Breakthrough Report

Florigen revisited: proteins of the FT/CETS/PEBP/PKIP/YbhB family may be the enzymes of small molecule metabolism

Short title: Florigen-family proteins may be enzymes

Olga Tsoy^{1,4} and Arcady Mushegian^{2,3*}

¹ Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich (TUM), 3, Maximus-von-Imhof-Forum, Freising, 85354, Germany

² Molecular and Cellular Biology Division, National Science Foundation, 2415 Eisenhower Avenue, Alexandria, Virginia 22314, USA

³ Clare Hall College, University of Cambridge, Cambridge CB3 9AL, United Kingdom

⁴ current address: Chair of Computational Systems Biology, University of Hamburg, Notkestrasse, 9, 22607, Hamburg, Germany

* corresponding author: mushegian2@gmail.com

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is Arcady Mushegian (mushegian2@gmail.com).

Abstract

Flowering signals are sensed in plant leaves and transmitted to the shoot apical meristems, where the formation of flowers is initiated. Searches for a diffusible hormone-like signaling entity (“florigen”) went on for many decades, until in the 1990s a product of plant gene *FT* was identified as the key component of florigen, based on genetic evidence and protein localization studies. Sequence homologs of FT protein are found throughout prokaryotes and eukaryotes; some eukaryotic family members appear to bind phospholipids or interact with the components of the signal transduction cascades. We studied molecular features of the FT homologs in prokaryotes and analyzed their genome context, to find tentative evidence connecting the bacterial family members with small molecule metabolism, often involving sugar- or ribonucleoside-containing substrates. Most FT homologs share a constellation of five charged residues, three of which, i.e., two histidines and an aspartic acid, circumfere the rim of a well-defined cavity on the protein surface. We argue that this conserved feature is more likely to be an enzymatic active center than a catalytically inactive ligand-binding site. We propose that most of FT-related proteins are enzymes operating on small diffusible molecules, which may constitute an overlooked essential ingredient of the florigen signal.

Introduction

Flower production in plants occurs in response to the environmental cues – most importantly, changes in the day length. The role of photoperiodicity in all living forms from bacteria to higher eukaryotes is well established, and it has been shown in numerous studies that the perception of the photoperiodic signal in plants occurs primarily in the leaves. Flowers, however, are formed mostly by shoot apical meristems (SAM), which are typically shielded from direct light, and the mechanisms which convey the flowering signal from leaves to SAMs have been a matter of speculation and investigation for most of the 20th century.

Plant physiologists have established, by the 1930s, that angiosperms generally fall in three categories, i.e., long-day plants, in which blooming is turned on by day lengthening / night shortening; short-day plants, which initiate blooming upon day shortening / night lengthening; and day-neutral plants, which bloom in response to the cues other than day length (Garner and Allard, 1920; Knott, 1934). Experiments on flowering in partially shaded plants and in grafts between light-induced and uninduced plants suggested the existence of a flower-promoting diffusible substance; the idea may have been first proposed by Julius Sachs (Sachs, 1865), but was consolidated in the modern form by M. Chailakhyan (1902-1991), who proposed the term “florigen” as the name for the flower-inducing chemical entity (Chailakhyan, 1936). The early research on the photoperiodic signal sensing in leaves is reviewed in (Zeevaart, 2006; Kobayashi and Weigel, 2007), and an account of Chailakhyan's remarkable life and research work can be found in (Romanov, 2012).

The transfer of a flowering signal from leaves to the shoot apex has been studied over the years in many species of flowering plants, and general similarity of its properties to those of small molecules was noted; for example, the florigen fraction appeared to move in phloem at the speed comparable to that of other plant assimilates (King et al., 1968). It has been established also that grafts and extracts of induced plants could transfer flowering ability not only in the uninduced vegetatively developing plants of the same species, but sometimes also in other species or genera, suggesting the evolutionary conservation of the signaling pathways (Zeevaart, 1976).

Despite general acceptance of the idea of florigen as a conserved hormone-like substance, the attempts at the isolation and chemical characterization of the responsible entity did not succeed for several decades. Reviewing the state of the affairs in 1976, J.A.D. Zeevaart listed the

factors that have been tested unsuccessfully for the ability to induce plant flowering; the list of failed candidates included sugars, amino acids, sterols, gibberellins, salicylic acid, ethylene, cytokinins, the photosynthetic capacity, initiation of protein synthesis, and others (Zeevaart, 1976). Several hypotheses have been put forward to explain the difficulties of identifying florigen: perhaps it was not one molecule but several distinct hormones or metabolites, acting jointly in a specific succession, or as a mix with specific ratio of components; or, possibly, the identity of florigen was masked by simultaneous presence of flowering inhibitors in the same samples (Lang et al., 1977); or, maybe, flowering was caused by propagation of electric, rather than chemical, signals between leaves and meristems, such as plasma membrane potential (Penel et al., 1985).

Despite all the work to test these and other hypotheses, there was little progress in biochemical characterization of florigen until early 1990s, when the search for a flower-inducing activity started employing the tools of molecular biology. For example, in what may have been the last published study by Chailakhyan, a radiolabeled protein band of ~27 kDa was observed in the induced, but not in naïve, leaves of *Rudbeckia*, followed by accumulation of the similar-sized band in the shoot apex tissues (Milyaeva et al., 1991); the work is largely unavailable to the Western reader because of the temporary interruption of translation and indexing of Russian-language journals upon the collapse of the Soviet Union. Around the same time, it has been shown (Takeba et al., 1990) that aqueous extracts of *Lemna*, *Pharbitis* and *Brassica* contained a flower-inducing fraction dominated by a 20-30 kDa protein band; the authors noted, however, that the florigen activity was resistant to proteinase K digestion that removed the protein, perhaps suggesting a role for an associated small molecule.

Nearly simultaneously, the results of genetic screens for *A.thaliana* mutants with late flowering phenotypes were published (Koornneef et al., 1991). This was a watershed moment in the studies of the molecular determinants of flowering initiation, and the important discoveries that ensued in the following three decades have been reviewed extensively (Andrés and Coupland, 2012; Lifschitz et al., 2014; Kinoshita and Richter, 2020). The state of the knowledge may be summarized as follows.

Flowering in *Arabidopsis* is LD-dependent, and is enabled through the circadian clock-controlled transcriptional co-regulator *CONSTANS* (*CO*) and its target *FLOWERING LOCUS T*. The protein product of the latter gene has emerged as the integrator of the environmental inputs,

relaying these signals into the gene regulatory network that controls flowering. FT protein is produced in the phloem companion cells of the leaves, enters the phloem sieve elements and is transported from the leaves to the base of shoot apical meristem, where it has been detected experimentally. Considerable evidence exists that the *FT* gene is not expressed in SAM above its basal portion, and that the FT protein may move from cell to cell in plants. This is a key set of properties expected of florigen. A strong genetic and transgenic evidence suggests that FT is required for activating floral meristem identity genes and flowering time controllers, such as MADS-box transcription factors SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), FRUITFULL (FUL/AGL8) and AGL24, and ultimately the master regulator of flower development LEAFY (LFY). Orthologs of FT, such as SINGLE FLOWER TRUSS (SFT) in tomato, HEADING DATE 3a (HD3a) in rice, and their counterparts in many other species (sometimes called CETS proteins in the plant context, after the names of several better-studied representatives from *Antirrhinum*, *Arabidopsis* and *Lycopersicon*) share many of these properties with FT, with some variation in the precise wiring of the transcriptional networks. The paralogs of *FT*, found in varying numbers in all examined angiosperm species, also tend to be involved in flowering control, acting either synergistically or antagonistically with *FT*, though additional or exclusive roles in other aspects of plant development have been shown for a subset of them (Abelenda et al., 2019; Ahn et al., 2006; Conti and Bradley, 2007; Danilevskaya et al., 2011; Jin et al., 2021; Krieger et al., 2010; Pin and Nilsson, 2012; Pnueli et al., 2001; Tsuji et al., 2011).

Despite all effort, FT protein has not been detected in phloem-free tips of SAM in any plant, only in the basal cells of the meristem zones. Protein localization studies at a single-cell resolution in vivo remain challenging, and it is not clear whether FT actually reaches the cells where the meristem identity genes are expressed. The argument that FT does just that – as opposed, for example, to activating an additional low molecular weight messenger – has been made in the literature, but the evidence is indirect, showing for example that some engineered fusions of FT to green fluorescent protein-based reporters are retained in the phloem at the bottom of the SAM zone, and in those cases they do not complementation the *ft* recessive mutants (Corbesier et al., 2007). A study in rice (Tamaki et al., 2007) has made a more direct claim that Hd3a, the protein ortholog of FT in rice, does reach the tip of the SAM to induce flowering. However, their Fig. 3, cited as the key supporting evidence, does not seem to show the protein reporter activity at the tip, where it is supposed to be. Therefore it remains possible that

FT is but one component of the flower induction signal, and that a small molecule, possibly activated by or acting synergistically with FT, is also involved (Lifschitz and Eshed, 2006; Andrés and Coupland, 2012).

