*, 293–304, doi:10.1111/tpj.13200.
12. Itkin, M.; Heinig, U.; Tzfadia, O.; Bhide, A.J.; Shinde, B.; Cárdenas, P.D.; Bocobza, S.E.; Unger, T.; Malitsky, S.; Finkers, R.; et al. Biosynthesis of Antinutritional Alkaloids in Solanaceous Crops Is Mediated by Clustered Genes. *Sci.* **2013**, *341*, 175–179, doi:10.1126/science.1240230.

13. Matsuba, Y.; Nguyen, T.T.; Wiegert, K.; Falara, V.; Gonzales-Vigil, E.; Leong, B.; Schäfer, P.; Kudrna, D.; A Wing, R.; Bolger, A.M.; et al. Evolution of a Complex Locus for Terpene Biosynthesis in *Solanum*[W][OPEN]. *Plant Cell* **2013**, *25*, 2022–2036, doi:10.1105/tpc.113.111013.
14. Qi, X.; Bakht, S.; Leggett, M.; Maxwell, C.; Melton, R.; Osbourn, A. A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8233–8238.
15. Field, B.; Fiston-Lavier, A.-S.; Kemen, A.; Geisler, K.; Quesneville, H.; Osbourn, A.E. Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 16116–16121.
16. Nützmann, H.-W.; Osbourn, A. Regulation of metabolic gene clusters in *Arabidopsis thaliana*. *New Phytol.* **2014**, *205*, 503–510, doi:10.1111/nph.13189.
17. Liu, Z.; Duran, H.G.S.; Harnvanichvech, Y.; Stephenson, M.; Schranz, M.E.; Nelson, D.; Medema, M.H.; Osbourn, A. Drivers of metabolic diversification: how dynamic genomic neighbourhoods generate new biosynthetic pathways in the Brassicaceae. *New Phytol.* **2019**, doi:10.1111/nph.16338.
18. Krokida, A.; Delis, C.; Geisler, K.; Garagounis, C.; Tsikou, D.; Peña-Rodríguez, L.M.; Katsarou, D.; Field, B.; Osbourn, A.E.; Papadopoulou, K.K. A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis. *New Phytol.* **2013**, *200*, 675–690, doi:10.1111/nph.12414.
19. Boycheva, S.; Daviet, L.; Wolfender, J.-L.; Fitzpatrick, T.B. The rise of operon-like gene clusters in plants. *Trends Plant Sci.* **2014**, *19*, 447–459, doi:10.1016/j.tplants.2014.01.013.
20. Kliebenstein, D.J. Genetic Control of Natural Variation in *Arabidopsis* Glucosinolate Accumulation. *Plant Physiol.* **2001**, *126*, 811–825, doi:10.1104/pp.126.2.811.
21. Kliebenstein, D.J.; Lambrix, V.M.; Reichelt, M.; Gershenzon, J.; Mitchell-Olds, T. Gene Duplication in the Diversification of Secondary Metabolism: Tandem 2-Oxoglutarate-Dependent Dioxygenases Control Glucosinolate Biosynthesis in *Arabidopsis*. *Plant Cell* **2001**, *13*, 681–693, doi:10.1105/tpc.13.3.681.
22. Petersen, A.; Hansen, L.G.; Mirza, N.; Crocoll, C.; Mirza, O.A.; Halkier, B.A. Changing substrate specificity and iteration of amino acid chain elongation in glucosinolate biosynthesis through targeted mutagenesis of *Arabidopsis* methylthioalkylmalate synthase *Biosci. Rep.* **2019**, *39*, 39, doi:10.1042/bsr20190446.
23. Tohge, T.; Wendenburg, R.; Ishihara, H.; Nakabayashi, R.; Watanabe, M.; Suplice, R.; Hoefgen, R.; Takayama, H.; Saito, K.; Stitt, M.; et al. Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nat. Commun.* **2016**, *7*, 12399, doi:10.1038/ncomms12399.
24. Zhang, J. Evolution by gene duplication: an update. *Trends Ecol. Evol.* **2003**, *18*, 292–298, doi:10.1016/s0169-5347(03)00033-8.
