REVIEW

Florigen and its homologs of FT/CETS/PEBP/RKIP/YbhB family may be the enzymes of small molecule metabolism: review of the evidence

Olga Tsoy[1,2] and Arcady Mushegian[3,4]*

Abstract

Background: 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 a product of plant gene *FT* was identified as the key component of florigen in the 1990s, based on the analysis of mutants, genetic complementation evidence, and protein and RNA 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. Most FT homologs are known to share a constellation of five charged residues, three of which, i.e., two histidines and an aspartic acid, are located at the rim of a well-defined cavity on the protein surface.

Results: We studied molecular features of the FT homologs in prokaryotes and analyzed their genome context, to find tentative evidence connecting the bacterial FT homologs with small molecule metabolism, often involving substrates that contain sugar or ribonucleoside moieties. We argue that the unifying feature of this protein family, i.e., a set of charged residues conserved at the sequence and structural levels, is more likely to be an enzymatic active center than a catalytically inert ligand-binding site.

Conclusions: We propose that most of FT-related proteins are enzymes operating on small diffusible molecules. Those metabolites may constitute an overlooked essential ingredient of the florigen signal.

Keywords: Florigen, Flowering, Phosphatidylethanolamine-binding protein, Raf-kinase interacting protein, YbhB, Tautomycin

Background

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 numerous studies have shown that the perception of the photoperiodic signal in plants occurs primarily in the leaves. Flowers, however, are formed mostly by shoot apical meristems (SAM) that are typically shielded from direct light, and the mechanisms conveying the flowering signal from leaves to SAMs have been a matter of speculation and investigation for most of the twentieth 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 [32, 58]. Experiments on floral induction in partially shaded plants and in grafts between light-induced and uninduced plants suggested

*Correspondence: mushegian2@gmail.com
[3] Molecular and Cellular Biology Division, National Science Foundation, 2415 Eisenhower Avenue, Alexandria, Virginia 22314, USA
Full list of author information is available at the end of the article

the existence of a diffusible substance promoting floral transition; the idea may have been first proposed by Julius Sachs [95], 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 [21]. The review of early research on the photoperiodic signal sensing in leaves can be found in Zeevaart [123] and Kobayashi and Weigel [60], and an account of Chailakhyan's remarkable life and scientific work has been given in Romanov [93].

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 rate comparable to that of other plant assimilates [54]. It has been established also that grafts and extracts of induced plants could transfer flowering ability not only to the uninduced vegetatively developing plants of the same species, but sometimes also to other species or genera, suggesting the evolutionary conservation of the signaling pathways [122].

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 yielded no results for several decades. A review of state of the affairs in 1976 listed the factors that have been tested unsuccessfully for the ability to induce floral transition; among the failed candidates there were sugars, amino acids, sterols, gibberellins, salicylic acid, ethylene, cytokinins, the photosynthetic capacity, initiation of protein synthesis, and others [122]. Several hypotheses were 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 [63]; or, maybe, flowering was caused by propagating a signal that spread from leaves to meristems not in chemical, but in electric form, such as plasma membrane potential [84].

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 a similar-sized band in the shoot apex tissues ([76]; 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 [107] 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 unaffected by proteinase K digestion that removed the protein, perhaps suggesting a role for an associated small molecule.

Nearly simultaneously, the results of genetic screens for *Arabidopsis thaliana* mutants with late flowering phenotypes were published [61]. This was a watershed moment in the studies of the molecular determinants of flowering initiation, and important discoveries ensued in the following three decades have been. The state of the knowledge may be outlined as follows (summarized from the following reviews – [6, 47, 49, 56, 69], which can be consulted for additional details and for the timeline).

Flowering in Arabidopsis is long day-dependent, and is enabled through the circadian clock-controlled transcriptional co-regulator *CONSTANS (CO)* and its target *FLOWERING LOCUS T (FT)*. The protein product of the latter gene has emerged as the integrator of the environmental inputs, relaying these signals into the gene regulatory networks that control flowering. The FT protein of 19.8 kDa (176 amino acids) is produced in the phloem companion cells of the leaves, enters phloem sieve elements and is transported from leaves to the base of shoot apical meristem, where it has been detected experimentally. Considerable evidence exists that the *FT* gene is expressed in SAM mostly or only in 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 the expression of floral meristem identity and flowering time genes, 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 other plant 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 homologs of *FT*, found in varying numbers in all examined angiosperm species, also tend to be involved in flowering control, many of them acting synergistically, redundantly or antagonistically with *FT*; perhaps the best-studied antagonistic pair is Arabidopsis gene *FT* and its paralogous gene *TFL1* and the counterparts of those

two genes in other flowering plants similarly have opposing effects on plant development [4, 59, 62, 68, 73, 89, 112, 119]. FT and its homologs have been also implicated in other aspects of plant organogenesis, such as seed germination in bamboo and tuber formation in potato, as well as in a growing variety of physiological processes, such as vacuolar sorting, stomata opening, and sink-source regulation [3, 24, 26, 49, 52, 55, 87, 103].

Despite all the effort to pinpoint its precise location, FT protein has not been unambiguously detected in phloem-free tips of SAM in any plant, only in the basal portions 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 bottom half of the SAM zone, and in those cases they do not complement the *ft* recessive mutants [25]. A study in rice [108] has made a more direct claim, i.e., that Hd3a, the rice protein orthologous to FT, does reach the tip of the SAM to induce flowering. However, their Fig. 3E, F, M, and N, cited as the key supporting evidence, does not seem to show the protein reporter activity at the tip, where it is supposed to be – only in the bottom halves of the SAM zone, just like in the case of Fig. 2G in Corbesier et al. [25]. More recent analysis used a sensitive fluorescent assay to report that FT is found in the basal as well as the apical part of SAM [2], though their Fig. 2 again shows a fluorescence-negative zone at the extreme apex. In fact, it has been hypothesized that, contrary to a naïve expectation, the restriction of FT to the basal portions of SAMs may be a tightly regulated, TFL1-mediated event, actually beneficial for maintaining the supply of undifferentiated stem cells in the more apical parts [13, 120]. Be it as it may, 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 [6, 68].

Another gene of Arabidopsis known to be involved in flowering control, *FD*, has been identified in a robust genetic screen 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 [1]. The *FD* gene encodes a protein from the bZIP-type transcription factor family, which physically interacts with the FT protein in yeast two-hybrid system, in the bimolecular fluorescence complementation assays in vivo, and apparently in the affinity-purified protein complexes [1, 118]. In rice, the FT ortholog HD3a does not interact with the bZIP factors directly, but co-precipitates with the FD ortholog OsFD1 in a tripartite complex that also includes a scaffolding 14–3-3 protein [109]. Largely on the basis of this evidence, 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, to the DNA regions located near many of the genes involved in flowering control [124].