Another gene of *Arabidopsis* involved in flowering control, *FD*, has been identified as a recessive suppressor of *FT* – the ability of overexpressed *FT* to induce precocious flowering is impaired in plants with the lowered production of *FD* (Abe et al., 2005). The *FD* gene encodes a bZIP-type transcription factor, which interacts with the FT protein in yeast two-hybrid system, in the bi-molecular fluorescence complementation assays *in vivo*, and apparently in the affinity-purified protein complexes (Abe et al., 2005; Wigge, 2005). In rice, the FT ortholog HD3a does not interact with the bZIP factors directly, but the two proteins co-precipitate in a tripartite complex that also includes a scaffolding 14-3-3 protein (Taoka et al., 2011). These interaction data are the major part of the evidence on the basis of which FT has been assigned the molecular function of a transcriptional co-activator, proposed to form a putative multisubunit complex that binds to the regulatory regions and controls the expression of the flowering identity genes. A corroboration of these protein interaction experiments is provided by the analysis of gene expression and ChIP-Seq data, which reveal that tagged FT binds, apparently mostly in the *FD*-dependent manner, near many of the genes involved in flowering control (Zhu et al., 2021).

Analysis of the amino acid sequence has shown that FT protein belongs to a widely conserved sequence family, members of which are encoded by genomes of many bacteria, archaea and nearly all eukaryotes. The founding member of the family, isolated from bovine brain, is a cytoplasmic protein that can bind *in vitro* to many low molecular weight compounds of different chemical structure, including certain phospholipids (Bernier and Jollès, 1984). The name phosphatidylethanolamine-binding protein (PEBP) became attached to the family, though functional relevance of phosphatidylethanolamine binding has not been demonstrated for any homolog of this protein in any species. The genes or protein products of FT/PEBP family turn up in many genetic screens and binding assays. This has resulted in a variety of putative properties assigned to these proteins, including, in addition to phospholipid binding in plants and animals, also inhibition of carboxypeptidase Y and regulation of Ras GTPase in yeast, inhibition of Raf-1 kinase in mammals (hence an alternative name of the protein family, RKIP), suppression of transepithelial migration of the mammalian host neutrophils by a secreted, plasmid-encoded

PEBP homolog from uropathogenic *E.coli* (Lau et al., 2014), and an uncharacterized role in the modification of polyketide chains in *Streptomyces* (Li et al., 2008, 2009).

Structural studies of crystallized FT/PEBP-like proteins from diverse species of bacteria and eukaryotes have revealed a distinct spatial fold, with a structural core dominated by two beta-sheets and superficially resembling an immunoglobulin-like arrangement (entry 11.1.7 in the ECOD database (Cheng et al., 2015)). A prominent structural feature of this family is an evolutionary conserved cavity on the surface of the molecule. The cavity sometimes accommodates anions included in the crystallization media, including the phosphoryl moiety of the co-crystallized phosphoethanolamine, but is neither hydrophobic nor large enough to bind lipids; on the other hand, several sites suitable for binding hydrophobic ligands have been inferred on the molecule by in silico studies, but those sites tend to be located in the least conserved regions of the molecule (Roderick et al., 2002; Ahn et al., 2006; Shemon et al., 2010; Rudnitskaya et al., 2012; Guo et al., 2018; Nakamura et al., 2019). Curiously, the site of interaction between FT and FD in *A.thaliana* has not been structurally characterized.

In this work, we argue that the patterns of sequence and structure conservation in the family, as well as the genomic context of the FT/PEBP proteins in prokaryotes, are best compatible with the idea that these proteins are enzymes involved in production, attachment or removal of low molecular weight ligands, and that some of such small molecules could play a role in induction of flowering in angiosperms. Thus, rather than consisting only of the FT protein itself, florigen's active moiety may comprise a low molecular weight product of an enzymatic activity of FT.

Results

Sequence conservation in the FT homologs throughout the Tree of Life suggests shared ancestry and common molecular function

We have collected the homologs of plant FT proteins by the PSI-BLAST searches of the NCBI NR database, focusing mostly on completely sequenced genomes. We also consulted the NCBI COG resource that annotates conserved orthologous genes in bacteria and archaea. In the following, unless specified otherwise, we refer to the prokaryotic orthologs from this family as YbhB proteins, after the name of the chromosomal gene in *E.coli*, and use some of the FT/CETS/PEBP/PKIP name aliases for eukaryotic homologs. Complete genomes of the

unicellular and multicellular eukaryotes tend to encode at least one, or commonly more than one, FT homolog; a rare exception are some parasitic eukaryotes with reduced genomes, such as microsporidia, which do not appear to have genes from this family. In bacteria and archaea, the distribution of YbhB orthologs is more complex. YbhB genes (NCBI COG1881) are found in almost all major lineages of bacteria and archaea. Among the clades of archaea, only methanogenic *Euryarchaeota* appear to lack the YbhB homologs, and among bacteria, only the species in phylae *Firmicutes*, *Mollicutes* and in the order *Spirochaetales* tend to be YbhB-free. In most other clades of bacteria and archaea, between 30 and 80% of all species encode YbhB homologs. All told, there are 782 copies of YbhB family proteins found in 594 bacterial and archaeal genomes out of the 1309 genomes in the 2020 release of the COG database. YbhB homologs are also encoded by the genomes of many giant DNA viruses from the *Megaviricetes* class and from some related lineages.

We produced a multiple sequence alignment of the representative proteins from many of these clades and inferred a phylogenetic tree of these sequences. The alignment is shown in Figure 1, and the tree in the Newick format is available as the Supplemental Data Set 1. Many of the branches in the phylogenetic tree are long, and the position of a few clades is not well-resolved. The partitioning of prokaryotic sequences in the tree, however, shows close correspondence to the phylogenetic positions of the species in which they are found; the statistically supported clades generally do not mix bacterial and archaeal representatives, and it is only occasionally that a sequence from a phylogenetically distant bacterium is found within a group of more closely related homologs, as in the case of one sequence from Gram-negative proteobacterium *Helicobacter pylori* nested inside a group of YbhB-like sequences from Gram-positive bacteria. All this indicates that the evolution of the YbhB family in prokaryotes must have included the episodes of rapid evolution and differential gene loss, while the horizontal gene transfer events may have been relatively rare in this family.

The eukaryotic sequences are separated from bacterial and archaeal sequences by a long internal branch in the unrooted tree. Within eukaryotes, nearly all plant FT/CETS homologs segregate as one clade, and fungal/metazoan homologs represent another assembly, occasionally intermingled with the representatives of unicellular eukaryotes; evolution and specialization of FT homologs in different plant lineages has been extensively discussed, e.g., (Klinteräs et al., 2012; Karlgren et al., 2011; Jin et al., 2021). Within fungal and metazoan proteins, a distinct

clade includes FT/PEBP homologs detected recently in an unexpected biological context, namely as the constituents of mitochondrial ribosomal large subunit, where they are known as mitochondrial ribosomal proteins (Mrp) L35/L38. These FT/PEBP/YbhB homologs have distinct N-terminal sequence extensions, found in one gene product within each completely sequenced animal, fungal and protist genome (many of animal and fungal genomes also include additional, extension-free homologs). This domain is not included in the alignment in Figure 1; none of the plant FT homologs appear to contain such a region.

Taken together, these data suggest the ubiquity of the FT/CETS/PEBP/PKIP/YbhB-like proteins throughout the evolution of life and a single origin of all eukaryotic homologs from a prokaryotic ancestor. It is likely also that all or most of those proteins share an ancient conserved molecular function, reflected in the nearly-universal sequence conservation of several key amino acid residues and the common structural scaffold of the entire protein family (Figure 1, and see the next section).

Structural analysis of PEBP/PKIP/FT/YbhB proteins supports the hypothesis of an enzymatic function

A common biochemical function of the entire PEBP/YbhB protein family is further suggested by mapping of the conserved and variable sequence elements of the family onto the available three-dimensional structures (Figures 1 and 2). There are five nearly-invariant polar amino acid residues in the alignment – two aspartic acids located at the C-terminus of strand 3, two histidines found at the N-termini of strands 4 and 6, and an arginine that adjoins the second conserved histidine. Substitutions in these positions in prokaryotes are rare, and multiple replacements are seen mostly in the eukaryotic mitochondrial ribosomal proteins L35/L38, which is the clade that is most likely to have acquired a new molecular function during evolution. There are also several highly conserved residues, in particular prolines, in the loops around these signature amino acids (Figure 1). The five most conserved residues are also found close to each other in the spatial structures of the proteins (Figure 2).

Both histidines, the first of the two aspartates, and the proline-rich loop before the strand 6 form the rim of a hydrophilic hole, which is the largest cavity conserved in all structurally characterized proteins in the family. The second of the conserved aspartates forms forms three

hydrogen bonds with the side chain of the conserved arginine, constituting a conserved “shoulder” on the surface of the molecule (Figure 2).

The polar cavity on the surface of all PEBP/PKIP/FT/YbhB protein molecules has been interpreted as a site that binds a ligand containing a polar group, such as phosphate; many of the described or suspected interactions of eukaryotic PEBP/PKIP/FT with phospholipids and phosphoproteins have been rationalized in this way. Recently, a report that FT from *Arabidopsis* may bind phosphatidylcholine (PC) more strongly and more specifically than other lipids has prompted the analysis of high-resolution structures of FT crystallized in the presence of PC. Unexpectedly, in the crystals obtained under various conditions no electron density for PC could be detected, and computational docking suggested four potential binding sites for PC, all located far from the conserved polar cavity (Nakamura et al., 2014a, 2014b). It is also notable that the electrostatic calculations suggest that the overall charge of the cavity, at least in vacuum, is strongly negative, making it perhaps not so conducive for direct interactions with phosphate, whereas the charge on the “shoulder” may be more positive (Figure 2B). On the whole, the attempts to relate the known biological properties of the FT homologs to their sequence and structure still do not provide a satisfactory explanation of the role of the most prominent, conserved elements of the family.

