Journal of Experimental Botany, Vol. 68, No. 15 pp. 4029–4043, 2017
doi:10.1093/jxb/erx152 Advance Access publication 18 May 2017

REVIEW PAPER

Tying the knot: occurrence and possible significance of gene fusions in plant metabolism and beyond

Jillian M. Hagel and Peter J. Facchini*

Department of Biological Sciences, University of Calgary, 2500 University Dr N.W., Alberta T2N 1N4, Canada

* Correspondence: pfacchin@ucalgary.ca

Received 15 December 2016; Editorial decision 11 April 2017; Accepted 11 April 2017

Editor: Alain Goossens, VIB-UGent

Abstract

Gene fusions have recently attracted attention especially in the field of plant specialized metabolism. The occurrence of a gene fusion, in which originally separate gene products are combined into a single polypeptide, often corresponds to the functional association of individual components within a single metabolic pathway. Examples include gene fusions implicated in benzylisoquinoline alkaloid (BIA), terpenoid, and amino acid biosynthetic pathways, in which distinct domains within a fusion catalyze consecutive, yet independent reactions. Both genomic and transcriptional mechanisms result in the fusion of gene products, which can include partial or complete domain repeats and extensive domain shuffling as evident in the BIA biosynthetic enzyme norcoclaurine synthase. Artificial gene fusions are commonly deployed in attempts to engineer new or improved pathways in plants or microorganisms, based on the premise that fusions are advantageous. However, a survey of functionally characterized fusions in microbial systems shows that the functional impact of fused gene products is not straightforward. For example, whereas enzyme fusions might facilitate the metabolic channeling of unstable intermediates, this channeling can also occur between tightly associated independent enzymes. The frequent occurrence of both fused and unfused enzymes in plant and microbial metabolism adds additional complexity, in terms of both pathway functionality and evolution.

Key words: Benzylisoquinoline alkaloid metabolism, metabolic engineering, microbial metabolism, plant metabolism, synthetic biology.

Introduction

The increasing availability of genomic sequence data has expedited investigations into the origin and evolution of genes. Several well-established mechanisms are known to govern the formation of new genes, including exon shuffling, gene duplication, horizontal gene transfer, gene fusion and fission, and the *de novo* recruitment from non-coding DNA (Long *et al.*, 2003; Espinosa-Cantú *et al.*, 2015; McLysaght and Guerzoni, 2015). Examples of neofunctionalization

Abbreviations: ADH, arogenate dehydrogenase; ADT, arogenate dehydratase; AS, abietadiene synthase; ASα/ASβ or ASα-ASβ, anthranilate synthase α and β domains; AK, aspartate kinase; AT, acyltransferase; BAS, benzalacetone synthase; BIA, benzylisoquinoline alkaloid; CBM, carbohydrate-binding motif/domain; 4CL, 4-coumaroyl-CoA ligase; CM, chorismate mutase; CPR, cytochrome P450 reductase; CPS, copalyl-labdadienyl diphosphate synthase; CYP, cytochrome P450; DAPA-AT, diaminopelargonic acid aminotransferase; DAH7PS, 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase; DHAPAT, dihydroxyacetone phosphate synthase; DHFR, dihydrofolate reductase; DHQD, dehydroquinate dehydratase; DRS, 1,2-dehydroreticuline synthase; DRR, 1,2-dehydroreticuline reductase; DTBS, dethiobiotin synthetase; DTC1, *syn*-pimara-7,15-diene synthase; γ-ECL, γ-glutamylcysteine ligase; FAR, fatty acid reductase; GS, glutathione synthetase; HSD, homoserine dehydrogenase; IGPS, indole-3-glycerol phosphate synthase; KS, kaurene synthase; LCYB, lycopene β-cyclase; LCYE, lycopene ε-cyclase; LHC, light-harvesting complex; MsrA/B, methionine sulfoxide reductases A/B; NCS, norcoclaurine synthase; PAI, phosphoribosyl anthranilate isomerase;P5CDH, Δ^1-pyrroline-5-carboxylate dehydrogenase; PRODH, proline dehydrogenase; REPI, reticuline epimerase; SHD, shikimate dehydrogenase; TS, thymidylate synthase; TSα/TSβ or TSα-TSβ, tryptophan synthase α and β domains; WS, wax synthase.

within the context of specialized metabolite biosynthesis have been established for model crops, such as tomato (Tomato Genome Consortium, 2012) and tobacco (Sierro *et al.*, 2014). A variety of post-transcriptional mechanisms, such as alternative splicing modulated by long non-coding RNA, also contribute to the emergence of new proteins (Bardou *et al.*, 2014). Gene fusion and/or fission events account for ~0.5% of all prokaryotic genes, and the vast majority of characterized gene fusions occur in bacteria and fungi. As such, current opinions surrounding the impact of protein fusion at the molecular and evolutionary levels derive primarily from the microbial realm. Although our current appreciation of the organization of plant genomes, including the clustering and fusion of genes, is less extensive compared with microorganisms, impressive advances have recently been reported. For example, the unexpected occurrence of gene clustering has been extensively documented in plant specialized metabolism with respect to biosynthetic genes involved in the formation of benzylisoquinoline alkaloids (BIAs), terpenoids, cyanogenic glycosides, and benzoxazinoids (reviewed by Boycheva *et al.*, 2014; Nützmann and Osbourn, 2014). Recently, genes encoding norcoclaurine synthase (NCS), the central enzyme in BIA metabolism, were found clustered in sacred lotus (*Nelumbo nucifera*), a basal eudicot (Vimolmangkang *et al.*, 2016). Since gene fusions can arguably be viewed as an extreme outcome of clustering, their occurrence and participation in plant metabolic pathways is not surprising. Similar to the purported hypotheses concerning the formation and advantages of gene clusters, gene fusions can provide insight into the mechanism and physiological function of component proteins (Bradbury *et al.*, 2013). In this review, we aim to (i) describe known gene fusions implicated in plant metabolism, particularly those involved in specialized biosynthetic pathways and amino acid metabolism; (ii) overview the possible mechanisms responsible for the occurrence of gene fusions; (iii) explore the impact of fusion events on metabolism using well-established microbial models; and (iv) summarize the use of fusions in biotechnology.

Gene fusions in plant metabolism

Gene fusions are particularly relevant to the functioning of metabolic pathways. Enzymes participating in the biosynthesis of a related group of metabolites are often co-regulated, and frequently display cellular co-localization. For example, many enzymes involved in BIA biosynthesis are co-ordinately regulated at the transcriptional level (Zulak *et al.*, 2007; Angelova *et al.*, 2010), and occur in distinct cell types. In opium poppy, most BIA biosynthetic enzymes are associated with the cortical endoplasmic reticulum (ER) of sieve elements, whereas some downstream enzymes occur predominantly in adjacent cells known as laticifers (Hagel and Facchini, 2013). By default, the fusion of two enzymes participating in a single metabolic pathway ensures co-expression and co-localization. Although gene fusions arguably contribute to pathway functionality in some of the examples described below, the consequences of many fusion events are presently unknown.

Norcoclaurine synthase

Most plant gene fusions have been discovered using a combination of traditional biochemical applications and functional genomics investigations based on established transcriptome databases. In some cases, evidence of gene fusion emerges long after an enzyme is purified and characterized at the molecular level. For example, although norcoclaurine synthase (NCS) activity was first characterized >15 years ago (Samanani and Facchini, 2001) and the corresponding gene was isolated 3 years later (Samanani *et al.*, 2004), *NCS* gene fusions in opium poppy (*Papaver somniferum*) and related plants were recognized only recently (Li *et al.*, 2016). NCS catalyzes the first committed step in BIA metabolism, condensing dopamine and 4-hydroxyphenylacetaldehyde to form the central pathway intermediate norcoclaurine (Fig. 1). A series of *O*- and *N*-methylations, and an aromatic ring hydroxylation, convert norcoclaurine to (*S*)-reticuline, a key branch point intermediate in the biosynthesis of several BIAs, including the analgesic morphine, the antitumor agent noscapine, and the antimicrobial sanguinarine (Hagel and Facchini, 2013).

Analysis of the first available NCS sequence from meadow rue (*Thalictrum flavum*) showed no significant homology to other enzymes, but rather to Bet v 1 allergen and pathogenesis-related (PR)10 proteins. The physiological functions of Bet v 1 and PR10 proteins are largely unknown despite their ubiquitous occurrence in plants. The origin of BIA metabolism is probably monophyletic, at least in members of the order Ranunculales (Liscombe *et al.*, 2005), which facilitates BLAST (Basic Local Alignment Search Tool; Altschul *et al.*, 1990) surveys of sequence libraries for orthologous NCS transcripts in related BIA-producing plants. The cognate amino acid sequence for functional NCS orthologs from a variety of plant species displayed considerable dissimilarity (Lee and Facchini, 2010). For example, NCS from opium poppy and meadow rue share <40% amino acid identity. Further analysis of transcriptome data and empirical cDNAs obtained from several species showed evidence of extensive domain shuffling and DNA fusion events affecting NCS transcripts within the Papaveraceae (Li *et al.*, 2016). Phylogenetic analysis of >40 NCS sequences from multiple species in the order Ranunculales primarily revealed evolutionary, rather than functional, relationships and did not show patterns related to protein domain architecture. Activity was confirmed for six NCS homologs, expressed as recombinant proteins in bacteria (*Escherichia coli*) and yeast (*Saccharomyces cerevisiae*), drawn from several species in the family, including enzymes containing two, three, or four tandem repeated domains (Li *et al.*, 2016). Significantly, multiple catalytic domains appeared to correlate with a proportional increase in catalytic efficiency. This result supports a recently proposed hypothesis that multiple-domain fusions in NCS could represent an evolutionary mechanism to enhance the performance of a relatively inefficient enzyme (Lichman *et al.*, 2015) that could be recalcitrant to an improvement in catalytic efficiency via typical evolutionary mechanisms (e.g. point mutations). Additional biochemical evidence, perhaps derived from a more extensive study of natural and artificial fusion products, could determine