The emerging picture of the FT role in transcriptional control of gene expression is complicated by the sub-cellular localization studies of FT and its homologs. FT appears to be transported into the nucleus in a FD-dependent manner, but remains largely cytoplasmic in *fd* plants [2]. On the other hand, its best-studied paralog TFL1 in Arabidopsis has been localized to plasma membrane, tonoplast, and dense vesicles in situ, and to the membranes of fractionated protoplasts, with direct implications in protein trafficking to the storage vacuoles [103], whereas the ortholog of TF in potato, StSP6A, has also been seen in association with membranes [3].
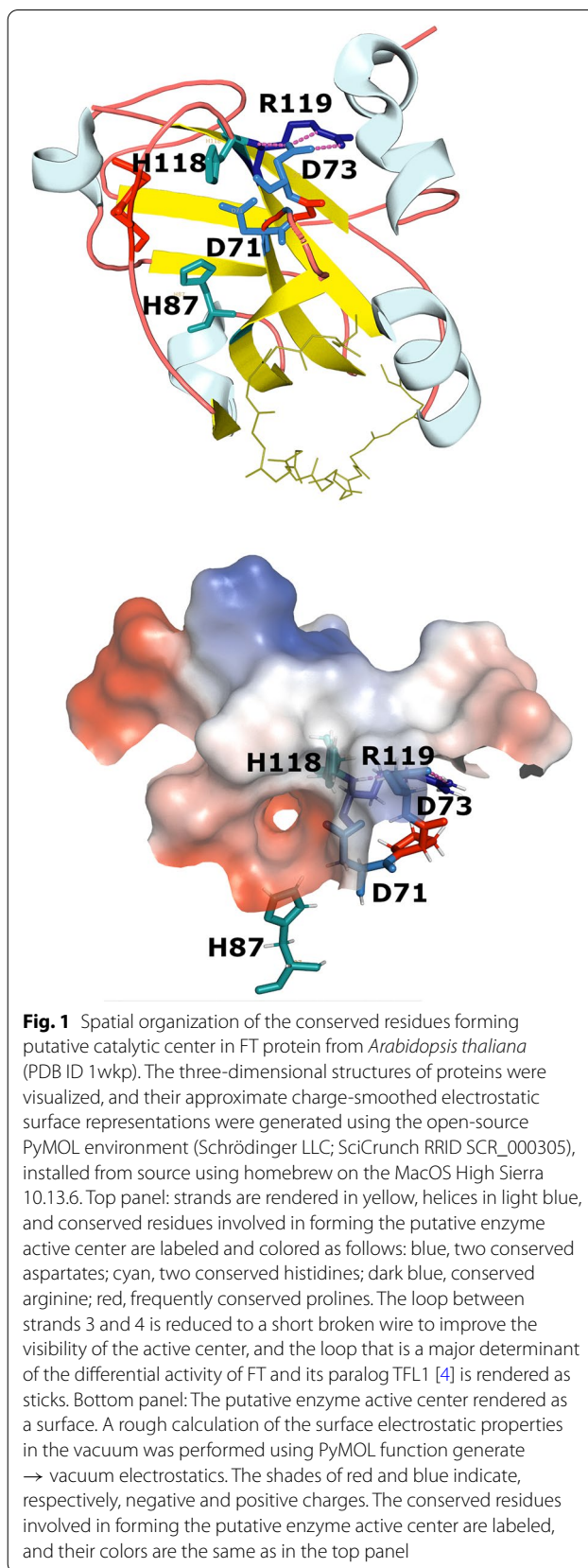
Analysis of the amino acid sequence has shown that FT protein belongs to a widely conserved sequence family, members of which are encoded by genomes of many bacteria, archaea and nearly all eukaryotes. The founding member of the family, isolated from bovine brain, is a hydrophilic, cytoplasmic protein that can bind in vitro to many low molecular weight compounds of different chemical structure, including certain phospholipids [15]. 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 a surprisingly large variety of genetic screens and binding assays. This has resulted in a long list of putative properties assigned to these proteins, including, in addition to in vitro phospholipid binding in plant and animal homologs, also inhibition of carboxypeptidase Y in vitro and regulation of Ras GTPase in vivo by the yeast homolog Tfs1 [16, 36]; inhibition of Raf-1 kinase in mammals – hence an alternative name of the protein family, Raf-Kinase Inhibitor Protein, or RKIP [121]; suppression of trans-epithelial migration of the mammalian host neutrophiles by YbcL, a PEBP homolog from uropathogenic *E.coli* [64]; and an uncharacterized role in the modification of polyketide chains in *Streptomyces* [66, 67]. In a continuation of the theme of diverse cellular locations in the family members, one of the FT co-orthologs in mammalian cells, PEBP1,
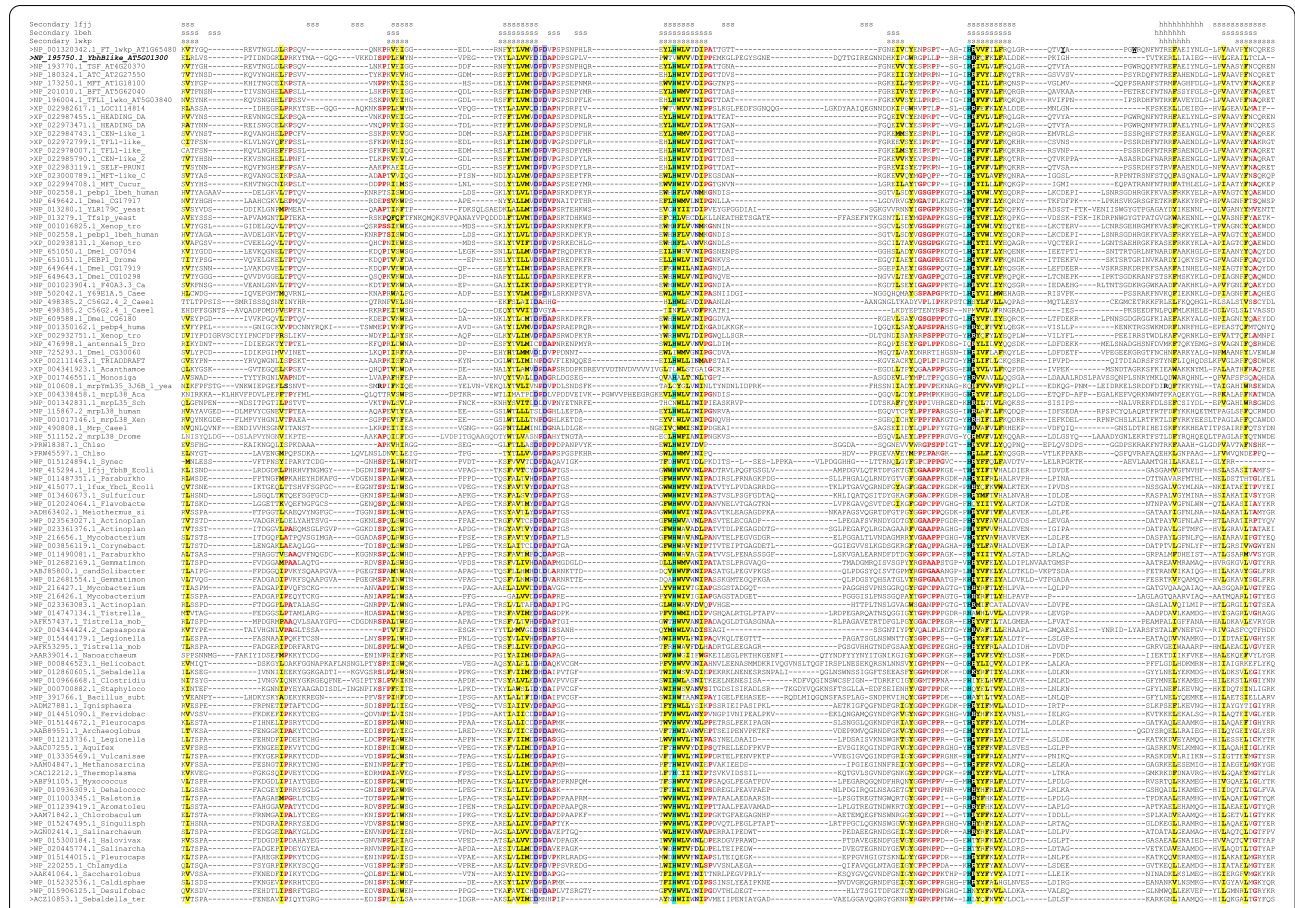
directly interacts with some isoforms of membrane-localized 15-lipoxygenase and controls their function [115], whereas at least three animal and one bacterial homolog, i.e., mammalian PEBP4, a predicted odorant-binding protein in fruit fly, a putative venom gland constituent in gall wasp, and aforementioned YbcL, are either known or predicted to be secreted extracellularly [19, 40, 64, 86].

Structural studies of FT/PEBP-like proteins from diverse species of bacteria and eukaryotes, conducted by X-ray crystallography and nuclear magnetic resonance approaches [8, 11, 12, 29, 37, 77, 90, 99, 100, 110] have revealed a distinct spatial fold, with a structural core dominated by two beta-sheets. Structurally, the fold superficially resembles an immunoglobulin-like arrangement but is not specifically related to any other fold (entry 11.1.7 in the ECOD database; Cheng et al., [23]). A prominent structural feature of this family is an evolutionary conserved depression on the surface of the molecule; the rim of the cavity is lined by three of the residues conserved in the entire family, i.e., an aspartic acid located at the C-terminus of strand 3 (in the FT protein of Arabidopsis, it is denoted D71 [42]) and two histidines found at the N-termini of strands 4 and 6, denoted H87 and H118. The quintet of the most conserved residues is completed by another aspartic acid close to the first one (D73), and an arginine that adjoins the second conserved histidine (R119) and bonds via a salt bridge with D73, forming a "shoulder" next to the cavity [99]. The spatial configuration of these residues and select other conserved features in Arobidopsis FT protein are rendered in Fig. 1.

In the known crystal structures, the cavity sometimes accommodates anions included in the media, but it 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 [4, 39, 83, 92, 94, 101]. Curiously, the site of interaction between FT and FD in *A.thaliana* has not been structurally characterized.

In this work, we highlight the patterns of sequence and structure conservation in the family and present the results of the analysis of genomic context of the FT/PEBP proteins in prokaryotes. We argue that the total evidence is best compatible with the idea that these proteins are not only (or even perhaps not at all) transcriptional co-activators, but enzymes involved in production, conjugation, or removal of low molecular weight ligands.


Fig. 1 Spatial organization of the conserved residues forming putative catalytic center in FT protein from *Arabidopsis thaliana* (PDB ID 1wkp). The three-dimensional structures of proteins were visualized, and their 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. Top panel: 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 active center, and the loop that is a major determinant of the differential activity of FT and its paralog TFL1 [4] is rendered as sticks. Bottom panel: The putative enzyme active center rendered as a surface. A rough calculation of the surface electrostatic properties in the vacuum was performed using PyMOL function generate → vacuum electrostatics. The shades of red and blue indicate, respectively, negative and positive charges. The conserved residues involved in forming the putative enzyme active center are labeled, and their colors are the same as in the top panel

Fig. 2 Multiple sequence alignment of select members of FT/CETS/PEBP/RKIP/YbhB family. Sequence identifiers in GenBank or PDB are shown before each sequence. For the arabidopsis proteins, gene product names and AGI locus codes are also shown. In the Secondary structure lines above the alignment, s stands for a beta-strand and h stands for an alpha-helix. Conserved hydrophobic residues (I, L, M, V, F, Y, W) are indicated by yellow shading, conserved small-side-chain or turn/kink-prone residues (A, G, S, P) are indicated by bold red type, and the conserved constellation of charged residues discussed in the text are marked as follows: gray-shaded blue type, two conserved aspartates; cyan shading, two conserved histidines; and black-shaded white type, conserved arginine. In the sequence XP_024370999.1, the bold underlined X marks the position of a low-complexity sequence insertion that most likely represents a falsely predicted exon

