Title: Genomic insights into the evolution of plant chemical defense

Authors:

Chloe P. Drummond*[1], Tanya Renner[1]

[1]*The Pennsylvania State University, Department of Entomology, 501 ASI Building University Park, PA 16802, USA*

*Author for correspondence:

Chloe P. Drummond

Address: The Pennsylvania State University, Department of Entomology,

501 ASI Building, University Park, PA 16802, USA

Phone: +1-814-865-1895

E-mail: czd429@psu.edu

Abstract

Plant trait evolution can be impacted by common mechanisms of genome evolution, including whole genome and small-scale duplication, rearrangement, and selective pressures. With the increasing accessibility of genome sequencing for non-model species, comparative studies of trait evolution among closely related or divergent lineages have supported investigations into plant chemical defense. Plant defensive compounds include major chemical classes such as terpenoids, alkaloids, and phenolics, and are used in primary and secondary plant functions. These include promotion of plant health, facilitation of pollination, defense against pathogens, and responses to a rapidly changing climate. We discuss mechanisms of genome evolution and use examples from recent studies to impress a stronger understanding of the link between genotype and phenotype as it relates to the evolution of plant chemical defense. We conclude with considerations for how to leverage genomics, transcriptomics, metabolomics, and functional assays for studying the emergence and evolution of chemical defense systems.

Abbreviations: Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), whole genome duplication (WGD).

Introduction

Plant chemical defense compounds are important for primary and secondary function and are also known to serve a variety of important roles, including pollinator attraction [1], herbivore and pathogen defense [2], and response to abiotic stress [3]. Some are thought to have been maintained due to increased fitness during a historical change in climate [e.g., 4,5]. Once present, some are believed to have evolved in concert with insects resulting in a diversity of compounds in plants [6–8]. Certain plant lineages feature certain biochemical classes due to co-evolutionary arms races with insects (e.g., butterflies and glucosinolates in *Brassica* plants [6], parsnip webworm and furanocoumarins in parsnip [8]), however the specific biochemicals used are not necessarily the same across species within a lineage. The evolutionary and ecological significance of plant chemical defense compounds necessitates investigation into their evolution. A stronger understanding of the relationship between genotype and phenotype is required to address the evolution of these important chemical defense compounds. It is becoming clearer that establishing how genome evolution impacts the evolution of these chemical defense compounds is integral to illuminating this relationship.

Foundational to genomic analysis of plant chemical defense evolution is an understanding of metabolite biosynthesis and characterization of genes underlying these pathways. Recent genomic studies have leveraged our understanding of plant biosynthetic pathways to target key gene families for comparative analyses, resulting in robust hypotheses for how genomic evolution (e.g., gene family expansions and genomic rearrangements) has influenced chemical defense evolution in certain lineages. For example, studies have revealed patterns in genomic evolution between lineages and related those patterns to the evolution of biosynthesis pathways (e.g., identifying lineage-specific, local duplication in an important

biosynthesis gene family). These hypotheses for how genomic evolution have influenced chemical defense evolution can (and should) be robustly honed, however, with the addition of transcriptomic and metabolomic data, as well as functional assays (Fig. 1). More recently, comparative transcriptomic analyses have been used to identify genes involved in chemical defense and localize their expression. In a similar way, comparative metabolomics has allowed for the identification and localization of metabolite profiles. Because such analyses enable identification of candidate genes, they have the potential to reveal whether genomic evolution has influenced plant adaptations specifically related to chemical defense. Finally, enzymatic assays have, perhaps most importantly, been used to assess protein function and help to corroborate the role of candidate genes or isoforms.

In this review, we discuss common mechanisms of plant genome architecture evolution, highlight recent studies that advance understanding of the effect of such mechanisms on the evolution of plant defensive chemicals (e.g., terpenoids, alkaloids, and phenolics), and discuss relevant methodological approaches. We do not attempt to address the effects of small-scale genomic mutation, such as allelic divergence within a lineage or post-transcriptional evolution (e.g., alternative splicing) as they relate to the evolution of plant chemical defense, nor do we attempt to address genome evolution induced by parasitism. Figure 1 reviews current multi-omic methods to investigate trait evolution from a genomic evolution perspective and is referenced in the following discussions.

We highlight at least three major classes of defensive chemicals: terpenoids, which are found commonly across nearly all plants and considered primary metabolites (e.g., abscisic acid, gibberellins, brassinosteroids, carotenoids, chlorophyll), though some are thought to be more specialized for interaction with biotic and abiotic stress (e.g., nepetalactone, menthol, taxol) [9];

alkaloids, which have been extensively studied in Solanaceae species [e.g., 10–12] for plant-insect interactions, are common stimulants (e.g., caffeine in coffee, tobacco, opium in poppy, and cocaine from coca), and whose mechanisms of toxicity are varied, including enzymatic alterations and inhibition of DNA synthesis and repair, and central nervous system alteration [13]; and phenolics, which are produced in plants in response to biotic and abiotic stress, are important in plant development (e.g., pigmentation), defense against pathogens, and defense against ultraviolet radiation [14]. Uncovering the genomic mechanisms underlying the evolution of defense compounds in different plant lineages is one step toward understanding the link between genotype and phenotype as it relates to plant chemical defense and the complex role of these metabolites in interactions with insects, ecological adaptations, and potential production of these compounds for human use.

Genomic architecture and the evolution of plant chemical defense

Whole genome duplication

Ancient whole genome duplications (WGD) occurred at the origin of angiosperms, the core-eudicots, and monocots [15–20]. Polyploids are thought to establish due to increased fitness in harsh environmental conditions [21]. Although fractionation may occur after WGD, whereby homeologous genomic regions undergo gene loss and diploidization, syntenic fingerprints of these ancient duplication events can still be found in the genomes of extant angiosperm lineages [22–24]. Many plant lineages have also undergone recent WGD via allo- or autopolyploidization [25]. The post-WGD process of neofunctionalization can enable new gene functions to arise, sometimes causing new phenotypes [26–28]. Figure 1A(1-4;6-7) depicts some of the genomic analyses that can be done to investigate the effects of WGD on trait evolution.

116 Post-WGD, evolutionary pressures can affect subgenomes differently and lead to
117 differential roles of subgenomes in the evolution of a trait. For example, in *Brassica juncea*
118 (Chinese mustard, Brassicaceae), there are two deletions, one in each of the two subgenomes,
119 with conserved variation between oil-use and vegetable-use varieties that are associated with
120 genes involved in abiotic stress response (*TGA1* and *HSP20*) [29]. In this case, mutations in both
121 subgenomes may have led to differential phenotypes in varieties selected for different features.
122 In another example, while structural variations are significantly more frequent in *B. juncea*
123 subgenome B than in subgenome A, GWAS analysis shows that two loci containing orthologs of
124 *MYB28*, a regulatory gene involved in glucosinolate biosynthesis, are associated with higher
125 glucosinolate content and are both found on subgenome A. This case reveals a potential
126 differential role of the subgenomes in expression of glucosinolates, which are selected for and
127 against in vegetable and oil-seed varieties, respectively, but are also important in herbivore and
128 pathogen defense [30,31].
129 Patterns of gene retention following genome multiplication can signify the importance of
130 multiplication events as they relate to the evolution of a particular phenotype. For example, while
131 the genus *Lavandula* (lavender, Lamiaceae) underwent two lineage-specific genome
132 duplications, genes retained following these duplication events were enriched for molecular
133 functions directly related to terpenoid biosynthesis, which may have been advantageous for
134 coping with the changing Mediterranean environment [4,32,33]. Similarly, in *Camellia* (tea,
135 Ericaceae), which shares a WGD event with 17 other families in the order Ericales, one, eight,
136 and four duplicated genes related to caffeine, catechin, and theanine biosynthesis respectively,
137 were retained post WGD. The duplicated gene copies were up-expressed in various tissues and
138 under different temperature treatments, suggesting the importance of both copies in biosynthesis

of these compounds. In rhododendron and persimmon, however, which do not produce caffeine or theanine, but that share the WGD event, the caffeine-related gene duplication was not retained in either of the species, only three and one catechin-related gene duplications were retained respectively, and only one theanine-related gene duplication was retained in rhododendron. [34]. These genes may perform different functions in rhododendron and persimmon, and differential retention may have played a role in the evolution of caffeine biosynthesis in tea, which is important for tea flavor and may play a role in pollinator interactions [35].

