Annual Review of Plant Biology

Plant GATA Factors: Their Biology, Phylogeny, and Phylogenomics

Claus Schwechheimer,[1] Peter Michael Schröder,[1] and Crysten E. Blaby-Haas[2,3]

[1]School of Life Sciences, Plant Systems Biology, Technical University of Munich, Freising, Germany; email: claus.schwechheimer@wzw.tum.de

[2]Biology Department, Brookhaven National Laboratory, Upton, New York, USA; email: cblaby@bnl.gov

[3]Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, New York, USA

Annu. Rev. Plant Biol. 2022. 73:123–48

First published as a Review in Advance on February 7, 2022

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

https://doi.org/10.1146/annurev-arplant-072221-092913

Keywords

Arabidopsis thaliana, evolution, GATA factors, phylogenomics, rice

Abstract

GATA factors are evolutionarily conserved transcription factors that are found in animals, fungi, and plants. Compared to that of animals, the size of the plant GATA family is increased. In angiosperms, four main GATA classes and seven structural subfamilies can be defined. In recent years, knowledge about the biological role and regulation of plant GATAs has substantially improved. Individual family members have been implicated in the regulation of photomorphogenic growth, chlorophyll biosynthesis, chloroplast development, photosynthesis, and stomata formation, as well as root, leaf, and flower development. In this review, we summarize the current knowledge of plant GATA factors. Using phylogenomic analysis, we trace the evolutionary origin of the GATA classes in the green lineage and examine their relationship to animal and fungal GATAs. Finally, we speculate about a possible conservation of GATA-regulated functions across the animal, fungal, and plant kingdoms.