Survey and summary of the computational evidence
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 searching the NCBI NR protein database, or databases restricted by taxon (e.g., viruses), focusing on completely sequenced genomes. The database searches were done using PSI-BLAST program with standard settings [5], mostly in June 2021. We also consulted the NCBI COG resource that annotates conserved orthologous genes in bacteria and archaea [31]. In the following, unless specified otherwise, we refer to the prokaryotic FT homologs as YbhB proteins, after the chromosomal locus in *E.coli* that has orthologs in many other bacteria, and use some of the FT/CETS/PEBP/RKIP 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. 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 phylae *Firmicutes*, *Mollicutes* and the order *Spirochaetales* mostly contain species that are YbhB-free. In other clades of bacteria and archaea, between 30 and 80% of all species encode YbhB-family proteins. All told, 782 copies of YbhB-family proteins are found in 594 bacterial and archaeal genomes out of the 1309 genomes in the 2020 release of the COG database [31]. YbhB homologs are also encoded by the genomes of many giant DNA
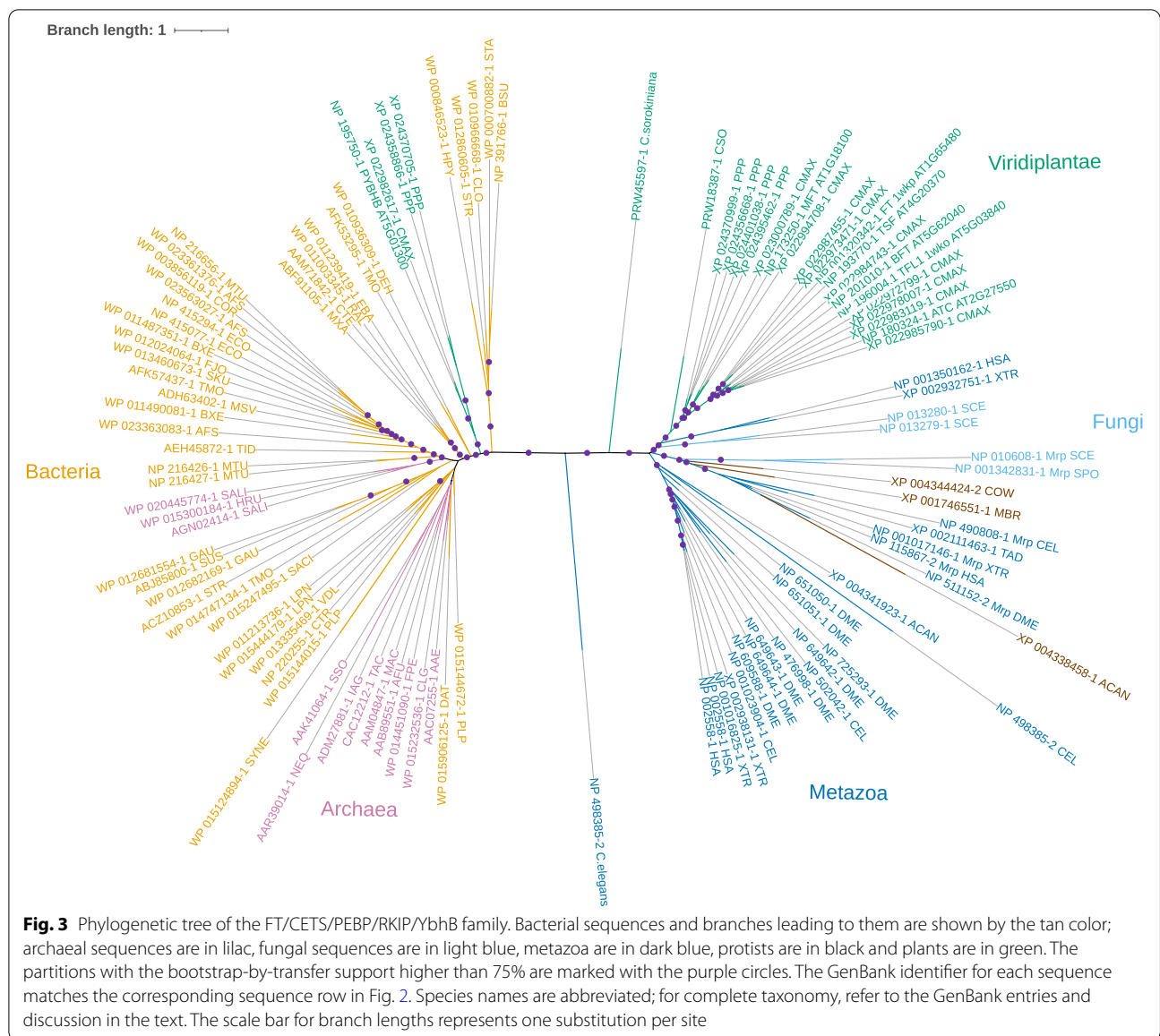
Fig. 3 Phylogenetic tree of the FT/CETS/PEBP/RKIP/YbhB family. Bacterial sequences and branches leading to them are shown by the tan color; archaeal sequences are in lilac, fungal sequences are in light blue, metazoa are in dark blue, protists are in black and plants are in green. The partitions with the bootstrap-by-transfer support higher than 75% are marked with the purple circles. The GenBank identifier for each sequence matches the corresponding sequence row in Fig. 2. Species names are abbreviated; for complete taxonomy, refer to the GenBank entries and discussion in the text. The scale bar for branch lengths represents one substitution per site

viruses from the class *Megaviricetes* 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. Amino acid sequences were aligned using the program MUSCLE v. 3.8.31 [27], and the phylogenetic trees were constructed using the PhyML maximum likelihood approach [38] available through the Booster server at Pasteur Institute (booster.pasteur.fr) and the Galaxy FastTree workflow [72]. The validity of the tree partitions was assessed bootstrap-by-transfer approach with 200 replicates, which has been reported to have higher resolution than the traditional bootstrap and to induce fewer falsely

supported branches [65]. The sequence alignment is shown in Fig. 2, the tree is shown in Fig. 3, and the source phylogeny in the Newick format is available as Supplemental Data Set 1.

The phylogeny that we obtained is dominated by a deep split between prokaryotic and eukaryotic sequences; the long internal branch separating those two groups is broken only by two eukaryotic outliers, a fast-evolving homolog from *C.elegans* and a sequence from green alga *Chlorella sorokiniana*. Within eukaryotes, plant FT/CETS homologs segregate as one clade, and fungal/metazoan homologs form another assembly, occasionally intermingled with the representatives of unicellular eukaryotes. A well-defined clade of fungal and metazoan

proteins 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 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 homologs that do not contain such an extension). This domain is not included in the alignment in Fig. 2; none of the plant FT homologs appear to contain such a region.

The six well-known paralogs in *A.thaliana* are FLOW-ERING LOCUS T (FT; AT1G65480), TWIN SISTER OF FT (TSF AT4G20370), TERMINAL FLOWER 1 (TFL1; AT5G03840), BROTHER OF FT AND TFL1 (BFT; AT5G62040), MOTHER OF FT AND TFL1 (MFT; AT1G18100), and *Arabidopsis Thaliana* CENTRORA-DIALIS (ATC; AT2G27550). These sequences, and their counterparts in other species, show the expected topology, with the MFT branch splitting off the common angiosperm stem first, and the FT/TFL branch diversifying later, with additional variations within angiosperms [14, 22, 51, 117]. The sequences from more primitive plants included in the alignment, i.e., moss *Physcomitrium patiens* (former *Physcomitrella*, recently re-absorbed into a larger genus – [75]) and another representative from *Chlorella sorokiniana*, predictably form their own deep clades near the common root of Viridiplantae. Evolution and functional specialization of FT homologs in different lineages of Viridiplantae, including early branching ones, have been extensively reviewed elsewhere [41, 57, 70].

Within the prokaryotic partition of the tree, the phylogeny is less well-resolved; it includes many deep-branching clades, which are often supported statistically but their position relative to other such clades is unclear. As a rule, however, these clades do not mix bacterial and archaeal representatives. On the other hand, on several occasions a group of homologs from closely related bacterial species would include also a sequence from a phylogenetically distant bacterium; in just one example, a sequence from Gram-negative proteobacterium *Helicobater pylori*, WP000846523.1, resides within a group of YbhB-like sequences from Gram-positive bacteria (Fig. 3). This indicates that the evolution of the YbhB family in bacteria must have included occasional horizontal gene transfer events.

The trend discussed above, i.e., the partitioning of all FT/YbhB homologs into the eukaryotic and prokaryotic tribes in the tree, is broken in the case concerning a neglected, uncharacterized member of the FT family in arabidopsis, AT5G01300. The expression of this poorly studied protein in plants and its apparent close similarity to bacterial YbhB may have been first discussed by Schieffer [98], and the expression of an orthologous gene in *Brassica*, as well as its large evolutionary distance from the other homologs, were recently noted in Sheng et al. [102]. In our tree, the sequence of AT5G01300, together with its homologs from another angiosperm and a moss, were clearly nested, with strong statistical support, within the prokaryotic portion (Fig. 3). Database searches and preliminary sequence analysis indicate that nearly all sequenced plant genomes encode such bacteria-like orthologs of AT5G01300 (unpublished data). We propose to call this 7th member of the FT family in Arabidopsis *PYBHB* (for plant YbhB homolog). A more detailed analysis of the phylogenetic origin of PYBHB would require a denser sampling of green plant lineages, as well as of their immediate algal ancestors and bacteria; such an analysis would be of great interest but is beyond the scope of this work.

Taken together, these data suggest, first and foremost, that FT/CETS/PEBP/RKIP/YbhB-like proteins are ubiquitous throughout the evolution of life, and that nearly all eukaryotic homologs are likely to have a single origin in a prokaryotic ancestor, with a possible secondary retention of a bacterial-like PYBHB clade in plants. Moreover, nearly-universal sequence conservation of several key amino acid residues and the common structural scaffold of the entire protein family (Figs. 1 and 2, and see the next section) suggest that most of those proteins share an ancient conserved molecular function.

Sequence-structure analysis of PEBP/RKIP/FT/YbhB proteins supports the hypothesis of a universally conserved function

A common biochemical function of the entire PEBP/YbhB protein family is further suggested by delineation of the conserved and variable sequence elements of the family (Fig. 2). As has been noticed before [11, 12, 99, 100], there are five nearly-invariant polar amino acid residues in the alignment. Several additional highly conserved residues, in particular prolines, are found around these signature amino acids (Figs. 1 and 2). Evolutionary substitutions in the five polar residues are exceedingly rare. Interestingly, multiple replacements in those five sites are observed in just two groups of the PEBP/RKIP/FT/YbhB proteins. One of such groups is the clade of animal and fungal mitochondrial ribosomal proteins L35/L38, which must have acquired a new molecular function during evolution. The other group comprises the plant PYBHB proteins mentioned above. It is tempting to speculate that the PYBHB functions are likewise derived, possibly plant-specific, and could be related to the organelle function; PYBHB sequences, however, do not bear obvious signals for sorting to chloroplasts or mitochondria, and do not seem to

cluster with the groups of bacteria that are close to the putative bacterial ancestors of plant organelles (unpublished data).

High-resolution spatial structures of many FT/PEBP/RKIP/FT/YbhB homologs from many diverse species have been obtained (see references above). Mapping the conserved sequence features onto these structures suggests that they are clustered in space, with three of the five most conserved residues located around the rim of the main cavity on the surface of the molecule, and the remaining two forming a salt bridge just above this rim (Fig. 1). The polar hole on the surface of the PEBP/RKIP/FT/YbhB protein molecules binds some of the polar ligands added to the crystallization media, such as cacodylate, phosphotyrosine, phosphoethanolamine, or (4S)-3-[(E)-1-Oxo-2-butenyl]-4-(phenylmethyl)-2-oxazolidinone, a small molecule that disrupts mammalian RKIP interaction with protein kinase Raf. On the other hand, when phosphatidylethanolamine or other polar lipids were included in the media, they were not seen in the electron density. For example, a recent report that FT from arabidopsis may bind phosphatidylcholine (PC) more strongly and more specifically than other lipids [81] has prompted the analysis of high-resolution structures of FT crystallized in the presence of PC. The crystals were obtained in a variety of conditions, but PC molecule could not be located in any of them; computational docking tentatively suggested four binding sites for PC, all located far from the conserved polar cavity [81, 82]. 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 conducive for direct interactions with phosphate, whereas the charge on the "shoulder" may be more positive (Fig. 1).