Genome linkage mapping assigns subgenomes to known progenitors of a polyploid, which can be useful for assessing the evolution of a trait when genetic constituents of each progenitor are required for the new trait [5]. For example, GWAS analysis identified two candidate loci responsible for the cyanogenesis phenotype in polyploid *Trifolium repens* (white clover, Fabaceae): one corresponding to the known *Ac/ac* gene cluster that controls the presence of cyanogenic glucosides, and one corresponding to the known *Li/li* gene cluster that controls the presence of their hydrolyzing enzyme, linamarase [5]. The dominant alleles of both loci are required for the cyanogenesis phenotype because the recessive alleles are deletions of the genes. Through genetic mapping, the GWAS loci containing *Ac/ac* and *Li/li* were found in the progenitor *T. occidentale* and *T. pallescens* subgenomes respectively. In addition, the sequence of the *Ac/ac* locus of *T. repens* shared more similarity with *T. occidentale* than *T. pallescens.* Although the GWAS locus containing Li/li locus was placed in the *T. pallescens* subgenome, the Li/li sequence was not found in the *T. pallescens* genome. The authors suggest that the genotype of the sequenced individual was li/li and thus missing the locus, or that present-day *T. pallescens* has completely lost the Li/li locus.This example illustrates a dual inheritance of the cyanogenic trait from non-cyanogenic progenitors via allopolyploidy.

Because the order and clustering of genes required for a certain phenotype can be retained after WGD, genome duplication events that distinguish lineages can be used to estimate the relative timing of the development of a phenotype. For example, the evolution of the iridoid pathway in *Nepeta* (catnip, Lamiaceae) seems to predate a *Nepeta*-specific WGD event, based on syntenic clustering of non-homologous iridoid biosynthesis genes (*ISY*, *NEPS*, and *MLPL*) in *N. cataria* (a tetraploid with 2 clusters) and *N. mussarii* (a diploid with 1 cluster) [36]. This suggests that iridoids, important for plant defense and multi-species interactions, evolved via a conserved iridoid biosynthesis pathway in this group [37].

Local gene duplication and loss

Small-scale duplications, including local or tandem gene duplications, occur frequently within plant genomes [38–40]. These small-scale duplication events can arise from transposable elements (TEs), slipped strand mispairing, or unequal crossing over during meiosis, and can account for gene family expansions within lineages. Local gene loss may occur via TEs that interrupt a gene or repress expression, slipped strand mispairing that excises DNA, or through pseudogenization via accumulation of mutations in a gene that result in nonsense mutations or frameshifts. It is possible that gene loss is more commonly facilitated by fractionation, or DNA excision, rather than gene-by-gene pseudogenization of formerly functioning genes [41]. The fate of genes post-small-scale duplication mirror that of genes post-WGD, where processes such as neofunctionalization can promote new gene function, and thus play an important role in trait evolution. In addition, a co-regulated tandem array can impact levels of gene expression and influence trait evolution.

To investigate the role of small-scale duplications in trait evolution, lineages with or without a trait can be investigated for gene family expansions or contractions (Fig. 1A(1)). In addition, whether local duplications are shared or lineage-specific can inform whether the evolution of a trait is conserved or is evolving in a lineage-specific manner. For example, lineage-specific evolution in alkaloid biosynthesis seems to have played a major role in *Zanthoxylum* (Sichuan pepper, Rutaceae), which may use alkaloids for insect defense [42]. The Sichuan pepper genome is composed of over 50% transposable elements (TEs) (1.72Gb out of the reported 2.63Gb assembly length) and 16,796 in-tact long terminal repeats (LTRs) were identified in Sichuan pepper compared to 371 in the close relative *Citrus sinensis.* 2,816 protein-coding genes were inserted into gene regions or 2kb flanking regions by long terminal repeats (LTRs) and the protein-coding genes are enriched for functions such as “defense response”, “stilbene biosynthetic process”, and “coumarin biosynthetic process”. This suggests that TEs might play an important role in the expansion of genes used for chemical defense functions in Sichuan pepper. In addition, key candidate genes for GX-50 biosynthesis (*TYDC*, *3OHase*, *PAL*, *OMT*, and *BAHD-AT*) and sanshool biosynthesis (*BCAD*, *SCPL-AT*, and *FAD*) are expanded in the Sichuan pepper genome compared to citrus relatives. Additionally, enriched functions of Sichuan pepper-specific gene families and gene family expansions suggest the importance of local duplications on the evolution of secondary metabolite biosynthesis in the genus. For example, genes from families specific to *Zanthoxylum* are enriched for KEGG pathways related to “plant-pathogen interaction” and significantly expanded gene families are enriched for GO terms including stress resistance related to “defense response” and biosynthetic processes related to alkaloids, stilbenes, and coumarins.

In *Scutellaria* (skullcaps, Lamiaceae), key elements of flavonoid biosynthesis seem to be conserved within the genus, with some possible lineage-specific evolution [43]. For example, *Scutellaria*-specific genes are enriched for domains related to secondary metabolite biosynthesis, such as cytochrome P450s and O-methyltransferase, perhaps signifying the important role of secondary metabolite biosynthesis in the genus. In addition, tandem expansions of flavonoid biosynthesis genes that function early in the pathway occurred after speciation of two *Scutellaria* species (*PAL* and *CHS*, and *4CL* in *S. baicalensis* and *S. barbata* respectively) suggesting that the flavonoid biosynthesis pathway has evolved in a lineage-specific manner in this genus. The *CYP* gene family, including *CYP82D1-9*, which catalyzes the formation of baicalein and scutellarein, is tandemly duplicated in both species, suggesting conservation of this biosynthesis pathway. Finally, evolution of flavone biosynthesis is potentially conserved between the two species, evidenced by a duplication of *4CLL*, which enables biosynthesis of 4′-deoxyflavones, occurring prior to the *S. baicalensis* and *S. barbata* speciation event, and a tandem duplication of a flavone biosynthesis gene *FNSII1-FNSII2* found in both species.

In a final example, *Rubus chingii* (Fu-pen-zi, Rosaceae) produces abundant hydrolyzable tannins (HTs), which contribute to biotic and abiotic stress response. In contrast, its relative *Malus x domestica* (apple, Rosaceae), does not produce abundant HTs. A collinear tandem duplication of three genes involved in HT biosynthesis or degradation (*CXE*, *UGT*, and *SCPL*) were found in *R. chingii* with 11, eight, and six copies of *CXE*, *UGT*, and *SCPL*, respectively [44]. The region of this tandem array is found syntenically in the apple genome on four chromosomes. Interestingly, key *CXE* family genes (*TA*s) are lost in the apple genome, which may have resulted in a lack of HTs, but the low levels of HTs produced in apple may be the result of the homologous expansion of this tandem array.