Contents

1. GATA FACTORS

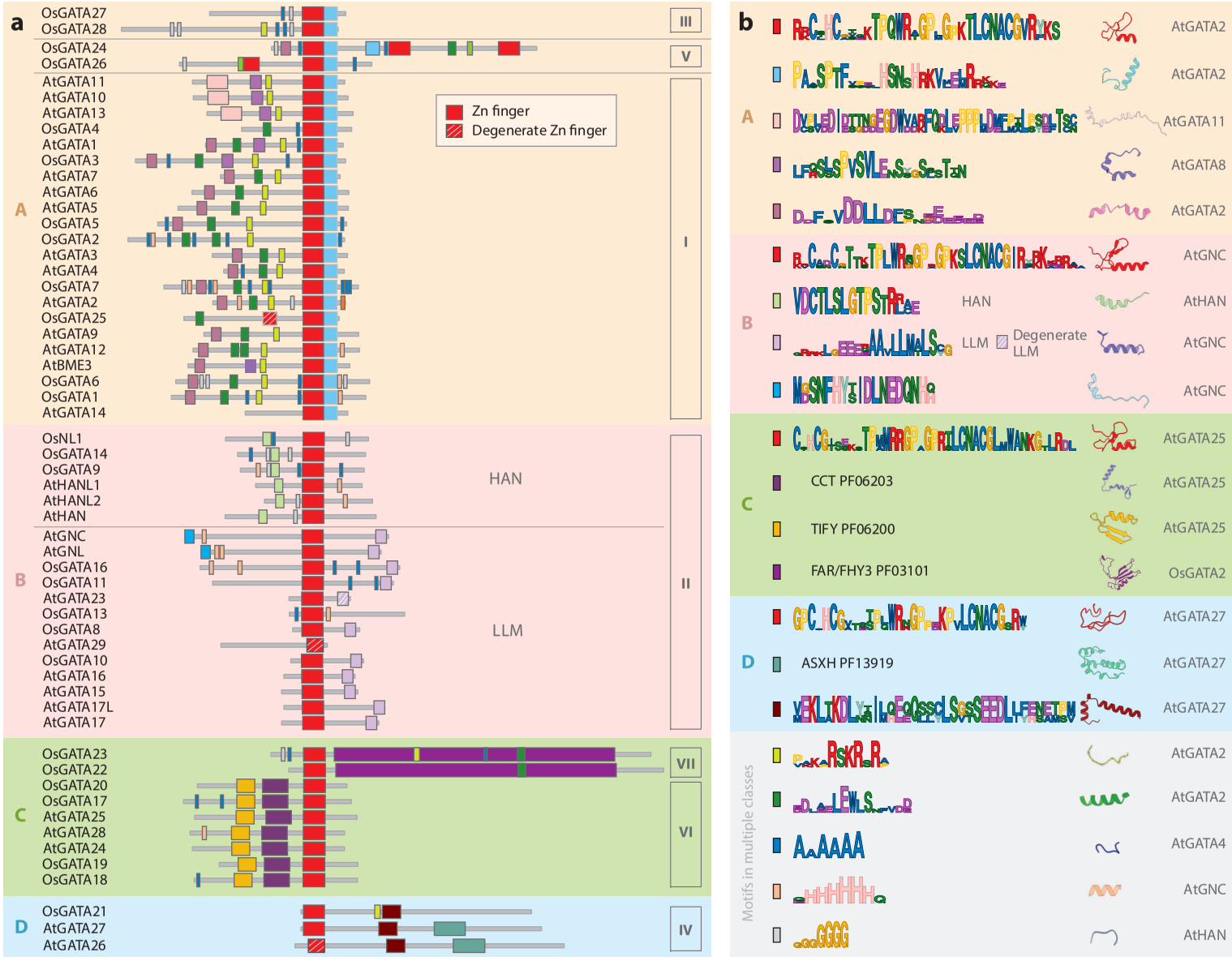
GATA factors are transcriptional regulators present in animals, plants, and fungi, which recognize the DNA sequence W-G-A-T-A-R through a single type IV zinc finger (79, 88). The GATA families from animals and yeasts are comparatively small. Six GATA transcription factors can be identified in human (60), five in *Drosophila melanogaster* (31), and ten in *Saccharomyces cerevisiae* (78). Human GATAs are required for essential developmental processes such as the differentiation of the haematopoietic and central nervous systems (HsGATA-1–HsGATA-3) or of embryonic stem cells and cardiovascular embryogenesis (HsGATA-4–HsGATA-6) (12, 54). Some vertebrate GATAs have two zinc fingers, and only the C-terminal zinc finger is involved in DNA binding while the N-terminal zinc finger modulates DNA-binding specificity or mediates the interaction with other proteins (104). Biological functions and the molecular mode of action of GATAs are surprisingly conserved between humans and *Drosophila* (16). Fungal GATAs, in contrast, have distinct functions in the regulation of nitrogen and carbon metabolism and siderophore biosynthesis as well as in light-regulated growth (5, 18, 92, 107).

GATA: a transcription factor family defined by their type IV zinc finger domain

Zinc finger: a zinc-binding protein fold that, in the case of GATA factors, serves in DNA binding

Siderophore: a metal-complexing compound produced by many fungi to bind iron

In the green lineage, angiosperm genomes encode multiple proteins containing GATA-type zinc fingers that can be subdivided based on the zinc finger domain into four main classes, class A–class D, and into seven subfamilies based on the presence of additional domains (32, 88) (**Figure 1**). Generally, plant GATAs contain one zinc finger with two CX_2C motifs interspaced by a 17–20-amino-acid-long loop ($CX_2CX_{17–20}CX_2C$) that is followed by a highly basic region (88) (**Figure 1**). Structurally, the GATA domain is composed of two antiparallel β-sheets followed by an α-helix and a nonstructured basic tail (**Figure 1**). All domains, including the GATA domain



(*Caption appears on following page*)

Figure 1 (*Figure appears on preceding page*)

Conserved motifs and domains identified in *Arabidopsis* and rice GATA factors. (*a*) Schematic representation of the domain architecture and motifs present in *Arabidopsis* and rice GATA factors. The proteins were analyzed with MEME (6) and grouped in classes A–D based on their GATA DNA-binding domain and subfamilies I to VII based on their structural organization. LLM is a leucine-leucine-methionine-containing motif; HAN is a short motif first identified in *Arabidopsis* HANABA TARANU. (*b*) Motifs identified in GATA factors, as shown in panel *a*, together with a structural prediction from selected GATA factors as specified (**https://robetta.bakerlab.org**). All motifs were identified using the listed proteins, except for the given zinc finger sequence logos, which were identified using the class A, B, C, or D sequences from the phylogenetic tree in **Figure 2**. To aid in visualizing sequence conservation, amino acids not included in the regular expression representation of each motif were removed from the sequence logo. The sequence logos of domains from the Pfam database (**http://pfam.xfam.org/**) are not shown. Abbreviations: MEME, Multiple expectation maximizations for motif elicitation; Zn, zinc.

itself, may potentially engage in protein–protein interactions, but such interactions have not been described for plant GATAs.