A phenotyping experiment on a panel of mutagenized versions of FT [42] has revealed a differential effect of substitutions in the five conserved charged residues. Out of 14 missense mutations in these residues, six did not alter the early-flowering phenotype of the positive control – the wild-type plant in which the FT transgene was overexpressed – and eight caused a dominant-negative effect of flowering delay compared to the wild type (Supplementary Fig. 4 in [42]). Altering a charge in one or more of residues D71, D73, H118 and R119 appears to be particularly effective in switching the phenotype from early to late flowering. This clearly suggests that the charge distribution on the surface of the FT protein, at the cavity rim itself as well as on its adjoining shoulder, is important; even so, we have no mechanistic explanation for the role of those most prominently conserved elements in carrying out the function of FT and its homologs, either in plants or in other species.

In this survey, we are interested whether these signature elements of sequence and structure may suggest a biochemical function of the FT/YbhB homologs. The Mechanism and Catalytic Site Atlas (M-CSA) resource contains annotated information about the amino acid determinants of catalysis in conserved enzyme families [91]. A web interface allows users to specify a set of conserved residues anywhere in the molecule and search the database with such a signature. We queried M-CSA and identified 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 searches can be reproduced at https://www.ebi.ac.uk/thornton-srv/m-csa/browse/ by selecting the set of conserved residues). The specific three-dimensional configuration of these residues in PEBP/RKIP/FT/YbhB proteins appears to be quite unique, however, as judged by the analysis with the PINTS program, which compares similar spatial arrangements of key residues in non-homologous proteins [104]. Nonetheless, it is clear that a combination of two histidines, two aspartates and an arginine has been repeatedly utilized in evolution to build enzymatic active centers, enabling many kinds of catalytic conversions.

These observations are in sharp contrast to what is known about patterns of sequence conservation in non-catalytic ligand-binding protein domains. We collected all domains in PFAM 33.1 [28], 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. For each of those domains, we downloaded the curated seed alignment from PFAM and used the Skylign server [116] to build HMM profiles and to get the information count for the residues of interest. 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 Data Set 2). Thus, the known conserved ligand-binding domains do not utilize the combination of two histidines and an aspartic acid residue for their non-catalytic interactions with small molecules. Taken

together, these investigations of the conserved features in PEBP/FT/YbhB proteins suggest that they are more likely to be a part of the catalytic center than to serve solely as a binding interface.

Genomic context of FT homologs 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). The first kind of linkage we studied was the information on domain fusions.

It is quite common for two protein domains to exist as separate genes in some species, but to be fused into one gene encoding a multidomain protein in others. Such translational fusions, especially those that are evolutionary conserved, are strongly enriched in proteins that work in concert with one another [45, 105]. Analysis of YbhB homologs in bacteria and archaea shows that they are most commonly encoded as stand-alone open reading frames and form domain fusions infrequently. In actinobacteria, however, the YbhB homologs are frequently found as the C-terminal portions of longer proteins, fused to 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, followed by the FN3 repeat region, a putative glucose/sorbosone dehydrogenase (GSDH) region with predicted beta-propeller structure, 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 [79], and it is conceivable that the C-terminal YbhB homology domains are involved in the transformations of GSDH substrates or products (or, possibly, in the metabolism of its cofactor).

Table 1 Genome-context information suggesting functional links between YbhB homologs and various enzymes of biosynthesis and salvage of small molecules

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 fusions	Glucose-sorbosone dehydrogenase, (Pectate Lyase-like carbohydrate-binding module, FN3-like domain) (COG1881, COG2133)	*Actinobacteria, Gammaproteobacteria*	Carbohydrate transport and metabolism / Hydrolase activity, acting on glycosyl bonds
Operons	Dialkylmaleic anhydride synthesis and conjugation module of tautomycin/tautomycetin biosynthesis operon	*Actinobacteria*	// Five-carbon substrate (a pentose?)
Phyletic vectors (prokaryotes)	Acyl-CoA synthetase, NDP forming (COG1042).	220 species	Energy production and conversion / ATP-binding, N-acetyltransferase activity // Nucleoside-containing substrate and product
Phyletic vectors (prokaryotes)	YbaR/Trm112 activator of RNA and protein methyltransferases; COG2835	233 species	Translation, ribosomal structure and genesis // RNA modification
Phyletic vectors (eukaryotes)	NUDT3 and other NUDIX hydrolases	*Mammalia*	// Hydrolases preferring pyrophosphate-containing substrates (nucleoside phosphates or phospholipids)
Shared putative regulatory motifs	RlmA 23S rRNA m(1)G745 methyltransferase (COG2226)	*Enterobacteriaceae*	Coenzyme transport and metabolism / rRNA base methylation
Shared putative regulatory motifs	YobB putative carbon-nitrogen hydrolase family protein (COG0388)	*Enterobacteriaceae*	Energy production and conversion / Nitrogen compound metabolic process
Shared putative regulatory motifs	AdrB c-di-GMP phosphodiesterase (COG2200)	*Enterobacteriaceae*	Signal transduction mechanisms / Cellular response to DNA damage stimulus // Nucleoside substrate
Integrated evidence	RlhA 23S rRNA 5-hydroxycytidine C2501 synthase (COG0826)	n/a	Signal transduction mechanisms / rRNA processing, rRNA modification

See text for a more detailed characterization of each putative functional link

We then surveyed the databases of conserved genomic neighborhoods and scanned the literature for the evidence of conservation of FT/YbhB chromosomal neighbors in different genomes. As with protein fusions, the FT/YbhB family genes are rarely found within conserved positional associations, such as operons, with other ORFs. In two species of *Streptomyces*, however, YbhB homologs are located within 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, form a subsystem that is 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 five-carbon compound, through the succession of the steps that are incompletely understood [66, 67]; it is plausible that the products of *ttnL* and *ttmL* play a role in the utilization of such a compound – perhaps a sugar or its derivative.

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 [33, 35]. We analyzed phyletic vectors using the psi-square program [34] with default settings and vector similarity measured using Pearson correlation-based distance. The vector of YbhB (COG1881) was used as the query, and the NCBI COG database release of 2014 was used as the search space (digitized phyletic vector information was not available for the 2021 COGs release at the time of writing). After retaining the matching vectors of comparable cardinality, i.e., the genes present in 300–400 species, to avoid spurious matching to nearly-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 acetyl-CoA + ADP + Pi $\rightleftharpoons$ acetate + ATP + CoA [80]. The roles of the bacterial homologs of this enzyme is less clear, as some of them appear to be catalytically inactive and possibly play auxiliary roles in the acylation and deacylation of proteins [111]. 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 [113].

One more way to predict functional linkages between gene products in bacteria is to examine their putative regulatory regions, which tend to be located in the intergenic spacers, usually to the 5′ ends of the open reading frames or operons they regulate. We performed an analysis of conserved upstream regions of *ybhB* genes, using the complete bacterial genomes extracted from Ensembl Bacteria database (release 47) [43]. The regulatory regions of homologous genes were aligned with the MUSCLE program, the most conservative regions were chosen to build a positional weight matrix by the SignalX routine of the Genome Explorer program [78], and the genomes were scanned by Genome Explorer with the threshold equal to the lowest score in the training set, i.e., in the *ybhB* homologous regions. Only some species in the family *Enterobacteriaceae* had a detectable conserved site upstream of the y*bhB* open reading frame. Unlike most known binding sites for the specialized transcription factors, this site lacked palindromic structure and has a consensus sequence TACACTT. Scanning 41 *Enterobacteriaceae* species from 14 genera with a probabilistic model of the site, we identified additional intergenic regions where the variants of these sites occur (Table 1, and Supplemental Data Set 3). Three genes were found to contain a highly similar conserved upstream site in 16–17 species, representing 8 genera; these were *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 hypothesized to represent a modified form of the canonical -10 element sequence TATAATT [44], though 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.

We also have consulted several databases of gene essentiality, as well as curated databases of pre-computed functional linkages between genes in various model organisms. The bacterial fitness database [88] reports a moderate reduction in mobility for an *E.coli* mutant with transposon-tagged YbcL, a prophage-encoded paralog of YbhB, whereas the Database of Essential Genes [71] identifies the YbhB counterpart in *H.pylori* as essential; in neither of this cases is there any information about possible molecular mechanisms. The FunCoup database, which uses naïve Bayesian approach to integrate

information on interactions and functions from 10 different genomic and proteomic measurement spaces [85], shows a small network of interacting proteins in *E.coli*, consisting of two chaperones, i.e., a heat shock 70-family DnaK and a protease Lon, as well as YbhB and RlhA. Interactions with chaperones are frequently observed in the protein interaction data and may reflect non-specific cellular proteostasis needs and a broad clientele of many chaperone systems; the connection to the other protein in the group, RhlA, may be more revealing, as it appears to be a component of the 5-hydroxycytidine synthase enzymatic complex, involved in modification of 23S rRNA [53]. Finally, the analysis of phyletic patterns across many eukaryotes, collected in PhyloGene database [96] 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 [74].

Conclusions

"The burden of proof should be proportional to the strangeness of the facts."

George Flournoy. Translated from Des Indes à la Planète Mars: Étude sur un Cas de Somnambulisme Avec Glossolalie.

"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.

In this survey, we reviewed, by necessity briefly, the strong genetic evidence of the key role of the FT/CETS family in the floral induction in plants. We also argue that, while the evidence of genetic interactions of *FT* gene with other genes in the floral induction pathways is not in doubt, the molecular function of the FT proteins in flowering may have been misunderstood. Indeed, the analysis of literature, as well as many orthogonal computational experiments presented here, suggest that the proteins from the FT/CETS/PEBP/RKIP/YbhB family, which are ubiquitous not only in plants, but also in animals, fungi, protists, prokaryotes and giant viruses, might be catalytic subunits of the enzymes involved in biochemical transformations of small molecules, in addition to, or even

instead of, 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 as a transcriptional co-activator promoting floral induction, 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 any evidence that these functions may be conserved in eukaryotes, including plants?"

The first of these questions can be tentatively answered by examination of the genomic context of FT/YbhB homologs. Though no single gene could be repeatedly linked to YbhB by multiple approaches, a trend emerges when the genomic context data are considered jointly (Table 1). The evidence appears to point towards functional linkages of YbhB to sugar and/or ribonucleoside modifications, implying that FT/YbhB may be involved in the metabolism of a monosaccharide such as ribose or another pentose, or perhaps their nucleotide-like derivative. Relatedly, FT proteins could be involved in phospholipid metabolic pathways, which include lipid-nucleotide conjugates as key intermediates [48].

