# Defense-related phenylpropanoid biosynthetic gene clusters in rice

Guy Polturak and Anne Osbourn

Department of Biochemistry and Metabolism, John Innes Centre, Norwich, United Kingdom

Email: guy.polturak@jic.ac.uk; anne.osbourn@jic.ac.uk

Since colonising land, plants have undergone massive metabolic diversification. This is reflected by the numbers of metabolites that they collectively produce (estimated to be $10^5$ to $10^6$) [1,2]; also, by our increasing awareness of the thousands of uncharacterised genes within plant genomes with predicted roles in specialised metabolism (approximately 20%) [3]. The ~ 200,000 plant metabolites that have been reported thus far from nature [4] are likely to represent just the tip of the iceberg. An enormous amount of metabolic 'dark matter' encoded within the DNA instruction manual of the Plant Kingdom therefore remains to be discovered.

## Why do plants make so many structurally diverse metabolites?

Hartmann commented in his 2007 review [5] that 'In the 1950s, secondary metabolites were regarded as metabolic waste or detoxification products'. In 2006 Firn and Jones [6] proposed the Screening Hypothesis, which was based on the premise that '...potent biomolecular activity is an inherently rare property for any chemical structure to possess...' and that '...organisms have to generate substantial chemical diversity for a few compounds to have any likelihood of possessing biomolecular activity'. Pichersky et al [7] challenged this view, pointing out that 'Basic evolutionary theory posits that selection operates on existing fitness and not on future potential….Given that new alleles arise by random mutations and therefore initially their frequency in the population is low, their frequency would be unlikely to increase and reach fixation in the population if they did not confer a selective advantage and particularly if they cost the plant unnecessary expenditure of energy.' The production of plant specialised metabolites, many of which are structurally elaborate, will inevitably require a considerable investment of energy and thus would be expected to serve a function.

It is now clear that the specialised chemicals that plants produce do indeed make important contributions to the survival of plants in their natural environments, and that the emergence of different types of chemistries in different plant lineages is likely to

reflect adaptation to particular niches and associated selection pressures [1,8]. The ecological functions of plant specialised metabolites include acting as attractants for pollinators and seed dispersal agents, providing protection against herbivores, pests and pathogens, and restricting the growth of neighbouring plants (allelopathy) [1,8]. In some cases, the ability to make the same or a very similar type of compound has arisen independently in taxonomically remote groups of plants (convergent evolution), suggestive of selection for a particular type of bioactive with a conserved function. The occurrence of caffeine, for example, in coffee (family *Rubiaceae*), tea (family *Theaceae*) and other unrelated plants, is a result of such convergent evolution [9].

In other cases, metabolic diversification may provide a rich chemical 'language' for interaction with microbial communities, enabling highly specific communication and recognition. For example, Arabidopsis produces a complex palette of Arabidopsis-specific triterpenes that sculpt the root microbiome by variously attracting or deterring different types of soil microbes, so shaping the root microbiome, presumably to the benefit of the plant [10]. Inversely, the rhizosphere microbiome modulates production and secretion of specialised metabolites from plant roots [11].

**Gene clustering in plant specialised metabolism**

The formation of plant specialised metabolites typically involves a series of enzyme-mediated chemical reactions that together constitute biosynthetic pathways. It follows that such pathways are encoded by multiple genes (often 3-10 or more) [12]. It is well known that the genes for some well characterised plant natural product pathways are dispersed throughout the genome, examples including anthocyanins, carotenoids and glucosinolates [12]. However, in 1997 Frey et al. [13] published a paper reporting a gene cluster for the biosynthesis of the hydroxamic acid 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA) in maize. Since then, >30 biosynthetic gene clusters for diverse classes of plant specialised metabolites, including terpenes, alkaloids, cyanogenic glucosides, fatty acids, polyketides and acyl sugars have been reported from a variety of different plant species, including monocots, eudicots and lower plants [14]. The largest of these is the avenacin cluster, which contains 12 genes and is required the synthesis of antimicrobial defence compounds in oat roots [15]. In many cases, the products of these clustered pathways are known to have beneficial roles in interactions between plants and their environments, for example by providing protection against biotic/abiotic stresses such as pathogens, pests, herbivores,

neighbouring plants, and drought stress, and also in shaping the plant microbiome [14].