25. Alseekh, S.; Fernie, A.R. Metabolomics 20 years on: what have we learned and what hurdles remain? *Plant J.* **2018**, *94*, 933–942, doi:10.1111/tpj.13950.
26. Wang, S.; Alseekh, S.; Fernie, A.R.; Luo, J. The Structure and Function of Major Plant Metabolite Modifications. *Mol. Plant* **2019**, *12*, 899–919, doi:10.1016/j.molp.2019.06.001.
27. Panchy, N.; Lehti-Shiu, M.; Shiu, S.-H. Evolution of Gene Duplication in Plants. *Plant Physiol.* **2016**, *171*, 2294–2316, doi:10.1104/pp.16.00523.
28. Töpfer, N.; Fuchs, L.-M.; Aharoni, A. The PhytoClust tool for metabolic gene clusters discovery in plant genomes. *Nucleic Acids Res.* **2017**, *45*, 7049–7063, doi:10.1093/nar/gkx404.
29. A Kautsar, S.; Duran, H.G.S.; Blin, K.; Osbourn, A.; Medema, M.H. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res.* **2017**, *45*, W55–W63, doi:10.1093/nar/gkx305.
30. Williams, E.J.; Bowles, D.J. Coexpression of Neighboring Genes in the Genome of *Arabidopsis thaliana*. *Genome Res.* **2004**, *14*, 1060–1067, doi:10.1101/gr.2131104.
31. Medema, M.H.; Osbourn, A. Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways. *Nat. Prod. Rep.* **2016**, *33*, 951–962, doi:10.1039/c6np00035e.
32. Zhu, G.; Wang, S.; Huang, Z.; Zhang, S.; Liao, Q.; Zhang, C.-Z.; Lin, T.; Qin, M.; Peng, M.; Yang, C.; et al. Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell* **2018**, *172*, 249–261.e12, doi:10.1016/j.cell.2017.12.019.
33. Tohge, T.; De Souza, L.P.; Fernie, A.R. On the natural diversity of phenylacylated-flavonoid and their in planta function under conditions of stress. *Phytochem. Rev.* **2017**, *17*, 279–290, doi:10.1007/s11101-017-9531-3.

34. Jungblut, T.P.; Schnitzler, J.-P.; Heller, W.; Sandermann, H.; Hertkorn, N.; Szymczak, W.; Metzger, J.W. Structures of UV-B Induced Sunscreen Pigments of the Scots Pine (*Pinus sylvestris* L.). *Angew. Chem. Int. Ed.* **1995**, *34*, 312–314, doi:10.1002/anie.199503121.
35. Peng, M.; Shahzad, R.; Gul, A.; Subthain, H.; Shen, S.; Lei, L.; Zheng, Z.; Zhou, J.; Lu, D.; Wang, S.; et al. Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nat. Commun.* **2017**, *8*, 1975, doi:10.1038/s41467-017-02168-x.
36. Zeng, X.; Yuan, H.; Dong, X.; Peng, M.; Jing, X.; Xu, Q.; Tang, T.; Wang, Y.; Zha, S.; Gao, M.; et al. Genome-wide Dissection of Co-selected UV-B Responsive Pathways in the UV-B Adaptation of Qingke. *Mol. Plant* **2019**, *13*, 112–127, doi:10.1016/j.molp.2019.10.009.
37. Alseekh, S.; De Souza, L.P.; Benina, M.; Fernie, A.R. The style and substance of plant flavonoid decoration; towards defining both structure and function. *Phytochemistry* **2020**, *174*, 112347, doi:10.1016/j.phytochem.2020.112347.
38. Luo, J.; Nishiyama, Y.; Fuell, C.; Taguchi, G.; Elliott, K.; Hill, L.; Tanaka, Y.; Kitayama, M.; Yamazaki, M.; Bailey, P.; et al. Convergent evolution in the BAHD family of acyl transferases: identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. *Plant J.* **2007**, *50*, 678–695, doi:10.1111/j.1365-313x.2007.03079.x.