The answer to the other question posed above, i.e., whether the putative molecular function have been preserved throughout the evolution of the FT/CETS/PEBP/RKIP/YbhB family, appears to be clearly "yes". There is a striking pattern of sequence conservation and spatial juxtaposition of the key charged residues in the family at a long phylogenetic span (Figs. 1 and 2), with only two narrow clades experiencing major disruption in those positions (Figs. 2 and 3). One of those clades comprises proteins with a changed function (mitochondrial ribosomal components), and the function of the other, which includes a divergent FT homolog, is enigmatic.

Bioinformatic analysis suggests that the conserved sets of two histidines and an aspartic acid clustered in space are frequent in the enzymatic active centers but are not found in the non-catalytic ligand-binding domains. In the same vein, broad occurrence of a protein family in bacteria, archaea, giant viruses and nearly all eukaryotes may not be unusual for a metabolic enzyme, but would be a rare occurrence for transcriptional regulators, as they are typically not shared between bacteria and eukaryotes [9].

Several recent observations on the role of sugar and lipid metabolism genes in floral induction may be relevant. For example, StSP6A protein, the potato ortholog of FT, is co-expressed with a sugar transporter SWEET11 in the stolon apical and sub-apical meristems during tuber induction, and the products of two proteins interact

physically in a heterologous split-ubiquitin assay in yeast cells [3]. In arabidopsis, a related sugar transporter, SWEET10, appears to be regulated by FT at the transcriptional level, though the data on physical interactions in that case have not been reported [7]. Also, manipulations of the biosynthetic enzymes producing trehalose-6-phosphate in arabidopsis have shown that floral response and shoot branching require this phosphosugar, which apparently acts through the FT pathway [30, 114]. Genetic evidence has also suggested the involvement of phosphorylethanolamine cytidylyltransferase (*PECT1* gene product) in flowering in arabidopsis [106].

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 [50], whereas another class, gibberellins, are synthesized by transforming the pentose skeleton generated in the 1-deoxy-D-xylulose 5-phosphate pathway [97]. Recent studies have significantly expanded the repertoire of linear and cyclic oligonucleotides that serve as essential signaling messengers in bacteria and animals [17, 18], suggesting that some nucleoside derivatives with regulatory properties may remain undiscovered in plants.

A critic might point out that bioinformatic evidence of the enzymatic function of FT and their homologs presented in this manuscript is too circumstantial – some would say, strange and/or speculative – and that it has not been corroborated by direct wet-laboratory experiments. However, exactly the same can be said about the experimental wet-lab support for the role of FT as a transcriptional co-activator – a hypothesis that is based on several lines of suggestive evidence, but is unmoored from the hard 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 [10, 20, 46], it seems timely and urgent to test whether the FT homologs in various organisms may have an enzymatic activity.

One experimental approach would be to compare small-molecule metabolite profiles in the *ybhB* knockout mutants and identify compounds that are in a shortage or excess in the mutant compared to the wild type. Bacterial model systems may be perhaps most amenable to such an analysis, as *ybhB* homologs tend to be single-copy in bacteria and are dispensable for growth in laboratory culture in most tested strains. Metabolomics-based identification of putative products or substrates of YbhB homologs would be the first step in determining the putative enzymatic activity that we predict to be conserved in most members of this protein family, and may lead the way to the identification of the small-molecule component of the plant flowering hormone.

Abbreviations
COG: Clusters of Orthologous Groups; GSDH: glucose/sorbosone dehydrogenase; PC: phosphatidylcholine; PEBP: phosphatidylethanolamine-binding protein; RKIP: Raf-1 kinase-inhibiting protein; SAM: shoot apical meristems.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12870-022-03432-z.

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

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

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

Declarations

Ethics approval and consent to participate
N/A

Consent for publication
N/A

Competing interests
The authors declare that they do not have competing interests. AM is a program director at the National Science Foundation, an agency of the U.S. Government; 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.

Author details
[1]Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich (TUM), 3, Maximus-von-Imhof-Forum, 85354 Freising, Germany. [2]Current address: Chair of Computational Systems Biology, University of Hamburg, Notkestrasse, 9, 22607 Hamburg, Germany. [3]Molecular and Cellular Biology Division, National Science Foundation, 2415 Eisenhower Avenue, Alexandria, Virginia 22314, USA. [4]Clare Hall College, University of Cambridge, Cambridge CB3 9AL, UK.

References

1. Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science. 2005;309:1052–6. https://doi.org/10.1126/science.1115983. PMID: 16099979.

2. Abe M, Kosaka S, Shibuta M, Nagata K, Uemura T, Nakano A, Kaya H. Transient activity of the florigen complex during the floral transition in *Arabidopsis thaliana*. Development. 2019;146;dev171504. https://doi.org/10.1242/dev.171504. PMID: 30940631.

3. Abelenda JA, Bergonzi S, Oortwijn M, Sonnewald S, Du M, Visser RGF, Sonnewald U, Bachem CWB. Source-sink regulation is mediated by interaction of an FT homolog with a SWEET protein in potato. Current Biol. 2019;29:1178-1186.e6. https://doi.org/10.1016/j.cub.2019.02.018. Epub 2019 Mar 21. PMID: 30905604.

4. Ahn JH, Miller D, Winter VJ, Banfield MJ, Lee JH, Yoo SY, Henz SR, Brady RL, Weigel D. A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. EMBO J. 2006;25:605–14. https://doi.org/10.1038/sj.emboj.7600950. Epub 2006 Jan 19. PMID: 16424903; PMCID: PMC1383534.

5. Altschul SF, Madden TI, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402. https://doi.org/10.1093/nar/25.17.3389. PMID: 9254694; PMCID: PMC146917.

6. Andrés F, Coupland G. The genetic basis of flowering responses to seasonal cues. Nat Rev Genet. 2012;13:627–39. https://doi.org/10.1038/nrg3291. PMID: 22898651.

7. Andrés F, Kinoshita A, Kalluri N, Fernández V, Falavigna VS, Cruz TMD, Jang S, Chiba Y, Seo M, Mettler-Altmann T, Huettel B, Coupland G. The sugar transporter SWEET10 acts downstream of FLOWERING LOCUS T during floral transition of *Arabidopsis thaliana*. BMC Plant Biol. 2020;20:53. https://doi.org/10.1186/s12870-020-2266-0. PMID: 32013867; PMCID: PMC6998834.

8. Arakaki T, Neely H, Boni E, Mueller N, Buckner FS, Van Voorhis WC, Lauricella A, DeTitta G, Luft J, Hol WG, Merritt EA. The structure of *Plasmodium vivax* phosphatidylethanolamine-binding protein suggests a functional motif containing a left-handed helix. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2007;63:178–82. https://doi.org/10.1107/S1744309107007580. PMID: 17329808; PMCID: PMC2330187.

9. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol. 2004;14:283–91. https://doi.org/10.1016/j.sbi.2004.05.004. PMID: 15193307.

10. Baker PA, Fritz SC, Dick CW, Eckert AJ, Horton BK, Manzoni S, et al. The emerging field of geogenomics: constraining geological problems with genetic data. Earth Sci Rev. 2014;135:38–47. https://doi.org/10.1016/j.earscirev.2014.04.001.

11. Banfield MJ, Barker JJ, Perry AC, Brady RL. Function from structure? The crystal structure of human phosphatidylethanolamine-binding protein suggests a role in membrane signal transduction. Structure. 1998;6:1245–54. https://doi.org/10.1016/s0969-2126(98)00125-7. PMID: 9782050.

12. Banfield MJ, Brady RL. The structure of *Antirrhinum* centroradialis protein (CEN) suggests a role as a kinase regulator. J Mol Biol. 2000;297:1159–70. https://doi.org/10.1006/jmbi.2000.3619. PMID: 10764580.

13. Baumann K, Venail J, Berbel A, Domenech MJ, Money T, Conti L, Hanzawa Y, Madueno F, Bradley D. Changing the spatial pattern of TFL1 expression reveals its key role in the shoot meristem in controlling Arabidopsis flowering architecture. J Exp Bot. 2015;66:4769–80. https://doi.org/10.1093/jxb/erv247. PMID: 26019254; PMCID: PMC4507777.

14. Bennett T, Dixon LE. Asymmetric expansions of FT and TFL1 lineages characterize differential evolution of the EuPEBP family in the major angiosperm lineages. BMC Biol. 2021;19:181. https://doi.org/10.1186/s12915-021-01128-8. PMID: 34465318; PMCID: PMC8408984.

15. Bernier I, Jollès P. Purification and characterization of a basic 23 kDa cytosolic protein from bovine brain. Biochim Biophys Acta. 1984;790:174–81. https://doi.org/10.1016/0167-4838(84)90221-8. PMID: 6435678.

16. Bruun AW, Svendsen I, Sørensen SO, Kielland-Brandt MC, Winther JR. A high-affinity inhibitor of yeast carboxypeptidase Y is encoded by TFS1 and shows homology to a family of lipid binding proteins. Biochemistry. 1998;37(10):3351–7. https://doi.org/10.1021/bi971286w. PMID: 9521655.

17. Burroughs AM, Aravind L. Identification of uncharacterized components of prokaryotic immune systems and their diverse eukaryotic reformulations. J Bacteriol. 2020;202:e00365–20. https://doi.org/10.1128/JB.00365-20. PMID: 32868406; PMCID: PMC7685563.

18. Burroughs AM, Zhang D, Schäffer DE, Iyer LM, Aravind L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res. 2015;43:10633–54. https://doi.org/10.1093/nar/gkv1267. Epub 2015 Nov 20. PMID: 26590262; PMCID: PMC4678834.

19. Cambier S, Ginis O, Moreau SJM, Gayral P, Hearn J, Stone GN, Giron D, Huguet E, Drezen JM. Gall wasp transcriptomes unravel potential effectors involved in molecular dialogues with oak and rose. Front Physiol. 2019;10:926. https://doi.org/10.3389/fphys.2019.00926. PMID: 31396099; PMCID: PMC6667641.

20. Carr SM, Marshall HD, Wareham T, Craig D. The big ORF theory. In emerging trends in computational biology, bioinformatics, and systems biology (Morgan Kaufmann) 2015. pp. 265–274. eBook ISBN 9780128026465.