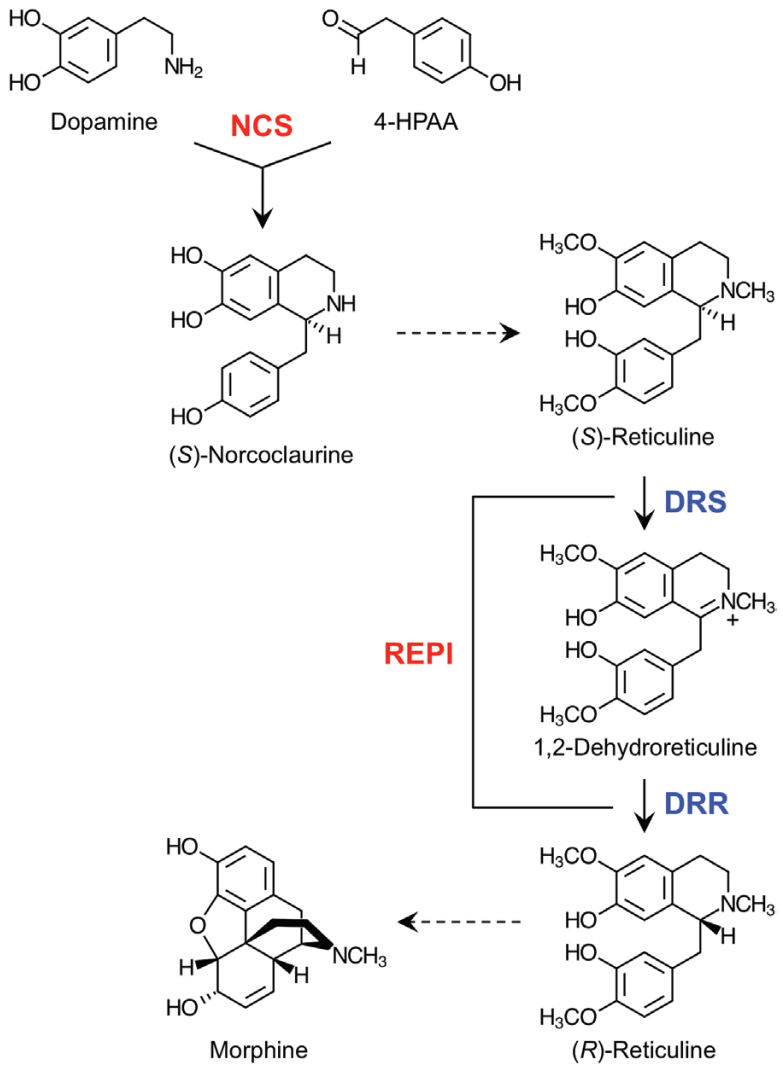

Fig. 1. Two known benzylisoquinoline alkaloid (BIA) biosynthetic enzymes, norcoclaurine synthase (NCS) and reticuline epimerase (REPI), are the result of gene fusion events and operate at key junctions in the pathway (indicated in red). The first committed step catalyzed by NCS yields the central pathway intermediate (*S*)-norcoclaurine. (*S*)-Norcoclaurine is converted via a series of transformations to (*S*)-reticuline, a key branch-point intermediate in the formation of several different BIA structural types, including the morphinan alkaloids. The gateway to morphine biosynthesis in opium poppy involves the stereochemical inversion of reticuline from the *S* to the *R* enantiomer, a two-step process catalyzed by the cytochrome P450 (CYP)–aldo-keto reductase (AKR) fusion enzyme REPI. The CYP domain of REPI acts as 1,2-dehydroreticuline synthase (DRS) and the AKR domain acts as 1,2-dehydroreticuline reductase (DRR). The DRS and DRR reactions are shown in blue. Dotted lines indicate multiple enzymatic steps.

whether domain fusions contribute to enhanced enzyme performance. A domain-oriented analysis of eight NCS variants from opium poppy and Persian poppy (*Papaver bracteatum*), a close relative accumulating the morphinan alkaloid thebaine, was performed using RADAR (Rapid Automatic Detection and Alignment of Repeats in protein sequences; http://www.ebi.ac.uk/Tools/pfa/radar/), visual inspection, and subsequent sequence alignment (Fig. 2). NCS variants could be divided into 'active' domains (i.e. those containing catalytic residues; Li *et al.*, 2016), which occur in each variant as single or multiple (up to four) consecutive repeats. All but one variant (PbNCS2) possessed a conserved C-terminal domain, although the significance of this domain is not known and PbNCS2 has not been tested for NCS activity. The repetition of domains was incomplete in some variants. For example, ~20 amino acids from the N-terminus of PsNCS1 and PsNCS2 lacked the full complement of A3 residues, but aligned well with the C-terminal portions of A1, A2, A3, and A4 (Fig. 2). Green fluorescent protein (GFP) fusion-based localization of PsNCS1 (Hagel and Facchini, 2012) and PsNCS2 (Lee and Facchini, 2010) in cultured opium poppy cells identified a putative ER localization signal, putatively in the N-terminal regions. An equivalent N-terminal sequence is not found in PsNCS3, suggesting a distinct cellular localization for this quadruple catalytic domain variant.

Genomic sequence analysis is critical to understanding the evolution of *NCS* orthologs. Current perspectives regarding domain shuffling and tandem fusions in NCS rely solely on sequences derived from mature transcripts, precluding any meaningful insight into whether the fusions are genomic DNA based or transcriptionally mediated. The presence of a single intron was noted in five linked *NCS* orthologs in the sacred lotus (*Nelumbo nucifera*) genome (Vimolmangkang *et al.*, 2016). The genomic organization of *NCS* genes is expected to

be more complex in members of the Papaveraceae, including opium poppy and Persian poppy, based on the occurrence of tandem repeats and shuffled domains represented in the transcriptome databases of all available species within the family.

Reticuline epimerase

In opium poppy and other morphinan alkaloid-producing plants, such as Persian poppy, (*S*)-reticuline undergoes stereochemical inversion to (*R*)-reticuline as a gateway reaction owing to the stereospecificity of salutaridine synthase, which accepts only the *R* conformer (Gesell *et al.*, 2009). The *S* to *R* epimerization of reticuline is catalyzed by reticuline epimerase (REPI), the enzymatic product of another gene fusion event in opium poppy (Farrow *et al.*, 2015; Winzer *et al.*, 2015). PCR-derived genomic sequence data from opium poppy showed the occurrence of six exons and five introns in the *REPI* gene (Fig. 3) (Farrow *et al.*, 2015). In opium poppy and Persian poppy, REPI consists of two discrete domains: an N-terminal cytochrome P450 (CYP82Y2) and a C-terminal aldo-keto reductase (AKR). The CYP82Y2 domain catalyzes the oxidation of (*S*)-reticuline to 1,2-dehydroreticuline, a reaction referred to as 1,2-dehydroreticuline synthase (DRS) activity (Fig. 1). The AKR domain subsequently catalyzes the reduction of 1,2-dehydroreticuline to (*R*)-reticuline, a conversion referred to as 1,2-dehydroreticuline reductase (DRR) activity. The two REPI domains can be expressed independently as recombinant, functional DRS and DRR enzymes. Moreover, DRS and DRR occur naturally as separate enzymes in *Papaver* species, such as the common field poppy (*Papaver rhoeas*), that are not known to produce morphinan alkaloids. Alignment of *DRS* and *DRR* genes from *P. rhoeas* with the *REPI* gene from opium poppy reveals a strong conservation of exon–intron structure, and shows that the DRS and DRR domains of REPI contain the full sequence complement found in naturally independent DRS and DRR enzymes (Fig. 3). Furthermore, the alignment highlights a short (12 amino acid) peptide separating the two REPI domains. The lack of an intron between the fused *DRS* and *DRR* genes in the opium poppy genome, together with the occurrence of complete and structurally conserved DRS and DRR domains, suggests that the fusion was facilitated by a DNA-based mechanism involving two tightly linked ancestral genes. If, indeed, *DRS* and *DRR* occurred as separate and fully functional genes in an ancestral species prior to the emergence of the *REPI* fusion in opium poppy, the *DRR* promoter sequence and *DRS* stop codon must have been deleted at an intermediate evolutionary stage. Alternatively, the occurrence

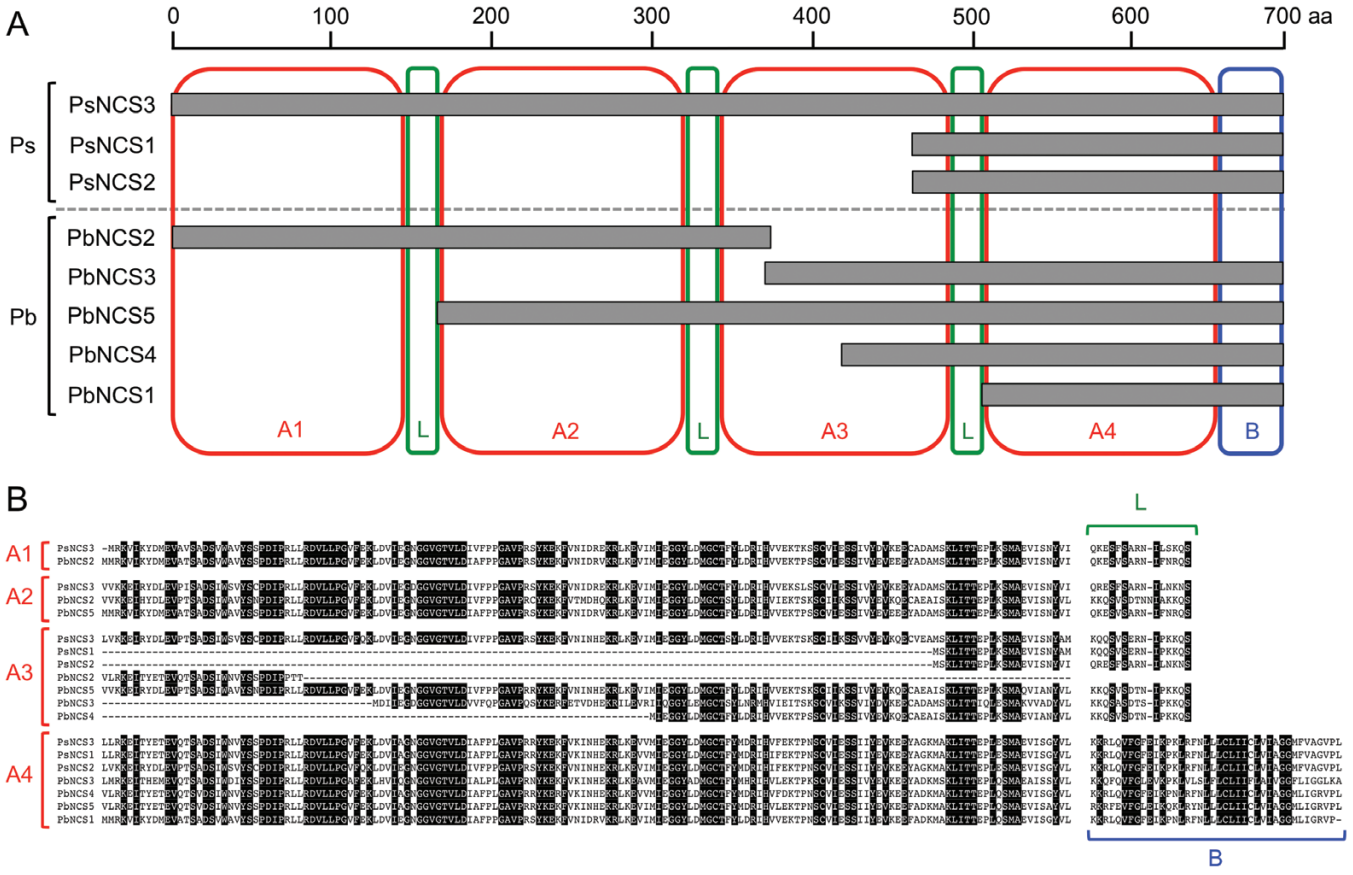

Fig. 2. (A) Repeated segments in NCS polypeptides (gray boxes) from opium poppy (*Papaver somniferum*) and Persian poppy (*P. bracteatum*) were identified using RADAR (http://www.ebi.ac.uk/Tools/pfa/radar/) combined with visual inspection. Red boxes indicate repeated regions (A1–A4) that contain catalytically active residues (Li *et al.*, 2016; Ilari *et al.*, 2009). Green boxes show 'linker' regions (L), and the blue box identifies a conserved C-terminal domain. (B) Amino acid alignment of repeated domains was performed using CLUSTALW followed by manual rearrangement. Absolutely conserved residues across all polypeptides in A, L, or B domains are highlighted in black. Abbreviations: NCS, norcoclaurine synthase; Ps, *Papaver somniferum*; Pb, *Papaver bracteatum*; aa, amino acid.