We suggest that these signature elements of sequence and structure are indicative of an enzymatic function of the FT/YbhB homologs. The search of Mechanism and Catalytic Site Atlas (M-CSA) resource, which contains annotated information about the amino acid determinants of catalysis in conserved enzyme families, identifies 21 families of enzymes that have two His, two Asp and one Arg in their active centers. These families represent all 7 of the top-level Enzyme Classification classes of activities, i.e., oxidoreductases, transferases, hydrolases, lyases, isomerases, ligases and translocases. If the selection criteria are relaxed to only two histidines and one aspartic acid surrounding the hydrophilic hole, the search retrieves 105 families, corresponding to almost 11% of the 964 entries in the database. The particular three-dimensional configuration of these residues in PEBP/PKIP/FT/YbhB proteins appears to be quite unique, as judged by the analysis with the PINTS program, which compares similar spatial arrangements of key residues in non-homologous proteins (Stark and Russell, 2003). Nonetheless, it is clear that a combination of two histidines, two aspartates and a lysine has been

repeatedly utilized in evolution to build enzymatic active centers, enabling catalytic conversions of different kinds.

These observations are in sharp contrast to what is known about sequence conservation in non-catalytic ligand-binding protein domains. A collection of such domains was extracted from the PFAM database, using keywords “ligand+bind” and “phosphate-binding”, resulting in 1044 conserved domains. Removal of the clearly annotated enzymatic domains that bind phosphate in their active centers, such as the protein kinases or ATPases, and selection of a non-redundant set of sequence families, produces 651 putative non-catalytic ligand-binding domains. Examination of the sequence profiles of these families suggests that the majority of the conserved positions in these domains tend to be occupied by non-polar residues, although conserved charged residues, such as lysine, histidine, arginine, glutamate or aspartate, are also found occasionally. We recorded the identity of all charged residues that were characterized by the information content of more than 50% of the maximum possible value for their position, and found only 25 domains that had more than two conserved charged residues; none of those domains simultaneously had two conserved histidines and a conserved aspartate (Supplemental File 2). Thus, the known conserved ligand-binding domains do not appear to use the combination of two histidines and an aspartic acid residue for their non-catalytic interactions with small molecules, suggesting again that the conserved feature in PEBP/FT/YbhB proteins is more likely to be a part of the catalytic center than to serve solely as a binding interface.

Genomic context of FT homologs in bacteria and yeast suggests connections with small molecule metabolism

We took advantage of the broad taxonomic distribution of the YbhB homologs and tried to infer the putative functional linkages for these gene products on the basis of their genome context (Table 1). One kind of a linkage between proteins may be revealed when two protein domains exist as separate open reading frames in some species, but are fused into multidomain proteins in others; such translational fusions, especially those that are evolutionary conserved, are enriched in proteins that work in concert with one another (Huynen et al., 2000; Suhre and Claverie, 2004). The YbhB homologs in bacteria and archaea are most commonly encoded as stand-alone open reading frames and form domain fusions infrequently. In actinobacteria, however, they are found in the C-terminal portions of longer proteins, which also contain the

modules implicated in carbohydrate metabolism. For example, protein WP_014179790.1 in *Streptomyces* sp. consists of the N-terminal pectate lyase-like carbohydrate-binding module (domain ID in NCBI CDD database cd04082), followed by the FN3 repeat region (cd00063), a putative glucose/sorbose dehydrogenase (GSDH) region with predicted beta-propeller structure (cd268170), and finally the C-terminal YbhB homology domain. This theme is partially preserved, albeit with domain rearrangement, in some species of evolutionarily distant proteobacteria, where the order of domains is GSDH-YbhB, or occasionally GSDH-FN3-YbhB; examples include WP_014747134.1 in an alphaproteobacterium *Tistrella* and proteins in gammaproteobacteria, such as WP_096298086.1 in *Luteimonas*, PYD93448.1 in *Pseudomonas syringae* pv. *pisi*, or WP_145513070.1 in *Xantomonas perforans*. GSDH enzymes utilize quinone cofactors to convert hexoses and their derivatives into a variety of products in bacteria (Miyazaki et al., 2006).

We then surveyed the databases of conserved genomic neighborhoods and scanned the literature for the evidence about the conservation of FT/YbhB chromosomal neighbors in different genomes. As with the protein fusions, the FT/YbhB family genes are not commonly found within predicted operons or have conserved positional associations with other ORFs. In two species of *Streptomyces*, however, YbhB homologs are located within the large (20 genes) operons responsible for biosynthesis of polyketide-based small molecules, i.e., tautomycin in *S. spiroverticillatus* and tautomycetin in *S. griseochromogenes*. These polyketides are of pharmacological interest, because they are potent inhibitors of mammalian protein phosphatases. The YbhB family genes, called, respectively, *ttnL* and *ttmL*, together with the adjacent seven open reading frames, are required to manufacture a rare dialkylmaleic anhydride moiety of tautomycin and tautomycetin and to conjugate it to the polyketide backbone, produced by the remaining genes in the same operon. Dialkylmaleic anhydride is required for the activity of the mature product, and is made *de novo* from propionate and an unidentified C5 compound, through the succession of the steps that are incompletely understood (Li et al., 2008, 2009).

Another approach of connecting genes into functionally related groups is to analyze their phyletic vectors, i.e., the representations of the presences and absences of their orthologous genes in different genomes (Glazko and Mushegian, 2004; Glazko et al., 2006). We have searched the space of phyletic vectors associated with the 2014 release of NCBI COGs, using the vector of YbhB (COG1881) as the query. After retaining the matching vectors with comparable

cardinality (i.e., the genes present in 300-400 species, to avoid spurious matching to almost-ubiquitous proteins), we found two COGs that were most often co-inherited by the same genomes as COG1881. One of those, COG1042, is annotated as Acyl-CoA synthetase (NDP forming); it is found in 358 species, 220 of which encode both COG1881 and COG1042. The enzyme, best studied in hyperthermophilic archaea and protists, is involved in the substrate-level phosphorylation, by the equation $\text{acetyl-CoA} + \text{ADP} + \text{Pi} \rightleftharpoons \text{acetate} + \text{ATP} + \text{CoA}$ (Musfeldt and Schönheit, 2002). The roles of the bacterial homologs is less clear, as some of them appear to be catalytically inactive and possibly play auxiliary roles in the acylation and deacylation of proteins (Thao and Escalante-Semerena, 2012). The other gene with matching phyletic pattern, COG2835, annotated as “uncharacterized conserved protein YbaR, Trm112 family”, is found in 390 genomes, 233 of which also encode COG1881. YbaR/Trm112 family encodes activators of several methyltransferases involved in modification of rRNA, tRNA and peptide release factors (van Tran et al., 2018)

One more way to predict functional linkages between gene products is to examine their putative regulatory regions, which in bacteria tend to be located in the intergenic regions, usually to the 5' ends of the open reading frames or operons they regulate. We performed such analysis of YbhB homologs in several bacterial taxa and detected a conserved site upstream of the YbhB open reading frame in the family *Enterobacteriaceae*. Unlike most known binding sites for the specialized transcription factors, this site lacks palindromic structure and has a consensus sequence TACACTT. Scanning 41 *Enterobacteriaceae* species from 14 genera with a probabilistic model of the site identifies additional intergenic regions where the variants of these sites occur. (Table 1, and Supplemental File 3). Three genes that were found to contain a highly similar conserved upstream site in 16-17 species, representing 8 genera are *rlmA* (23S rRNA m(1)G745 methyltransferase), *yobB* (putative carbon-nitrogen hydrolase family protein), and *adrB* (c-di-GMP phosphodiesterase). Sites matching this consensus in *E.coli* have been noticed before and predicted to represent a modified form of the canonical -10 element sequence TATAATT (Huerta and Collado-Vides, 2003); it is located in an appropriate position relative to the ORFs starts, but the regions in which we found this upstream element do not appear to have the recognizable -35 sequence nearby. The significance of the shared TACACTT element thus remains unclear, though a common regulation mechanism for genes that share this site is a possibility.

Finally, we have consulted several curated databases of pre-computed functional linkages. PEBP/FT/YbhB homologs are not well-represented among the modules reported by those databases, but two observations were of interest. The FunCoup database, which uses naïve Bayesian approach to integrate the information from 10 different genomic and proteomic measurement spaces (Persson et al., 2021), shows a small network of interacting proteins in *E.coli*, which includes two chaperones, i.e., a heat shock 70-family DnaK and a protease Lon, as well as YbhB and YbcL. Interactions with chaperones, which are frequently observed in the protein interaction data, may reflect cellular proteostasis needs and a broad clientele of many chaperone systems, but are often not particularly informative about the protein function. The connection to the other protein in the group, YbcL, may be more revealing: that protein, recently renamed RlhA, appears to be a component of the 5-hydroxycytidine synthase enzymatic complex involved in modification of 23S rRNA (Kimura et al., 2017). And the analysis of phyletic patterns across many eukaryotes, collected in PhyloGene database (Sadreyev et al., 2015) has revealed one highly-scoring match co-inherited with human PEBP1 gene, i.e., NUDT3, encoding an enzyme from the Nudix class. Nudix proteins are hydrolases noted for the affinity to the pyrophosphate moieties in their substrates, often lipid, nucleoside or oligonucleotide derivatives (McLennan, 2006).

Discussion

“Circumstantial evidence is a very tricky thing,” answered Holmes thoughtfully. “It may seem to point very straight to one thing, but if you shift your own point of view a little, you may find it pointing in an equally uncompromising manner to something entirely different.”