21. Chailakhyan M. New facts supporting the hormonal theory of plant development. Dokl. Akad. Nauk SSSR. 1936;4:77–81.

22. Chardon F, Damerval C. Phylogenomic analysis of the PEBP gene family in cereals. J Mol Evol. 2005;61:579–90. https://doi.org/10.1007/s00239-004-0179-4. PMID: 16170456.

23. Cheng H, Liao Y, Schaeffer RD, Grishin NV. Manual classification strategies in the ECOD database. Proteins. 2015;83:1238–51. https://doi.org/10.1002/prot.24818. Epub 2015 May 8. PMID: 25917548; PMCID: PMC4624060.

24. Conti L, Bradley D. TERMINAL FLOWER1 Is a mobile signal controlling *Arabidopsis* architecture. Plant Cell. 2007;19:767–78. https://doi.org/10.1105/tpc.106.049767. Epub 2007 Mar 16.

25. Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, et al. FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science. 2007;316:1030–3.https://doi.org/10.1126/science.1141752. Epub 2007 Apr 19. PMID: 17446353.

26. Danilevskaya ON, Meng X, McGonigle B, Muszynski MG. Beyond flowering time: pleiotropic function of the maize flowering hormone florigen. Plant Signal Behav. 2011;6:1267–70. https://doi.org/10.4161/psb.6.9.16423. PMID: 21847027; PMCID: PMC3258048.

27. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32;1792–7. https://doi.org/10.1093/nar/gkh340. PMID: 15034147; PMCID: PMC390337.

28. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47:D427–D432. https://doi.org/10.1093/nar/gky995. PMID: 30357350; PMCID: PMC6324024.

29. Eulenburg G, Higman VA, Diehl A, Wilmanns M, Holton SJ. Structural and biochemical characterization of Rv2140c, a phosphatidylethanolamine-binding protein from *mycobacterium tuberculosis*. FEBS Lett. 2013;587:2936–42. https://doi.org/10.1016/j.febslet.2013.07.038. Epub 2013 Jul 29. PMID: 23907008.

30. Fichtner F, Barbier FF, Annunziata MG, Feil R, Olas JJ, Mueller-Roeber B, Stitt M, Beveridge CA, Lunn JE. Regulation of shoot branching in arabidopsis by trehalose 6-phosphate. New Phytol. 2021;229:2135–51. https://doi.org/10.1111/nph.17006. Epub 2020 Nov 25. PMID: 33068448.

31. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. 2021;49:D274–D281. https://doi.org/10.1093/nar/gkaa1018. PMID: 33167031; PMCID: PMC7778934.

32. Garner WW, Allard HA. Effect of the relative length of day and night and other factors of the environment on growth and reporduction in plants. Mon Weather Rev. 1920;48:415–5.

33. Glazko, G.V., and Mushegian, A.R. Detection of evolutionarily stable fragments of cellular pathways by hierarchical clustering of phyletic patterns. Genome Biol. 2004;5: R32. https://doi.org/10.1186/gb-2004-5-5-r32. Epub 2004 Apr 27. PMID: 15128446; PMCID: PMC416468.

34. Glazko, G., Gordon, A., and Mushegian, A. The choice of optimal distance measure in genome-wide datasets. Bioinformatics 2005;21 Suppl 3: iii3-11. https://doi.org/10.1093/bioinformatics/bti1201. PMID: 16306389.

35. Glazko, G., Coleman, M., and Mushegian, A. Similarity searches in genome-wide numerical data sets. Biol. Direct 2006;1: 13. https://doi.org/10.1186/1745-6150-1-13. PMID: 16734895; PMCID: PMC1489924.

36. Gombault, A., Warringer, J., Caesar, R., Godin, F., Vallée, B., Doudeau, M., Chautard, H., Blomberg, A., and Bénédetti, H. A phenotypic study of TFS1 mutants differentially altered in the inhibition of Ira2p or CPY. FEMS Yeast Res 2009;9: 867–874. https://doi.org/10.1111/j.1567-1364.2009.00535.x. Epub 2009 May 30. PMID: 19552705.

37. Granovsky, A.E., Clark, M.C., McElheny, D., Heil, G., Hong, J., Liu, X., Kim, Y., Joachimiak, G., Joachimiak, A., Koide, S., and Rosner MR. Raf kinase inhibitory protein function is regulated via a flexible pocket and novel phosphorylation-dependent mechanism. Mol Cell Biol. 2009;29: 1306–1320. https://doi.org/10.1128/MCB.01271-08. PMID: 19103740; PMCID: PMC2643833.

38. Guindon, S., and Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2003;52: 696–704. https://doi.org/10.1080/10635150390235520. PMID: 14530136.

39. Guo, C., Chang, T., Sun, T., Wu, Z., Dai, Y., Yao, H., and Lin, D. Anti-leprosy drug Clofazimine binds to human Raf1 kinase inhibitory protein and enhances ERK phosphorylation. Acta Biochim Biophys Sin Shanghai 2018;50: 1062–1067. https://doi.org/10.1093/abbs/gmy095. PMID: 30137201.

40. He, H., Liu, D., Lin, H., Jiang, S., Ying, Y., Chun, S., Deng, H., Zaia, J., Wen, R., and Luo, Z. Phosphatidylethanolamine binding protein 4 (PEBP4) is a secreted protein and has multiple functions. Biochim Biophys Acta. 2016;1863: 1682–1689. https://doi.org/10.1016/j.bbamcr.2016.03.022. Epub 2016 Mar 28. PMID: 27033522; PMCID: PMC5336313.

41. Hedman, H., Källman, T., and Lagercrantz, U. Early evolution of the MFT-like gene family in plants. Plant Mol Biol 2009;70: 359–369. https://doi.org/10.1007/s11103-009-9478-x. Epub 2009 Mar 14. PMID: 19288213.

42. Ho WW, Weigel D. Structural features determining flower-promoting activity of Arabidopsis FLOWERING LOCUS T. Plant Cell. 2014;26:552–64 https://doi.org/10.1105/tpc.113.115220.

43. 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. Ensembl Genomes 2020-enabling non-vertebrate genomic research. Nucleic Acids Res 2020;48: D689–D695. https://doi.org/10.1093/nar/gkz890. PMID: 31598706; PMCID: PMC6943047.

44. Huerta, A.M., and Collado-Vides, J. Sigma70 promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. J Mol Biol 2003;333: 261–278. https://doi.org/10.1016/j.jmb.2003.07.017. PMID: 14529615.

45. Huynen, M., Snel, B., Lathe, W., and Bork, P. Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome Res 2000;10: 1204–1210. https://doi.org/10.1101/gr.10.8.1204. PMID: 10958638; PMCID: PMC310926.

46. Iyer, L.M., Aravind, L., Bork, P., Hofmann, K., Mushegian, A.R., Zhulin, I.B., and Koonin, E.V. Quod erat demonstrandum? The mystery of experimental validation of apparently erroneous computational analyses of protein sequences. Genome Biol. 2001;2: RESEARCH0051. https://doi.org/10.1186/gb-2001-2-12-research0051. Epub 2001 Nov 13. PMID: 11790254; PMCID: PMC64836.

47. Izawa, T. What is going on with the hormonal control of flowering in plants? Plant J 2021;105:431–445. https://doi.org/10.1111/tpj.15036. Epub 2020 Dec 2. PMID: 33111430.

48. Jennings, W., and Epand, R.M. CDP-diacylglycerol, a critical intermediate in lipid metabolism. Chem Phys Lipids 2020;230: 104914. https://doi.org/10.1016/j.chemphyslip.2020.104914. Epub 2020 Apr 28. PMID: 32360136.

49. Jin, S., Nasim, Z., Susila, H., and Ahn, J.H. Evolution and functional diversification of FLOWERING LOCUS T/TERMINAL FLOWER 1 family genes in plants. Semin Cell Dev Biol 2021;109: 20–30. https://doi.org/10.1016/j.semcdb.2020.05.007. PMID: 32507412.

50. Kakimoto, T. Biosynthesis of cytokinins. J Plant Res 2003;116: 233–239. https://doi.org/10.1007/s10265-003-0095-5. Epub 2003 Apr 29. PMID: 12721785.

51. Karlgren, A., Gyllenstrand, N., Källman, T., Sundström, J.F., Moore, D., Lascoux, M., and Lagercrantz, U. Evolution of the PEBP gene family in plants: functional diversification in seed plant evolution. Plant Physiol. 2011;156: 1967–1977. https://doi.org/10.1104/pp.111.176206. Epub 2011 Jun 3. PMID: 21642442; PMCID: PMC3149940.

52. Khosa, J., Bellinazzo, F., Kamenetsky Goldstein, R., Macknight, R., and Immink, R.G.H. PHOSPHATIDYLETHANOLAMINE-BINDING PROTEINS: the conductors of dual reproduction in plants with vegetative storage organs. J Exp Bot 2021;72: 2845–2856. https://doi.org/10.1093/jxb/erab064. PMID: 33606013.

53. Kimura, S., Sakai, Y., Ishiguro, K., and Suzuki, T. Biogenesis and iron-dependency of ribosomal RNA hydroxylation. Nucleic Acids Research 2017;45: 12974–12986. https://doi.org/10.1093/nar/gkx969. PMID: 29069499; PMCID: PMC5727448.

54. King RW, Evans LT, Wardlaw IF. Translocation of the floral stimulus in *Pharbitis nil* in relation to that of assimilates. Zeitschrift Fur Pflanzenphysiologie. 1968;59:S377–88.

55. Kinoshita, T., Ono, N., Hayashi, Y., Morimoto, S., Nakamura, S., Soda, M., Kato, Y., Ohnishi, M., Nakano, T., Inoue, S., and Shimazaki, K. FLOWERING LOCUS T regulates stomatal opening. Curr Biol 2011;21: 1232–1238. https://doi.org/10.1016/j.cub.2011.06.025. PMID: 21737277.

56. Kinoshita, A., and Richter, R. Genetic and molecular basis of floral induction in *Arabidopsis thaliana*. J Experimental Botany 2020;71: 2490–2504. https://doi.org/10.1093/jxb/eraa057. PMID: 32067033; PMCID: PMC7210760.

57. Klintenäs, M., Pin, P.A., Benlloch, R., Ingvarsson, P.K., and Nilsson, O. Analysis of conifer FLOWERING LOCUS T/TERMINAL FLOWER1-like genes provides evidence for dramatic biochemical evolution in the angiosperm FT lineage. New Phytol 2012;196: 1260–1273. https://doi.org/10.1111/j.1469-8137.2012.04332.x. Epub 2012 Sep 28. PMID: 23020222.