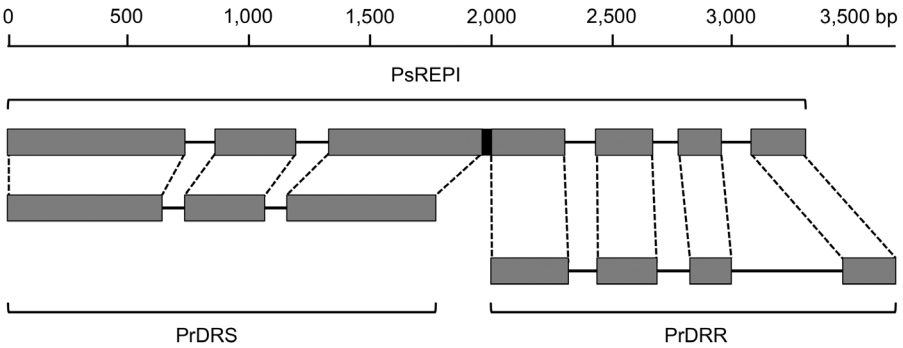

Fig. 3. Structure of opium poppy (*Papaver somniferum*) and field poppy (*Papaver rhoeas*) genes encoding reticuline epimerase (REPI). The gene fusion in opium poppy (*PsREPI*) encodes 1,2-dehydroreticuline synthase (*DRS*) and 1,2-dehydroreticuline reductase (*DRR*) domains fused by a short, intron-free linker. In contrast, the orthologous field poppy genes (*PrDRS* and *PrDRR*) occur separately, but each shows a conserved structure compared with corresponding domains in *PsREPI*. Exons and introns are represented as gray boxes and solid black lines, respectively. The linker region in PsREPI is highlighted as a black box. Dotted lines represent conserved exons in the DRS and DRR domains of orthologous opium poppy and field poppy genes. REPI protein sequences have been reported previously (Farrow *et al.*, 2015).

of *REPI* in opium poppy could pre-date a fission event, leading to the formation of discrete *DRS* and *DRR* genes found in other *Papaver* species, such as field poppy. Owing to the lack of assembled genomes, the *REPI* gene copy number in opium poppy or related species is not known, which hinders efforts to investigate the evolutionary origins of *REPI* as a fused gene. Cytological evidence has suggested the allopolyploid species *Papaver setigerum* (2*n*=44), which morphologically resembles *P. somniferum* (2*n*=22) and accumulates opiate alkaloids, as the wild ancestor of opium poppy (Levy and Milo, 1998). If domesticated opium poppy actually emerged via diploidization from an allopolyploid ancestor, evolutionary comparisons with related species with respect to gene copy numbers and DNA-level rearrangement should provide considerable insight into gene fusion processes.

Lycopene cyclases

The possibility that a single, two-domain enzyme could predate the emergence of discrete, independent enzymes is well documented in plant metabolism. A recent example includes carotenoid biosynthetic enzymes in green algae (Blatt *et al.*, 2015) (Table 1). The production of carotenoids in plants involves two enzymes, lycopene β-cyclase (LCYB) and lycopene ε-cyclase (LCYE). The two lycopene cyclases are closely related and probably arose via ancient gene duplication. Whereas LCYB and LCYE occur as separate enzymes in plants, prasinophyte algae of the order Mamiellales exhibit the single gene *LCY*, which is comprised of LCYB, LCYE, and LHC (light-harvesting complex) domains. It was proposed that *LCY* in green algae originated as the result of an unequal crossover event, a phenomenon regarded as the main mechanism of gene duplication (Graur and Li, 2000). An unequal crossover event might have led directly to an in-frame fusion of genes encoding duplicated lycopene cyclase domains, which could represent the ancestral state of LCYB and LCYE in plants. Subsequent gene fission of the ancestral fusion would probably have occurred only after the two domains evolved their distinct lycopene cyclase activities. In support of this hypothesis, the Mamiellales are one of the most basal clades in the green algae lineage (Turmel *et al.*, 2009; Marin and Melkonian, 2010). Unlike many bifunctional fusion products that catalyze sequential reactions within a single pathway, the domains of LCY compete for a single substrate, and constitute a gateway for channeling carotenoids into one of two biosynthetic routes. The LCYB domain converts lycopene to γ-carotene, a precursor to β-carotene, zeaxanthin, violaxanthin, and neoxanthin, whereas the LCYE domain converts lycopene to δ-carotene, a precursor to α-carotene and lutein. Possible selection pressures toward the separation of lycopene cyclase enzymes in higher plants include the ability to control independently expression levels of genes encoding LCYB and LCYE. The relative demand for α- and β-carotenoids varies depending on the photoacclimation state of a plant (Ballottari *et al.* 2007). By individually tuning gene expression levels, substrate flow between the two branch pathways can be more precisely regulated.

Bifunctional diterpene synthases

A similar and equally complex case involving gene duplication, and fusion and fission processes has emerged within the field of terpenoid metabolic biochemistry. Bifunctional diterpene synthases catalyze both class II and class I cyclization reactions, a process otherwise requiring two independent diterpene cyclase enzymes. All terpene synthases can be grouped into class II or class I based on the reaction mechanism and the products formed (Chen *et al.*, 2011; Zi *et al.*, 2014). Class II enzymes catalyze protonation-induced cyclization without alteration of the allylic diphosphate ester bond in the substrate, whereas class I enzymes ionize the prenyl diphosphate to generate one or more carbocation intermediates. These two enzyme classes are best illustrated in the context of gibberellin biosynthesis. The class II enzyme CPS (copalyl-labdadienyl diphosphate synthase) acts on GGPP (genanylgeranyl diphosphate) as the substrate to produce an *ent*-CPP (copalyl/labdadienyl diphosphate) intermediate, followed by the action of the class I enzyme KS (kaurene synthase) yielding *ent*-kaurene. These two reactions mark the beginning of gibberellin biosynthesis in bacteria, fungi, and plants, and CPS/KS sequence analysis has suggested a common ancestry for the cognate genes (Morrone *et al.*, 2009)

Table 1. *Examples of gene fusion products involved in metabolism in plants and algae*

Fusion	Domains	Occurrence	Significance	Pathway	Reference(s)
DHQD–SHD	DHQD+SHD	Algae; plants	DHQD and SHD are independent in bacteria, fused in plants and algae, and part of the pentafunctional, multienzyme fusion AROM in fungi and apicomplexans	Shikimate	Richards *et al.* (2006); Maeda and Dudareva (2012)
ADH–ADT	ADH+ADT	Stramenopiles (*Nannochloropsis oceanica*)	Independent in bacteria and plants; divert common substrate to either tyrosine or phenylalanine	Tyrosine and phenylalanine	Vieler *et al.* (2012)
ASα–ASβ	ASα+ASβ	Stramenopiles; some bacteria	Independent enzymes form a heteromer in plants; catalyze a single reaction	Tryptophan	Jiroutová ***et al.*** (2007); Ashenafi *et al.* (2015)
IGPS–PAI	IGPS+PAI	Stramenopiles; some fungi	Independent enzymes in bacteria and plants	Tryptophan	Jiroutová et al. (2007)
TSα–TSβ	TSα+TSβ	Most stramenopiles; some fungi	Independent enzymes in bacteria and plants	Tryptophan	Jiroutová ***et al.*** (2007)
AK–HSD	AK+HSD	Higher plants; some bacteria and algae	Monomeric AK, HSD, and fused AK–HSD all catalyze committed steps Each is allosterically regulated	Aspartate-derived amino acids	Schroeder *et al.* (2010); Vieler *et al.* (2012); Clark and Lu (2015)
DAPA-AT–DTBS	DATA-AT+DTBS	Plants; most fungi	Fusion enzyme encoded by adjacent genes in Arabidopsis Expressed via both single and chimeric mRNAs through alternative splicing	Biotin	Muralla *et al.* (2008); Cobessi *et al.* (2012)
LCY	LCYB+LCYE+LHC	Green algae	LCYP and LCYE are separate enzymes in plants and produce different carotene isomers Product stoichiometry is dependent on the linker	Carotenoids	Blatt *et al.* (2015)
DTC1[a]	CPS-type+KS-type domains	Moss	Catalyzed by independent enzymes in rice	Momilactone	Okada *et al.* (2016)
CPS–KS[a]	CPS-type+KS-type domains	Moss; fungi	Higher plant CPS and KS probably derived from duplication and subfunctionalization of an ancestral CPS–KS fusion	Gibberellin	Hayashi *et al.* (2006); Zi *et al.* (2014)
AS[a]	CPS-type+KS-type domains	Grand fir	CPS-type and KS-type domains are catalytically independent, but structurally interdependent Angiosperm diterpene synthases show loss of activity in one or the other domain	Labdane diterpenoids	Chen *et al.* (2011); Zhou *et al.* (2012); Zi *et al.* (2014)
NCS	NCS domains	Ranunculales	NCS activity despite substantial domain shuffling	BIA	Li *et al.* (2016)
REPI	DRS+DRR	Opium poppy; Persian poppy	Role of fusion unknown Could ensure co-expression or co-localization	BIA	Farrow *et al.* (2015); Winzer *et al.* (2015

[a] Members of the bifunctional diterpene synthase enzyme family containing class II (CPS) and class I (KS) terpene synthase (TPS) domains.

(Table 1). In bacteria, CPS and KS are independent enzymes, but occur as bifunctional CPS/KS diterpene synthases in the moss *Phycomitrella patens* and in fungi. In higher plants, CPS and KS again occur as separate enzymes. These observations have led to the establishment of the hypothesis that gibberellin biosynthesis may have originally relied on a bifunctional CPS/KS in the early stages of terrestrial plant evolution, arising from the fusion of CPS and KS enzymes of bacterial origin (Zi *et al.*, 2014). Subsequent gene duplication and subfunctionalization would have established the multitude of independent CPS and KS enzymes currently found in vascular plants.

Other bifunctional diterpene synthases have been identified in mosses and gymnosperms, but not in angiosperms. Prominent examples include several enzymes contributing to specialized metabolism including abietadiene synthase (AS) (Table 1), which contributes to resin acid biosynthesis in conifers (Vogel *et al.*, 1996), and *syn*-pimara-7,15-diene synthase (DTC1), which participates in momilactone biosynthesis in the moss *Hypnum plumaeforme* (Okada *et al.*, 2016). When AS was first cloned from grand fir (*Abies grandis*) >20 years ago, its evolutionary relationship with monofunctional CPS and KS enzymes was not initially apparent owing to substantial sequence divergence. Unlike

the additive fusion of complete CYP and AKR domains in REPI, AS is not comprised of discrete, 'full-complement' CPS and KS domains. Rather, AS is similar in overall length to either CPS or KS independent enzymes (Zi *et al.*, 2014). Kinetic analysis and structural determination of AS from grand fir have revealed that the spatially distinct class II and class I active sites are structurally interdependent, yet catalytically independent (Peters *et al.*, 2001; Zhou *et al.*, 2012). AS and other bifunctional terpene synthases highlight the impact of gene duplication, fusion, and sequence alteration processes within the evolution of plant specialized metabolic pathways.