Arthur Conan Doyle. *The Boscombe Valley Mystery*.

We present the computational evidence indicating that the molecular function of the FT proteins in flowering, as well as the biochemical roles of the FT/CETS/PEBP/PKIP/YbhB homologs, could have been misunderstood: the proteins from this family might be the enzymes involved in the biochemical transformation of small molecules, either instead of, or in addition to, their postulated role of being stoichiometric subunits within plant transcription complexes.

Much of the analysis reported here comes from the examination of prokaryotic genome sequences. Obviously, studies in archaea and bacteria cannot be expected to validate the role of FT in the transcriptional control of flowering in plants, as prokaryotes lack most of the plant signal transduction systems and downstream effectors of flowering. Instead, we asked two different questions, i.e., “What can be deduced about the molecular functions of the FT orthologs in prokaryotes?” and “Is there evidence that these functions may be conserved in eukaryotes, including plants?”

The first of these questions can be tentatively answered by taking into account the genomic context of FT/YbhB homologs hinting at their possible functional connections to enzymes involved in specific metabolic pathways. Though no single gene could be linked to YbhB by several orthogonal approaches, a trend emerges when these data are considered jointly (Table 1). The evidence appears to point towards functional linkages of YbhB to sugar and ribonucleotide/RNA modifications, implying that FT/YbhB may be involved in the metabolism of a monosaccharide such as ribose or another pentose, or perhaps their nucleoside-like derivative. Relatedly, FT proteins could be involved in phospholipid metabolism that proceeds through lipid conjugates with nucleotides; indeed, recent genetic evidence has suggested the involvement of phosphorylethanolamine cytidylyltransferase (*PECT1* gene product) in flowering in Arabidopsis plants (Susila et al., 2021). Finally, FT proteins could have a role in attaching or removing some low molecular weight moieties from larger molecules, such as proteins or nucleic acids; this could explain the association of FT and its homologs with nuclear proteins seen in plants and other eukaryotes.

The answer to the other question posed above, i.e., whether the putative enzymatic function could have been preserved through the evolution of the FT/CETS/PEBP/PKIP/YbhB family, appears to be clearly in the affirmative; the pattern of sequence conservation and spatial juxtaposition of the key charged residues in the family at a long phylogenetic span is striking (Figures 1 and 2), even though this constellation of conserved residues has not been definitely mapped to a specific biological function thus far. Bioinformatic analysis suggests that the conserved sets of two histidines and an aspartic acid clustered in space are frequent in the enzyme active centers but virtually never are found in the non-catalytic ligand-binding domains. Finally, broad occurrence of the protein family in bacteria, archaea, giant viruses and nearly all eukaryotes would not be unusual for an enzyme, especially one involved in metabolism or

signaling, but would be a much rarer occurrence for transcriptional regulators, as they are typically not shared between bacteria and eukaryotes (Babu et al., 2004).

There is an ample precedent of utilization of sugars, nucleotides and products of RNA breakdown for plant hormone biosynthesis; one class of plant hormones, cytokinins, are purine derivatives that can be produced either by isoprenylation of adenosine phosphate or by tRNA degradation (Kakimoto, 2003), whereas another class, gibberellins, are synthesized by transforming the pentose skeleton generated in the 1-deoxy-D-xylulose 5-phosphate pathway (Salazar-Cerezo et al., 2018). Recent studies have significantly expanded the repertoire of linear and cyclic oligonucleotides that serve as essential signaling messengers in bacteria and animals (Burroughs et al., 2015; Burroughs and Aravind, 2020), again suggesting that some nucleoside derivatives with regulatory properties may remain undiscovered in plants.

A critic might point out that the bioinformatic evidence of the enzymatic function of FT and their homologs presented in this manuscript, let alone of the proposed involvement of FT in specific biochemical pathways, is too circumstantial – some would say, speculative – and thus far has not been corroborated by the wet-laboratory experiments. However, exactly the same can be said about the experimental support for the role of FT as a transcriptional co-activator – a hypothesis that has not been corroborated by the computational evidence and is quite unmoored from the facts of comparative genomics. In the spirit of considering the entire corpus of the available data, and using reciprocal illumination from different classes of evidence to improve the precision of scientific hypotheses (Iyer et al., 2001; Baker et al., 2014; Carr et al., 2015), it seems timely and urgent to start testing the hypothesis of a metabolic role of the FT homologs in various organisms. Comparison of metabolic profiles in the wild-type organisms and in FT knock-down mutants could be a first step in determining the identity of the substrates and products of the putative enzymatic activity that we predict to be conserved in most members of this protein family.

Methods

Sequence homologs were collected using the PSI-BLAST program (Altschul et al., 1997) with standard settings, run either against the NCBI NR protein sequence database or the databases restricted by taxon (e.g., viruses). Nucleotide and amino acid sequences were aligned

using the program MUSCLE v. 3.8.31 (Edgar, 2004). Distribution of the YbhB homologs in bacterial and archaeal genomes was studied using the web interface for the 2020 release of NCBI COG database (Galperin et al., 2021). The phylogenetic trees were constructed by PhyML program implemented in Phylogeny.fr server (Dereeper et al., 2008), with 1000 bootstrap replicates and approximate Likelihood Ratio Test. The three-dimensional structures of proteins were visualized and approximate, charge-smoothed electrostatic surface representations were generated using the open-source PyMOL environment (Schrödinger LLC; SciCrunch RRID SCR_000305), installed from source using homebrew on the MacOS High Sierra 10.13.6. Gene adjacency on the chromosomes and protein domain fusions were analyzed by visual inspection of the sequence records in GenBank. Phyletic vectors were analyzed using the psi-square program with default settings and vector similarity measured using Pearson correlation-based distance (Glazko et al., 2005); the NCBI COG database release of 2014 was used as the search space, because the digitized phyletic vector information was not available for the 2020 release at the time of writing. For promoter analysis, the complete bacterial genomes were downloaded from Ensembl Bacteria database (release 47) (Howe et al., 2020) in March 2020; if several strains were available for a species, one random strain has been chosen. The regulatory regions of homologous genes were aligned with MUSCLE and the most conservative regions were chosen to build a positional weight matrix (PWM) by the SignaX routine of the Genome Explorer program (Mironov et al., 2000). The genome scan was performed by Genome Explorer with the threshold set equal to the lowest score in the training set, i.e., in the set of *ybhB* homologous regions. The usage of specific amino acid residues by the enzymatic active centers was performed with the aid of Mechanism and Catalytic Site Atlas (M-CSA) online resource, using the Web interface to select the residues of interest (Ribeiro et al., 2018). The usage of specific amino acid residues by ligand-binding domains was studied by collecting all domains in PFAM 33.1 (El-Gebali et al., 2019) with keywords “ligand bind” or “phosphate-binding”. The known enzymatic domains were manually excluded based on their description, and for each remaining domain, the curated seed alignment has been downloaded from PFAM and used as an input for the Skyline server (Wheeler et al., 2014) to build HMM profiles and to get the information count for the residues of interest.

Supplemental Data

Supplemental Data Set 1. Newick-formatted phylogenetic tree of FT/YbhB homologs in bacteria, archaea and eukaryotes.

Supplemental Data Set 2. List of ligand-binding domains in PFAM and their usage of conserved amino acid residues

Supplemental Data Set 3. Occurrences of the motif with consensus sequence TACACTT in *Enterobacteriaceae*.

Acknowledgements and Authors' Contributions

OT is supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grant 01ZX1908A). AM is a program director at the National Science Foundation, an agency of the U.S. Government; his work was supported by the National Science Foundation's Independent Research/Development and Long-Term Professional Development programs, but the statements and opinions expressed herein are made in the personal capacity and do not constitute the endorsement by NSF or the government of the United States.

AM designed the research; OT and AM performed research; OT and AM wrote the paper.

543 References

- Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* 309, 1052–1056.
- Abelenda, J.A., Bergonzi, S., Oortwijn, M., Sonnewald, S., Du, M., Visser, R.G.F., Sonnewald, U., and Bachem, C.W.B. (2019). Source-Sink Regulation Is Mediated by Interaction of an FT Homolog with a SWEET Protein in Potato. *Current Biology* 29, 1178–1186.e6.
- Ahn, J.H., Miller, D., Winter, V.J., Banfield, M.J., Lee, J.H., Yoo, S.Y., Henz, S.R., Brady, R.L., and Weigel, D. (2006). A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *EMBO J* 25, 605–614.
- Altschul, S.F., Madden, T.I., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402.
- Andrés, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. *Nat Rev Genet* 13, 627–639.
- Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., and Teichmann, S.A. (2004). Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* 14, 283–291.
- Baker, P.A., Fritz, S.C., Dick, C.W., Eckert, A.J., Horton, B.K., Manzoni, S., Ribas, C.C., Garzone, C.N., and Battisti, D.S. (2014). The emerging field of geogenomics: Constraining geological problems with genetic data. *Earth-Science Reviews* 135, 38–47.
- Bernier, I., and Jollès, P. (1984). Purification and characterization of a basic 23 kDa cytosolic protein from bovine brain. *Biochim. Biophys. Acta* 790, 174–181.
- Burroughs, A.M., and Aravind, L. (2020). Identification of Uncharacterized Components of Prokaryotic Immune Systems and Their Diverse Eukaryotic Reformulations. *J Bacteriol* 202.
- Burroughs, A.M., Zhang, D., Schäffer, D.E., Iyer, L.M., and Aravind, L. (2015). Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res* 43, 10633–10654.
- Carr, S.M., Marshall, H.D., Wareham, T., and Craig, D. (2015). The Big ORF Theory. In *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology*, (Elsevier), pp. 265–274.