58. Knott, J.E. Effect of a localized photoperiod on spinach. In Proc Amer Soc Hortic Sci. 1934; 152–154.

59. Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., and Araki, T. A pair of related genes with antagonistic roles in mediating flowering signals. Science 1999;286: 1960–1962. https://doi.org/10.1126/science.286.5446.1960. PMID: 10583960.

60. Kobayashi, Y., and Weigel, D. Move on up, it's time for change--mobile signals controlling photoperiod-dependent flowering. Genes Dev 2007;21: 2371–2384. https://doi.org/10.1101/gad.1589007. PMID: 17908925.

61. Koornneef M, Hanhart CJ, van der Veen JH. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. Mol Gen Genet. 1991;229:57–66. 1896021. https://doi.org/10.1007/BF00264213.

62. Krieger, U., Lippman, Z.B., and Zamir, D. The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nat Genet 2010;42: 459–463. https://doi.org/10.1038/ng.550. Epub 2010 Mar 28. PMID: 20348958.

63. Lang, A., Chailakhyan, M.K., and Frolova, I.A. 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. 1977;74: 2412–2416. https://doi.org/10.1073/pnas.74.6.2412. PMID: 16592404; PMCID: PMC432182.

64. Lau, M.E., Danka, E.S., Tiemann, K.M., and Hunstad, D.A. Bacterial lysis liberates the neutrophil migration suppressor YbcL from the periplasm of uropathogenic *Escherichia coli*. Infect. Immun. 2014;82: 4921–4930. https://doi.org/10.1128/IAI.01838-14. Epub 2014 Sep 2. PMID: 25183735; PMCID: PMC4249291.

65. Lemoine, F., Domelevo Entfellner, J.B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., and Gascuel, O. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature. 2018;556: 452–456. https://doi.org/10.1038/s41586-018-0043-0. PMID: 29670290; PMCID: PMC6030568.

66. Li, W., Ju, J., Rajski, S.R., Osada, H., and Shen, B. Characterization of the tautomycin biosynthetic gene cluster from *Streptomyces spiroverticillatus* unveiling new insights into dialkylmaleic anhydride and polyketide biosynthesis. J. Biol. Chem. 2008;283: 28607–28617. https://doi.org/10.1074/jbc.M804279200. Epub 2008 Aug 15. PMID: 18708355; PMCID: PMC2568922.

67. Li, W., Luo, Y., Ju, J., Rajski, S.R., Osada, H., and Shen, B. Characterization of the tautomycetin biosynthetic gene cluster from *Streptomyces griseochromogenes* provides new insight into dialkylmaleic anhydride biosynthesis. J. Nat. Prod. 2009;72: 450–459. https://doi.org/10.1021/np8007478. PMID: 19191560; PMCID: PMC2967020.

68. Lifschitz, E., and Eshed, Y. Universal florigenic signals triggered by FT homologues regulate growth and flowering cycles in perennial day-neutral tomato. J Exp Bot 2006;57: 3405–3414. https://doi.org/10.1093/jxb/erl106. Epub 2006 Sep 27. PMID: 17005925.

69. Lifschitz, E., Ayre, B.G., and Eshed, Y. Florigen and anti-florigen – a systemic mechanism for coordinating growth and termination in flowering plants. Front Plant Sci. 2014;5: 465. https://doi.org/10.3389/fpls.2014.00465. PMID: 25278944; PMCID: PMC4165217.

70. Liu, Y.Y., Yang, K.Z., Wei, X.X., Wang, X.Q. Revisiting the phosphatidylethanolamine-binding protein (PEBP) gene family reveals cryptic FLOWERING LOCUS T gene homologs in gymnosperms and sheds new light on functional evolution. New Phytol 2016;212: 730–744. https://doi.org/10.1111/nph.14066. Epub 2016 Jul 4. PMID: 27375201.

71. Luo H., Lin, Y., Gao F., Zhang C.-T., and Zhang, R. DEG 10, an update of the Database of Essential Genes that includes both protein-coding genes and non-coding genomic elements. Nucleic Acids Research 2014;42: D574-D580. https://doi.org/10.1093/nar/gkt1131. Epub 2013 Nov 15. PMID: 24243843; PMCID: PMC3965060.

72. Mareuil, F., Doppelt-Azeroual, O. and Ménager, H. A public galaxy platform at Pasteur used as an execution engine for web services [version 1; not peer reviewed]. F1000Research 2017;6: 1030 (poster) https://doi.org/10.7490/f1000research.1114334.1.

73. Matsoukas, I.G., Massiah, A.J., and Thomas, B. Florigenic and antiflorigenic signaling in plants. Plant Cell Physiol 2012;53: 1827–1842. https://doi.org/10.1093/pcp/pcs130. PMID: 23008422.

74. McLennan, A.G. The Nudix hydrolase superfamily. Cell Mol Life Sci 2006;63: 123–143. https://doi.org/10.1007/s00018-005-5386-7. PMID: 16378245.

75. Medina R, Johnson MG, Liu Y, Wickett NJ, Shaw AJ, Goffinet B. Phylogenomic delineation of *Physcomitrium* (Bryophyta: Funariaceae) based on targeted sequencing of nuclear exons and their flanking regions rejects the retention of *Physcomitrella*, *Physcomitridium* and *Aphanorrhegma*. Jnl of Systematics Evolution. 2019;57:404–17. https://doi.org/10.1111/jse.12516.

76. Milyaeva, E.L., Golyanovskaya, S.A., Aksenova, N.P., and Chailakhyan, M.Kh. . Changes of protein composition in leaves and stem apices of bicolored coneflower with transition to flowering. Dokl. Akad. Nauk SSSR (bot. Sciences) 1991;11–15.

77. Mima J, Hayashida M, Fujii T, Narita Y, Hayashi R, Ueda M, et al. Structure of the carboxypeptidase Y inhibitor IC in complex with the cognate proteinase reveals a novel mode of the proteinase-protein inhibitor interaction. J Mol Biol. 2005;346:1323–34. https://doi.org/10.1016/j.jmb.2004.12.051 Epub 2005 Jan 25. PMID: 15713484.

78. Mironov AA, Vinokurova NP, Gelfand MS. Software for analysis of bacterial genomes. Mol Biol. 2000;34:222–31.

79. Miyazaki, T., Sugisawa, T., and Hoshino, T. Pyrroloquinoline quinone-dependent dehydrogenases from *Ketogulonicigenium vulgare* catalyze the direct conversion of L-sorbosone to L-ascorbic acid. AEM 2006;72: 1487–1495. https://doi.org/10.1128/AEM.72.2.1487-1495.2006. PMID: 16461703; PMCID: PMC1392885.

80. Musfeldt, M., and Schönheit, P. 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*. J Bacteriology 2002;184: 636–644. https://doi.org/10.1128/JB.184.3.636-644.2002. PMID: 11790732; PMCID: PMC139507.

81. Nakamura, Y., Andrés, F., Kanehara, K., Liu, Y., Dörmann, P., and Coupland, G. Arabidopsis florigen FT binds to diurnally oscillating phospholipids that accelerate flowering. Nat Commun 2014a;5: 3553. https://doi.org/10.1038/ncomms4553. PMID: 24698997; PMCID: PMC3988816.

82. Nakamura, Y., Andrés, F., Kanehara, K., Liu, Y., Coupland, G., and Dörmann, P. Diurnal and circadian expression profiles of glycerolipid biosynthetic genes in Arabidopsis. Plant Signal Behav 2014b;9: e29715. https://doi.org/10.4161/psb.29715. PMID: 25763705; PMCID: PMC4205134.

83. Nakamura, Y., Lin, Y.-C., Watanabe, S., Liu, Y.-C., Katsuyama, K., Kanehara, K., and Inaba, K. High-resolution crystal structure of Arabidopsis FLOWERING LOCUS T illuminates its phospholipid-binding site in flowering. IScience 2019;21: 577–586. https://doi.org/10.1016/j.isci.2019.10.045. Epub 2019 Oct 26. PMID: 31726375; PMCID: PMC6854094.

84. Penel C, Gaspae T, Greppin H. Rapid interorgan communications in higher plants with special reference to flowering. Biol Plant. 1985;27:334–8. https://doi.org/10.1007/BF02879873.

85. Persson, E., Castresana-Aguirre, M., Buzzao, D., Guala, D., and Sonnhammer, E.L.L. FunCoup 5: functional association networks in all domains of life, supporting directed links and tissue-specificity. J Mol Biol 2021;433: 166835. https://doi.org/10.1016/j.jmb.2021.166835. Epub 2021 Feb 2. PMID: 33539890.

86. Pikielny, C.W., Hasan, G., Rouyer, F., and Rosbash, M. Members of a family of Drosophila putative odorant-binding proteins are expressed in different subsets of olfactory hairs. Neuron. 1994;12: 35–49. https://doi.org/10.1016/0896-6273(94)90150-3. PMID: 7545907.

87. Pin, P.A., and Nilsson, O. The multifaceted roles of FLOWERING LOCUS T in plant development: FT, a multifunctional protein. Plant Cell Environ 2012;35: 1742–1755. https://doi.org/10.1111/j.1365-3040.2012.02558.x. Epub 2012 Jul 15. PMID: 22697796.

88. Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Liu, H., Kuehl, J.V., Melnyk, R.A., Lamson, J.S., Suh, Y., et al. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature. 2018;557: 503–509. https://doi.org/10.1038/s41586-018-0124-0. Epub 2018 May 16. PMID: 29769716.

89. Ratcliffe OJ, Amaya I, Vincent CA Rothstein S, Carpenter R, Coen ES, Bradley DJ. A common mechanism controls the life cycle and architecture of plants. Development. 1998;125: 1609–1615. https://doi.org/10.1242/dev.125.9.1609 PMID: 9521899.

90. Rautureau, G.J., Vovelle, F., Schoentgen, F., Decoville, M., Locker, D., Damblon, C., and Jouvensal, L. NMR structure of a phosphatidyl-ethanolamine binding protein from *Drosophila*. Proteins. 2010;78: 1606–1610. https://doi.org/10.1002/prot.22682. PMID: 20131378.

91. Ribeiro, A.J.M., Holliday, G.L., Furnham, N., Tyzack, J.D., Ferris, K., and Thornton, J.M. Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. Nucleic Acids Res 2018;46: D618–D623. https://doi.org/10.1093/nar/gkx1012. PMID: 29106569; PMCID: PMC5753290.

