HHS Public Access

Author manuscript

New Phytol. Author manuscript; available in PMC 2021 August 01.

Published in final edited form as:

New Phytol. 2020 August ; 227(4): 992–994. doi:10.1111/nph.16631.

Doing the gene shuffle to close synteny: Dynamic assembly of biosynthetic gene clusters

Reuben J. Peters

Roy J. Carver Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, IA 50011, U.S.A.

In the original definition, synteny implied physical co-localization of genetic loci – i.e., genes – as demonstrated by genetic linkage. This applies to the emerging realization that plants appear to have shuffled their genomes to assemble biosynthetic gene clusters (BGCs) for more specialized metabolism (Nutzmann *et al.*, 2018). However, this assembly process is not well understood. Indeed, the evolutionary timescale over which these BGCs can be assembled, or persist, has not been widely explored. In this issue of *New Phytologist* Liu et al. (pp. 1109–1123) have carried out such analysis of a particular type of BGC in the Brassicaceae. Specifically, by anchoring their studies with oxidosqualene cyclases, the authors are then limited to examining triterpenoid metabolism. Nevertheless, within this context comprehensive bioinformatic analyses were carried out, along with biochemical characterization of selected BGCs. The results strongly indicate that at least these Brassicaceae triterpenoid BGCs exhibit a surprisingly dynamic nature, with many clearly having arisen since the evolutionary divergence of the constituent species examined here. In particular, while synteny is evident, the co-clustered genes generally are differentially ordered and, although often containing members of the same enzymatic sub-families, phylogenetically distinct. Distinct triterpenoid natural product outcome was shown for two such BGCs by recombinant biochemical characterization. By contrast, others of these BGCs appear to have persisted, with their orthologous function demonstrated by biochemical characterization for one pair, implying consistent selective pressure for retention of the resulting triterpenoids, with strong biological activity recently shown for at least the thalianol product of the pair of BGCs examined here (Huang *et al.*, 2019). Accordingly, these BGCs are readily assembled, yet also can be retained, indicating more than just genomic instability underlies their appearance. While the results reported here strictly pertain to Brassicaceae triterpenoid BGCs, it seems most likely that such dynamic gene shuffling applies to plant BGCs more broadly, providing wider import.

As noted by Liu et al., the repeated appearance of genes from the same enzymatic sub-families indicates that these provide “a core palette of triterpene-diversifying enzymes” that have been shuffled in such BGC assembly, and that these provide “natural combinatorics of enzyme families, which can be mimicked using synthetic biology to engineer diverse bioactive molecules”. Even beyond those enzymatic families highlighted here, it should be noted that other plant families may have utilized distinct enzymatic (sub-)families to

rjpeters@iastate.edu.

generate similar chemical diversity. For example, while Liu et al. suggest that the CYP716A sub-family of cytochrome P450 (CYP) mono-oxygenases appears only rarely, their own data indicates that members of this sub-family do exhibit a significant association with oxidosqualene cyclases. Thus, given the already noted frequent use of the CYP716 family in generating triterpenoid diversity (Miettinen *et al.*, 2017), it can be hypothesized that members of this CYP family might be even more prevalent in triterpenoid BGCs in other plant families and, hence, should be included in the "core palette". More generally, previous work on gene pairing of CYPs and terpene synthases more broadly has highlighted a number of other CYP families (Boutanaev *et al.*, 2015), which can then be similarly hypothesized to serve in the "core palette" for diversification of smaller terpenoid natural products and, presumably, drawn upon for such combinatorial biosynthetic efforts as well.