The understanding of the biological function of plant GATAs has greatly improved over the past decade. In this review, we summarize the current knowledge on the four classes of angiosperm GATA factors. We further use presently available genomic data for a phylogenomic analysis of GATA evolution in the green lineage. On this basis, we finally discuss the evolution of GATA-regulated biological processes in plants, animals, and fungi in a comparative manner.

2. ANGIOSPERM GATA FACTORS: THEIR PHYLOGENY AND FUNCTION

Angiosperm GATA factors can be subdivided into four classes based on their zinc finger domain (32, 88) (**Figures 1** and **2**). The *Arabidopsis thaliana* genome encodes 30 GATA factors that divide into 2 larger classes, A (14 members) and B (11 members), and 2 smaller classes, C (3 members) and D (2 members) (32, 88). Although the number of GATA factors varies between individual species, the relative size of each of the 4 classes appears to remain fairly constant; e.g., the rice genome contains 11 class A, 8 class B, 6 class C, and 1 class D GATAs (**Figures 1** and **2**). While analyses of some vascular plant species have yielded in part larger and in part smaller GATA families, it cannot be excluded that the quality and completeness of the genome assembly or structural annotations limited the analysis, particularly in the cases where the GATA family was found to be comparatively small (48, 88).

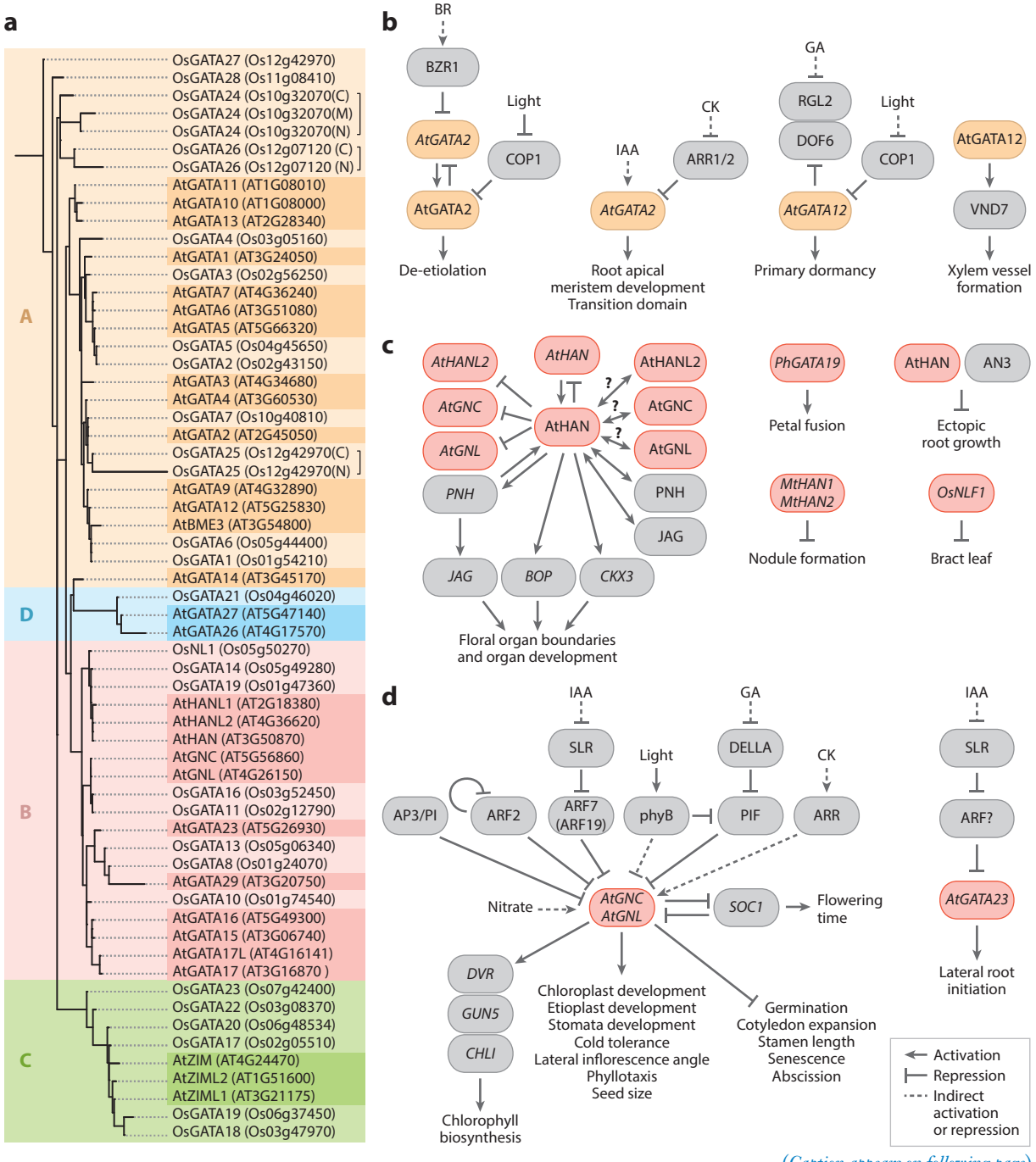
Phylogenomics: an approach for improving protein function predictions by considering how protein sequences are related to one another