Dethiobiotin synthetase–diaminopelargonic acid aminotransferase

Diaminopelargonic acid aminotransferase (DAPA-AT) and dethiobiotin synthetase (DTBS) catalyze the antepenultimate and penultimate steps of biotin biosynthesis, respectively. Despite the essential cellular functions of biotin (vitamin B8), its *de novo* production is restricted to bacteria, a few fungi, and plants. In prokaryotes, DAPA-AT and DTBS occur as separate enzymes encoded by distinct genes. In contrast, plants and fungi possess a single, bifunctional fusion enzyme (Table 1). Genetic studies on *Arabidopsis thaliana* have revealed that fusion arises from the non-stop transcription of adjacent tandem genes (*BIO1* and *BIO3*), presumably derived from prokaryote ancestral genes (Muralla *et al.*, 2008). *BIO1* and *BIO3* define a single locus, and are expressed as both single and fused chimeric BIO3–BIO1 transcripts through alternative splicing events. Among the different splice variants are a BIO3–BIO1 monocistronic fusion encoding the bifunctional enzyme, and a bicistronic variant with distinct, yet overlapping reading frames. Although a fusion event between prokaryotic ancestral genes is thought to have occurred early in the evolution of eukaryotes, the ability to produce a bicistronic transcript through differential splicing is probably a more recent evolutionary event restricted to a few angiosperm species. The bicistronic variant of *A. thaliana*, which includes 10 additional nucleotides that introduce a premature stop codon, is theoretically capable of producing separate DAPA-AT and DTBS proteins, although it was demonstrated that the fused enzyme is the sole protein product of the *BIO3–BIO1* locus (Cobessi *et al.* 2012).

Biochemical and kinetic analyses provide support for substrate channeling within the DAPA-AT–DTBS fusion enzyme. Diaminopelargonic acid (DAPA), the product of the DAPA-AT domain, is not released into the bulk solution during the reaction, but instead is directly transferred to the DTBS domain for subsequent turnover to dethiobiotin (DTB). The fusion exhibits kinetic co-operativity with respect to all substrates, suggesting substrate-induced conformational changes and allosteric interdependency of the two domains (Cobessi *et al.*, 2012). Such kinetic behavior has not been reported for the monofunctional bacterial enzymes, although a tight association and substrate channeling between monomeric DAPA-AT and DTBS enzymes is possible.

Enzymes of aromatic amino acid metabolism

The ability to produce aromatic amino acids is common to bacteria, ascomycete fungi, certain apicomplexans, plants, and various other taxa such as ciliates and diatoms. Aromatic amino acid biosynthesis begins with the formation of chorismate via the seven-step shikimate pathway, followed by branch pathways leading to tryptophan, phenylalanine, and/or tyrosine. The inheritance pattern of shikimate pathway enzymes in eukaryotes, which originated in various lineages of bacteria, is complicated by multiple horizontal gene transfer, endosymbiosis, gene fusion, gene loss, and gene replacement events (Richards *et al.*, 2006). This complexity is reflected in the diversity of enzyme arrangements among different taxa. For example, in various fungi, apicomplexans, and ciliates, five of seven shikimate pathway enzymes are fused into a single polypeptide given the abbreviation AROM. Conversely, only two of these seven enzymes are fused in plants and algae [i.e. 3-dehydroquinate dehydratase (DHQD) and shikimate dehydrogenase (SHD); Table 1] (Maeda and Dudareva, 2012). In bacteria, none of the shikimate enzymes is fused. It has been proposed that the *arom* gene fusion occurred early in the evolution of eukaryotes, and that the fusion was lost in plants and algae as the result of gene replacement mediated by multiple horizontal gene transfer events from prokaryotes, and endosymbiosis (Richards *et al.*, 2006). Metabolism beyond the shikimate pathway is similarly divergent among different taxa. For example, several enzymes (or enzyme domains) appear as fusions in stramenopile algae (e.g. *Nannochloropsis oceanica*; Vieler *et al.*, 2012), certain bacteria, and fungi, in both tryptophan and phenylalanine/tyrosine pathways (Table 1). Conversely, such fusions are not present in plants.

In plants, the selection of some fusions (DHQD–SHD) and the purported loss of others (AROM) raise questions about the general benefits, or drawbacks, of gene fusions. The structure of *A. thaliana* DHQD–SHD shows that the active sites of DHQD and SHD face each other in close proximity, which probably facilitates optimal concentration of the 3-dehydroshikimate intermediate (Singh and Christendat, 2006). However, similarly close interactions that promote substrate channeling and/or shielding from the bulk solution are often possible between separate enzymes (as described in the section below concerning the role of gene fusions in microorganisms). The interaction between bacterial DHQD and SHD enzymes is not known (Mir *et al.*, 2015). An obvious advantage to the recruitment of unfused enzymes is the opportunity for co-ordinated, yet independent, regulation. For example, although fused in some organisms, anthranilate synthase (ASα/ASβ) in plants consists of two independent subunits, the larger ASα and smaller ASβ proteins, which combine to form a heterodimer or heterotetramer (Maeda and Dudareva, 2012). ASα is encoded by two genes, one expressed constitutively and the other regulated by development, wounding, and pathogen challenge, whereas ASβ is encoded by a single gene. Inducible *ASα* and *ASβ* genes are co-ordinately regulated as a plant defense response (Niyogi *et al.*, 1993).

Aspartate kinase–homoserine dehydrogenases (AK–HSD)

In plants and microorganisms, aspartate is the common precursor of the essential amino acids lysine, threonine, methionine, and isoleucine (Jander and Joshi, 2009). The first and third steps toward the biosynthesis of three of these amino acids are catalyzed by aspartate kinase (AK) and homoserine dehydrogenase (HSD), respectively. In *A. thaliana*, a family of genes encodes three monofunctional AK enzymes and two bifunctional AK–HSD fusion proteins (Table 1). Both AK and AK–HSD enzymes are subject to feedback inhibition by their downstream products. For example, AKs are feedback inhibited by lysine, a mechanism mediated by two ACT domains at the C-terminus. In contrast, the activity of AK–HSDs is inhibited by threonine, owing to the presence of two ACT domains in the linker region spanning the AK and HSD domains (Paris *et al.*, 2003). Until more recently, monofunctional HSD enzymes were only reported in Gram-negative bacteria and yeast. The first plant monofunctional HSD enzyme was cloned from soybean and, unlike the AK–HSD fusion, was found to be threonine insensitive (Schroeder *et al.*, 2010).

How and why the AK–HSD fusion arose in amino acid metabolism has not been established. The prokaryotic occurrence of AK–HSD suggests an early physiological role for the fusion protein in bacteria that was subsequently inherited by plants and algae (Vieler *et al.*, 2012). However, AK and HSD have also been preserved as monomeric proteins, suggesting an essential role for independent enzymes in the tightly regulated pathway leading to lysine, threonine, methionine, and isoleucine. The fact that each variant is uniquely subject to feedback inhibition suggests that pathway regulation is an important factor underpinning the maintenance of all members of the gene set (Clark and Lu, 2015). Structural analyses of AK and HSD enzymes have focused on features contributing to allostery (Curien *et al.*, 2008), although an enzyme fusion has not yet been crystallized. The fact that the AK–HSD fusion catalyzes non-consecutive steps in amino acid metabolism would appear to preclude substrate channeling as a driving force behind selection of the gene fusion. Interestingly, the linker region between the AK and HSD domains is responsible for the regulatory effect of threonine in both bacterial and plant AK–HSD fusions, leading to speculation that the linker, and not the catalytic domains, is the driving force for selection of the gene fusion.

Gene fusion mechanisms

Processes resulting in chimeric proteins possessing 'fused' domains can largely be grouped into two categories: (i) genomic DNA-based gene fusion; and (ii) transcription-mediated gene fusion (Long *et al.*, 2003, 2013). In the case of genomic DNA-based gene fusions, gene duplication often provides the raw material for the emergence of new, chimeric, or shuffled genes. Examples of processes that lead to fused gene products are represented in Fig. 4. Gene duplication is a central mechanism in plants contributing to the formation of new genes, particularly in specialized metabolism (Ober, 2010). Of specific relevance to the formation of chimeric genes are unequal crossover events between sister chromatids, which occur during meiosis. In fact, this process is so common that 17% of all *A. thaliana* genes occur in tandem arrays thought to have originated via unequal crossover events (Arabidopsis Genome Initiative, 2000). Partial duplication of genes, leading to gene fission, can yield 'uncoupled' domains and potentially create entirely new chimeras when duplicated pieces are proximally juxtaposed. For example, a new protein could arise from a series of juxtaposed DNA segments that are read as exons (Fig. 4A, B) (Kaessmann, 2010). In some situations, juxtaposition is sufficiently 'tight' to yield in-frame chimeras, which are read as a single exon, or a mutation leads to fusion via the removal of a stop codon (Kumar and Rao, 2011).

Read-through transcription of neighboring tail-to-head tandem genes giving rise to chimeric mRNA is not uncommon in eukaryotes and is of potential significance in plants. Up to 110 genes in *A. thaliana* and 45 genes in rice could be transcribed as chimeric mRNA, potentially encoding chimeric or separate proteins (Shahmuradov *et al.*, 2010). Transcription-mediated gene fusion is also relevant to juxtaposed fission products, as alternative splicing might yield alternatively shuffled exons (Babushok *et al.*, 2007). The fact that NCS enzymes of the Papaveraceae show extensive repetition of complete or partial DNA segments suggests evolution involving (complete or partial) domain duplication followed by subsequent genomic DNA-based or transcription-mediated fusion processes.