With the increasing number of characterised gene clusters that produce terpenes, alkaloids and various other classes of specialised metabolites, one group remains conspicuous by its absence - the phenylpropanoids. This large and structurally diverse class of compounds includes flavonoids, lignins and a range of additional molecules, many of which are involved in plant defence [16]. The genes for some of the best-characterised phenylpropanoid pathways, such as those for flavonoid (e.g. anthocyanin) and lignin biosynthesis, are not organised in clusters in plant genomes, and until recently there have been no reports of biosynthetic gene clusters for this major family of plant natural products. Such lack of characterised phenylpropanoid gene clusters is unexpected, considering the diversity and ubiquity of these compounds in the plant kingdom, and the fact that *in silico* analyses of plant genomes have predicted occurrence of phenylpropanoid-related gene clusters in similar numbers to those for terpenes and alkaloids [17].

**The first phenylpropanoid biosynthetic gene clusters to be discovered in plants**

Stemming from previous work, in which a genome-wide association study was coupled with metabolomics analysis (mGWAS) in rice [18]**,** two papers now report the discovery of such clusters for phenolamides, a widely distributed group of phenylpropanoid derivatives. One of the clusters appears to be rice-specific and is implicated in the formation of hydroxycinnamoyl tyramine (HT) [19], while the other is conserved across other monocots and produces a different type of phenolamide, hydroxycinnamoyl putrescine (HP) [20] (Fig. 1). Both of these clustered pathways contribute to plant defense [19, 20].

HT and other hydroxycinnamic acid amides (HCAAs) are known to have roles in plant growth and development and in biotic/abiotic stress tolerance [21]. The first step in the biosynthesis of HT involves the decarboxylation of L-tyrosine to tyramine, a reaction catalysed by a pyridoxal 5’-phosphate (PLP)-dependent decarboxylase. PLP is supplied by pyridoxamine 5’-phosphate oxidases, which generate PLP from the precursor pyridoxine 5’-phosphate (PNP). Tyramine can then be acylated with one of several different hydroxycinnamoyl groups, to yield HT. This reaction is catalysed by BAHD *N*-acyltransferases, using hydroxycinnamoyl-coenzyme A thioesters as acyl

donors (e.g. cinnamoyl-CoA, *p*-coumaroyl-CoA, feruloyl-CoA, caffeoyl-CoA, sinapoyl-CoA).

Shen et al. [19] reasoned that by examining leaf tyramine levels across 533 previously genotyped rice accessions, they may be able to gain insights into the genetic control of natural variation in tyramine metabolism and hence HCAA biosynthesis. A metabolite-based genome wide association study (mGWAS) revealed a locus with the most significant associated single nucleotide polymorphisms (SNPs) on rice chromosome 10. A predicted decarboxylase gene (*OsTyDC1*) with high sequence similarity to decarboxylases involved in the biosynthesis of tyramine in other plant species was identified as a promising candidate. This gene was located in a candidate biosynthetic gene cluster containing genes for a predicted pyridoxamine 5'-phosphate oxidase gene (*OsPDX3*); also two previously characterised tryptamine/tyramine *N*-acyltransferase genes (*OsTHT1* and *OsTHT2*) implicated in the biosynthesis of aromatic amine conjugates (phenolamides) in rice [22].

Heterologous expression of OsPDX3 and OsTyDC1 in *Escherichia coli* followed by functional analysis, confirmed that the enzymes had activities consistent with a role in tyramine biosynthesis; OsTyDC1 decarboxylates L-tyrosine to tyramine, while OsPDX3 generates PLP, an essential co-factor for the OsTyDC1 enzyme (Fig. 1). Overexpression of OsTyDC1 in rice resulted in elevated tyramine levels, further supporting this conclusion. The occurrence of *OsPDX3* in the HT cluster is of special note, as it provides a first reported example for a plant biosynthetic gene cluster that contains a gene involved in generating an enzyme co-factor, rather than in chemical conversion of a pathway intermediate. Notably, although OsPDX3 provides the co-factor for OsTyDC1, the pairing of *PDX3* and the *THT* acyltransferases in *Poacea* genomes is presumed to have preceded the recruitment of *OsTyDC1* to the cluster [19]. Overexpression of OsTHT1 and OsTHT2 in rice or *Nicotiana benthamiana* resulted in accumulation of elevated levels of HT and hydroxycinnamoyl tryptamine, consistent with previous *in vitro* results [22]. Collectively these data suggest that *OsTyDC1* forms part of a four-gene cluster for the biosynthesis of HT and related compounds in rice. The four genes are co-expressed in the roots and flag leaves, and are also strongly induced by jasmonic acid treatment and pathogen challenge, demonstrating that they do indeed encode a functional pathway, and suggesting a role in defense response to biotic stress.