39. Fraser, C.M.; Thompson, M.G.; Shirley, A.M.; Ralph, J.; Schoenherr, J.A.; Sinlapadech, T.; Hall, M.C.; Chapple, C. Related *Arabidopsis* Serine Carboxypeptidase-Like Sinapoylglucose Acyltransferases Display Distinct But Overlapping Substrate Specificities1[OA]. *Plant Physiol.* **2007**, *144*, 1986–1999, doi:10.1104/pp.107.098970.
40. Stehle, F.; Brandt, W.; Stubbs, M.; Milkowski, C.; Strack, D. Sinapoyltransferases in the light of molecular evolution. *Phytochem.* **2009**, *70*, 1652–1662, doi:10.1016/j.phytochem.2009.07.023.
41. Liu, Y.; Zhou, B.; Qi, Y.-W.; Liu, C.; Liu, Z.; Ren, X. Biochemical and functional characterization of AcUFGT3a, a galactosyltransferase involved in anthocyanin biosynthesis in the red-fleshed kiwifruit (*Actinidia chinensis*). *Physiol. Plant.* **2017**, *162*, 409–426, doi:10.1111/ppl.12655.
42. Tohge, T.; Zhang, Y.; Peterrek, S.; Matros, A.; Rallapalli, G.; Tandrón, Y.A.; Butelli, E.; Kallam, K.; Hertkorn, N.; Mock, H.; et al. Ectopic expression of snapdragon transcription factors facilitates the identification of genes encoding enzymes of anthocyanin decoration in tomato. *Plant J.* **2015**, *83*, 686–704, doi:10.1111/tpj.12920.
43. Tohge, T.; Fernie, A.R. Annotation of Plant Gene Function via Combined Genomics, Metabolomics and Informatics. *J. Vis. Exp.* **2012**, e3487, doi:10.3791/3487.
44. Tohge, T.; Fernie, A.R. Co-expression and co-responses: within and beyond transcription. *Front. Plant Sci.* **2012**, *3*, 3, doi:10.3389/fpls.2012.00248.
45. Mutwil, M.; Klie, S.; Tohge, T.; Giorgi, F.M.; Wilkins, O.; Campbell, M.M.; Fernie, A.R.; Usadel, B.; Nikoloski, Z.; Persson, S. PlaNet: Combined Sequence and Expression Comparisons across Plant Networks Derived from Seven Species[W][OA]. *Plant Cell* **2011**, *23*, 895–910, doi:10.1105/tpc.111.083667.
46. Mutwil, M.; Øbro, J.; Willats, W.; Persson, S. GeneCAT—novel webtools that combine BLAST and co-expression analyses. *Nucleic Acids Res.* **2008**, *36*, W320–6, doi:10.1093/nar/gkn292.
47. Tohge, T.; Fernie, A.R. Lignin, mitochondrial family, and photorespiratory transporter classification as case studies in using co-expression, co-response, and protein locations to aid in identifying transport functions. *Front. Plant Sci.* **2014**, *5*, 5, doi:10.3389/fpls.2014.00075.
48. Araujo, W.; Tohge, T.; Ishizaki, K.; Leaver, C.J.; Fernie, A.R. Protein degradation – an alternative respiratory substrate for stressed plants. *Trends Plant Sci.* **2011**, *16*, 489–498, doi:10.1016/j.tplants.2011.05.008.
49. Wada, M.; Takahashi, H.; Amin, A.-U.-; Nakamura, K.; Hirai, M.Y.; Ohta, D.; Kanaya, S. Prediction of operon-like gene clusters in the *Arabidopsis thaliana* genome based on co-expression analysis of neighboring genes. *Gene* **2012**, *503*, 56–64, doi:10.1016/j.gene.2012.04.043.
50. Takos, A.; Knudsen, C.; Lai, D.; Kannangara, R.; Mikkelsen, L.; Motawia, M.S.; Olsen, C.E.; Sato, S.; Tabata, S.; Jørgensen, K.; et al. Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. *Plant J.* **2011**, *68*, 273–286, doi:10.1111/j.1365-313x.2011.04685.x.