- Chailakhyan, M. (1936). New facts supporting the hormonal theory of plant development. *Dokl. Akad. Nauk SSSR* 4, 77–81.
- Cheng, H., Liao, Y., Schaeffer, R.D., and Grishin, N.V. (2015). Manual classification strategies in the ECOD database: ECOD Manual Classification Strategies. *Proteins* 83, 1238–1251.
- Conti, L., and Bradley, D. (2007). TERMINAL FLOWER1 Is a Mobile Signal Controlling *Arabidopsis* Architecture. *Plant Cell* 19, 767–778.
- Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., et al. (2007). FT Protein Movement Contributes to Long-Distance Signaling in Floral Induction of *Arabidopsis*. *Science* 316, 1030–1033.
- Danilevskaya, O.N., Meng, X., McGonigle, B., and Muszynski, M.G. (2011). Beyond flowering time: pleiotropic function of the maize flowering hormone florigen. *Plant Signal Behav* 6, 1267–1270.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36, W465–469.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792–1797.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res* 47, D427–D432.
- Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Vera Alvarez, R., Landsman, D., and Koonin, E.V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* 49, D274–D281.
- Garner, W.W., and Allard, H.A. (1920). Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *Monthly Weather Review* 48, 415–415.
- Glazko, G.V., and Mushegian, A.R. (2004). Detection of evolutionarily stable fragments of cellular pathways by hierarchical clustering of phyletic patterns. *Genome Biol.* 5, R32.
- Glazko, G., Gordon, A., and Mushegian, A. (2005). The choice of optimal distance measure in genome-wide datasets. *Bioinformatics* 21 Suppl 3, iii3–11.

Glazko, G., Coleman, M., and Mushegian, A. (2006). Similarity searches in genome-wide numerical data sets. *Biol. Direct* 1, 13.

Guo, C., Chang, T., Sun, T., Wu, Z., Dai, Y., Yao, H., and Lin, D. (2018). Anti-leprosy drug Clofazimine binds to human Raf1 kinase inhibitory protein and enhances ERK phosphorylation. *Acta Biochim. Biophys. Sin. (Shanghai)* 50, 1062–1067.

Howe, K.L., Contreras-Moreira, B., De Silva, N., Maslen, G., Akanni, W., Allen, J., Alvarez-Jarreta, J., Barba, M., Bolser, D.M., Cambell, L., et al. (2020). Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Res* 48, D689–D695.

Huerta, A.M., and Collado-Vides, J. (2003). Sigma70 Promoters in *Escherichia coli*: Specific Transcription in Dense Regions of Overlapping Promoter-like Signals. *Journal of Molecular Biology* 333, 261–278.

Huynen, M., Snel, B., Lathe, W., and Bork, P. (2000). Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res* 10, 1204–1210.

Iyer, L.M., Aravind, L., Bork, P., Hofmann, K., Mushegian, A.R., Zhulin, I.B., and Koonin, E.V. (2001). Quod erat demonstrandum? The mystery of experimental validation of apparently erroneous computational analyses of protein sequences. *Genome Biol.* 2, RESEARCH0051.

Jin, S., Nasim, Z., Susila, H., and Ahn, J.H. (2021). Evolution and functional diversification of FLOWERING LOCUS T/TERMINAL FLOWER 1 family genes in plants. *Semin Cell Dev Biol* 109, 20–30.

Kakimoto, T. (2003). Biosynthesis of cytokinins. *J Plant Res* 116, 233–239.

Karlgren, A., Gyllenstrand, N., Källman, T., Sundström, J.F., Moore, D., Lascoux, M., and Lagercrantz, U. (2011). Evolution of the PEBP Gene Family in Plants: Functional Diversification in Seed Plant Evolution. *Plant Physiol.* 156, 1967–1977.

Kimura, S., Sakai, Y., Ishiguro, K., and Suzuki, T. (2017). Biogenesis and iron-dependency of ribosomal RNA hydroxylation. *Nucleic Acids Research* 45, 12974–12986.

King, R.W., Evans, L.T., and Wardlaw, I.F. (1968). Translocation of the floral stimulus in *Pharbitis nil* in relation to that of assimilates. *Zeitschrift Fur Pflanzenphysiologie* 59, S377-388.

Kinoshita, A., and Richter, R. (2020). Genetic and molecular basis of floral induction in *Arabidopsis thaliana*. *Journal of Experimental Botany* 71, 2490–2504.

Klintenäs, M., Pin, P.A., Benlloch, R., Ingvarsson, P.K., and Nilsson, O. (2012). Analysis of conifer FLOWERING LOCUS T/TERMINAL FLOWER1-like genes provides evidence for dramatic biochemical evolution in the angiosperm FT lineage. *New Phytol.* 196, 1260–1273.

Knott, J.E. (1934). Effect of a localized photoperiod on spinach. In *Proc Amer Soc Hortic Sci*, pp. 152–154.

Kobayashi, Y., and Weigel, D. (2007). Move on up, it's time for change--mobile signals controlling photoperiod-dependent flowering. *Genes Dev.* 21, 2371–2384.

Koornneef, M., Hanhart, C.J., and van der Veen, J.H. (1991). A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Molec. Gen. Genet.* 229, 57–66.

Krieger, U., Lippman, Z.B., and Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat. Genet.* 42, 459–463.

Lang, A., Chailakhyan, M.K., and Frolova, I.A. (1977). Promotion and inhibition of flower formation in a dayneutral plant in grafts with a short-day plant and a long-day plant. *Proc. Natl. Acad. Sci. U.S.A.* 74, 2412–2416.

Lau, M.E., Danka, E.S., Tiemann, K.M., and Hunstad, D.A. (2014). Bacterial lysis liberates the neutrophil migration suppressor YbcL from the periplasm of uropathogenic *Escherichia coli*. *Infect. Immun.* 82, 4921–4930.

Li, W., Ju, J., Rajski, S.R., Osada, H., and Shen, B. (2008). Characterization of the tautomycin biosynthetic gene cluster from *Streptomyces spiroverticillatus* unveiling new insights into dialkylmaleic anhydride and polyketide biosynthesis. *J. Biol. Chem.* 283, 28607–28617.

Li, W., Luo, Y., Ju, J., Rajski, S.R., Osada, H., and Shen, B. (2009). Characterization of the tautomycin biosynthetic gene cluster from *Streptomyces griseochromogenes* provides new insight into dialkylmaleic anhydride biosynthesis. *J. Nat. Prod.* 72, 450–459.

Lifschitz, E., and Eshed, Y. (2006). Universal florigenic signals triggered by FT homologues regulate growth and flowering cycles in perennial day-neutral tomato. *J. Exp. Bot.* 57, 3405–3414.

Lifschitz, E., Ayre, B.G., and Eshed, Y. (2014). Florigen and anti-florigen “a systemic mechanism for coordinating growth and termination in flowering plants. *Front. Plant Sci.* 5.

McLennan, A.G. (2006). The Nudix hydrolase superfamily. *Cell Mol Life Sci* 63, 123–143.

- Milyaeva, E.L., Golyanovskaya, S.A., Aksenova, N.P., and Chailakhyan, M.Kh. (1991). Changes of Protein Composition in Leaves and Stem Apices of Bicolored Coneflower with Transition to Flowering. *Dokl. Akad. Nauk SSSR (Bot. Sciences)* 11–15.
- Mironov, A.A., Vinokurova, N.P., and Gelfand, M.S. (2000). Software for analysis of bacterial genomes. *Mol Biol* 34, 222–231.
- Miyazaki, T., Sugisawa, T., and Hoshino, T. (2006). Pyrroloquinoline Quinone-Dependent Dehydrogenases from *Ketogulonicigenium vulgare* Catalyze the Direct Conversion of l-Sorbose to l-Ascorbic Acid. *AEM* 72, 1487–1495.
- Musfeldt, M., and Schönheit, P. (2002). Novel Type of ADP-Forming Acetyl Coenzyme A Synthetase in Hyperthermophilic Archaea: Heterologous Expression and Characterization of Isoenzymes from the Sulfate Reducer *Archaeoglobus fulgidus* and the Methanogen *Methanococcus jannaschii*. *JB* 184, 636–644.
- Nakamura, Y., Andrés, F., Kanehara, K., Liu, Y., Dörmann, P., and Coupland, G. (2014a). *Arabidopsis* florigen FT binds to diurnally oscillating phospholipids that accelerate flowering. *Nat Commun* 5, 3553.
- Nakamura, Y., Andrés, F., Kanehara, K., Liu, Y., Coupland, G., and Dörmann, P. (2014b). Diurnal and circadian expression profiles of glycerolipid biosynthetic genes in *Arabidopsis*. *Plant Signal Behav* 9, e29715.
- Nakamura, Y., Lin, Y.-C., Watanabe, S., Liu, Y.-C., Katsuyama, K., Kanehara, K., and Inaba, K. (2019). High-Resolution Crystal Structure of *Arabidopsis* FLOWERING LOCUS T Illuminates Its Phospholipid-Binding Site in Flowering. *IScience* 21, 577–586.
- Penel, C., Gaspae, Th., and Greppin, H. (1985). Rapid interorgan communications in higher plants with special reference to flowering. *Biol Plant* 27, 334–338.
- Persson, E., Castresana-Aguirre, M., Buzzao, D., Guala, D., and Sonnhammer, E.L.L. (2021). FunCoup 5: Functional Association Networks in All Domains of Life, Supporting Directed Links and Tissue-Specificity. *J Mol Biol* 166835.
- Pin, P.A., and Nilsson, O. (2012). The multifaceted roles of FLOWERING LOCUS T in plant development: FT, a multifunctional protein. *Plant, Cell & Environment* 35, 1742–1755.
- Pnueli, L., Gutfinger, T., Hareven, D., Ben-Naim, O., Ron, N., Adir, N., and Lifschitz, E. (2001). Tomato SP-Interacting Proteins Define a Conserved Signaling System That Regulates Shoot Architecture and Flowering. *Plant Cell* 13, 2687–2702.