Tyramine content was not correlated with variation in *OsTyDC1* gene expression, nor with the transcript levels or enzyme activities of *OsTHT1* or *OsTHT2.* Shen and co-authors therefore speculate that *OsTyDC1* and the *OsTHTs* may act in concert to regulate variation in tyramine content, a hypothesis supported by the demonstration that the gene expression ratio of *OsTyDC1* over *OsTHT1* (but not *OsTHT2*) was significantly correlated with tyramine content. Transgenic rice lines accumulating high levels of tyramine have a dwarf phenotype because of effects on cell division [23]. Based on this observation, Shen et al. [19] suggest that tyramine may be a toxic intermediate, and that the balance of expression levels of *OsTyDC1* and *OsTHT1* may mitigate against self-poisoning. It is also possible that the stoichiometry of the protein products of these genes may need to be maintained in order to avoid disruption of a potential metabolon [24, 25]. It has previously been shown for several other plant biosynthetic gene clusters that disruption of cluster function can result in accumulation of toxic intermediates, which may in part explain the need for co-inheritance of full pathway gene sets [12,26,27]. The Shen et al. paper does not, however, allude to any negative consequences of elevated levels of tyramine as a consequence of overexpression of *OsTyDC1.*

*OsTHT1* and *OsTHT2*, a tandem pair of *N*-acyltransferase genes located on rice chromosome 10, were previously implicated by mGWAS and linkage analysis in determining levels of the phenolamide coumaroyl serotonin in the grain [22]. The current study was restricted to analysis of the intermediate tyramine, rather than the phenolamide pathway end-products. Further work is needed to integrate the outputs of these studies and to confirm whether the gene cluster reported by Shen et al. underlies natural variation in phenolamide content.

**The various roles of phenolamides in plant defense**

The HP biosynthetic gene cluster reported by Fang et al [20] consists of three genes encoding an ornithine decarboxylase (*OsODC*) and two tandem duplicate putrescine hydroxycinnamoyl acyltransferases (*OsPHT3* and *OsPHT4*). Both the HT and HP clusters thus notably include a PLP-dependent decarboxylase and a pair of BAHD-family acyltransferases. Both clusters are also induced by pathogen challenge. However, the HT and HP clusters (found on chromosomes 10 and 9, respectively) produce different types of phenolamide products, which seemingly contribute to disease resistance in different ways [19,20]. In the case of the HT cluster,

overexpression or knock-out of the *OsTHT1* and *OsTHT2* genes resulted in altered resistance to the bacterial pathogen *Xanthomonas oryzae* pathovar *oryzae* (*Xoo*) and the fungal pathogen *Magnaporthe oryzae*, consistent with a role for HT in disease resistance in rice. *In vitro* experiments showed that the two representative HTs tested, *p*-coumaroyl-tyramine and feruloyl-tyramine, inhibited *Xoo* growth and *M. oryzae* conidial germination, suggesting a direct role for these compounds in plant defense (i.e. as phytoalexins) [19]. Conversely, over-expression of genes from the HP cluster in rice resulted in cell death, together with enhanced resistance to *M. oryzae* [20]. Interestingly, a transcription factor that negatively regulates cell death (APIP5) binds to the promoter of the *OsPHT4* gene, repressing its transcription. Silencing of *APIP5* thus induces *OsPHT4* expression and HP accumulation [20]. *APIP5* is located on chromosome 6 and does not form part of the HP cluster. Collectively, these results suggest that phenolamides may have various functions in plant protection, from direct inhibition of pathogen growth to defense-related signalling.