51. Obayashi, T.; Kinoshita, K.; Nakai, K.; Shibaoka, M.; Hayashi, S.; Saeki, M.; Shibata, D.; Saito, K.; Ohta, H. ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in *Arabidopsis*. *Nucleic Acids Res.* **2006**, *35*, D863–D869, doi:10.1093/nar/gkl783.

52. Obayashi, T.; Hayashi, S.; Saeki, M.; Ohta, H.; Kinoshita, K. ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucleic Acids Res.* **2008**, *37*, D987–D991, doi:10.1093/nar/gkn807.
53. Zhang, Z.; Xu, L. Arabidopsis BRASSINOSTEROID INACTIVATOR2 is a typical BAHD acyltransferase involved in brassinosteroid homeostasis. *J. Exp. Bot.* **2018**, *69*, 1925–1941, doi:10.1093/jxb/ery057.
54. Gupta, O.P.; Karkute, S.; Banerjee, S.; Meena, N.L.; Dahuja, A. Contemporary Understanding of miRNA-Based Regulation of Secondary Metabolites Biosynthesis in Plants. *Front. Plant Sci.* **2017**, *8*, 82, doi:10.3389/fpls.2017.00374.
55. Fan, R.; Li, Y.; Li, C.; Zhang, Y. Differential microRNA Analysis of Glandular Trichomes and Young Leaves in *Xanthium strumarium* L. Reveals Their Putative Roles in Regulating Terpenoid Biosynthesis. *PLOS ONE* **2015**, *10*, 17, doi:10.1371/journal.pone.0139002.
56. Cui, L.; Shan, J.-X.; Shi, M.; Gao, J.-P.; Lin, H.-X. ThemIR156-SPL9-DFR pathway coordinates the relationship between development and abiotic stress tolerance in plants. *Plant J.* **2014**, *80*, 1108–1117, doi:10.1111/tj.12712.
57. Rokas, A.; Wisecaver, J.H.; Lind, A. The birth, evolution and death of metabolic gene clusters in fungi. *Nat. Rev. Genet.* **2018**, *16*, 731–744, doi:10.1038/s41579-018-0075-3.
58. Fernie, A.R.; Tohge, T. Location, location, location—no more! The unravelling of chromatin remodeling regulatory aspects of plant metabolic gene clusters. *New Phytol.* **2015**, *205*, 458–460, doi:10.1111/nph.13214.
59. Nützmann, H.-W.; Osbourn, A. Gene clustering in plant specialized metabolism. *Curr. Opin. Biotechnol.* **2014**, *26*, 91–99, doi:10.1016/j.copbio.2013.10.009.
60. Meneghini, M.D.; Wu, M.; Madhani, H.D. Conserved Histone Variant H2A.Z Protects Euchromatin from the Ectopic Spread of Silent Heterochromatin. *Cell* **2003**, *112*, 725–736, doi:10.1016/s0092-8674(03)00123-5.
61. Sarnowska, E.A.; Rolicka, A.; Bucior, E.; Cwiek, P.; Tohge, T.; Fernie, A.R.; Jikumaru, Y.; Kamiya, Y.; Franzen, R.; Schmelzer, E.; et al. DELLA-Interacting SWI3C Core Subunit of Switch/Sucrose Nonfermenting Chromatin Remodeling Complex Modulates Gibberellin Responses and Hormonal Cross Talk in Arabidopsis1[W]. *Plant Physiol.* **2013**, *163*, 305–317, doi:10.1104/pp.113.223933.
62. Sarnowska, E.; Gratkowska, D.M.; Sacharowski, S.P.; Cwiek, P.; Tohge, T.; Fernie, A.R.; Siedlecki, J.A.; Koncz, C.; Sarnowski, T.J. The Role of SWI/SNF Chromatin Remodeling Complexes in Hormone Crosstalk. *Trends Plant Sci.* **2016**, *21*, 594–608, doi:10.1016/j.tplants.2016.01.017.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).