Genomic rearrangements and transcriptional regulation

In addition to local duplications, genome rearrangements can occur in plants in the form of chromosomal rearrangements during polyploidization [45] or movement of co-adapted loci into colocalized gene clusters [46–48]. Metabolic gene clusters are physically clustered genes that may include one or more operons that act together in metabolite biosynthesis. The formation of these clusters is hypothesized to be due to selective pressure for coinheritance, where colocalization reduces the likelihood of loss of important individual genes during recombination [47,49]. Another hypothesis for the formation of metabolic gene clusters is the efficiency and likelihood of complete co-expression of genes required for metabolite biosynthesis [47,50]. A hypothesis for the maintenance of intact metabolic gene clusters is that there is a strong selective pressure to reduce toxic metabolite intermediates in a biosynthesis pathway that can occur when a cluster is no longer intact (e.g., disrupted by mutation) [47,51,52]. Because metabolic gene clusters and neofunctionalization of tandem duplications are often co-regulated, genomic arrangement through synteny or collinearity can influence the evolution of a trait (Fig. 1A(5)) [53].

For example, consistent with findings in other species [54–56], terpenoid biosynthesis genes are physically clustered in lavender and some clusters fall into the same co-expression networks, suggesting coinheritance and co-regulation of terpenoid biosynthesis [4]. This might promote terpenoid production in the genus, while potentially providing the benefit of less toxic intermediates [47]. In another example, like other vascular species such as rice and barnyard grass, *Calohypnum plumiforme* (bryophyte in Hypnaceae) produces momilactones, which are diterpenoids used in pathogen defense and allelopathic interactions. A biosynthesis gene cluster

(BGC) of important genes in momilactone biosynthesis (two cytochrome P450s, one *CpDTC1*/*HpDTC1* and one "dehydrogenase momilactone A synthase") was found in *Calohypnum* and induced upon stress exposure [57]. When compared with other plant genomes, this BGC was only found in the rice and barnyard grass, but they were not in syntenic regions. This study suggests not only the importance of BGCs in momilactone biosynthesis, but also presents a case of independent evolution of a BGC.

A final example of genomic rearrangement as it relates to chemical defense evolution comes from the post-WGD fission and fusion events and formation of a benzylisoquinoline alkaloid BGC of 15 genes in the genus *Papaver* (poppy, Papaveraceae) [58]. Poppy produces the benzylisoquinoline alkaloid compounds morphinan (morphine) and noscapine in response to mechanical damage, and these alkaloids share a biosynthesis pathway that branches to produce each compound [59]. *Papaver somniferum* and *P. setigerum* are sister to *P. rhoeas*, and the two species share a WGD and produce relatively higher levels of morphinan and noscapine than *P. rhoeas.* A model of chromosomal fission and fusion events post-WGD reveals that the genes around the chromosomal rearrangement breakpoints are enriched for KEGG pathways related to isoquinoline and indole alkaloid biosynthesis. This suggests that the shared WGD event and its subsequent genomic rearrangements may have influenced the co-regulation of genes involved in chemical defense evolution. The formation of a benzylisoquinoline alkaloid BGC that is shared between *P. somniferum* and *P. setigerum* and not present in *P. rhoeas* is another example of the influence of genomic rearrangement on chemical defense evolution. Genes in the BGC exhibit higher gene expression than their ancestral copies, suggesting that the formation of the BGC has increased benzylisoquinoline alkaloid expression within poppy. Based on syntenic analysis of each of the three species, the STORR gene, which is a fusion of two genes and is involved in

morphinan biosynthesis, was present in the two BGC-containing species as the result of a translocation event. In the BGC-containing species, the post-donor locus is syntenic with the pre-donor locus but does not contain the two non-adjacent STORR genes, and the post-recipient locus is syntenic with the pre-recipient locus but contains the fused STORR gene. This is another example of the influence of post-WGD rearrangement on the evolution of chemical defense. In addition, the authors suggest that the STORR gene fusion prevents accumulation of toxic intermediates. The remainder of the genes in the BGC may have been incorporated via non-tandem small-scale duplication based on the lack of synteny or co-localization of the genes and their original copies. However, the authors caution this interpretation, citing the possibility of tandem duplication with subsequent deletion. This evolutionary analysis and additional tests of gene expression and gene regulation reveal that overall, the evolution of this BGC was critical to the evolution benzylisoquinoline alkaloid biosynthesis in poppy.

Co-option and independent evolution

When genes with a pre-existing function are recruited for a new function, this is known as co-option. Gene duplications, whether via WGD or small-scale duplications, are thought to facilitate co-option [60,61]. Through this process, similarly to neofunctionalization as described above, newly duplicated gene copies can be released from selection pressure, allowing for the fixation of mutations that lead to the emergence of modified or new biological pathways or traits [62,63]. An important evolutionary pattern is one in which modified or new phenotypes evolve independently in distant lineages. While the terms parallel and convergent evolution remain contentious, a developmental biology understanding is that they represent phenotypes that evolve

297 via the same or different genetic and regulatory pathways, respectively [64–66]. The genomic
298 mechanism of convergence via co-option shapes the patterns of trait evolution found in plants.
299 An example of co-option as it relates to chemical defense evolution comes from another
300 *Trifolium repens* example. Its progenitor, *T. occidentale*, has the *Ac/ac* locus, which controls
301 presence of cyanogenic glucosides, but lacks the *Li/li* locus, which controls the presence of their
302 hydrolyzing enzyme, linamarase [5]. This suggests that *T. occidentale* uses cyanogenic
303 glucosides for other metabolic functions and perhaps the *Ac/ac* locus and cyanogenic glucosides
304 were co-opted in for chemical defense in the presence of the *Li/li* locus in *T. repens*.
305 An example of convergence as it relates to chemical defense evolution comes from
306 *Hypericum perforatum* (St. John's Wort, Hypericaceae) in the biosynthesis of hyperforin, a
307 polycyclic polyprenylated acylphloroglucinol (PPAP) that has thus far been identified only in
308 this genus, is likely used for plant defense, and has antidepressant activity [67–69]. Two BGCs
309 identified in *H. perforatum* contain copies of genes confirmed to be involved in biosynthesis of
310 the hyperforin precursor phloroisobutyrophenon (PIBP) [67]. The two BGCs have different
311 expression and localization profiles and might be regulated for different functions or contribute
312 to different combinations of PPAP compounds. Syntenic and substitution rate divergence time
313 analyses revealed that BGC1 and BGC2 evolved via different duplications and genomic
314 rearrangements, and that while BGC1 is likely shared across the *Hypericum* order Malpighiales,
315 the formation of BGC2 is more recent and is likely only shared by a few species of *Hypericum.*
316 This points to potential independent evolution of PPAP biosynthesis within Malpighiales given
317 the lineage-specific pathway found in *Hypericum.* Specifically, the evolutionary model of BGC1
318 is either a shared origin of a two-gene cluster between the *Hypericum* order Malpighiales and the
319 *Arabidopsis* order Brassicales or independent evolution of the two-gene cluster in these orders.

This is followed by recruitment of two additional genes in the common ancestor of Malpighiales. An enzymatically active syntenic homolog of BGC1 in *Mesua ferrea* (ironwood, Calophyllaceae), a Malpighiales relative that also produces PPAPs, points to this recruitment in the common ancestor of Malpighiales. Additional evidence for this timing is the syntenic homologs of BGC1 in non-PPAP producing Malpighiales relatives that contain combinations of the same genes in BGC1, but only one or the other of two required genes for PPAP biosynthesis. The presence of these clustered genes across Malpighiales suggests that it evolved in a common ancestor and has since undergone lineage-specific gene loss or duplication. The evolutionary model of BGC2 is a co-occurring duplication of one region of BGC2 containing one gene of the cluster, and a duplication of the region of BGC1 containing the remaining genes, followed by genomic rearrangement. These co-occurring duplications occurred after the split between *Mesua* and *Hypericum*, thus suggesting potential convergent evolution within *Hypericum* of PPAP function and biosynthesis.