Mechanisms leading to fusions between non-homologous DNA segments have been explored in the context of biosynthetic gene clusters. Horizontal gene transfer is the primary driver of gene cluster (i.e. operon) formation in microorganisms (Jensen, 2016). In contrast, plant gene clusters appear to arise *de novo* via a common ancestor, or via translocation and/or gene duplication followed by neofunctionalization and subfunctionalization (Boycheva *et al.*, 2014). So far, all reported plant gene clusters are restricted to specialized metabolism, including BIA, terpenoid, cyanogenic glucoside, and benzoxazinoid biosynthetic pathways (Nützmann and Osbourn, 2014). Of potential significance is the presence of neighboring transposable elements and retrotransposons in the vicinity of the gene clusters in several species, including *A. thaliana*, Japanese honeysuckle (*Locinera japonica*), and opium poppy (Field *et al.*, 2011; Winzer *et al.*, 2012; Krokida *et al.*, 2013). Some clusters exhibit subtelomeric positioning or are found near the centromere, both of which are genomic locations that experience higher rates of recombination, translocation, and rearrangement (Boycheva *et al.*, 2014). Translocation events could facilitate the juxtaposition or insertion of non-homologous DNA segments, forming clusters of closely linked or fused genes (Fig. 4B). For example, DNA transposition might have contributed to the fusion of genes encoding the non-homologous P450 and AKR domains of REPI. A complete opium poppy genome could shed light on whether the *REPI* gene is part of a cluster, and whether

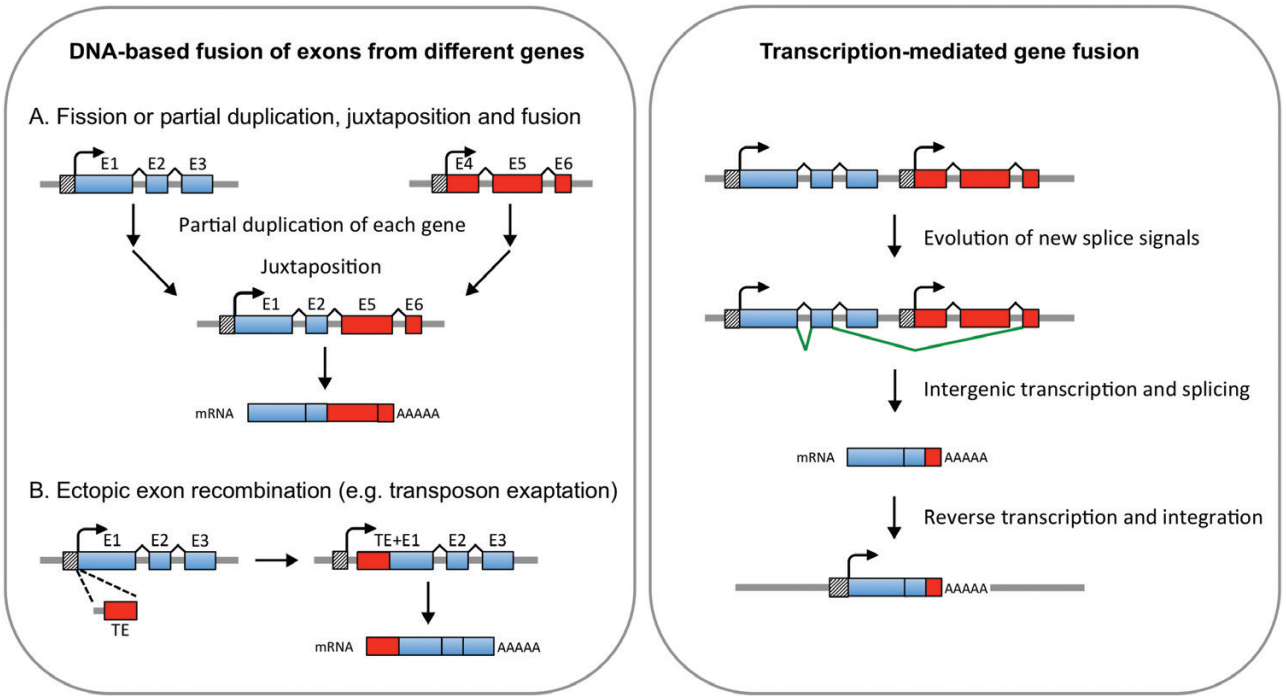

Fig. 4. Examples of DNA and transcription-mediated gene fusion mechanisms. DNA-mediated gene fusion (left panel): (A) partial duplication of two genes is followed by juxtaposition of the duplicated segments within the genome, which can yield a chimeric gene via the evolution of novel splicing signals and/or transcription termination sites; (B) ectopic exon recombination illustrated using the example of transposon exaptation (Rutter *et al.*, 2012), which is established in plants (Donoghue *et al.*, 2011). Transposon exaptation yields novel functional genes via insertion of mobile genetic elements into exons (in this case, E1) (Rutter *et al.*, 2012). Transcription-mediated gene fusion (right panel): the evolution of new splicing signals and transcriptional read-through from an upstream gene can yield new chimeric mRNAs. Chimeric mRNAs can be reverse transcribed to produce chimeric retrogenes, which can acquire or evolve promoters in their 5' upstream region. Hashed boxes indicate promoter regions. Bent arrows designate transcription start sites. Gray lines indicate genomic sequence. Blue and red shading represent exons or domains derived from different sources (e.g. genes or mobile elements) that are ultimately fused. Abbreviations: E, exon; TE, transposable element. Based on Kaessmann (2010), Long *et al.* (2013), and Donoghue *et al.* (2011).

it is located in a genomic region with features suggesting a relatively high rate of recombination.

Functional significance of gene fusions: lessons from microorganisms

As gene fusions continue to emerge within the field of plant metabolism, several hypotheses regarding their functional significance have been suggested. A popular theory is that domain fusion strongly implies functional interaction within shared metabolic pathways (Bradbury *et al.*, 2013). Obviously, fusions ensure the co-expression of otherwise independently regulated genes, and the systemic and cellular co-localization of cognate gene products. However, the impact of fusion on the kinetic performance of constituent enzyme domains, compared with the kinetic performance of the unfused components, is not well investigated in plants. In contrast, enzyme fusions are common in microorganisms, and several have been extensively studied at the biochemical level (Table 2). Insights drawn from microbial protein fusions provide a framework to consider the potential significance of fusions in plants, especially in terms of substrate channeling, altered substrate binding, catalytic efficiency, and other kinetic parameters.

In some cases, direct empirical comparisons of microbial enzymes have shown the enhanced performance of fused enzymes relative to unfused counterparts. One example involves methionine sulfoxide reductases A (MsrA) and B (MsrB), which exist as the fusion MsrAB in the bacterium *Treponema denticola*, but occur separately in other organisms (Table 2). Increased catalytic efficiency was determined as the result of fusing independent MsrA and MsrB domains, but the improvement was dependent on the specific linker region found in *T. denticola* MsrAB (Han *et al.*, 2016). As discussed earlier, a critical role for the linker region was also reported for the LCY fusion in green algae (Table 1). Remarkably, it was shown that product stoichiometry could be modulated by modifying the linker region of the LCY protein fusion, or by truncating the C-terminal LHC domain (Blatt *et al.*, 2015). In some systems, the presence of one domain appears to influence the substrate binding properties of the other, possibly through interdomain interactions. The bacterial fusion protein GshF, comprised of γ-ECL (γ-glutamylcysteine ligase) and GS (glutathione synthetase) domains, completes glutathione biosynthesis in a single step. Kinetic studies have shown that the GS domain, which catalyzes the second half of the two-part reaction, binds exogenously added glycine substrate at physiologically irrelevant (i.e. extremely high) K_m values. Conversely, when glycine is provided through its fusion partner γ-ECL, the K_m of the GS domain is dramatically reduced (Vergauwen *et al.*, 2006). Independent GS enzymes found in other microbial species bind glycine

Table 2. *Examples of gene fusion products in microbial metabolism*

Fusion	Domains	Occurrence	Significance	Pathway	Reference(s)
PutA	PRODH+P5CDH	Fusion in Gram-negative bacteria Separate in other bacteria and eukaryotes	Substrate channeling occurs in both fused and unfused enzymes	Proline metabolism	Moxley *et al.* (2014); Sanyal *et al.* (2015)
Bifunctional DAH7PS	DAH7PS+CM	*Geobacillus* spp.	Fusion imparts allosteric regulation in chorismate biosynthesis	Shikimate metabolism	Nazmi *et al.* (2016)
MsrAB	MsrA+MsrB	*Treponema denticola*	Increased catalytic efficiency compared with unfused enzymes Linker region crucial	Protein repair	Han *et al.* (2016)
CYP170	CYP170A+terpene synthase	*Streptomyces* spp.	Terpene synthase domain functional only in some fusions	Albaflavenone metabolism	Moody *et al.* (2012)
GshF	γ-ECL+GS	*Streptococcus agalactiae* *Pasteurella multocida*	GS domain mediates dimer formation Reduced K_m for GS domain in fusion	Glutathione metabolism	Stout *et al.* (2012)
SACTE_2871	Dioxygenase+CBM	*Streptomyces* spp. SirexAA-E	CBM domain localizes the catalytic domain near lignin substrate	Lignin degradation	Bianchetti *et al.* (2013)
FARAT	FAR+DHAPAT	*Tetrahymena thermophila*	Fusion catalyzes two parallel conversions providing both substrates for a third reaction	Fatty acid metabolism	Dittrich-Domergue *et al.* (2014)
TS–DHFR	TS+DHFR	*Toxoplasma gondii*	Interdomain interaction Substrate channeling and bulk solution shielding	Folate metabolism	Sharma *et al.* (2013)

with low K_m values without the need for an interacting partner (Janowiak and Griffith, 2005; Janowiak *et al.*, 2006; Vergauwen *et al.*, 2006). Unlike the microbial GshF fusion, the K_m of isolated DRS for (*S*)-reticuline was not dramatically different compared with the K_m of DRS as part of REPI in opium poppy, although the apparent V_{max} was somewhat lower for the fusion (Farrow *et al.*, 2015) (Fig. 1). The K_m and V_{max} of DRR for 1,2-dehydroreticuline were not compared with equivalent REPI parameters.

Although the functional or physiological significance of the REPI fusion is not known, the linked domains potentially co-ordinate or shield the intermediate 1,2-dehydroreticuline from the bulk solution, thereby preventing enamine formation (Farrow *et al.*, 2015). Direct evidence of substrate channeling between fused domains was shown for the bifunctional enzyme thymidylate synthase–dihydrofolate reductase (TS–DHFR) (Sharma *et al.*, 2013) (Table 2). TS and DHFR occur as separate enzymes in most species, but are fused in the parasitic protozoan *Toxoplasma gondii*. Structural analysis of TS–DHFR revealed that dihydrofolate intermediate is directly transferred between TS and DHFR active sites without entering the bulk solution. This direct transfer of substrate between active sites and shielding from bulk solution was noted in the case of the *A. thaliana* fusion DAPA-AT–DTBS (Table 1) (Cobessi *et al.*, 2012). While it is tempting to draw conclusions from such studies regarding the significance of fusion, it is important to exercise caution in the absence of a proper functional comparison of fused and unfused components. Importantly, substrate channeling is also known to occur between tightly associated, but distinct monomeric proteins. For example, channeling of a reactive intermediate acetaldehyde occurs in the microbial DmpFG enzyme, a heterodimer comprised of unfused aldolase (DmpG) and dehydrogenase (DmpF) domains (Smith *et al.*, 2012). Moreover, channeling occurs between the unfused monomer domains in the aldolase–dehydrogenase complex BphI–BphJ (Carere *et al.*, 2011). The question of whether fusion provides added benefits in terms of substrate channeling was addressed directly in a recent study involving proline metabolic enzymes. Proline dehydrogenase (PRODH) and Δ^1-pyrroline-5-carboxylate dehydrogenase (P5CDH) catalyze successive oxidations converting proline to glutamate. In most organisms, PRODH and P5CDH occur as independent proteins, but these enzymes are fused as PutA in Gram-negative bacteria. Available crystal structures (Srivastava *et al.*, 2010; Singh *et al.*, 2014) together with steady-state and transient kinetic data (Moxley *et al.*, 2014) have established the mechanism underpinning substrate channeling in PutA. However, it was also shown that monomeric, unfused PRODH and P5CDH enzymes from *Thermus thermophilus* utilize an analogous channeling mechanism (Sanyal *et al.*, 2015). This finding is in line with the 'Rosetta Stone hypothesis', which uses gene fusion events to predict protein–protein interactions (Enright *et al.*, 1999; Marcotte *et al.*, 1999), and raises doubts about enzyme–enzyme fusions as a requirement for, or enhancement of, substrate channeling.