Chromatin immunoprecipitation followed by quantitative polymerase chain reaction (ChIP-qPCR): a method for the detection of DNA elements bound by a transcription factor

2.1. Class A GATA Factors

The 14 *A. thaliana* class A-GATAs have a single zinc finger with 18 amino acids in the zinc finger loop, which is N-terminally flanked by a highly acidic conserved patch as well as a basic patch of unknown biological function (**Figures 1** and **2*a***). The rice family comprises 12 members and includes 3 GATAs with multiple zinc fingers (**Figure 2*a***).

The most comprehensive studies are of *AtGATA2*, a key regulator of the crosstalk between brassinosteroid (BR) signaling and photomorphogenesis, two processes controlling hypocotyl elongation and skotomorphogenic development in *A. thaliana* seedlings (62) (**Figure 2*b***). *AtGATA2* expression is reduced after BR treatment and elevated in BR biosynthesis and signaling mutants (62). *AtGATA2* overexpression induced photomorphogenic growth in the dark, and *AtGATA2* suppression reduced hypocotyl elongation in seedlings grown in the light (62). *AtGATA2* overexpression resulted in the activation and repression of close to 3,000 genes (62). These transcriptome changes could be correlated with photomorphogenic growth or BR signaling. A few direct target genes of AtGATA2 were identified by chromatin immunoprecipitation followed by quantitative polymerase chain reaction (ChIP-qPCR) (62). Overexpression of the closely



(Caption appears on following page)

Figure 2 (*Figure appears on preceding page*)