A final example of convergence as it relates to metabolite evolution comes from the blood-red nectar pigments found in the gecko-pollinated *Nesocodon mauritianus* (Campanulaceae) and hummingbird-visited *Jaltomata herrerae* (Solanaceae). The red coloration is derived from an alkaloid called nesocodin. Two of the enzymes used in its synthesis and identified in the nectars of *N. mauritianus* and *J. herrerae* (carbonic anhydrases and alcohol oxidases) have low sequence similarity between the two plant species (~42% and ~21% identity, respectively). There are also more closely related homologs of the carbonic anhydrases elsewhere in each others' genomes, suggesting that each species uses a different copy [70]. In addition, the alcohol oxidases found in the nectar from each species are not from the same enzyme family (GMC flavonenzyme oxidoreductase in *N. mauritianus* and berberine-bridge

family within the flavin adenine dinucleotide/flavin mononucleotide (FAD/FMN)-containing dehydrogenase superfamily in *J. herrerae*). These lines of evidence suggest that the two species have converged on this phenotype under their own selective pressures.

Tests of genomic evolution related to chemical defense

Some of the processes highlighted above, such as whole genome duplication and gene family expansion and loss, do not necessarily result in evolution of a trait. For example, a functional enrichment analysis that suggests a biological activity associated with a gene expansion [e.g., 42,43] can only serve as hypotheses of gene activity and function. Additional tests of gene activity and function should be conducted to make further assessments of genomic evolution of a trait, such as whether a lineage-specific gene family expansion contains a candidate gene copy known to be involved in trait expression. In the context of trait evolution, comparative transcriptomics is used to identify copies of genes or gene networks that are upregulated and their location, and thus identify candidate genes/networks for trait expression (Fig. 1B). In this same context of trait evolution, comparative metabolomics is used to identify the location and quantity of metabolites related to a trait of interest, thus corroborating the hypotheses of candidate genes/networks involved in trait expression (Fig. 1C). Importantly, mismatches between gene upregulation and metabolite presence or quantity can illuminate an incorrect hypothesis about which genes or gene families are involved in trait expression. Enzymatic analysis can test the activity of a protein from a candidate gene to further corroborate that gene's involvement in trait expression (Fig. 1D). Finally, selection tests can be conducted on gene family phylogenies to assess whether positive or purifying selection has contributed to the

365 evolution of lineage-specific, local expansions or candidate genes related to the trait of interest
366 (Fig. 1A(1)).

367 For example, in *Lavandula*, expression of terpenoid biosynthesis genes generally
368 coincides with the presence of terpenoids in the same tissues, revealing candidate genes for
369 terpenoid biosynthesis [4]. Most gene copies of expanded terpenoid biosynthesis gene families
370 such as terpenoid synthases, which includes *TPS-b* responsible for monoterpene biosynthesis, are
371 highly expressed in the glandular trichomes where the volatile terpenoids for essential oils are
372 produced. In addition, genes whose expression was positively correlated with the presence of
373 linalool, linalyl acetate, and lavandulyl acetate, the primary terpenoids in lavender flowers, were
374 mostly found in flowers and glandular trichomes. In another example, the expression of *LaAAT*
375 and quantity of lavendulyl acetate coordinately fluctuated across flower development.

376 In an example from *Nepeta* [36], candidate genes responsible for the biosynthesis of 8OG
377 (*GES*, *G8H*, and *HGO*), the iridoid precursor, are expressed across tissues in *Nepeta*, but very
378 lowly expressed in *Hyssopus*, which aligns with the lack of iridoids in *Hyssopus*. In addition,
379 expression levels of *NEPS* and *MLPL* (both involved in iridoid biosynthesis) were correlated
380 with tested enzymatic activity in *Nepeta* accessions with distinct nepetalactone stereo-
381 chemotypes, suggesting that specific *NEPS* and *MLPL* genes are responsible for creating each of
382 the nepetalactone stereoisomers. Finally, iridoid evolution in *Nepeta* is described by an ancestral
383 duplication in *PRISE* (progesterone 5β-reductase/iridoid synthase (ISY) family), which had only
384 minor ISY enzymatic activity, followed by positive selection that formed functioning ISY
385 enzymes. *PRISE* and *NEPS* phylogenetic dating and concurrent timing of positive selection in
386 *ISY* and diversification of *NEPS* suggest that the evolution of their catalytic activity was in
387 concert.

In *Zanthoxylum*, candidate genes involved in GX-50 and sanshool alkaloid biosynthesis were identified in the husk, given the correlation of the expression of alkaloid biosynthesis genes and the presence of alkaloids in that tissue [42]. In one example, the husk had the highest GX-50 content and highly expressed members of five GX-50 gene families that belong to a single co-expression module. In another example, the husk had the highest content of hydroxy-β-sanshool, which is converted into hydroxy-α-sanshool, the compound known for its numbing property, and highly expressed 18 copies of *BCAD*, a gene family involved in sanshool biosynthesis. Interestingly, these *BCADs* and one copy of *SCPL-AT*, another gene family involved in sanshool biosynthesis, were in the same co-expression module that is closely related to GX-50 biosynthesis, suggesting possible co-expression of the two alkaloid families. Husk-specificity of alkaloid metabolites and alkaloid biosynthesis gene expression suggests that this tissue played a role, perhaps via insect interactions, in the evolution of these compounds in Sichuan pepper.

In an example of flavonoid biosynthesis evolution in *Scutellaria*, tissue-specific metabolomics and transcriptomics identified the location of metabolites and candidate genes involved in flavonoid biosynthesis, while misalignment in these data established a hypothesis of functional divergence between *S. baicalensis* and *S. barbata* [43]. Duplication of *4CLL*, which enables biosynthesis of 4′-deoxyflavones, occurred prior to the *S. baicalensis* and *S. barbata* speciation event. One of the copies of the ancestral *4CLL* duplication is not expressed in *S. baicalensis* or *S. barbata*, suggesting that the duplication enabled the inherited biosynthesis of 4′-deoxyflavones. In addition, copies of the scutellarein biosynthesis gene, *C4H*, found early in the pathway were identified as candidate genes for producing scutellarin, the glycoside of scutellarein, in the stem, leaf, and flower in both species. Copies of the flavonoid biosynthesis genes *CHS* and *CYP450* were identified as candidate genes for producing baicalein, norwogonin,

411 wogonin, and their glycosides in the roots of both species. Expression of tandemly duplicated
412 *CHS* genes specific to *S. baicalensis* supported the hypothesis that flavonoid biosynthesis in *S.*
413 *baicalensis* is affected by the evolution of tandem arrays. Surprisingly, however, in *S. barbata*
414 expression levels of *CYP82D1* and *CYP82D2* misaligned with the metabolite profile, which
415 suggested functional divergence of hydroxylation and evolutionary divergence in the flavonoid
416 pathway between the species. Finally, Ka/Ks values between orthologous gene pairs between the
417 two species indicates purifying selection, suggesting conservation in flavone biosynthesis in
418 *Scutellaria*.
419
420 **Future considerations**
421 Recent studies have creatively and elegantly pushed the limits of identifying the genomic
422 fingerprints of plant chemical defense evolution [e.g., 4,5,29,34,36,42–44,57,58,67,70–80].
423 Multiple mechanisms of genomic evolution, alongside selective pressures from the important
424 role that these compounds play in primary and ecological functions, work in concert to produce
425 the evolutionary patterns of plant chemical defense observed. Identifying candidate mechanisms
426 of genomic evolution is only the first step in developing a chemical defense evolution hypothesis
427 (Fig. 1A). Recent studies have combined genome, transcriptome, metabolome, and functional
428 enzymatic data to further corroborate and test hypotheses (Fig. 1).
429 Perhaps not surprisingly, the interplay of biological and chemical analysis is integral to
430 fully understanding the biological system of chemical defense (where, how, and when of defense
431 compounds) and uncovering the evolutionary pathway (where, how, and when of genes, their
432 regulation, and selective pressures) that led to a lineage's current system. Namely, methods in
433 functional genomics, including enzymatic assays and gene knockouts of candidate genes integral