Future work aimed at elucidating substrate channeling and the possible roles of metabolons in plant metabolism should enlist insights available from both fusions and, when possible, monomeric enzymes. A generally overlooked feature

in microbial systems is the potential role of fusions in co-localization. For example, the DRS component of REPI ensures localization to the ER membrane by virtue of its CYP N-terminal endomembrane anchor. A landmark study of metabolite channeling via a dynamic metabolon in sorghum (Laursen *et al.*, 2016) highlights the key anchoring and recruitment roles played by CYPs. The four-enzyme metabolon in sorghum is involved in the formation of the defensive compound dhurrin and, although it does not include fused enzymes, the CYPs form membrane-bound hetero- and homo-oligomers that recruit a soluble glucosyltransferase. In the case of opium poppy, the REPI fusion could conceivably represent an alternative solution for the recruitment of a soluble enzyme (i.e. DRR) to the membrane. Beyond potential for substrate channeling and co-localization, a fusion could offer reaction-specific advantages, particularly with respect to enzymes catalyzing reversible reactions. For example, it was suggested that the DAPA-AT–DTBS fusion in *A. thaliana* could serve to promote the 'forward' direction of the DAPA-AT, as the product of DAPA-AT is channeled to its fusion partner DTBS instead of acted upon by DAPA-AT in the reverse direction (Cobessi *et al.*, 2012). This potential advantage is also relevant to REPI, as DRR catalyzes a reversible reaction (Farrow *et al.*, 2015) (Fig. 1). Isolated DRR converts 1,2-dehydroreticuline to (*R*)-reticuline, but also catalyzes the reverse reaction forming 1,2-dehydroreticuline from (*R*)-reticuline. Conversely, when DRR is fused to DRS as REPI, (*R*)-reticuline is not accepted as a substrate, highlighting a significant and potentially relevant change in enzyme function. This feature could be further explored with additional kinetic analyses of REPI, DRS, and DRR, and by comparing the metabolite profiles of (*S*)-reticuline, (*R*)-reticuline, and 1,2-dehydroreticuline in REPI-expressing plants with profiles of plants harboring unfused DRS and DRR enzymes.

Biotechnological applications of gene fusions

Cases of perceived improvement in enzyme kinetic features have prompted the application of gene fusions in biotechnology, particularly in the fields of plant and microbial metabolic engineering. The emergence of synthetic biology platforms in microorganisms and plants offers new and/or improved opportunities for the commercial production of natural or semi-synthetic products (Wurtzel and Kutchan, 2016). Metabolic engineering is also a popular tool for the improvement of crop plants in terms of resistance to environmental stress. The use of gene fusions in biotechnology is associated with both advantages and disadvantages. One advantage is that the host organism needs only to be transformed once, rather than twice, with a genetic construct encoding heterologous proteins. This feature is attractive when working with a plant species recalcitrant to genetic transformation, or in cases where plant breeding is required subsequent to the transformation event. Rice plants transformed with a fusion of two bacterial trehalose biosynthetic genes *otsA* and *otsB* were reported to display less photo-oxidative damage and a more favorable mineral balance compared with wild-type plants under salt, drought, and low temperature stress conditions (Garg *et al.*, 2002) (Table 3). Although a direct comparison was not made with rice co-transformed or successively transformed with independent *otsA* and *otsB* genes, the *otsA–otsB* fusion produced in *E. coli* exhibited a higher catalytic efficiency *in vitro* compared with the unfused components (Seo *et al.*, 2000). Increased salt tolerance has also been achieved in perennial ryegrass (*Lolium perenne*) through transformation with a plant gene fusion (Cen *et al.*, 2016), and expression of a microbial fusion in geranium (*Pelargonium* sp. *Frensham*) chloroplasts enhanced the plants' ability to assimilate and phytoremediate formaldehyde (Song *et al.*, 2010).

Table 3. *Examples of the application of gene fusion products in biotechnology*

Fusion	Host	Result	References
4-Coumaroyl-CoA ligase (4CL)[a]+stilbene synthase (STS)[a]	Yeast *E. coli*	Resveratrol production higher compared with unfused enzyme expression (yeast) Not compared in *E. coli*[b]	Zhang *et al.* (2006, 2015)
Drought and Salt Tolerance (DST)[a]+SRDX domain[a]	Perennial ryegrass	Gene expression suppression Salt tolerance compared with non-transgenics	Cen *et al.* (2016)
Isoflavone synthase (IFS)[a]+P450 reductase Rhf-RED	*E. coli*	P450 activity preserved[b]	Schückel *et al.* (2012)
Various P450s[a]+CPR1[a]	*E. coli*	P450 activity preserved[b]	Schückel *et al.* (2012)
Trehalose-6-phosphate synthase (TPS; *OtsA*)+trehalose-6-phosphate phosphatase (TPP; *OtsB*)	Rice	Trehalose accumulation Improved abiotic stress tolerance[c]	Garg *et al.* (2002)
Fatty acid reductase (FAR)+wax synthase (WS)	Tobacco	Increased wax ester production Altered wax composition Same result using unfused enzymes	Aslan *et al.* (2014, 2015)
4-Coumaroyl-CoA ligase (4CL)[a]+benzalacetone synthase (BAS)[a]	Yeast	Improved raspberry ketone production over unfused enzymes	Lee *et al.* (2016)
3-Hexulose-6-phosphate synthase (HPS)+6-phosphate-3-hexuloisomerase (PHI)	Geranium	Enhanced ability to assimilate formaldehyde[b]	Song *et al.* (2010)

[a] Plant enzyme.
[b] No comparison versus unfused enzymes.
[c] Fusion shows improved performance *in vitro*.

An additional benefit of using gene fusions in metabolic engineering applications is that the two (or more) protein domains are implicitly co-localized within the host organism. The host organism might not recognize localization information contained on a foreign protein, or possess the ideal cellular or subcellular anatomy for adequate enzyme performance. For example, bacteria do not possess the endomembrane structure needed for optimal performance of ER-bound plant CYPs, particularly as these CYPs require a second ER-bound enzyme, cytochrome P450 reductase (CPR), for efficient electron transfer. Together with other modifications, gene fusion technology has enabled plant CYP expression in *E. coli* through in-frame coupling with a microbial or plant CPR (Schückel *et al.*, 2012) (Table 3). Combining microbial and plant enzymes illustrates a third benefit to the application of gene fusions; the ability to create novel enzymes that might not exist in nature. A recent study involving the heterologous production of raspberry ketone in baker's yeast (*S. cerevisiae*) empirically and extensively evaluated different gene fusion possibilities to obtain the optimal enzyme–linker–enzyme combination (Lee *et al.*, 2016). 4-Coumaroyl-CoA ligase (4CL) and benzalacetone synthase (BAS) were fused in two different orientations (i.e. 4CL–BAS or BAS–4CL) with either flexible or rigid linkers. Additionally, the performance of two different 4CL variants was compared. This mix-and-match approach resulted in a fusion that increased yields >5-fold compared with native monofunctional enzymes.

The factors enhancing fusion performance relative to monofunctional enzymes in heterologous systems, when an empirical comparison is actually performed, are rarely explored, arguably because biotechnological achievements are driven by outcomes rather than curiosity. For example, as part of an engineering project aimed at resveratrol production in alternative systems, yeast expressing a fusion between 4CL and stilbene synthase (STS) was shown dramatically to outperform strains that co-expressed 4CL and STS as independent enzymes (Zhang *et al.*, 2006). The result was not pursued further, although it was noted that protein–protein interactions could increase metabolic channeling or place enzyme active sites in close proximity. A largely overlooked disadvantage of using fusions in biotechnology is that the non-natural linkage of two otherwise active enzymes could have detrimental effects on functionality via unforeseen problems in protein folding and/or stability, antagonistic domain–domain interactions, inadequate substrate transfer, or negative impact on substrate binding. Unfortunately, failed artificial gene fusion outcomes are usually not reported. Considering the evolutionary time scale required to select for active fusions in plants and microorganisms, and the many gene fusion or shuffling 'failures' found in nature, such as catalytically active domains fused to inactive domains (Compaan and Ellington, 2003; Hiltunen *et al.*, 2012; Moody *et al.*, 2012; Zi *et al.* 2014) and the numerous inactive NCS fusions (Li *et al.*, 2016) (Table 1), gene fusion engineering is a risky endeavor. Even if the fusion does not result in a loss of enzyme function, a significant functional improvement compared with the unfused components may not result. Work involving increased wax ester production in tobacco using bacterial fatty acid reductase (FAR) and wax synthase (WS) genes showed that transient expression in leaves of a gene fusion encoding FAR–WS was functionally equivalent to the outcome using unfused genes (Aslan *et al.*, 2014) (Table 3). Although gene fusion remains a common strategy in attempts to debottleneck engineered metabolic pathways in plants and microorganisms (Elleuche, 2015; Yu *et al.*, 2015), success is sporadic and must be appropriately measured against suitable controls. In the future, new genome editing technologies could be used to test hypotheses surrounding the contribution—or lack thereof—of fusions. For example, in plant species where sufficient genomic data are available, single-domain enzymes could be replaced with gene fusions and impacts on metabolism could be studied.

Conclusions and perspectives

Gene fusion events generated a plethora of protein chimeras in plants, microbes, and other organisms. Dozens of papers report the occurrence and function of gene fusions in the context of microbial metabolism. In contrast, reports of gene fusions relevant to plant metabolism are limited. This disparity probably reflects a knowledge gap concerning plant gene fusions, rather than a reduced occurrence or functional importance in metabolism compared with microbes. Regardless of whether such events are the result of DNA shuffling at the genomic level, or by alternative splicing and similar transcription-based mechanisms, new arrangements of existing protein domains emerge. A review of gene fusion would not be complete without a consideration of gene duplication and gene clustering, and the possible roles of these features in predictions of enzyme function within metabolic pathways. Significantly, all known gene clusters in plants encompass components that participate in specialized metabolism (Boycheva *et al.*, 2014; Nützmann and Osbourn, 2014). Although this is not strictly true in the case of gene fusions, it is remarkable that at least two gene fusion events are relevant to BIA metabolism alone. Moreover, gene fusion, fission, subfunctionalization, and neofunctionalization processes have contributed to the enormous diversity of terpene synthases found in plants (Zi *et al.*, 2014). This review is largely restricted to more discernible or established cases of gene fusion, such as bifunctional enzymes with two distinct active sites bearing evident relationships to extant monofunctional enzymes. However, protein domains are known to correlate with exon structures (Liu and Grigoriev, 2004), and singular domains can merge with a variety of different partners yielding functionally unrelated enzymes that are not considered bifunctional (Derouiche *et al.*, 2016), and whose genetic origins as fusions are not obvious. It is likely that many multidomain enzymes, particularly those sharing domain-level homology with different enzyme types, arose from some combination of gene fusion, fission, duplication, or rearrangement. Whether this evolutionary mechanism can be mimicked to engineer plant pathways more efficiently in microorganisms, or to generate more robust plants, appears presently a matter of luck.