Phylogenetic organization and biological functions of land plant GATA factors. (*a*) Phylogenetic tree of GATA factors from the land plants *Arabidopsis thaliana* (At; *dark shading*) and rice [*Oryza sativa* (Os); *light shading*]. The zinc finger regions defined by PF00320 were extracted from *Arabidopsis* and rice GATA factors, aligned with the NCBI COBALT alignment tool (77), and used to build a phylogenetic tree with W-IQ-TREE (102). The tree was visualized with iTOL (56). (*b–d*) Schematic representation of prominent roles of GATA factors in plant growth and development. Genes that are transcriptionally regulated are shown in italics. Abbreviations: AN3, ANGUSTIFOLIA3; AP3, APETALA3; ARF, AUXIN RESPONSE FACTOR; ARR, ARABIDOPSIS RESPONSE REGULATOR; BOP, BLADE-ON-PETIOLE; BR, brassinosteroid; BZR1, BRASSINAZOLE RESISTANT1; CK, cytokinin; CKX3, CYTOKININ OXIDASE3; COP1, CONSTITUTIVELY MORPHOGENIC1; DOF6, DNA BINDING ONE ZINC FINGER6; DVR, 3,8-divinyl protochlorophyllide a 8-vinyl reductase; GA, gibberellin; GUN5, GENOME UNCOUPLED5; IAA, auxin (indole-3-acetic acid); iTOL, Interactive Tree of Life; JAG, JAGGED; NCBI, National Center for Biotechnology Information; phyB, phytochrome B; PI, PISTILLATA; PIF, PHYTOCHROME-INTERACTING FACTOR; PNH, PINHEAD; RGL2, RGA-LIKE2; SLR, SOLITARY ROOT; SOC1, SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1; VND7, VASCULAR RELATED NAC DOMAIN7.

related *AtGATA4* produced similar, albeit weaker, phenotypes (62). *AtGATA2* thus promotes photomorphogenic seedling development, and *AtGATA4* may play a similar, but less prominent, role (62). The levels of AtGATA2 protein increase after light treatment while *AtGATA2* transcript levels are reduced, suggesting a direct feedback regulation of AtGATA2 on its own transcription, which is supported by ChIP-qPCR results.

The ubiquitin pathway is also involved in the regulation of GATA2 levels. In the dark, degradation of AtGATA2 is likely mediated by the photomorphogenesis regulator and E3 ligase CONSTITUTIVELY MORPHOGENIC1 (COP1), which can ubiquitinate GATA2 in vitro (62) (**Figure 2*b***). BRs repress *AtGATA2* transcription by the BR-activated response regulator BRASSINAZOLE RESISTANT1 (BZR1) (62). Knockdown and knockout of the putative rice ortholog *OsGATA7* in japonica rice result in decreased height and leaf inclination and reduced primary branch height, as well as reduced grain number and weight (117). In line with the observations made in *Arabidopsis*, the plant architecture of these lines resembles weak BR-deficient or BR-insensitive mutants (117).

In the root, *AtGATA2* is specifically expressed in the transition domain between the meristematic proliferation domain and the elongation zone (46) (**Figure 2*b***). *AtGATA2* expression in the transition domain is increased after auxin treatment and dependent on functional auxin transport (46). *AtGATA2* overexpression strongly disturbs root meristem development and, among single and double mutants of *AtGATA2*, *AtGATA4*, or *AtGATA12* (AT5G25830), *gata2 gata12* double mutants have a significantly shorter root with smaller proliferation and transition domains (46). *AtGATA12* and *AtGATA5* can also regulate the expression of *VASCULAR RELATED NAC DOMAIN7* (*VND7*), a transcription factor required for xylem vessel formation (24) (**Figure 2*b***). Expression analyses revealed their transcription in differentiating protoxylem vessel elements and in putative precursors of metaxylem vessel elements located between two protoxylem vessels (24). Concomitant with the proposed role of these GATAs upstream of *VND7*, overexpression of *AtGATA12*, but not of *AtGATA5*, induced the ectopic formation of lignified xylem vessel–like cells with thickened secondary cell walls, albeit with a comparatively low frequency (24). *AtGATA2* and *AtGATA12* are thus important regulators of root development and differentiation.

Transgenic *Arabidopsis* lines overexpressing fusions between AtGATA4 and the SUPERMAN repression domain X (SRDX) have increased shoot biomass under two different nitrogen growth regimes, which was attributed to an improved nitrogen use efficiency (97). *AtGATA4* expression is enhanced after ammonium treatment in the root (97), and *AtGATA4-SRDX* overexpression mainly affected root morphology in that the roots were shorter and had a reduced number of lateral roots than the wild type (97).