434 to plant biosynthesis pathways are required to test hypotheses [e.g., 36,44,70,72]. This next step
435 of functional analysis has the significant potential to define the relationship between genotype
436 and phenotype, as well as improve understanding of how chemical defense systems emerge and
437 evolve (Fig. 1D).
438 Comparisons between studies can generate hypotheses of shared or unique mechanisms
439 of genome evolution across plant lineages that contribute to chemical defense evolution [e.g.,
440 9,81,82]. However, future studies that investigate distantly related species with similar
441 biochemical profiles [70,83] but in a genomic context [e.g., 36] could draw more robust, direct
442 comparisons across plants by testing questions of parallel or convergent evolution. For example,
443 these studies could test whether the same genes or BGCs have been co-opted for biosynthesis
444 function. In addition, to fully understand plant chemical defense evolution from a genomic
445 perspective, the role of selection on genome evolution and vice versa needs to be uncovered.
446 This may require interdisciplinary studies between evolution and ecology whereby the
447 hypotheses of how genomic evolution has influenced the evolution of metabolite biosynthesis
448 are tested within an ecological context. For example, sister species that occupy divergent
449 ecological niches should be compared for differences in patterns of genomic evolution. In a
450 similar way, recent phylogenetic diversification within a genus, correlated with shifts in
451 biochemical expression or regulation and shifts in ecological context can also point to the role of
452 selection in metabolite biosynthesis evolution [e.g., 84] (Fig. 1A(6)). In instances where
453 polyploid plants can be bred, direct experimentation of the effects of polyploidy on functional
454 traits and fitness can be done [e.g., 85]. More feasibly, positive selection tests on candidate genes
455 involved in metabolite biosynthesis have shown whether selective pressures have played a part in
456 the evolution of metabolite biosynthesis [e.g., 36,43] (Fig. 1A(1)). However, these tests are often

done in the context of distant relatives. If these tests are carried out in a comparative way between closely related species occupying divergent niches, this may provide more robust insight into what ecological factor(s) contribute to an identified selective pressure and perhaps led to adaptation to biotic or abiotic conditions. This comparative investigation would be bolstered with evidence of adaptation from ecological common garden studies that compare fitness under different environmental conditions. These future analyses would enrich understanding of the reciprocal or cyclical impacts of genome evolution on adaptation and selection on genome evolution.

Acknowledgements

This work is supported by AFRI Post Doctoral Fellowships grant no. 2020-67012-31896/project accession no. 1023201 from the USDA National Institute of Food and Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture. This material is based upon work supported by the National Science Foundation under Grant No. NSF DEB 2030871 and 2112586. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

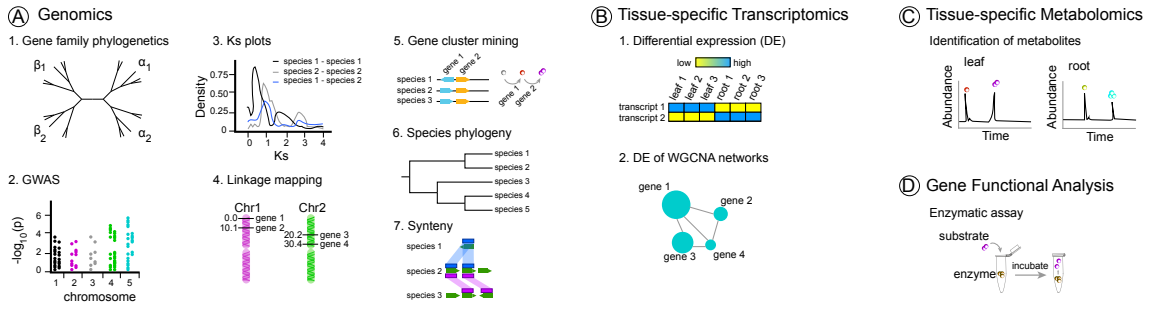

Figure 1. A workflow of investigating secondary metabolite evolution using genomics, transcriptomics, metabolomics, and gene functional analysis. **A. Genomics**: A1. Gene Family Phylogeny: a. Test gene family expansions and/or contractions; b. (with B1 and/or B2) Map DE genes or networks (these are putative candidate genes); c. Identify potential independent evolution of putative candidate genes; d. Date the evolution of gene families to test if important families evolved concurrently; e. Ka/Ks to test for positive selection in branches leading to candidate genes; f. (with D1) reconstruct ancestral sequences to test chronology and evolution of enzymatic activity. A2. GWAS: a. Identify candidate loci associated with a polymorphic phenotype. A3. Ks Plots: a. Identify WGD events; b. Date specific duplications of interest to either pre- or post-WGD. A4: Linkage Mapping: a. (If find WGD in A3) Identify parental inheritance of relevant genomic material. A5: Gene Cluster Mining: a. Identify biosynthesis gene clusters (BGCs); b. (With B1 and/or B2) Confirm putative cis-regulation of BGCs; c. (With A7) Identify whether BGCs are shared (ancestral/syntenic) or lineage specific. A6: Species Phylogeny: a. (If find WGD in A3) Map WGD events; b. (If find expansions and/or contractions in A1) Map change in expansions and/or contractions of gene families. A7: Synteny: a. Identify shared (syntenic or small-scale and syntenic) vs. lineage specific (only small-scale) duplications. **B: Tissue-Specific Transcriptomics**: B1. DE: a. Identify where secondary metabolite biosynthesis occurs (can combine with B2 and/or C1). B2. DE of WGCNA: a. Identify which genes are co-expressed; b. Identify where biosynthesis occurs (can combine with B1 and/or B1). **C: Tissue-Specific Metabolomics**: C1. Identification of Metabolites: a. Identify where secondary metabolite biosynthesis occurs (can combine with B1 and/or B2); b. (with B1 and/or B2) Identify potential functional divergence of genes or gene networks based on mismatches in metabolite and transcriptome profiles. **D: Gene Functional Analysis**: D1. Enzymatic Assay: a. Confirm function of candidate genes or BGCs; b. (with A1e) Confirm function of genes under selection.

References

1. Dudareva N, Pichersky E: **Biochemical and Molecular Genetic Aspects of Floral Scents**. *Plant Physiology* 2000, **122**:627–634.
2. Bennett RN, Wallsgrove RM: **Secondary metabolites in plant defence mechanisms**. *New Phytologist* 1994, **127**:617–633.
3. Akula R, Ravishankar GA: **Influence of abiotic stress signals on secondary metabolites in plants**. *Plant Signaling & Behavior* 2011, **6**:1720–1731.
4. *Li J, Wang Y, Dong Y, Zhang W, Wang D, Bai H, Li K, Li H, Shi L: **The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpenoid biosynthesis**. *Horticulture Research* 2021, **8**:53.