**Strategies for cluster and pathway discovery**

The physical clustering of genes for metabolic pathways in plant genomes is likely to have arisen in response to particular selection pressures [12]. It follows, then, that the chemicals produced by such clusters are likely to confer important plant traits. Algorithms to predict candidate biosynthetic gene clusters in plant genomes (e.g. plantiSMASH, PhytoClust) have been developed, inspired by approaches used in microbes [3]. These algorithms rely on pre-defined, curated profile hidden Markov models (pHMMs) to identify genomic loci encoding multiple different families of enzymes associated with natural product biosynthesis. Shen et al. [19] comment that their HT cluster was not predicted by these tools, and advocate the advantages of the mGWAS approach as a strategy to identify new clusters. However, optimisation of search parameters of the genome mining algorithms (for example, by lowering the threshold to include candidate clusters containing genes for only two different enzyme classes, and/or by expanding the gene family types incorporated in the algorithm), would also inevitably enable the two new phenolamide clusters to be detected. In fact, the overall function of the HT gene cluster could already be predicted based on known gene functions, as Shen et al. [19] demonstrate. The power of the genome mining algorithms will continue to increase as more examples of plant biosynthetic gene clusters emerge.

The real strength of the mGWAS approach lies not so much in finding gene clusters, but in identifying genes (biosynthetic or otherwise) that modulate natural variation in metabolite content. The application of untargeted metabolomics in combination with GWAS, in order to map natural variation in the entire metabolome and understand the underlying genetic basis, represents a major challenge. Untargeted studies of this kind will be subject to experimental variation, generate a vast body of information, and be very challenging to interpret. Here, going beyond the state of the art will be necessary to drive functional insights [4]. Many phenotypic screens (metabolite-based or otherwise, e.g. disease resistance) are likely to be highly variable, and so as a general point it will be important to ensure that the read-outs chosen for GWAS analysis are as robust and reproducible as possible, in order to secure the most effective route to elucidating the genetic basis of key traits of interest. Despite these challenges, the Shen et al. and Fang et al. studies successfully demonstrate the utilisation of mGWAS for identification of candidate genes, eventually leading to elucidation of new biosynthetic pathways. They further shed light on the contribution of the phenolamide pathways that they discover to disease resistance in rice.

## Conclusion

In summary, the discovery of the phenolamide clusters in rice represents an important step forward in understanding metabolic diversity in this important food crop, and provides new insights into the roles of small molecules in plant defense. Knowledge of the types of enzymes involved in the biosynthesis of these compounds will aid refinement of genome mining tools, since pHMMs for the relevant enzyme families can be incorporated into the search algorithms. It will be interesting in the future to see whether more examples of phenylpropanoid biosynthetic gene clusters emerge as genetics and genomics-driven investigations continue to uncover the vast hidden iceberg of plant specialised metabolism,

## Conflict of interest

The authors declare that they have no conflict of interest.

## Acknowledgements

G.P. is supported by a Royal Society Kohn International Fellowship (NIF\R1\180677) and a Marie Skłodowska-Curie Individual Fellowship (838242). A.O.'s lab is supported by the Biological Sciences Research Council (BBSRC)-funded Institute Strategic Programme Grant 'Molecules from Nature' (BB/P012523/1) and the John Innes Foundation.

## References

[1] Dixon RA, Strack D. Phytochemistry meets genome analysis, and beyond. Phytochemistry 2003; 62:815-816.

[2] Afendi FM, Okada T, Yamazaki M, et al. KNApSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. Plant and Cell Physiology 2012; 53:e1-e1

[3] Owen C, Patron NJ, Huang A, et al. Harnessing plant metabolic diversity. Curr. Opin. Chem Biol. 2017; 40:24-30.

[4] Fang C, Fernie AR, Luo J. Exploring the diversity of plant metabolism. Trends Plant Sci. 2019; 24:83-98.

[5] Hartmann T. From waste products to ecochemicals: Fifty years research of plant secondary metabolism. Phytochemistry 2007; 68:2831-46.

[6] Firn RD, Jones CG. Do we need a new hypothesis to explain plant VOC emissions? Trends Plant Sci. 2006; 11:112–13.

[7] Pichersky E, Sharkey TD, Gershenzon J. Plant volatiles: a lack of function or a lack of knowledge? Trends Plant Sci. 2006; 11:421.

[8] Weng J-K. The evolutionary paths towards complexity: a metabolic perspective. New Phytol 2014, 201:1141-49.

[9] Pichersky E, Lewinsohn E. Convergent evolution in plant specialized metabolism. Annual review of plant biology 2011; 62:549-566.

[10] Huang AC, Jiang T, Liu Y-X, et al. A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. Science 2019; 364.