AtGATA12 has also been implicated in dormancy control (**Figure 2*b***). *AtGATA12* expression is differentially regulated by gibberellin (GA), which likely acts through the GA pathway-regulatory

DELLA protein RGA-LIKE2 (RGL2) and the transcription factor DNA BINDING ONE ZINC FINGER6 (DOF6) (86). *AtGATA12* is highly expressed in pollen grains, but its expression is diminished in germinating pollen and pollen tubes (57). In this context, *AtGATA12* was described as a pollen-specific GATA factor that, together with BZR1, may regulate the expression of hydroxyproline-rich protein genes (57).

A further link to dormancy control comes through *AtGATA8*, designated *BLUE MICROPYLAR END3* (*BME3*), which accumulates in the embryonic axis of cold-treated seeds, as suggested by the expression of the β-glucuronidase (GUS) reporter in an enhancer trap line (58). A *bme3* loss-of-function mutant has increased dormancy and delayed germination after cold treatment (58). A decrease in the expression of *GA3 OXIDASE* and *GA20 OXIDASE* genes in the *bme3* mutant, together with the expression of these genes following the cold-induction of *BME3*, suggests a role for BME3 in GA biosynthesis regulation (58). BME3 is also a predicted regulator of the auxin-BR–blue light interplay since it shows increased expression after auxin or BR treatments and reduced expression in light-grown seedlings (17). *bme3* mutants display altered expression of marker genes for the three signaling pathways, and dark-grown seedlings have a slightly reduced hypocotyl when compared to the wild type (17).

In tomato, class A GATA *SlGATA17* is upregulated after drought stress. *SlGATA17* overexpression lines were subsequently found to be more drought tolerant than the wild type, which correlated with increased activity of the phenylpropanoid biosynthesis pathway (118).

Angiosperms: the flowering plants comprising approximately 300,000 species

HAN domain: a characteristic domain of unknown biochemical function present at the N terminus of some B-class GATAs

LLM domain: a characteristic domain of unknown biochemical function present at the C terminus of some B-class GATAs

2.2. Class B GATA Factors

Among all four GATA families, class B GATA (B-GATA) factors from vascular plants are presently the best-studied GATAs. The *Arabidopsis* and rice genomes encode 11 and 8 B-GATAs, respectively. B-GATAs contain a single zinc finger with 18 amino acids in the zinc finger loop and, in angiosperms, class B can be subdivided into HAN-domain- or LLM-domain-containing B-GATAs (**Figure 1**). The HAN domain is located N-terminally to the zinc finger and named after its occurrence in the floral development regulator HANABA TARANU (AtHAN) (119). The LLM domain has its name from the conservation of leucine-leucine-methionine at the proteins' C termini (90). The HAN and LLM domains are specific to plant B-GATA factors, do not occur in any other proteins, and have no known biochemical function.

2.2.1. Biological functions of HAN-domain B-GATAs. The *Arabidopsis* genome encodes three, presumably functionally redundant, HAN-domain B-GATAs: AtHAN (AtGATA18), AtHANL1 (HAN-LIKE1; AtGATA20), and AtHANL2 (AtGATA19) (73, 119). Transcriptomics data suggest that AtHAN negatively regulates its own expression, as well as the expression of *AtHAN-LIKE2* and the LLM-domain B-GATAs *AtGNC* (*GATA, NITRATE-INDUCIBLE CARBON METABOLISM INVOLVED*; AtGATA21) and *AtGNL/CGA1* (*GNC-LIKE/CYTOKININ-RESPONSIVE GATA FACTOR1*; AtGATA22) (116) (**Figure 2*c***). Similar negative feedback loops have also been reported among the *Arabidopsis* LLM-domain B-GATA genes (84). The initial finding that HAN- and LLM-domain B-GATAs physically interact could not be reproduced (84, 116).