The authors compare the genome of lavender to those of four additional mint family species and five outgroups to investigate terpenoid biosynthesis evolution in the mint family. The authors show that lavender has undergone lineage-specific evolution of terpenoid biosynthesis-related genes. They also utilize metabolomic and transcriptomic data and discover that terpenoid biosynthesis genes are physically clustered in lavender and some clusters fall into the same co-expression networks, suggesting coinheritance and co-regulation of terpenoid biosynthesis. This might promote terpenoid production in the genus, while potentially providing the benefit of less toxic intermediates.

5. Olsen KM, Goad DM, Wright SJ, Dutta ML, Myers SR, Small LL, Li LF: **Dual-species origin of an adaptive chemical defense polymorphism**. *New Phytologist* 2021, **232**:1477–1487.
6. Ehrlich PR, Raven PH: **Butterflies and Plants: A Study in Coevolution**. *Evolution (N Y)* 1964, **18**:586–608.
7. Janz N: **Ehrlich and Raven revisited: mechanisms underlying codiversification of plants and enemies**. *Annual Review of Ecology, Evolution, and Systematics* 2011, **42**:71–89.
8. Berenbaum M: **Coumarins and caterpillars: a case for coevolution**. *Evolution (N Y)* 1983, **37**:163–179.
9. Zhou F, Pichersky E: **More is better: the diversity of terpene metabolism in plants**. *Current Opinion in Plant Biology* 2020, **55**:1–10.
10. Shonle I, Bergelson J: **Evolutionary ecology of the tropane alkaloids of Daturs stramonium L. (Solanaceae)**. *Evolution (N Y)* 2000, **54**:778–788.
11. Hlywka JJ, Stephenson^ GR, Sears MK, Yada RY: **Effects of Insect Damage on Glycoalkaloid Content in Potatoes (Solanum tuberosum)**. *J Agric Food Chem* 1994, **42**:2545–2550.
12. Kirwa HK, Murungi LK, Beck JJ, Torto B: **Elicitation of Differential Responses in the Root-Knot Nematode Meloidogyne incognita to Tomato Root Exudate Cytokinin, Flavonoids, and Alkaloids**. *Journal of Agricultural and Food Chemistry* 2018, **66**:11291–11300.
13. Matsuura HN, Fett-Neto AG: **Plant Alkaloids: Main Features, Toxicity, and Mechanisms of Action**. In *Plant Toxins*. . Springer Netherlands; 2015:1–15.
14. Cheynier V, Comte G, Davies KM, Lattanzio V, Martens S: **Plant phenolics: Recent advances on their biosynthesis, genetics, and ecophysiology**. *Plant Physiology and Biochemistry* 2013, **72**:1–20.
15. Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, et al.: **One thousand plant transcriptomes and the phylogenomics of green plants**. *Nature* 2019, **574**:679–685.
16. Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, et al.: **A genome triplication associated with early diversification of the core eudicots**. *Genome Biology* 2012, **13**:R3.
17. Vekemans D, Proost S, Vanneste K, Coenen H, Viaene T, Ruelens P, Maere S, van de Peer Y, Geuten K: **Gamma Paleohexaploidy in the Stem Lineage of Core Eudicots: Significance for MADS-Box Gene and Species Diversification**. *Molecular Biology and Evolution* 2012, **29**:3793–3806.

18. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al.: **Ancestral polyploidy in seed plants and angiosperms**. *Nature* 2011, **473**:97–100.
19. Tang H, Bowers JE, Wang X, Paterson AH: **Angiosperm genome comparisons reveal early polyploidy in the monocot lineage**. *Proceedings of the National Academy of Sciences* 2010, **107**:472–477.
20. Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J: **Widespread Whole Genome Duplications Contribute to Genome Complexity and Species Diversity in Angiosperms**. *Molecular Plant* 2018, **11**:414–428.
21. van de Peer Y, Ashman T-L, Soltis PS, Soltis DE: **Polyploidy: an evolutionary and ecological force in stressful times**. *The Plant Cell* 2021, **33**:11–26.
22. Wolfe KH: **Yesterday's polyploids and the mystery of diploidization**. *Nature Reviews Genetics* 2001, **2**:333–341.
23. Wendel JF, Lisch D, Hu G, Mason AS: **The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation**. *Current Opinion in Genetics & Development* 2018, **49**:1–7.
24. Freeling M, Scanlon MJ, Fowler JE: **Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences**. *Current Opinion in Genetics & Development* 2015, **35**:110–118.
25. Wendel JF, Jackson SA, Meyers BC, Wing RA: **Evolution of plant genome architecture**. *Genome Biology* 2016, **17**:37.
26. Conant GC, Wolfe KH: **Turning a hobby into a job: How duplicated genes find new functions**. *Nature Reviews Genetics* 2008, **9**:938–950.
27. de Smet R, van de Peer Y: **Redundancy and rewiring of genetic networks following genome-wide duplication events**. *Current Opinion in Plant Biology* 2012, **15**:168–176.
28. Taylor JS, Raes J: **Duplication and Divergence: The Evolution of New Genes and Old Ideas**. *Annual Review of Genetics* 2004, **38**:615–643.
29. **Yang J, Wang J, Li Z, Li X, He Z, Zhang L, Sha T, Lyu X, Chen S, Gu Y, et al.: **Genomic signatures of vegetable and oilseed allopolyploid Brassica juncea and genetic loci controlling the accumulation of glucosinolates**. *Plant Biotechnology Journal* 2021, **19**:2619–2628.
The authors investigate a polyploid species, *Brassica juncea*, and show how subgenomes can evolve under different evolutionary pressures post-WGD. They show mutations in each of the two subgenomes, with conserved variation between oil-use and vegetable-use varieties. These mutations may have led to the observed differential phenotypes. They also use GWAS analysis correlating loci with glucosinolate content to show that there is a potential differential role of the subgenomes in expression of glucosinolates.
30. Kim JH, Jander G: **Myzus persicae (green peach aphid) feeding on Arabidopsis induces the formation of a deterrent indole glucosinolate**. *The Plant Journal* 2007, **49**:1008–1019.
31. Bednarek P, Piślewska-Bednarek M, Svatoš A, Schneider B, Doubský J, Mansurova M, Humphry M, Consonni C, Panstruga R, Sanchez-Vallet A, et al.: **A Glucosinolate Metabolism Pathway in Living Plant Cells Mediates Broad-Spectrum Antifungal Defense**. *Science (1979)* 2009, **323**:101–106.
32. Lebel S, Schellenbaum P, Walter B, Maillot P: **Characterisation of the Vitis vinifera PR10 multigene family.** *BMC Plant Biol* 2010, **10**:184.