[11] Korenblum E, Dong Y, Szymanski J, et al. Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. PNAS 2020; 117:3874-3883.

[12] Nutzmann H-W, Scazzocchio C, Osbourn, A. Metabolic gene clusters in eukaryotes. Ann. Rev. Genet 2018; 52:159-183.

[13] Frey M, Chomet P, Glawischnig E, et al. Analysis of a chemical plant defense mechanism in grasses. Science 1997; 277:696-9.

[14] Polturak G, Osbourn A. The emerging role of biosynthetic gene clusters in plant defense and plant interactions. PLoS Pathogens 2021; 17:e1009698.

[15] Li Y, Leveau A, Zhao Q, et al. Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. Nat. Comm. 2021; 12:1-13.

[16] Dixon RA, Paiva NL. Stress-induced phenylpropanoid metabolism. Plant Cell 1995; 7:1085-97.

[17] Schläpfer P, Zhang P, Wang C, et al. Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. Plant Physiology 2017; 173:2041-2059.

[18] Chen W, Gao Y, Xie X, et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nat. Genet. 2014; 46:714-721.

[19] Shen S, Peng M, Fang H, et al. An *Oryza*-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. Sci. Bull. 2021.

[20] Fang H, Shen S, Wang D, et al. A monocot-specific hydroxycinnamoylputrescine gene cluster contributes to immunity and cell death in rice. Sci. Bull. 2021.

[21] Walters DR. Polyamines and plant disease. Phytochemistry 2003; 64:97-107.

[22] Peng M, Gao Y, Chen W, et al. Evolutionarily distinct BAHD *N*-acyltransferases are responsible for natural variation of aromatic amine conjugates in rice. Plant Cell 2016; 28:1533-50.

[23] Kim YS, Park S, Kang K, et al. Tyramine accumulation in rice cells caused a dwarf phenotype via reduced cell division. Planta 2011; 233:251-60.

[24] Jorgensen K, Rasmussen AV, Morant M, et al. Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Curr. Opin. Plant Biol. 2005; 8:280–91.

[25] Nakayama T, Takahashi S, Waki T. Formation of flavonoid metabolons: Functional significance of protein-protein interactions and impact on flavonoid chemodiversity. Front. Plant Sci. 2019; 10:821.

[26] Chu H-Y, Wegel E, Osbourn A. From hormones to secondary metabolism: The emergence of secondary metabolic gene clusters in plants. Plant J. 2011; 66: 66-79.

[27] Takos AM, Rook F. Why biosynthetic genes for chemical defense compounds cluster. Trends Plant Sci. 2012; 17:383-88.

[28] Kawahara Y, de la Bastide M, Hamilton JP, et al. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice 2013; 6:1-10.