Arabidopsis han mutants were independently identified as mutants with altered floral organ identity, altered embryo patterning, and narrow leaf morphology (41, 73, 119). In the shoot, *han* mutants have small, flat shoot meristems; reduced floral organ numbers across all four whorls; and fused sepals (119). *AtHAN* is expressed in the boundary between the shoot meristem and newly initiated organs and in the boundaries between different floral organ whorls (119). The meristem regulators *CLAVATA1* (*CLV1*), *CLV2*, and *CLV3* have strong genetic interactions with *AtHAN*, and,

Bract leaf: a modified or specialized leaf supporting flowers or inflorescences that is different from foliage leaves

Neofunctionalization: the process by which, after gene duplication, one of the duplicated genes evolves a new function compared to the ancestral function

conversely, *AtHAN* is ectopically expressed in *clv* mutants (119). *AtHAN* may control *WUSCHEL* (*WUS*) expression, which regulates meristem size by restricting the expression of CLV3, the ligand for the CLV1 and CLV2 receptors. During floral development, AtHAN activates the expression of *PINHEAD* (*PNH*) and interacts with PNH protein (**Figure 2*c***). Together, AtHAN and PNH activate the expression of *JAGGED* (*JAG*) and *BLADE-ON-PETIOLE* (*BOP*) (21) (**Figure 2*c***). AtHAN further modulates cytokinin (CK) homeostasis in the boundary region by stimulating the expression of the *CYTOKININ OXIDASE3* (*CKX3*) (21) (**Figure 2*c***). Since *AtHAN* is also expressed early in provascular cells, the *han* mutant may, alternatively or additionally, be defective in the transport of nutrients or signals, which may prevent meristem growth and floral organ development (119). In *Petunia hybrida*, the downregulation of the HAN-like gene *PhGATA19* resulted in reduced fusion of the petals in this sympetalous species (83) (**Figure 2*c***). *AtHAN* overexpression alters cell division patterns and causes loss of meristem activity (119). *AtHAN* may thus act as a growth repressor at the boundaries in shoots and flowers (119).

A growth-repressive function is also observed in cotyledons where the loss of *AtHAN* function together with the transcriptional coactivator *ANGUSTIFOLIA3/GRF-INTERACTING FACTOR1* (*AN3/GIF1*) leads to ectopic root formation on cotyledons in the double mutant (47) (**Figure 2*c***). Further, loss of two HAN orthologous genes from the legume *Medicago truncatula* results in increased nodule formation (113) (**Figure 2*c***). In mutants of the orthologous HAN-domain GATAs *TASSEL SHEATH1* (*TSH1*) from maize, *NECK LEAF1* (*NL1*, OsGATA15) from rice, and *THIRD OUTER GLUME* (*TRD*) from barley, bract leaves occur (111). In the reference strain of each of these species, bract leaf formation is suppressed by the respective *GATA* gene, which should be the result of a grass-specific duplication and neofunctionalization of *HAN* (111). A HAN-domain mutation in one of the *tsh1* alleles indicates that the HAN domain is essential for the functionality of this GATA class (111). At least for the rice *nl1* mutant, delayed flowering, reduced panicle size with bracts, and abnormal upper internodes were reported as additional phenotypes (110). An independently isolated mutant of *OsGATA15*, *short and narrow flag leaf 1*, was characterized as a mutant with excessive vegetative growth with overgrown culms from internodes (37).

In the *Arabidopsis* embryo, *AtHAN* is required for the proper positioning of the proembryo boundary (73). *han* mutant embryos have vacuolated cells in the lower tier of the embryo and fewer suspensor cells (73). The expression domains of suspensor and lower-tier markers are shifted apically in globular-stage *han* embryos, and the presumed change in cell fate correlates, in the case of the suspensor marker–positive cells, with the cellular phenotype of the respective cells (73). Indicative of defects in proembryo boundary positioning, auxin distribution is shifted in *han* mutant embryos, which aligns with impaired embryonic root formation and is due to the inability to undergo an essential cell division of the hypophysis to form the quiescent center (QC) (73). *han* mutants are also defective in cotyledon initiation and growth; e.g., they initiate up to four cotyledons (119).

The *han* embryo phenotype may, at least in part, be explained by changes in auxin distribution due to altered PIN-FORMED (PIN) distribution. Auxin is transported within the plant by auxin transporters, including the auxin efflux carriers PIN1 and PIN7 (25). In wild-type embryos, auxin initially accumulates in the apical part of the embryo and then shifts to the suspensor preceding specification of the hypophysis. This shift in auxin distribution correlates with a shift in PIN7 polarity in the suspensor, as well as a shift of PIN1 distribution in the provascular cells of the proembryo (25). Both of the expression domains of *PIN1* and *PIN7* are shifted apically in *han* mutants (73). HAN-domain-containing B-GATAs may thus have a growth-repressive function, which may require the regulation of proper auxin distribution.