33. Stortenbeker N, Bemer M: **The SAUR gene family: the plant's toolbox for adaptation of growth and development**. *Journal of Experimental Botany* 2019, **70**:17–27.
34. *Wang Y, Chen F, Ma Y, Zhang T, Sun P, Lan M, Li F, Fang W: **An ancient whole-genome duplication event and its contribution to flavor compounds in the tea plant (Camellia sinensis)**. *Horticulture Research* 2021, **8**:1–12.
The authors demonstrate that the WGD event in *Camellia sinensis* whose origin is controversial, is shared by at least 17 other Ericales families. They also demonstrate the importance of this WGD in caffeine, catechin, and theanine biosynthesis evolution in tea, showing that unlike other species that experienced this WGD event, tea retained more duplicated genes related to secondary metabolite pathways and these genes are expressed in different tissues.
35. Wright GA, Baker DD, Palmer MJ, Stabler D, Mustard JA, Power EF, Borland AM, Stevenson PC: **Caffeine in Floral Nectar Enhances a Pollinator's Memory of Reward**. *Science (1979)* 2013, **339**:1202–1204.
36. **Lichman BR, Godden GT, Hamilton JP, Palmer L, Kamileen MO, Zhao D, Vaillancourt B, Wood JC, Sun M, Kinser TJ, et al.: **The evolutionary origins of the cat attractant nepetalactone in catnip**. *Science Advances* 2020, **6**:eaba0721.
The authors combined genomic, metabolomic, transcriptomic, enzymatic data as well as phylogenetic and selection analysis to generate a model for the evolution of volatile iridoids in *Nepeta*, which is the only genus in the subfamily Nepetoideae to produce nepetalactones. The final model describes a gene duplication of the *PRISE* ancestor, which had only minor ISY enzymatic activity, followed by positive selection that formed functioning ISY enzymes. *PRISE* and *NEPS* phylogenetic dating and the concurrent timing of positive selection identified in *ISY* and diversification of *NEPS* suggest that the evolution of their catalytic activity was in concert. Finally, the locations of gene clusters and inferred location of ancestral genes suggests that enzymatic activity of biosynthesis genes (including *ISY*) evolved first, followed by the formation of gene clusters, and then speciation of the two *Nepeta* species.
37. Biere A, Marak HB, van Damme JMM: **Plant chemical defense against herbivores and pathogens: generalized defense or trade-offs?** *Oecologia* 2004, **140**:430–441.
38. Defoort J, van de Peer Y, Carretero-Paulet L: **The evolution of gene duplicates in angiosperms and the impact of protein-protein interactions and the mechanism of duplication**. *Genome Biology and Evolution* 2019, **11**:2292–2305.
39. van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K: **The flowering world: a tale of duplications**. *Trends in Plant Science* 2009, **14**:680–688.
40. Maere S, de Bodt S, Raes J, Casneuf T, van Montagu M, Kuiper M, van de Peer Y: **Modeling gene and genome duplications in eukaryotes**. *Proceedings of the National Academy of Sciences* 2005, **102**:5454–5459.
41. Yu Z, Zheng C, Albert VA, Sankoff D: **Excision Dominates Pseudogenization During Fractionation After Whole Genome Duplication and in Gene Loss After Speciation in Plants**. *Frontiers in Genetics* 2020, **11**:1654.
42. *Wang M, Tong S, Ma T, Xi Z, Liu J: **Chromosome-level genome assembly of Sichuan pepper provides insights into apomixis, drought tolerance, and alkaloid biosynthesis**. *Molecular Ecology Resources* 2021, **21**:2533–2545.
The authors use comparative genomics, transcriptomics, and metabolomics to investigate drought tolerance, apomixis, and sanshool and wgx50/gx50 alkaloid biosynthesis in

Sichuan pepper. They show that a genus-specific WGD event in addition to expansion by long terminal repeats contributed to an 8-fold increase in genome size of *Zanthoxylum armatum*. Included in this genome size increase were genes involved in drought tolerance, suggesting the contribution of genome evolution to adaptation to arid conditions. With comparative transcriptomic and metabolomic data they were able to identify candidate genes involved in apomixis and biosynthesis of the alkaloids that produce the distinctive numbing or tingling sensation when eaten.

43. **Xu Z, Gao R, Pu X, Xu R, Wang J, Zheng S, Zeng Y, Chen J, He C, Song J: **Comparative genome analysis of Scutellaria baicalensis and Scutellaria barbata reveals the evolution of active flavonoid biosynthesis**. *Genomics, Proteomics & Bioinformatics* 2020, **18**:230–240.
The authors compared the genomes of two species of *Scutellaria* and nine additional angiosperms to address flavonoid biosynthesis evolution and illustrated the value of congeneric comparisons of species with similar trait profiles for identifying potential independent evolution at the species level. From comparative genomic analysis, they show that key elements of flavonoid biosynthesis seem to be conserved within the genus, with some possible lineage-specific evolution. The also present candidate genes based on the consistency of tissue metabolite profiles with upregulation of biosynthesis pathway genes in the same tissues. Surprisingly, however, in *S. barbata CYP82D1* was highly expressed in the stems and leaves and *CYP82D2* was lowly expressed throughout the various tissues analyzed, which misaligned with the metabolite profile, and suggests functional divergence of hydroxylation, and evolutionary divergence in the flavonoid pathway, between the species.

44. *Wang L, Lei T, Han G, Yue J, Zhang X, Yang Q, Ruan H, Gu C, Zhang Q, Qian T, et al.: **The chromosome-scale reference genome of Rubus chingii Hu provides insight into the biosynthetic pathway of hydrolyzable tannins**. *The Plant Journal* 2021, **107**:1466–1477.
The authors use genomics and functional analysis to investigate hydrolyzable tannin biosynthesis in *Rubus chingii* and relatives in the Rosaceae. The authors show that there were tandem expansions of *UGT*, *CXE*, and *SCPL* in *R. chingii*, and demonstrate that some copies in these expansions have enzymatic activity related to hydrolyzable tannin biosynthesis. Interestingly, they show that loss of a key biosynthesis gene in combination with the expansion of an important biosynthesis gene cluster may be involved in low levels of hydrolyzable tannins in the related apple.

45. Chen ZJ, Ni Z: **Mechanisms of genomic rearrangements and gene expression changes in plant polyploids**. *BioEssays* 2006, **28**:240–252.

46. Yeaman S: **Genomic rearrangements and the evolution of clusters of locally adaptive loci**. *Proceedings of the National Academy of Sciences* 2013, **110**:E1743–E1751.

47. Nützmann H-W, Scazzocchio C, Osbourn A: **Metabolic Gene Clusters in Eukaryotes**. *Annual Review of Genetics* 2018, **52**:159–183.

48. Polturak G, Osbourn A: **The emerging role of biosynthetic gene clusters in plant defense and plant interactions**. *PLOS Pathogens* 2021, **17**:e1009698.

49. Takos AM, Rook F: **Why biosynthetic genes for chemical defense compounds cluster**. *Trends in Plant Science* 2012, **17**:383–388.

50. Hurst LD, Pál C, Lercher MJ: **The evolutionary dynamics of eukaryotic gene order**. *Nature Reviews Genetics* 2004, **5**:299–310.