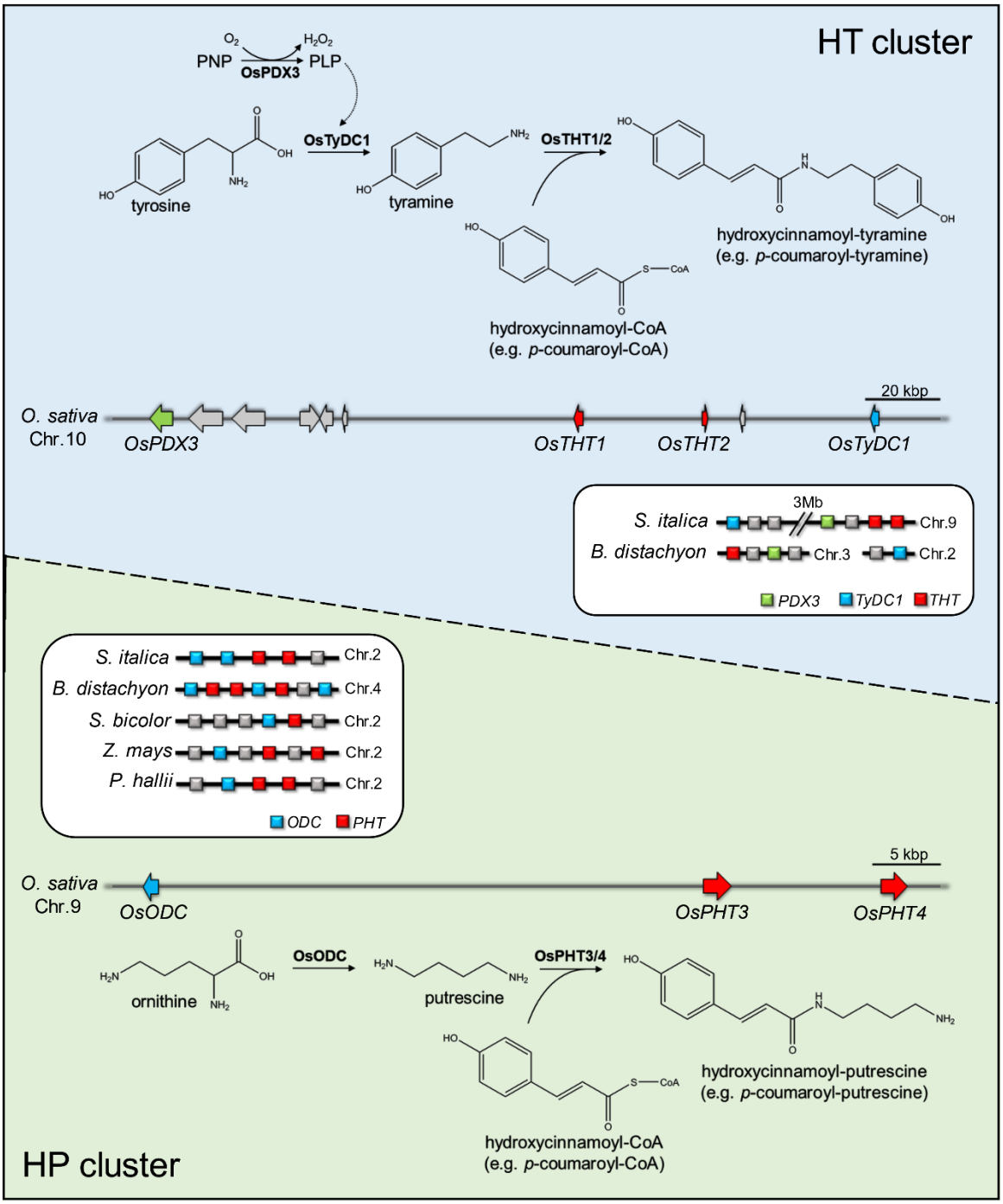
Dr Guy Polturak obtained his Ph.D from the Weizmann Institute of Science, Israel in 2016. He is a postdoctoral research fellow in Professor Anne Osbourn's group at the John Innes Centre, UK. His research interests are plant specialized metabolism and metabolic engineering.

314

Professor Anne Osbourn FRS received her B. Sc. In Botany from the University of Durham in 1982, and her Ph. D in Genetics from the University of Birmingham in 1985. She is a Group Leader at the John Innes Centre and has honorary professorships at the University of East Anglia and Shanghai Jiao Tong University. Her research focuses on plant natural products – biosynthesis, function, mechanisms of metabolic diversification, and metabolic engineering using synthetic biology approaches.

323

324




325

326 **Figure 1. Biosynthesis of hydroxycinnamoyl-tyramine (HT) and**
327 **hydroxycinnamoyl-putrescine (HP) in rice is driven by clustered genes**.
328 Illustrations of the rice HT and HP clusters are based on the *Oryza sativa Japonica*
329 group IRGSP-1.0 annotation [28]. Illustrations of homologous gene localization in
330 related *Poaceae* species are adapted from Shen et al. [19] and Fang et al. [20].
331 Os*TyDC1* homologs do not co-localize with *OsTHT1/2* and *OsPDX3* homologs in the
332 *Setaria italica* and *Brachypodium distachyon* genomes, while co-localization of one or

more copies of *OsODC* and *OsPHT3/4* homologs is conserved in various *Poaceae* genomes, including *Setaria italica*, *Brachypodium distachyon*, *Sorghum bicolor*, *Zea mays* and *Panicum hallii*. PNP, Pyridoxine 5′-phosphate; PLP, pyridoxal 5′-phosphate; TyDC, tyrosine decarboxylase; THT, tryptamine/tyramine hydroxycinnamoyl transferase; PDX, pyridoxamine 5'-phosphate oxidase; ODC, ornithine decarboxylase; PHT, putrescine hydroxycinnamoyl transferase.