Acknowledgements

This work was supported by financial contributions from a Natural Sciences and Engineering Research Council of Canada Discovery Grant, and from Epimeron Inc. PJF and JMH own stock in, serve on the Board of Directors of, and are provided compensation by Epimeron.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology **215,** 403–410.

Angelova S, Buchheim M, Frowitter D, Schierhorn A, Roos W. 2010. Overproduction of alkaloid phytoalexins in California poppy cells is associated with the co-expression of biosynthetic and stress-protective enzymes. Molecular Plant **3,** 927–939.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408,** 796–815.

Ashenafi M, Reddy PT, Parsons JF, Byrnes WM. 2015. The fused anthranilate synthase from *Streptomyces venezuelae* functions as a monomer. Molecular and Cellular Biochemistry **400,** 9–15.

Aslan S, Hofvander P, Dutta P, Sun C, Sitbon F. 2015. Increased production of wax esters in transgenic tobacco plants by expression of a fatty acid reductase:wax synthase gene fusion. Transgenic Research **24,** 945–953.

Aslan S, Sun C, Leonova S, Dutta P, Dörmann P, Domergue F, Stymne S, Hofvander P. 2014. Wax esters of different compositions produced via engineering of leaf chloroplast metabolism in *Nicotiana benthamiana*. Metabolic Engineering **25,** 103–112.

Babushok DV, Ohshima K, Ostertag EM, Chen X, Wang Y, Mandal PK, Okada N, Abrams CS, Kazazian HH Jr. 2007. A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. Genome Research **17,** 1129–1138.

Ballottari M, Dall'Osto L, Morosinotto T, Bassi R. 2007. Contrasting behavior of higher plant photosystem I and II antenna systems during acclimation. Journal of Biological Chemistry **282,** 8947–8958.

Bardou F, Ariel F, Simpson CG, Romero-Barrios N, Laporte P, Balzergue S, Brown JW, Crespi M. 2014. Long noncoding RNA modulates alternative splicing regulators in *Arabidopsis*. Developmental Cell **30,** 166–176.

Bianchetti CM, Harmann CH, Takasuka TE, Hura GL, Dyer K, Fox BG. 2013. Fusion of dioxygenase and lignin-binding domains in a novel secreted enzyme from cellulolytic Streptomyces sp. SirexAA-E. Journal of Biological Chemistry **288,** 18574–18587.

Blatt A, Bauch ME, Pörschke Y, Lohr M. 2015. A lycopene β-cyclase/lycopene ε-cyclase/light-harvesting complex-fusion protein from the green alga *Ostreococcus lucimarinus* can be modified to produce α-carotene and β-carotene at different ratios. The Plant Journal **82,** 582–595.

Boycheva S, Daviet L, Wolfender JL, Fitzpatrick TB. 2014. The rise of operon-like gene clusters in plants. Trends in Plant Science **19,** 447–459.

Bradbury LM, Niehaus TD, Hanson AD. 2013. Comparative genomics approaches to understanding and manipulating plant metabolism. Current Opinion in Biotechnology **24,** 278–284.

Carere J, Baker P, Seah SY. 2011. Investigating the molecular determinants for substrate channeling in BphI–BphJ, an aldolase–dehydrogenase complex from the polychlorinated biphenyls degradation pathway. Biochemistry **50,** 8407–8416.

Cen H, Ye W, Liu Y, Li D, Wang K, Zhang W. 2016. Overexpression of a chimeric gene, OsDST–SRDX, improved salt tolerance of perennial ryegrass. Scientific Reports **6,** 27320.

Chen F, Tholl D, Bohlmann J, Pichersky E. 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. The Plant Journal **66,** 212–229.

Clark TJ, Lu Y. 2015. Analysis of loss-of-function mutants in aspartate kinase and homoserine dehydrogenase genes points to complexity in the regulation of aspartate-derived amino acid contents. Plant Physiology **168,** 1512–1526.

Cobessi D, Dumas R, Pautre V, Meinguet C, Ferrer JL, Alban C. 2012. Biochemical and structural characterization of the *Arabidopsis* bifunctional enzyme dethiobiotin synthetase–diaminopelargonic acid aminotransferase: evidence for substrate channeling in biotin synthesis. The Plant Cell **24,** 1608–1625.

Compaan DM, Ellington WR. 2003. Functional consequences of a gene duplication and fusion event in an arginine kinase. Journal of Experimental Biology **206,** 1545–1556.

Curien G, Biou V, Mas-Droux C, Robert-Genthon M, Ferrer JL, Dumas R. 2008. Amino acid biosynthesis: new architectures in allosteric enzymes. Plant Physiology and Biochemistry **46,** 325–339.

Derouiche A, Shi L, Kalantari A, Mijakovic I. 2016. Evolution and tinkering: what do a protein kinase, a transcriptional regulator and chromosome segregation/cell division proteins have in common? Current Genetics **62,** 67–70.

Dittrich-Domergue F, Joubès J, Moreau P, Lessire R, Stymne S, Domergue F. 2014. The bifunctional protein TtFARAT from *Tetrahymena thermophila* catalyzes the formation of both precursors required to initiate ether lipid biosynthesis. Journal of Biological Chemistry **289,** 21984–21994.

Donoghue MT, Keshavaiah C, Swamidatta SH, Spillane C. 2011. Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana*. BMC Evolutionary Biology **11,** 47.

Elleuche S. 2015. Bringing functions together with fusion enzymes—from nature's inventions to biotechnological applications. Applied Microbiology and Biotechnology **99,** 1545–1556.

Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA. 1999. Protein interaction maps for complete genomes based on gene fusion events. Nature **402,** 86–90.

Espinosa-Cantú A, Ascencio D, Barona-Gómez F, DeLuna A. 2015. Gene duplication and the evolution of moonlighting proteins. Frontiers in Genetics **6,** 227.

Farrow SC, Hagel JM, Beaudoin GA, Burns DC, Facchini PJ. 2015. Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. Nature Chemical Biology **11,** 728–732.

Field B, Fiston-Lavier A-S, Kemen A, Geisler K, Quesneville H, Osbourn AE. 2011. Formation of plant metabolic gene clusters within dynamic chromosomal regions. Proceedings of the National Academy of Sciences, USA **108,** 16116–16121.

Garg AK, Kim J-K, Owens TG, Ranwala AP, Choi YD, Kochian LV, Wu RJ. 2002. Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proceedings of the National Academy of Sciences, USA **99,** 15898–15903.

Gesell A, Rolf M, Ziegler J, Díaz Chávez ML, Huang FC, Kutchan TM. 2009. CYP719B1 is salutaridine synthase, the C–C phenol-coupling enzyme of morphine biosynthesis in opium poppy. Journal of Biological Chemistry **284,** 24432–24442.

Graur D, Li W-H. 2000. Fundamentals of molecular evolution, 2nd edn. Sunderland, MA: Sinauer Associates.

Hagel JM, Facchini PJ. 2012. Subcellular localization of sanguinarine biosynthetic enzymes in cultured opium poppy cells. In Vitro Cellular and Developmental Biology (Plant) **48,** 233–240.

Hagel JM, Facchini PJ. 2013. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. Plant and Cell Physiology **54,** 647–672.

Han AR, Kim MJ, Kwak GH, Son J, Hwang KY, Kim HY. 2016. Essential role of the linker region in the higher catalytic efficiency of a bifunctional MsrA–MsrB fusion protein. Biochemistry **55,** 5117–5127.

Hayashi K, Kawaide H, Notomi M, Sakigi Y, Matsuo A, Nozaki H. 2006. Identification and functional analysis of bifunctional ent-kaurene synthase from the moss *Physcomitrella patens*. FEBS Letters **580,** 6175–6181.

Hiltunen HM, Illarionov B, Hedtke B, Fischer M, Grimm B. 2012. *Arabidopsis* RIBA proteins: two out of three isoforms have lost their bifunctional activity in riboflavin biosynthesis. International Journal of Molecular Sciences **13,** 14086–14105.

Ilari A, Franceschini S, Bonamore A, Arenghi F, Botta B, Macone A, Pasquo A, Bellucci L, Boffi A. 2009. Structural basis of enzymatic (S)-norcoclaurine biosynthesis. Journal of Biological Chemistry **284,** 897–904.

Jander G, Joshi V. 2009. Aspartate-derived amino acid biosynthesis in *Arabidopsis thaliana*. The Arabidopsis Book **7,** e0121.

Janowiak BE, Griffith OW. 2005. Glutathione synthesis in *Streptococcus agalactiae*. One protein accounts for gamma-glutamylcysteine synthetase and glutathione synthetase activities. Journal of Biological Chemistry **280,** 11829–11839.

Janowiak BE, Hayward MA, Peterson FC, Volkman BF, Griffith OW. 2006. Gamma-glutamylcysteine synthetase–glutathione synthetase: domain structure and identification of residues important in substrate and glutathione binding. Biochemistry **45,** 10461–10473.

Jensen PR. 2016. Natural products and the gene cluster revolution. Trends in Microbiology **24,** 968–977.

Jiroutová K, Horák A, Bowler C, Oborník M. 2007. Tryptophan biosynthesis in stramenopiles: eukaryotic winners in the diatom complex chloroplast. Journal of Molecular Evolution **65,** 496–511.

Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Research **20,** 1313–1326.

Krokida A, Delis C, Geisler K, Garagounis C, Tsikou D, Peña-Rodríguez LM, Katsarou D, Field B, Osbourn AE, Papadopoulou KK. 2013. A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis. New Phytologist **200,** 675–690.

Kumar R, Rao DN. 2011. A nucleotide insertion between two adjacent methyltransferases in *Helicobacter pylori* results in a bifunctional DNA methyltransferase. Biochemical Journal **433,** 487–495.

Laursen T, Borch J, Knudsen C, *et al*. 2016. Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. Science **354,** 890–893.

Lee D, Lloyd ND, Pretorius IS, Borneman AR. 2016. Heterologous production of raspberry ketone in the wine yeast *Saccharomyces cerevisiae* via pathway engineering and synthetic enzyme fusion. Microbial Cell Factories **15,** 49.

Lee EJ, Facchini P. 2010. Norcoclaurine synthase is a member of the pathogenesis-related 10/Bet v1 protein family. The Plant Cell **22,** 3489–3503.

Levy A, Milo J. 1998. Genetics and breeding of *Papaver somniferum*. In: **Bernáth J**, ed. Poppy, the genus *Papaver*. Amsterdam: OPA (Overseas Publishers Association), 93–103.

Li J, Lee E-J, Chang L, Facchini PJ. 2016. Genes encoding norcoclaurine synthase occur as tandem fusions in the Papaveraceae. Scientific Reports **6,** e39256.

Lichman BR, Gershater MC, Lamming ED, Pesnot T, Sula A, Keep NH, Hailes HC, Ward JM. 2015. 'Dopamine-first' mechanism enables the rational engineering of the norcoclaurine synthase aldehyde activity profile. FEBS Journal **282,** 1137–1151.

Liscombe DK, MacLeod BP, Loukanina N, Nandi OI, Facchini PJ. 2005. Evidence for the monophyletic evolution of benzylisoquinoline alkaloid biosynthesis in angiosperms. Phytochemistry **66,** 2501–2520.