Ribeiro, A.J.M., Holliday, G.L., Furnham, N., Tyzack, J.D., Ferris, K., and Thornton, J.M. (2018). Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res* 46, D618–D623.

Roderick, S.L., Chan, W.W., Agate, D.S., Olsen, L.R., Vetting, M.W., Rajashankar, K.R., and Cohen, D.E. (2002). Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nat. Struct. Biol.* 9, 507–511.

Romanov, G.A. (2012). Mikhail Khristoforovich Chailakhyan: The fate of the scientist under the sign of florigen. *Russ J Plant Physiol* 59, 443–450.

Rudnitskaya, A.N., Eddy, N.A., Fenteany, G., and Gascón, J.A. (2012). Recognition and reactivity in the binding between Raf kinase inhibitor protein and its small-molecule inhibitor locostatin. *J Phys Chem B* 116, 10176–10181.

Sachs, J. (1865). Wirkung des Lichtes auf die Blütenbildung unter Vermittlung der Laubblätter. *Bot. Ztg* 23, 117.

Sadreyev, I.R., Ji, F., Cohen, E., Ruvkun, G., and Tabach, Y. (2015). PhyloGene server for identification and visualization of co-evolving proteins using normalized phylogenetic profiles. *Nucleic Acids Res* 43, W154-159.

Salazar-Cerezo, S., Martínez-Montiel, N., García-Sánchez, J., Pérez-Y-Terrón, R., and Martínez-Contreras, R.D. (2018). Gibberellin biosynthesis and metabolism: A convergent route for plants, fungi and bacteria. *Microbiol Res* 208, 85–98.

Shemon, A.N., Heil, G.L., Granovsky, A.E., Clark, M.M., McElheny, D., Chimon, A., Rosner, M.R., and Koide, S. (2010). Characterization of the Raf kinase inhibitory protein (RKIP) binding pocket: NMR-based screening identifies small-molecule ligands. *PLoS ONE* 5, e10479.

Stark, A., and Russell, R.B. (2003). Annotation in three dimensions. PINTS: Patterns in Non-homologous Tertiary Structures. *Nucleic Acids Res* 31, 3341–3344.

Suhre, K., and Claverie, J.-M. (2004). FusionDB: a database for in-depth analysis of prokaryotic gene fusion events. *Nucleic Acids Res* 32, D273-276.

Susila, H., Nasim, Z., Gawarecka, K., Jung, J.-Y., Jin, S., Youn, G., and Ahn, J.H. (2021). PHOSPHORYLETHANOLAMINE CYTIDYLYLTRANSFERASE 1 modulates flowering in a florigen-independent manner by regulating *SVP*. *Development* 148, dev193870.

Takeba, G., Nakajima, Y., Kozaki, A., Tanaka, O., and Kasai, Z. (1990). A Flower-Inducing Substance of High Molecular Weight from Higher Plants. *Plant Physiol.* *94*, 1677–1681.

Tamaki, S., Matsuo, S., Wong, H.L., Yokoi, S., and Shimamoto, K. (2007). Hd3a protein is a mobile flowering signal in rice. *Science* *316*, 1033–1036.

Taoka, K., Ohki, I., Tsuji, H., Furuita, K., Hayashi, K., Yanase, T., Yamaguchi, M., Nakashima, C., Purwestri, Y.A., Tamaki, S., et al. (2011). 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature* *476*, 332–335.

Thao, S., and Escalante-Semerena, J.C. (2012). A positive selection approach identifies residues important for folding of *Salmonella enterica* Pat, an N ϵ -lysine acetyltransferase that regulates central metabolism enzymes. *Research in Microbiology* *163*, 427–435.

Tsuji, H., Taoka, K., and Shimamoto, K. (2011). Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. *Curr. Opin. Plant Biol.* *14*, 45–52.

van Tran, N., Muller, L., Ross, R.L., Lestini, R., Ltoquart, J., Ulryck, N., Limbach, P.A., de Crcy-Lagard, V., Cianfrani, S., and Graille, M. (2018). Evolutionary insights into Trm112-methyltransferase holoenzymes involved in translation between archaea and eukaryotes. *Nucleic Acids Research* *46*, 8483–8499.

Wheeler, T.J., Clements, J., and Finn, R.D. (2014). Skyline: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* *15*, 7.

Wigge, P.A. (2005). Integration of Spatial and Temporal Information During Floral Induction in *Arabidopsis*. *Science* *309*, 1056–1059.

Zeevaart, J.A.D. (1976). Physiology of Flower Formation. *Annu. Rev. Plant. Physiol.* *27*, 321–348.

Zeevaart, J.A.D. (2006). Florigen Coming of Age after 70 Years. *Plant Cell* *18*, 1783–1789.

Zhu, Y., Klasfeld, S., and Wagner, D. (2021). Molecular regulation of plant developmental transitions and plant architecture via PEPB family proteins – an update on mechanism of action. *Journal of Experimental Botany* *72*, 2301–2311.

545

Type of contextual evidence	Functionally linked proteins / domains / COGs	Taxa in which these functions are putatively linked	COG functional categories / most relevant GO terms // Additional comments
Domain fusion	Glucose-sorbose dehydrogenase, (Pectate Lyase-like carbohydrate-binding module, FN3-like domain) (COG1881, COG2133)	<i>Actinobacteria</i> , <i>Gammaproteobacteria</i>	Carbohydrate transport and metabolism/Hydrolase activity, acting on glycosyl bonds
Operon	Dialkylmaleic anhydride synthesis and conjugation module of tautomycin/tautomycetin biosynthesis operon	<i>Actinobacteria</i>	// five-carbon substrate – a pentose?
Phyletic vector (prokaryotes)	Acyl-CoA synthetase, NDP forming (COG1042).	220 species	Energy production and conversion/ATP-binding, N-acetyltransferase activity // nucleoside-containing substrate
Phyletic vector (prokaryotes)	YbaR/Trm112 activator of RNA and protein methyltransferases; COG2835	233 species	Translation, ribosomal structure and genesis // RNA modification
Phyletic vector (eukaryotes)	NUDT3 and other NUDIX hydrolases	<i>Metazoa</i>	// Hydrolases preferring pyrophosphate-containing substrates (nucleoside phosphates or phospholipids)
Shared putative regulatory motif	RlmA 23S rRNA m(1)G745 methyltransferase (COG2226)	<i>Enterobacteriaceae</i>	Coenzyme transport and metabolism/rRNA base methylation
Shared putative regulatory motif	YobB putative carbon-nitrogen hydrolase family protein (COG0388)	<i>Enterobacteriaceae</i>	Energy production and conversion/Nitrogen compound metabolic process
Shared putative regulatory motif	AdrB c-di-GMP phosphodiesterase (COG2200)	<i>Enterobacteriaceae</i>	Signal transduction mechanisms/Cellular response to DNA damage stimulus
Integrated evidence	RlhA 23S rRNA 5-hydroxycytidine C2501 synthase (COG0826)	n/a	Signal transduction mechanisms/rRNA processing, rRNA modification

546

547 **Table 1. Genome-context information suggesting functional links between YbhB homologs and various enzymes of**
548 **biosynthesis and salvage of small molecules.** See text for a more detailed characterization of each putative functional link.

Figure 2. Spatial organization of the conserved residues forming putative catalytic center in FT protein from *Arabidopsis thaliana* (PDB ID 1wkp). Top pane, strands are rendered in yellow, helices in light blue, and conserved residues involved in forming the putative enzyme active center are labeled and colored as follows: blue, two conserved aspartates; cyan, two conserved histidines; dark blue, conserved arginine; red, frequently conserved prolines. The loop between strands 3 and 4 is reduced to a short broken wire to improve the visibility of the putative active center, and the loop that is a major determinant of the antagonistic activity between FT and its paralog TFL1 (Ahn et al., 2006) is rendered as sticks. Bottom pane, the putative enzyme active center rendered as a surface. Rough calculation of the surface electrostatic properties in the vacuum were performed using PyMOL generate → vacuum electrostatics function; the shades of red indicate negative charges, and the shades of blue indicate positive charges. The conserved residues involved in forming the putative enzyme active center are labeled and colored as in the top pane.