One limitation of the results reported by Liu et al. is that the BGC conservation demonstrated by biochemical characterization involves species from same genus (*Arabidopsis*), which means the timescale of such conservation is somewhat restricted. Fortunately, work on a BGC associated with production of the diterpenoid momilactones in the Poaceae plant family provides at least one example of conservation over much greater phylogenetic distance. Similar to the *Arabidopsis* triterpenoid BGC, conservation of not only gene content but also order was observed between the momilactone BGCs found in rice *(Oryza sativa*) and its wild relative *Oryza punctata*, both of which produce this diterpenoid, while the more distantly related *Oryza brachyantha* does not produce momilactones nor contain the corresponding BGC (Miyamoto *et al.*, 2016). More informatively, *Echinochloa crus-galli*, which falls into a different clade of the Poaceae than *Oryza*, nevertheless produces momilactones and contains an apparently orthologous BGC, although this has undergone minimization (i.e., loss of the local gene duplicates evident in *Oryza*), along with some change in gene order, yet phylogenetic analysis of the constituent genes indicates nearest homology to those from the *Oryza* BGC rather than with members of the relevant enzymatic families from more closely related plant species (Guo *et al.*, 2017). Given the phylogenetic distance between *E. crus-galli* and *Oryza* it can be hypothesized that this orthology may reflect an ancient hybridization event, presumably enabled by overlapping geographic range of the extant species, and subsequent natural introgression of the momilactone BGC into the *E. crus-galli* lineage, indicating that there was strong selective pressure for production of this particular diterpenoid. Intriguingly, while genomic cohesion of the momilactone BGC in *E. crus-galli* might simply reflect the synteny required for introgression, another CYP is now found in this BGC. It can be speculated that this might play a role in momilactone biosynthesis, as the biochemical activity observed with a member of the same CYP76M sub-family has been used to suggest a role in rice momilactone biosynthesis (Wang *et al.*, 2012a), but that then begs the question why this CYP is not present in the *Oryza* momilactone BGCs.

Answers to such questions may be found in the realization that plant BGCs assembly is not driven by the need for co-inheritance of the entire biosynthetic pathway as a discrete genetic locus. As noted above, the rice momilactone BGC is not "complete" in this sense. Instead, it has been hypothesized that BGC assembly in plants requires not only positive selection pressure provided by the bioactivity of the resulting natural product, but also negative selection against partial pathway inheritance (Takos & Rook, 2012), as the latter has long

been appreciated to drive genetic linkage (Fisher, 1930). Consistent with this hypothesis, deleterious effects from loss of genes encoding enzymes acting later in the biosynthetic process have been reported for at least a few plant BGCs, including that for momilactones in rice (Xu *et al.*, 2012; Kemen *et al.*, 2014). Strikingly, it appears that examination of momilactone biosynthesis will continue to be informative with regards to the effect that this particular diterpenoid biosynthetic pathway exerts on genome organization, as the early diverging land plant *Hypnum plumaeforme* produces momilactones and the only relevant gene identified to-date is clearly phylogenetic distinct (Okada *et al.*, 2016). Thus, it will be of interest to determine if this moss also contains a corresponding BGC, as this presumably would have had to arisen via convergent evolution, and investigation of which might then help clarify the role of negative selection pressure in such BGC assembly.

Finally, it should be noted that the essentially exclusive use of vertical gene transmission in plants imposes distinct parameters on their assembly of BGCs relative to the horizontal gene transfer that drives microbial BGC synteny. In particular, genetic linkage of enzymatic genes for more specialized metabolism to essential genes negates any need to 'recruit' the former to a BGC as co-inheritance is guaranteed. For example, biochemical characterization of the rice CYP701A sub-family, which arose from multiple tandem gene duplication events and, hence, form an array in the rice genome, not only verified the *ent*-kaurene oxidase activity of the gene already known to be required for gibberellin phytohormone biosynthesis (Sakamoto *et al.*, 2004), but also divergent activity for one of the paralogs indicative of a role in momilactone (Kitaoka *et al.*, 2015), as well as other more specialized diterpenoid metabolism (Wang *et al.*, 2012b). Such linked genes may then be absent from BGCs, but still fall into the "core palette" of diversifying/tailoring enzyme families and, of relevance to the contention of Liu et al., then be worth including in combinatorial biosynthetic efforts. By contrast, the CYP76M sub-family member suggested to play a role in rice momilactone biosynthesis is found within a different diterpenoid BGC, suggesting the potential for co-evolution of BGCs found in the same lineage. While the role of these CYPs outside the rice momilactone BGC remain to be proven, the incomplete nature of this BGC certainly highlights the need to look beyond BGCs for the relevant enzymatic genes in plant natural products biosynthesis. In some sense, this hypothesis is consistent with the dynamic nature of plant BGCs indicated by the analysis of Brassicaceae triterpenoid BGCs reported by Liu et al in this issue, as such rapid gene shuffling must be driven by strong selective pressure, potentially both positive and negative as described above, but yet these BGCs do not seem to necessarily need to be 'complete'. Accordingly, more work remains to be done to fully elucidate the impact that more specialized metabolism has had on plant genome organization, specifically the gene shuffling that leads to the close synteny of BGCs.