Liu M, Grigoriev A. 2004. Protein domains correlate strongly with exons in multiple eukaryotic genomes—evidence of exon shuffling? Trends in Genetics **20,** 399–403.

Long M, Betrán E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. Nature Reviews Genetics **4,** 865–875.

Long M, VanKuren NW, Chen S, Vibranovski MD. 2013. New gene evolution: little did we know. Annual Review of Genetics **47,** 307–333.

Maeda H, Dudareva N. 2012. The shikimate pathway and aromatic amino acid biosynthesis in plants. Annual Review of Plant Biology **63,** 73–105.

Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. 1999. Detecting protein function and protein–protein interactions from genome sequences. Science **285,** 751–753.

Marin B, Melkonian M. 2010. Molecular phylogeny and classification of the *Mamiellophyceae class*. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. Protist **161,** 304–336.

McLysaght A, Guerzoni D. 2015. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. Philosophical Transactions of the Royal Society B: Biological Sciences **370,** 20140332.

Mir R, Jallu S, Singh TP. 2015. The shikimate pathway: review of amino acid sequence, function and three-dimensional structures of the enzymes. Critical Reviews in Microbiology **41,** 172–189.

Morrone D, Chambers J, Lowry L, Kim G, Anterola A, Bender K, Peters RJ. 2009. Gibberellin biosynthesis in bacteria: separate ent-copalyl diphosphate and ent-kaurene synthases in *Bradyrhizobium japonicum*. FEBS Letters **583,** 475–480.

Moody SC, Zhao B, Lei L, Nelson DR, Mullins JG, Waterman MR, Kelly SL, Lamb DC. 2012. Investigating conservation of the albaflavenone biosynthetic pathway and CYP170 bifunctionality in streptomycetes. FEBS Journal **279,** 1640–1649.

Moxley MA, Sanyal N, Krishnan N, Tanner JJ, Becker DF. 2014. Evidence for hysteretic substrate channeling in the proline dehydrogenase and Δ1-pyrroline-5-carboxylate dehydrogenase coupled reaction of proline utilization A (PutA). Journal of Biological Chemistry **289,** 3639–3651.

Muralla R, Chen E, Sweeney C, Gray JA, Dickerman A, Nikolau BJ, Meinke D. 2008. A bifunctional locus (BIO3–BIO1) required for biotin biosynthesis in *Arabidopsis*. Plant Physiology **146,** 60–73.

Nazmi AR, Lang EJM, Bai Y, Allison TM, Othman MH, Panjikar S, Arcus VL, Parker EJ. 2016. Interdomain conformational changes provide allosteric regulation en route to chorismate. Journal of Biological Chemistry **291,** 21836–21847.

Niyogi KK, Last RL, Fink GR, Keith B. 1993. Suppressors of *trp1* fluorescence identify a new *Arabidopsis* gene, *TRP4*, encoding the anthranilate synthase beta subunit. The Plant Cell **5,** 1011–1027.

Nützmann HW, Osbourn A. 2014. Gene clustering in plant specialized metabolism. Current Opinion in Biotechnology **26,** 91–99.

Ober D. 2010. Gene duplications and the time thereafter—examples from plant secondary metabolism. Plant Biology **12,** 570–577.

Okada K, Kawaide H, Miyamoto K, *et al*. 2016. HpDTC1, a stress-inducible bifunctional diterpene cyclase involved in momilactone biosynthesis, functions in chemical defence in the moss *Hypnum plumaeforme*. Scientific Reports **6,** 25316.

Paris S, Viemon C, Curien G, Dumas R. 2003. Mechanism of control of *Arabidopsis thaliana* aspartate kinase–homoserine dehydrogenase by threonine. Journal of Biological Chemistry **278,** 5361–5366.

Peters RJ, Ravn MM, Coates RM, Croteau RB. 2001. Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. Journal of the American Chemical Society **123,** 8974–8978.

Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, McLeod R, Roberts CW. 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. Eukaryotic Cell **5,** 1517–1531.

Rutter MT, Cross KV, Van Woert PA. 2012. Birth, death and subfunctionalization in the *Arabidopsis* genome. Trends in Plant Science **17,** 204–212.

Samanani N, Facchini PJ. 2001. Isolation and partial characterization of norcoclaurine synthase, the first committed step in benzylisoquinoline alkaloid biosynthesis, from opium poppy. Planta **213,** 898–906.

Samanani N, Liscombe DK, Facchini PJ. 2004. Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis. The Plant Journal **40,** 302–313.

Sanyal N, Arentson BW, Luo M, Tanner JJ, Becker DF. 2015. First evidence for substrate channeling between proline catabolic enzymes: a validation of domain fusion analysis for predicting protein–protein interactions. Journal of Biological Chemistry **290,** 2225–2234.

Schroeder AC, Zhu C, Yanamadala SR, Cahoon RE, Arkus KA, Wachsstock L, Bleeke J, Krishnan HB, Jez JM. 2010. Threonine-insensitive homoserine dehydrogenase from soybean: genomic organization, kinetic mechanism, and in vivo activity. Journal of Biological Chemistry **285,** 827–834.

Schückel J, Rylott EL, Grogan G, Bruce NC. 2012. A gene-fusion approach to enabling plant cytochromes p450 for biocatalysis. ChemBioChem **13,** 2758–2763.

Seo HS, Koo YJ, Lim JY, Song JT, Kim CH, Kim JK, Lee JS, Choi YD. 2000. Characterization of a bifunctional enzyme fusion of trehalose-6-phosphate synthetase and trehalose-6-phosphate phosphatase of *Escherichia coli*. Applied and Environmental Microbiology **66,** 2484–2490.

Shahmuradov IA, Abdulazimova AU, Solovyev VV, Qamar R, Cholan SN, Aliyev JA. 2010. Mono- and bi-cistronic chimeric mRNAs in *Arabidopsis* and rice genomes. Applied and Computational Mathematics **9,** 66–81.

Sharma H, Landau MJ, Vargo MA, Spasov KA, Anderson KS. 2013. First three-dimensional structure of *Toxoplasma gondii* thymidylate synthase–dihydrofolate reductase: insights for catalysis, interdomain interactions, and substrate channeling. Biochemistry **52,** 7305–7317.

Sierro N, Battey JN, Ouadi S, Bakaher N, Bovet L, Willig A, Goepfert S, Peitsch MC, Ivanov NV. 2014. The tobacco genome sequence and its comparison with those of tomato and potato. Nature Communications **5,** 3833.

Singh H, Arentson BW, Becker DF, Tanner JJ. 2014. Structures of the PutA peripheral membrane flavoenzyme reveal a dynamic substrate-channeling tunnel and the quinone-binding site. Proceedings of the National Academy of Sciences, USA **111,** 3389–3394.

Singh SA, Christendat D. 2006. Structure of *Arabidopsis* dehydroquinate dehydratase–shikimate dehydrogenase and implications for metabolic channeling in the shikimate pathway. Biochemistry **45,** 7787–7796.

Smith NE, Vrielink A, Attwood PV, Corry B. 2012. Biological channeling of a reactive intermediate in the bifunctional enzyme DmpFG. Biophysical Journal **102,** 868–877.

Song Z, Orita I, Yin F, Yurimoto H, Kato N, Sakai Y, Izui K, Li K, Chen L. 2010. Overexpression of an HPS/PHI fusion enzyme from *Mycobacterium gastri* in chloroplasts of geranium enhances its ability to assimilate and phytoremediate formaldehyde. Biotechnology Letters **32,** 1541–1548.

Srivastava D, Schuermann JP, White TA, Krishnan N, Sanyal N, Hura GL, Tan A, Henzl MT, Becker DF, Tanner JJ. 2010. Crystal structure of the bifunctional proline utilization A flavoenzyme from *Bradyrhizobium japonicum*. Proceedings of the National Academy of Sciences, USA **107,** 2878–2883.

Stout J, De Vos D, Vergauwen B, Savvides SN. 2012. Glutathione biosynthesis in bacteria by bifunctional GshF is driven by a modular structure featuring a novel hybrid ATP-grasp fold. Journal of Molecular Biology **416,** 486–494.

Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485,** 635–641.

Turmel M, Gagnon MC, O'Kelly CJ, Otis C, Lemieux C. 2009. The chloroplast genomes of the green algae Pyramimonas, Monomastix, and Pycnococcus shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. Molecular Biology and Evolution **26,** 631–648.

Vergauwen B, De Vos D, Van Beeumen JJ. 2006. Characterization of the bifunctional gamma-glutamate-cysteine ligase/glutathione synthetase (GshF) of *Pasteurella multocida*. Journal of Biological Chemistry **281,** 4380–4394.

Vieler A, Wu G, Tsai CH, *et al*. 2012. Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. PLoS Genetics **8,** e1003064.

Vimolmangkang S, Deng X, Owiti A, Meelaph T, Ogutu C, Han Y. 2016. Evolutionary origin of the NCSI gene subfamily encoding norcoclaurine synthase is associated with the biosynthesis of benzylisoquinoline alkaloids in plants. Scientific Reports **6,** 26323.

Vogel BS, Wildung MR, Vogel G, Croteau R. 1996. Abietadiene synthase from grand fir (*Abies grandis*). cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase involved in resin acid biosynthesis. Journal of Biological Chemistry **271,** 23262–23268.

Winzer T, Gazda V, He Z, *et al*. 2012. A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science **336,** 1704–1708.

Winzer T, Kern M, King AJ, *et al*. 2015. Plant science. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. Science **349,** 309–312.

Wurtzel ET, Kutchan TM. 2016. Plant metabolism, the diverse chemistry set of the future. Science **353,** 1232–1236.

Yu K, Liu C, Kim BG, Lee DY. 2015. Synthetic fusion protein design and applications. Biotechnology Advances **33,** 155–164.

Zhang E, Guo X, Meng Z, Wang J, Sun J, Yao X, Xun H. 2015. Construction, expression, and characterization of *Arabidopsis thaliana* 4CL and *Arachis hypogaea* RS fusion gene 4CL::RS in *Escherichia coli*. World Journal of Microbiology amd Biotechnology **31,** 1379–1385.

Zhang Y, Li SZ, Li J, Pan X, Cahoon RE, Jaworski JG, Wang X, Jez JM, Chen F, Yu O. 2006. Using unnatural protein fusions to engineer resveratrol biosynthesis in yeast and mammalian cells. Journal of the American Chemical Society **128,** 13030–13031.

Zhou K, Gao Y, Hoy JA, Mann FM, Honzatko RB, Peters RJ. 2012. Insights into diterpene cyclization from structure of bifunctional abietadiene synthase from Abies grandis. Journal of Biological Chemistry **287,** 6840–6850.

Zi J, Mafu S, Peters RJ. 2014. To gibberellins and beyond! Surveying the evolution of (di)terpenoid metabolism. Annual Review of Plant Biology **65,** 259–286.

Zulak KG, Cornish A, Daskalchuk TE, Deyholos MK, Goodenowe DB, Gordon PM, Klassen D, Pelcher LE, Sensen CW, Facchini PJ. 2007. Gene transcript and metabolite profiling of elicitor-induced opium poppy cell cultures reveals the coordinate regulation of primary and secondary metabolism. Planta **225,** 1085–1106.