51. von Rad U, Hüttl R, Lottspeich F, Gierl A, Frey M: **Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize**. *The Plant Journal* 2002, **28**:633–642.
52. Itkin M, Rogachev I, Alkan N, Rosenberg T, Malitsky S, Masini L, Meir S, Iijima Y, Aoki K, de Vos R, et al.: **GLYCOALKALOID METABOLISM1 Is Required for Steroidal Alkaloid Glycosylation and Prevention of Phytotoxicity in Tomato**. *The Plant Cell* 2011, **23**:4507–4525.
53. Tohge T, Fernie AR: **Co-regulation of Clustered and Neo-functionalized Genes in Plant-Specialized Metabolism**. *Plants* 2020, **9**:622.
54. Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, Peters RJ, Osbourn A: **Investigation of terpene diversification across multiple sequenced plant genomes.** *Proc Natl Acad Sci U S A* 2015, **112**:E81-8.
55. Liu Z, Cheema J, Vigouroux M, Hill L, Reed J, Paajanen P, Yant L, Osbourn A: **Formation and diversification of a paradigm biosynthetic gene cluster in plants**. *Nature Communications* 2020, **11**:5354.
56. Shoji T, Kajikawa M, Hashimoto T: **Clustered transcription factor genes regulate nicotine biosynthesis in tobacco**. *The Plant Cell* 2010, **22**:3390–3409.
57. **Mao L, Kawaide H, Higuchi T, Chen M, Miyamoto K, Hirata Y, Kimura H, Miyazaki S, Teruya M, Fujiwara K, et al.: **Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants**. *Proceedings of the National Academy of Sciences* 2020, **117**:12472–12480.
The authors analyze the genome of the bryophyte *Calohypnum plumiforme* to understand the molecular and genomic underpinnings of momilactone biosynthesis, comparing this to other momilactone-producing plants, and using enzymatic assays to verify candidate genes involved in biosynthesis. They show that the *C. plumiforme* genome contains a biosynthesis gene cluster (BGC) of important genes in momilactone biosynthesis and show that it is induced upon stress exposure. They enzymatically test this BGC to confirm its function in momilactone biosynthesis. From comparisons with other momilactone-producing land plants, the authors present this as a case of convergent evolution of a BGC.
58. Yang X, Gao S, Guo L, Wang B, Jia Y, Zhou J, Che Y, Jia P, Lin J, Xu T, et al.: **Three chromosome-scale Papaver genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway**. *Nature Communications* 2021, **12**:6030.
59. Morimoto S, Suemori K, Moriwaki J, Taura F, Tanaka H, Aso M, Tanaka M, Suemune H, Shimohigashi Y, Shoyama Y: **Morphine Metabolism in the Opium Poppy and Its Possible Physiological Function**. *Journal of Biological Chemistry* 2001, **276**:38179–38184.
60. Susumu Ohno: *Evolution by Gene Duplication*. Springer-Verlag; 1970.
61. Wagner A: **The fate of duplicated genes: loss or new function?** *BioEssays* 1998, **20**:785–788.
62. Kimura M, King JL: **Fixation of a deleterious allele at one of two "duplicate" loci by mutation pressure and random drift.** *Proceedings of the National Academy of Sciences* 1979, **76**:2858–2861.
63. Walsh JB: **How often do duplicated genes evolve new functions?** *Genetics* 1995, **139**:421–428.

64. Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE: **(Non)Parallel Evolution**. *Annual Review of Ecology, Evolution, and Systematics* 2018, **49**:303–330.
65. Rosenblum EB, Parent CE, Brandt EE: **The Molecular Basis of Phenotypic Convergence**. *Annual Review of Ecology, Evolution, and Systematics* 2014, **45**:203–226.
66. Arendt J, Reznick D: **Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation?** *Trends in Ecology & Evolution* 2008, **23**:26–32.
67. Wu S, Malaco Morotti AL, Wang S, Wang Y, Xu X, Chen J, Wang G, Tatsis EC: **Convergent gene clusters underpin hyperforin biosynthesis in St John's wort**. *New Phytologist* 2022, doi:10.1111/nph.18138.
68. Beerhues L: **Hyperforin**. *Phytochemistry* 2006, **67**:2201–2207.
69. Yang X-W, Grossman RB, Xu G: **Research Progress of Polycyclic Polyprenylated Acylphloroglucinols**. *Chemical Reviews* 2018, **118**:3508–3558.
70. Roy R, Moreno N, Brockman SA, Kostanecki A, Zambre A, Holl C, Solhaug EM, Minami A, Snell-Rood EC, Hampton M, et al.: **Convergent evolution of a blood-red nectar pigment in vertebrate-pollinated flowers**. *Proceedings of the National Academy of Sciences* 2022, **119**:e2114420119.
71. Dong S, Liu M, Liu Y, Chen F, Yang T, Chen L, Zhang X, Guo X, Fang D, Li L, et al.: **The genome of Magnolia biondii Pamp. provides insights into the evolution of Magnoliales and biosynthesis of terpenoids**. *Horticulture Research* 2021, **8**:38.
72. Lai D, Maimann AB, Macea E, Ocampo CH, Cardona G, Pičmanová M, Darbani B, Olsen CE, Debouck D, Raatz B, et al.: **Biosynthesis of cyanogenic glucosides in Phaseolus lunatus and the evolution of oxime-based defenses**. *Plant Direct* 2020, **4**:e00244.
73. Pei L, Wang B, Ye J, Hu X, Fu L, Li K, Ni Z, Wang Z, Wei Y, Shi L, et al.: **Genome and transcriptome of Papaver somniferum Chinese landrace CHM indicates that massive genome expansion contributes to high benzylisoquinoline alkaloid biosynthesis**. *Horticulture Research* 2021, **8**:5.
74. Qin S, Wei K, Cui Z, Liang Y, Li M, Gu L, Yang C, Zhou X, Li L, Xu W, et al.: **Comparative genomics of Spatholobus suberectus and insight into flavonoid biosynthesis**. *Frontiers in Plant Science* 2020, **11**:1372.
75. Qin L, Hu Y, Wang J, Wang X, Zhao R, Shan H, Li K, Xu P, Wu H, Yan X, et al.: **Insights into angiosperm evolution, floral development and chemical biosynthesis from the Aristolochia fimbriata genome**. *Nature Plants* 2021, **7**:1239–1253.
76. Wu Z, Liu H, Zhan W, Yu Z, Qin E, Liu S, Yang T, Xiang N, Kudrna D, Chen Y, et al.: **The chromosome-scale reference genome of safflower (Carthamus tinctorius) provides insights into linoleic acid and flavonoid biosynthesis**. *Plant Biotechnology Journal* 2021, **19**:1725–1742.
77. Wang P, Luo Y, Huang J, Gao S, Zhu G, Dang Z, Gai J, Yang M, Zhu M, Zhang H, et al.: **The genome evolution and domestication of tropical fruit mango**. *Genome Biology* 2020, **21**:60.
78. Chen Y-C, Li Z, Zhao Y-X, Gao M, Wang J-Y, Liu K-W, Wang X, Wu L-W, Jiao Y-L, Xu Z-L, et al.: **The Litsea genome and the evolution of the laurel family**. *Nature Communications* 2020, **11**:1675.
79. Chen H, Köllner TG, Li G, Wei G, Chen X, Zeng D, Qian Q, Chen F: **Combinatorial evolution of a terpene synthase gene cluster explains terpene variations in Oryza**. *Plant Physiology* 2020, **182**:480–492.

80. Tu L, Su P, Zhang Z, Gao L, Wang J, Hu T, Zhou J, Zhang Y, Zhao Y, Liu Y, et al.: **Genome of Tripterygium wilfordii and identification of cytochrome P450 involved in triptolide biosynthesis**. *Nature Communications* 2020, **11**:971.
81. Meyer GW, Bahamon Naranjo MA, Widhalm JR: **Convergent evolution of plant specialized 1,4-naphthoquinones: metabolism, trafficking, and resistance to their allelopathic effects**. *Journal of Experimental Botany* 2021, **72**:167–176.
82. Wen W, Alseekh S, Fernie AR: **Conservation and diversification of flavonoid metabolism in the plant kingdom**. *Current Opinion in Plant Biology* 2020, **55**:100–108.
83. Huang R, O'Donnell AJ, Barboline JJ, Barkman TJ: **Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes**. *Proc Natl Acad Sci U S A* 2016, **113**:10613–10618.
84. Züst T, Strickler SR, Powell AF, Mabry ME, An H, Mirzaei M, York T, Holland CK, Kumar P, Erb M, et al.: **Independent evolution of ancestral and novel defenses in a genus of toxic plants (Erysimum, Brassicaceae)**. *Elife* 2020, **9**.
85. Wei N, Du Z, Liston A, Ashman T: **Genome duplication effects on functional traits and fitness are genetic context and species dependent: studies of synthetic polyploid Fragaria**. *American Journal of Botany* 2020, **107**:262–272.