Acknowledgements

This work was funded by a grant from the NIH (GM131885).

References

Boutanaev A, Moses T, Zi J, Nelson D, Mugford S, Peters RJ, Osbourn A. 2015 Investigation of terpene diversification across multiple sequenced plant genomes. Proc Natl Acad Sci U S A 112: E81–E88. [PubMed: 25502595]

Fisher RA. 1930 The Genetical Theory of Natural Selection. Oxford: Clarendon Press.

Guo L, Qiu J, Ye C, Jin G, Mao L, Zhang H, Yang X, Peng Q, Wang Y, Jia L, et al. 2017 Echinochloa crus-galli genome analysis provides insight into its adaptation and invasiveness as a weed. Nat Commun 8(1): 1031. [PubMed: 29044108]

Huang AC, Jiang T, Liu YX, Bai YC, Reed J, Qu B, Goossens A, Nutzmann HW, Bai Y, Osbourn A. 2019 A specialized metabolic network selectively modulates Arabidopsis root microbiota. Science 364(6440): eaau6389. [PubMed: 31073042]

Kemen AC, Honkanen S, Melton RE, Findlay KC, Mugford ST, Hayashi K, Haralampidis K, Rosser SJ, Osbourn A. 2014 Investigation of triterpene synthesis and regulation in oats reveals a role for beta-amyrin in determining root epidermal cell patterning. Proc Natl Acad Sci U S A 111(23): 8679–8684. [PubMed: 24912185]

Kitaoka N, Wu Y, Xu M, Peters RJ. 2015 Optimization of recombinant expression enables discovery of novel cytochrome P450 activity in rice diterpenoid biosynthesis. Appl. Microbiol. Biotechnol 99(18): 7549–7558. [PubMed: 25758958]

Miettinen K, Pollier J, Buyst D, Arendt P, Csuk R, Sommerwerk S, Moses T, Mertens J, Sonawane PD, Pauwels L, et al. 2017 The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. Nat Commun 8: 14153. [PubMed: 28165039]

Miyamoto K, Fujita M, Shenton MR, Akashi S, Sugawara C, Sakai A, Horie K, Hasegawa M, Kawaide H, Mitsuhashi W, et al. 2016 Evolutionary trajectory of phytoalexin biosynthetic gene clusters in rice. Plant J 87(3): 293–304. [PubMed: 27133567]

Nutzmann HW, Scazzocchio C, Osbourn A. 2018 Metabolic Gene Clusters in Eukaryotes. Annu Rev Genet 52: 159–183. [PubMed: 30183405]

Okada K, Kawaide H, Miyamoto K, Miyazaki S, Kainuma R, Kimura H, Fujiwara K, Natsume M, Nojiri H, Nakajima M, et al. 2016 HpDTC1, a Stress-Inducible Bifunctional Diterpene Cyclase Involved in Momilactone Biosynthesis, Functions in Chemical Defence in the Moss Hypnum plumaeforme. Sci Rep 6: 25316. [PubMed: 27137939]

Sakamoto T, Miura K, Itoh H, Tatsumi T, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Agrawal GK, Takeda S, Abe K, et al. 2004 An overview of gibberellin metabolism enzyme genes and their related mutants in rice. Plant Physiol. 134: 1642–1653. [PubMed: 15075394]

Takos AM, Rook F. 2012 Why biosynthetic genes for chemical defense compounds cluster. Trends Plant Sci 17(7): 383–388. [PubMed: 22609284]

Wang Q, Hillwig ML, Okada K, Yamazaki K, Wu Y, Swaminathan S, Yamane H, Peters RJ. 2012a Characterization of CYP76M5–8 indicates metabolic plasticity within a plant biosynthetic gene cluster. J. Biol. Chem 287(9): 6159–6168. [PubMed: 22215681]

Wang Q, Hillwig ML, Wu Y, Peters RJ. 2012b CYP701A8: A rice *ent*-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism. Plant Physiol. 158(3): 1418–1425. [PubMed: 22247270]

Xu M, Galhano R, Wiemann P, Bueno E, Tiernan M, Wu W, Chung IM, Gershenzon J, Tudzynski B, Sesma A, et al. 2012 Genetic evidence for natural product-mediated plant-plant allelopathy in rice (*Oryza sativa*). New Phytol. 193(3): 570–575. [PubMed: 22150231]