Parsed Citations

- Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. (2005).** FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* 309, 1052–1056.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Abelenda, J.A., Bergonzi, S., Oortwijn, M., Sonnewald, S., Du, M., Visser, R.G.F., Sonnewald, U., and Bachem, C.W.B. (2019).** Source-Sink Regulation Is Mediated by Interaction of an FT Homolog with a SWEET Protein in Potato. *Current Biology* 29, 1178–1186.e6.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Ahn, J.H., Miller, D., Winter, V.J., Banfield, M.J., Lee, J.H., Yoo, S.Y., Henz, S.R., Brady, R.L., and Weigel, D. (2006).** A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *EMBO J* 25, 605–614.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Altschul, S.F., Madden, T.I., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997).** Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Andrés, F., and Coupland, G. (2012).** The genetic basis of flowering responses to seasonal cues. *Nat Rev Genet* 13, 627–639.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., and Teichmann, S.A. (2004).** Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* 14, 283–291.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Baker, P.A., Fritz, S.C., Dick, C.W., Eckert, A.J., Horton, B.K., Manzoni, S., Ribas, C.C., Garzone, C.N., and Battisti, D.S. (2014).** The emerging field of geogenomics: Constraining geological problems with genetic data. *Earth-Science Reviews* 135, 38–47.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Bernier, I., and Jollès, P. (1984).** Purification and characterization of a basic 23 kDa cytosolic protein from bovine brain. *Biochim Biophys. Acta* 790, 174–181.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Burroughs, A.M., and Aravind, L. (2020).** Identification of Uncharacterized Components of Prokaryotic Immune Systems and Their Diverse Eukaryotic Reformulations. *J Bacteriol* 202.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Burroughs, A.M., Zhang, D., Schäffer, D.E., Iyer, L.M., and Aravind, L. (2015).** Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res* 43, 10633–10654.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Carr, S.M., Marshall, H.D., Wareham, T., and Craig, D. (2015).** The Big ORF Theory. In *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology*, (Elsevier), pp. 265–274.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Chailakhyan, M. (1936).** New facts supporting the hormonal theory of plant development. *Dokl. Akad. Nauk SSSR* 4, 77–81.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Cheng, H., Liao, Y., Schaeffer, R.D., and Grishin, N.V. (2015).** Manual classification strategies in the ECOD database: ECOD Manual Classification Strategies. *Proteins* 83, 1238–1251.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Conti, L., and Bradley, D. (2007).** TERMINAL FLOWER1 Is a Mobile Signal Controlling Arabidopsis Architecture. *Plant Cell* 19, 767–778.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., et al. (2007).** FT Protein Movement Contributes to Long-Distance Signaling in Floral Induction of Arabidopsis. *Science* 316, 1030–1033.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Danilevskaya, O.N., Meng, X., McGonigle, B., and Muszynski, M.G. (2011).** Beyond flowering time: pleiotropic function of the maize flowering hormone florigen. *Plant Signal Behav* 6, 1267–1270.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al. (2008).** Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36, W465–469.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Edgar, R.C. (2004).** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792–1797.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019).** The Pfam protein families database in 2019. *Nucleic Acids Res* 47, D427–D432.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Vera Alvarez, R., Landsman, D., and Koonin, E.V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* 49, D274–D281.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Garner, W.W., and Allard, H.A. (1920). Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *Monthly Weather Review* 48, 415–415.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Glazko, G.V., and Mushegian, A.R. (2004). Detection of evolutionarily stable fragments of cellular pathways by hierarchical clustering of phyletic patterns. *Genome Biol.* 5, R32.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Glazko, G., Gordon, A., and Mushegian, A. (2005). The choice of optimal distance measure in genome-wide datasets. *Bioinformatics* 21 Suppl 3, iii3–11.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Glazko, G., Coleman, M., and Mushegian, A. (2006). Similarity searches in genome-wide numerical data sets. *Biol. Direct* 1, 13.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Guo, C., Chang, T., Sun, T., Wu, Z., Dai, Y., Yao, H., and Lin, D. (2018). Anti-leprosy drug Clofazimine binds to human Raf1 kinase inhibitory protein and enhances ERK phosphorylation. *Acta Biochim. Biophys. Sin. (Shanghai)* 50, 1062–1067.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Howe, K.L., Contreras-Moreira, B., De Silva, N., Maslen, G., Akanni, W., Allen, J., Alvarez-Jarreta, J., Barba, M., Bolser, D.M., Cambell, L., et al. (2020). Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Res* 48, D689–D695.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Huerta, A.M., and Collado-Vides, J. (2003). Sigma70 Promoters in Escherichia coli: Specific Transcription in Dense Regions of Overlapping Promoter-like Signals. *Journal of Molecular Biology* 333, 261–278.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Huynen, M., Snel, B., Lathe, W., and Bork, P. (2000). Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res* 10, 1204–1210.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Iyer, L.M., Aravind, L., Bork, P., Hofmann, K., Mushegian, A.R., Zhulin, I.B., and Koonin, E.V. (2001). Quod erat demonstrandum? The mystery of experimental validation of apparently erroneous computational analyses of protein sequences. *Genome Biol.* 2, RESEARCH0051.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Jin, S., Nasim, Z., Susila, H., and Ahn, J.H. (2021). Evolution and functional diversification of FLOWERING LOCUS T/TERMINAL FLOWER 1 family genes in plants. *Semin Cell Dev Biol* 109, 20–30.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kakimoto, T. (2003). Biosynthesis of cytokinins. *J Plant Res* 116, 233–239.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Karlgrén, A., Gyllenstrand, N., Källman, T., Sundström, J.F., Moore, D., Lascoux, M., and Lagercrantz, U. (2011). Evolution of the PEBP Gene Family in Plants: Functional Diversification in Seed Plant Evolution. *Plant Physiol.* 156, 1967–1977.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kimura, S., Sakai, Y., Ishiguro, K., and Suzuki, T. (2017). Biogenesis and iron-dependency of ribosomal RNA hydroxylation. *Nucleic Acids Research* 45, 12974–12986.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

King, R.W., Evans, L.T., and Wardlaw, I.F. (1968). Translocation of the floral stimulus in *Pharbitis nil* in relation to that of assimilates. *Zeitschrift Fur Pflanzenphysiologie* 59, S377–388.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kinoshita, A., and Richter, R. (2020). Genetic and molecular basis of floral induction in *Arabidopsis thaliana*. *Journal of Experimental Botany* 71, 2490–2504.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Klinterås, M., Pin, P.A., Benlloch, R., Ingvarsson, P.K., and Nilsson, O. (2012). Analysis of conifer FLOWERING LOCUS T/TERMINAL FLOWER1-like genes provides evidence for dramatic biochemical evolution in the angiosperm FT lineage. *New Phytol.* 196, 1260–1273.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Knott, J.E. (1934). Effect of a localized photoperiod on spinach. In *Proc Amer Soc Hortic Sci*, pp. 152–154.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kobayashi, Y., and Weigel, D. (2007). Move on up, it's time for change--mobile signals controlling photoperiod-dependent flowering. *Genes Dev.* 21, 2371–2384.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Koornneef, M., Hanhart, C.J., and van der Veen, J.H. (1991). A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Molec. Gen. Genet.* 229, 57–66.

Google Scholar: [Author Only Title Only Author and Title](#)

Krieger, U., Lippman, Z.B., and Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat. Genet.* 42, 459–463.

Google Scholar: [Author Only Title Only Author and Title](#)

Lang, A., Chailakhyan, M.K., and Frolova, I.A. (1977). Promotion and inhibition of flower formation in a dayneutral plant in grafts with a short-day plant and a long-day plant. *Proc. Natl. Acad. Sci. U.S.A.* 74, 2412–2416.

Google Scholar: [Author Only Title Only Author and Title](#)

Lau, M.E., Danka, E.S., Tiemann, K.M., and Hunstad, D.A. (2014). Bacterial lysis liberates the neutrophil migration suppressor YbcL from the periplasm of uropathogenic *Escherichia coli*. *Infect. Immun.* 82, 4921–4930.

Google Scholar: [Author Only Title Only Author and Title](#)

Li, W., Ju, J., Rajski, S.R., Osada, H., and Shen, B. (2008). Characterization of the tautomycin biosynthetic gene cluster from *Streptomyces spiroverticillatus* unveiling new insights into dialkylmaleic anhydride and polyketide biosynthesis. *J. Biol. Chem.* 283, 28607–28617.

Google Scholar: [Author Only Title Only Author and Title](#)

Li, W., Luo, Y., Ju, J., Rajski, S.R., Osada, H., and Shen, B. (2009). Characterization of the tautomycin biosynthetic gene cluster from *Streptomyces griseochromogenes* provides new insight into dialkylmaleic anhydride biosynthesis. *J. Nat. Prod.* 72, 450–459.

Google Scholar: [Author Only Title Only Author and Title](#)

Lifschitz, E., and Eshed, Y. (2006). Universal florigenic signals triggered by FT homologues regulate growth and flowering cycles in perennial day-neutral tomato. *J. Exp. Bot.* 57, 3405–3414.

Google Scholar: [Author Only Title Only Author and Title](#)

Lifschitz, E., Ayre, B.G., and Eshed, Y. (2014). Florigen and anti-florigen are a systemic mechanism for coordinating growth and termination in flowering plants. *Front. Plant Sci.* 5.

Google Scholar: [Author Only Title Only Author and Title](#)

McLennan, A.G. (2006). The Nudix hydrolase superfamily. *Cell Mol Life Sci* 63, 123–143.

Google Scholar: [Author Only Title Only Author and Title](#)

Milyaeva, E.L., Golyanovskaya, S.A., Aksenova, N.P., and Chailakhyan, M.Kh. (1991). Changes of Protein Composition in Leaves and Stem Apices of Bicolored Coneflower with Transition to Flowering. *Dokl. Akad. Nauk SSSR (Bot. Sciences)* 11–15.

Google Scholar: [Author Only Title Only Author and Title](#)

Mironov, A.A., Vinokurova, N.P., and Gelfand, M.S. (2000). Software for analysis of bacterial genomes. *Mol Biol* 34, 222–231.

Google Scholar: [Author Only Title Only Author and Title](#)

Miyazaki, T., Sugisawa, T., and Hoshino, T. (2006). Pyrroloquinoline Quinone-Dependent Dehydrogenases from *Ketogulonicigenium vulgare* Catalyze the Direct Conversion of L-Sorbosone to L-Ascorbic Acid. *AEM* 72, 1487–1495.