92. Roderick, S.L., Chan, W.W., Agate, D.S., Olsen, L.R., Vetting, M.W., Rajashankar, K.R., and Cohen, D.E. Structure of human phosphatidylcholine transfer protein in complex with its ligand. Nat Struct Biol 2002;9: 507–511. https://doi.org/10.1038/nsb812. PMID: 12055623.

93. Romanov GA. Mikhail Khristoforovich Chailakhyan: the fate of the scientist under the sign of florigen. Russ J Plant Physiol. 2012;59:443–50. https://doi.org/10.1134/S1021443712040103.

94. Rudnitskaya, A.N., Eddy, N.A., Fenteany, G., and Gascón, J.A. Recognition and reactivity in the binding between Raf kinase inhibitor protein and its small-molecule inhibitor locostatin. J Phys Chem B 2012;116: 10176–10181. https://doi.org/10.1021/jp303140j. PMID: 22861375.

95. Sachs J. Wirkung des Lichtes auf die Blütenbildung unter Vermittlung der Laubblätter. Bot Ztg. 1865;23:117.

96. Sadreyev, I.R., Ji, F., Cohen, E., Ruvkun, G., and Tabach, Y. PhyloGene server for identification and visualization of co-evolving proteins using normalized phylogenetic profiles. Nucleic Acids Res 2015;43: W154–159. https://doi.org/10.1093/nar/gkv452. PMID: 25958392; PMCID: PMC4489270.

97. Salazar-Cerezo, S., Martínez-Montiel, N., García-Sánchez, J., Pérez-Y-Terrón, R., and Martínez-Contreras, R.D. Gibberellin biosynthesis and metabolism: a convergent route for plants, fungi and bacteria. Microbiol Res 2018;208: 85–98. https://doi.org/10.1016/j.micres.2018.01.010. Epub 2018 Feb 3. PMID: 29551215.

98. Schieffer, G.M. Extending the frontiers of mass spectrometric instrumentation and methods. Graduate theses and dissertations. 2010;11528. https://lib.dr.iastate.edu/etd/11528. https://doi.org/10.31274/etd-180810-155.

99. Serre, L., Vallée, B., Bureaud, N., Schoentgen, F., and Zelwer, C. Crystal structure of the phosphatidylethanolamine-binding protein from bovine brain: a novel structural class of phospholipid-binding proteins. Structure. 1998;6: 1255–1265. https://doi.org/10.1016/s0969-2126(98)00126-9. PMID: 9782057.

100. Serre L, Pereira de Jesus K, Zelwer C, Bureaud N, Schoentgen F, Bénédetti H. Crystal structures of YBHB and YBCL from *Escherichia coli*, two bacterial homologues to a Raf kinase inhibitor protein. J Mol Biol. 2001;310:617–34. https://doi.org/10.1006/jmbi.2001.4784. PMID: 11439028.

101. Shemon AN, Heil GL, Granovsky AE, Clark MM, McElheny D, Chimon A, Rosner MR, Koide S. Characterization of the Raf kinase inhibitory protein (RKIP) binding pocket: NMR-based screening identifies small-molecule ligands. PLoS ONE. 2010;5:e10479. https://doi.org/10.1371/journal.pone.0010479. PMID: 20463977; PMCID: PMC2864760.

102. Sheng X, Zhao Z, Wang J, Yu H, Shen Y, Gu H. Identification of *Brassica oleracea* orthologs of the PEBP family and their expression patterns in curd development and flowering in cauliflower. Biotechnology & Biotechnological Equipment. 2020;34:605–13. https://doi.org/10.1080/13102818.2020.1790418.

103. Sohn EJ, Rojas-Pierce M, Pan S, Carter C, Serrano-Mislata A, Madueño F, Rojo E, Surpin M, Raikhel NV. The shoot meristem identity gene TFL1 is involved in flower development and trafficking to the protein storage vacuole. Proc Natl Acad Sci U S A. 2007;104:18801–6. https://doi.org/10.1073/pnas.0708236104. PMID: 18003908; PMCID: PMC2141857.

104. Stark A, Russell RB. Annotation in three dimensions. PINTS: Patterns in Non-homologous Tertiary Structures. Nucleic Acids Res. 2003;31:3341–4. https://doi.org/10.1093/nar/gkg506. PMID: 12824322; PMCID: PMC168913.

105. Suhre K, Claverie JM. FusionDB: a database for in-depth analysis of prokaryotic gene fusion events. Nucleic Acids Res. 2004;32:D273–276. https://doi.org/10.1093/nar/gkh053. PMID: 14681411; PMCID: PMC308787.

106. Susila H, Nasim Z, Gawarecka K, Jung JY, Jin S, Youn G, Ahn JH. PHOSPHORYLETHANOLAMINE CYTIDYLYLTRANSFERASE 1 modulates flowering in a florigen-independent manner by regulating *SVP*. Development. 2021;148:dev193870. https://doi.org/10.1242/dev.193870. PMID: 33268452.

107. Takeba G, Nakajima Y, Kozaki A, Tanaka O, Kasai Z. A flower-inducing substance of high molecular weight from higher plants. Plant Physiol. 1990;94:1677–81. https://doi.org/10.1104/pp.94.4.1677. PMID: 16667901; PMCID: PMC1077437.

108. Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K. Hd3a protein is a mobile flowering signal in rice. Science. 2007;316:1033–6. https://doi.org/10.1126/science.1141753. Epub 2007 Apr 19. PMID: 17446351.

109. Taoka K, Ohki I, Tsuji H, Furuita K, Hayashi K, Yanase T, Yamaguchi M, Nakashima C, Purwestri YA, Tamaki S, et al. 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. Nature. 2011;476:332–5. https://doi.org/10.1038/nature10272. PMID: 21804566.

110. Tavel L, Jaquillard L, Karsisiotis AI, Saab F, Jouvensal L, Brans A, Delmas AF, Schoentgen F, Cadene M, Damblon C. Ligand binding study of human PEBP1/RKIP: interaction with nucleotides and Raf-1 peptides evidenced by NMR and mass spectrometry. PLoS One. 2012;7:e36187. https://doi.org/10.1371/journal.pone.0036187. Epub 2012 Apr 27. PMID: 22558375; PMCID: PMC3338619.

111. Thao S, Escalante-Semerena JC. A positive selection approach identifies residues important for folding of *Salmonella enterica* Pat, an Nε-lysine acetyltransferase that regulates central metabolism enzymes. Res Microbiol. 2012;163:427–35. https://doi.org/10.1016/j.resmic.2012.05.008. Epub 2012 Jun 4. PMID: 22677774; PMCID: PMC3432723.

112. Tsuji H, Taoka K, Shimamoto K. Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. Curr Opin Plant Biol. 2011;14:45–52. https://doi.org/10.1016/j.pbi.2010.08.016. Epub 2010 Sep 21. PMID: 20864385.

113. van Tran N, Muller L, Ross RL, Lestini R, Létoquart J, Ulryck N, Limbach PA, de Crécy-Lagard V, Cianférani S, Graille M. Evolutionary insights into Trm112-methyltransferase holoenzymes involved in translation between archaea and eukaryotes. Nucleic Acids Res. 2018;46:8483–99. https://doi.org/10.1093/nar/gky638. PMID: 30010922; PMCID: PMC6144793.

114. Wahl V, Ponnu J, Schlereth A, Arrivault S, Langenecker T, Franke A, Feil R, Lunn JE, Stitt M, Schmid M. Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. Science. 2013;339:704–7. https://doi.org/10.1126/science.1230406. PMID: 23393265.

115. Wenzel SE, Tyurina YY, Zhao J, St Croix CM, Dar HH, Mao G, Tyurin VA, Anthonymuthu TS, Kapralov AA, Amoscato AA, et al. PEBP1 wardens ferroptosis by enabling lipoxygenase generation of lipid death signals. Cell. 2017;171:628–641.e26. https://doi.org/10.1016/j.cell.2017.09.044. PMID: 29053969; PMCID: PMC5683852.

116. Wheeler TJ, Clements J, Finn RD. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinform. 2014;15:7. https://doi.org/10.1186/1471-2105-15-7. PMID: 24410852; PMCID: PMC3893531.

117. Wickland DP, Hanzawa Y. The FLOWERING LOCUS T/TERMINAL FLOWER 1 gene family: functional evolution and molecular mechanisms. Mol Plant. 2015;8:983–97. https://doi.org/10.1016/j.molp.2015.01.007. Epub 2015 Jan 15. PMID: 25598141.

118. Wigge PA. Integration of spatial and temporal information during floral induction in Arabidopsis. Science. 2005;309:1056–9. https://doi.org/10.1126/science.1114358. Erratum in: Science. 2006;312(5780):1600.

119. Yamaguchi A, Kobayashi Y, Goto K, Abe M, Araki T. TWIN SISTER OF FT (TSF) acts as a floral pathway integrator redundantly with FT. Plant Cell Physiol. 2005;46:1175–89. https://doi.org/10.1093/pcp/pci151. Epub 2005 Jun 11. PMID: 15951566.

120. Yant L, Mathieu J, Schmid M. Just say no: floral repressors help Arabidopsis bide the time. Curr Opin Plant Biol. 2009;12:580–6. https://doi.org/10.1016/j.pbi.2009.07.006. Epub 2009 Aug 18. PMID: 19695946.

121. Yeung K, Seitz T, Li S, Janosch P, McFerran B, Kaiser C, Fee F, Katsanakis KD, Rose DW, Mischak H, Sedivy JM, Kolch W. Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. Nature. 1999;401:173–7. https://doi.org/10.1038/43686. PMID: 10490027.

122. Zeevaart JAD. Physiology of flower formation. Annu Rev Plant Physiol. 1976;27:321–48. https://doi.org/10.1146/annurev.pp.27.060176.001541.

123. Zeevaart, J.A.D. Florigen coming of age after 70 years. Plant Cell. 2006;18:1783–9. https://doi.org/10.1105/tpc.106.043513. PMID: 16905662; PMCID: PMC1533981.

124. Zhu Y, Klasfeld S, Wagner D. Molecular regulation of plant developmental transitions and plant architecture via PEPB family proteins – an update on mechanism of action. J Exp Bot. 2021;72:2301–11. https://doi.org/10.1093/jxb/eraa598. PMID: 33449083.

Publisher's Note