Google Scholar: [Author Only Title Only Author and Title](#)

Musfeldt, M., and Schönheit, P. (2002). Novel Type of ADP-Forming Acetyl Coenzyme A Synthetase in Hyperthermophilic Archaea: Heterologous Expression and Characterization of Isoenzymes from the Sulfate Reducer *Archaeoglobus fulgidus* and the Methanogen *Methanococcus jannaschii*. *JB* 184, 636–644.

Google Scholar: [Author Only Title Only Author and Title](#)

Nakamura, Y., Andrés, F., Kanehara, K., Liu, Y., Dörmann, P., and Coupland, G. (2014a). *Arabidopsis* florigen FT binds to diurnally oscillating phospholipids that accelerate flowering. *Nat Commun* 5, 3553.

Google Scholar: [Author Only Title Only Author and Title](#)

Nakamura, Y., Andrés, F., Kanehara, K., Liu, Y., Coupland, G., and Dörmann, P. (2014b). Diurnal and circadian expression profiles of glycerolipid biosynthetic genes in *Arabidopsis*. *Plant Signal Behav* 9, e29715.

Google Scholar: [Author Only Title Only Author and Title](#)

Nakamura, Y., Lin, Y.-C., Watanabe, S., Liu, Y.-C., Katsuyama, K., Kanehara, K., and Inaba, K. (2019). High-Resolution Crystal Structure of *Arabidopsis* FLOWERING LOCUS T Illuminates Its Phospholipid-Binding Site in Flowering. *IScience* 21, 577–586.

Google Scholar: [Author Only Title Only Author and Title](#)

Penel, C., Gaspae, Th., and Greppin, H. (1985). Rapid interorgan communications in higher plants with special reference to flowering. *Biol Plant* 27, 334–338.

Google Scholar: [Author Only Title Only Author and Title](#)

Persson, E., Castresana-Aguirre, M., Buzzao, D., Guala, D., and Sonnhammer, E.L.L. (2021). FunCoup 5: Functional Association Networks in All Domains of Life, Supporting Directed Links and Tissue-Specificity. *J Mol Biol* 166835.

Google Scholar: [Author Only Title Only Author and Title](#)

Pin, P.A., and Nilsson, O. (2012). The multifaceted roles of FLOWERING LOCUS T in plant development: FT, a multifunctional protein.

Plant, Cell & Environment 35, 1742–1755.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Pnueli, L., Gutfinger, T., Hareven, D., Ben-Naim, O., Ron, N., Adir, N., and Lifschitz, E. (2001). Tomato SP-Interacting Proteins Define a Conserved Signaling System That Regulates Shoot Architecture and Flowering. *Plant Cell* 13, 2687–2702.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ribeiro, A.J.M., Holliday, G.L., Furnham, N., Tyzack, J.D., Ferris, K., and Thornton, J.M. (2018). Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res* 46, D618–D623.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Roderick, S.L., Chan, W.W., Agate, D.S., Olsen, L.R., Vetting, M.W., Rajashankar, K.R., and Cohen, D.E. (2002). Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nat. Struct. Biol.* 9, 507–511.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Romanov, G.A. (2012). Mikhail Khristoforovich Chailakhyan: The fate of the scientist under the sign of florigen. *Russ J Plant Physiol* 59, 443–450.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Rudnitskaya, A.N., Eddy, N.A., Fenteany, G., and Gascón, J.A. (2012). Recognition and reactivity in the binding between Raf kinase inhibitor protein and its small-molecule inhibitor locostatin. *J Phys Chem B* 116, 10176–10181.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Sachs, J. (1865). Wirkung des Lichtes auf die Blütenbildung unter Vermittlung der Laubblätter. *Bot. Ztg* 23, 117.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Sadreyev, I.R., Ji, F., Cohen, E., Ruvkun, G., and Tabach, Y. (2015). PhyloGene server for identification and visualization of co-evolving proteins using normalized phylogenetic profiles. *Nucleic Acids Res* 43, W154-159.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Salazar-Cerezo, S., Martínez-Montiel, N., García-Sánchez, J., Pérez-Y-Terrón, R., and Martínez-Contreras, R.D. (2018). Gibberellin biosynthesis and metabolism: A convergent route for plants, fungi and bacteria. *Microbiol Res* 208, 85–98.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Shemon, A.N., Heil, G.L., Granovsky, A.E., Clark, M.M., McElheny, D., Chimon, A., Rosner, M.R., and Koide, S. (2010). Characterization of the Raf kinase inhibitory protein (RKIP) binding pocket: NMR-based screening identifies small-molecule ligands. *PLoS ONE* 5, e10479.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Stark, A., and Russell, R.B. (2003). Annotation in three dimensions. PINTS: Patterns in Non-homologous Tertiary Structures. *Nucleic Acids Res* 31, 3341–3344.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Suhre, K., and Claverie, J.-M. (2004). FusionDB: a database for in-depth analysis of prokaryotic gene fusion events. *Nucleic Acids Res* 32, D273-276.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Susila, H., Nasim, Z., Gawarecka, K., Jung, J.-Y., Jin, S., Youn, G., and Ahn, J.H. (2021). PHOSPHORYLETHANOLAMINE CYTIDYLTRANSFERASE 1 modulates flowering in a florigen-independent manner by regulating SVP. *Development* 148, dev193870.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Takeba, G., Nakajima, Y., Kozaki, A., Tanaka, O., and Kasai, Z. (1990). A Flower-Inducing Substance of High Molecular Weight from Higher Plants. *Plant Physiol.* 94, 1677–1681.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Tamaki, S., Matsuo, S., Wong, H.L., Yokoi, S., and Shimamoto, K. (2007). Hd3a protein is a mobile flowering signal in rice. *Science* 316, 1033–1036.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Taoka, K., Ohki, I., Tsuji, H., Furuita, K., Hayashi, K., Yanase, T., Yamaguchi, M., Nakashima, C., Purwestri, Y.A., Tamaki, S., et al. (2011). 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature* 476, 332–335.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Thao, S., and Escalante-Semerena, J.C. (2012). A positive selection approach identifies residues important for folding of *Salmonella enterica* Pat, an N ϵ -lysine acetyltransferase that regulates central metabolism enzymes. *Research in Microbiology* 163, 427–435.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Tsuji, H., Taoka, K., and Shimamoto, K. (2011). Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. *Curr. Opin. Plant Biol.* 14, 45–52.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

van Tran, N., Muller, L., Ross, R.L., Lestini, R., Ltoquart, J., Ulryck, N., Limbach, P.A., de Crcy-Lagard, V., Cianfrani, S., and Graille, M. (2018). Evolutionary insights into Trm112-methyltransferase holoenzymes involved in translation between archaea and eukaryotes. *Nucleic Acids Research* 46, 8483–8499.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wheeler, T.J., Clements, J., and Finn, R.D. (2014). Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinformatics 15, 7.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wigge, P.A. (2005). Integration of Spatial and Temporal Information During Floral Induction in Arabidopsis. Science 309, 1056–1059.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zeevaart, J.A.D. (1976). Physiology of Flower Formation. Annu. Rev. Plant. Physiol. 27, 321–348.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zeevaart, J.A.D. (2006). Florigen Coming of Age after 70 Years. Plant Cell 18, 1783–1789.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zhu, Y., Klasfeld, S., and Wagner, D. (2021). Molecular regulation of plant developmental transitions and plant architecture via PEPB family proteins – an update on mechanism of action. Journal of Experimental Botany 72, 2301-2311.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Type of contextual evidence

Functionally linked proteins / domains / COGs

Taxa in which these functions are putatively linked

COG functional categories / most relevant GO terms // Additional comments

Domain fusion

Glucose-sorbose dehydrogenase, (Pectate Lyase-like carbohydrate-binding module, FN3-like domain) (COG1881, COG2133)

Actinobacteria, Gammaproteobacteria

Carbohydrate transport and metabolism/Hydrolase activity, acting on glycosyl bonds

Operon

Dialkylmaleic anhydride synthesis and conjugation module of tautomycin/tautomycin biosynthesis operon

Actinobacteria

// five-carbon substrate – a pentose?

Phyletic vector (prokaryotes)

Acyl-CoA synthetase, NDP forming (COG1042).

220 species

Energy production and conversion/ATP-binding, N-acetyltransferase activity // nucleoside-containing substrate

Phyletic vector (prokaryotes)

YbaR/Tsm12 activator of RNA and protein methyltransferases; COG2835

233 species

Translation, ribosomal structure and genesis // RNA modification

Phyletic vector (eukaryotes)

NUDT3 and other NUDIX hydrolases

Metazoa

// Hydrolases preferring pyrophosphate-containing substrates (nucleoside phosphates or phospholipids)

Shared putative regulatory motif

RlmA 23S rRNA m(1)G745 methyltransferase (COG2226)

Enterobacteriaceae

Coenzyme transport and metabolism/rRNA base methylation

Shared putative regulatory motif

YobB putative carbon-nitrogen hydrolase family protein (COG0388)

Enterobacteriaceae

Energy production and conversion/Nitrogen compound metabolic process

Shared putative regulatory motif

AdrB c-di-GMP phosphodiesterase (COG2200)

Enterobacteriaceae

Signal transduction mechanisms/Cellular response to DNA damage stimulus

Integrated evidence

RlhA23S rRNA 5-hydroxycytidine C2501 synthase (COG0826)

n/a

Signal transduction mechanisms/rRNA processing, rRNA modification