2.2.2. Biological functions of LLM-domain B-GATAs. *A. thaliana* encodes six LLM-domain GATAs with an AAXLLMXLSXG signature motif, of which *AtGNC* and *AtGNL/CGA1* are the most prominent members (9, 84, 90) (**Figures 1** and **2**). *AtGNC* was initially identified as a nitrate-regulated transcript and the paralogous *AtGNL* as a sucrose- and CK-regulated gene, hence its alternative name *A. thaliana CYTOKININ-RESPONSIVE GATA FACTOR1* (*AtCGA1*) (11, 72, 76, 93) (**Figure 2*c***). AtGNC and AtGNL have long N termini and thereby differ from the other, shorter family members AtGATA15, AtGATA16, AtGATA17, and AtGATA17-LIKE (AtGATA17L). Mutant and overexpression analyses suggest that all six family members have redundant biochemical and biological functions (9, 84) (**Figures 1** and **2**).

Chlorophyll: the green pigment of plants; a porphyrin with a Mg^{2+} cofactor

Heme: a porphyrin of hemoglobin with iron as a cofactor

The LLM domain, like the HAN domain, has no known biochemical function. However, complementation and overexpression lines with a mutation of the LLM domain display different phenotypes when compared to the respective lines overexpressing the wild-type gene (9). While lines expressing B-GATAs with or without the LLM domain have similar greening and germination phenotypes, overexpressors of the LLM-domain-mutated variants display changes in leaf shape and petiole length, and, most importantly, they have a strongly reduced number of differentially regulated genes when compared to overexpressors of the wild-type gene, suggesting that they may be functionally impaired (9). AtGATA23 is a B-GATA with a degenerate LLM domain that has been implicated in the initiation of lateral root formation (20). *AtGATA23* is specifically expressed in xylem pole pericycle cells before the first asymmetric division and controls lateral root founder cell identity downstream of an auxin-responsive module (20) (**Figure 2*d***).

Arabidopsis AtGNC originally attracted attention based on the light-green phenotype of the *gnc* mutant (11). While the loss of multiple LLM-domain B-GATA genes results in a greater decrease in greening, *LLM-domain B-GATA* overexpression induces enhanced greening throughout the plant, very prominently at the base of light-grown seedling hypocotyls (9, 84, 90) (**Figure 2*d***). Remarkably, there is a clear correlation between *LLM-domain B-GATA* gene dosage or gene expression, in the mutants and in the overexpression lines, and the extent of the greening phenotype and also of other quantitative phenotypes, such as lateral shoot angle, flowering time, and phyllotaxy (9, 84, 90) (**Figure 2*d***). The effects of *GATA* expression on the greening response are cell autonomous because grafting a dark-green overexpressor hypocotyl onto a wild-type hypocotyl does not promote greening in the latter (49). *AtGNL* overexpression can also lead to chlorophyll accumulation in the root, and primary root length, as well as lateral root numbers, is altered in *AtGNL* overexpressors and *gnc* mutants (28, 52). The regulation and role of LLM-domain B-GATAs are conserved in rice where RNA interference (RNAi) suppression and overexpression of rice *CGA1* (Os02g12790; *OsGATA11*) antagonistically regulate greening and chloroplast biogenesis but also starch production and plant architecture (42). Furthermore, overexpression of a poplar *AtGNC* ortholog, *PdGATA19/PdGNC*, enhanced chlorophyll content and photosynthesis (2). The loss-of-function mutant showed retarded growth and enhanced secondary xylem differentiation (2).

In flowers, *AtGNC* and *AtGNL* expression is negatively regulated by the floral development regulators PISTILLATA (PI) and APETALA3 (AP3), which directly bind the promoters of the two *GATA* genes (66) (**Figure 2*d***). Negative regulation of these GATAs and, consequently, greening may be part of the mechanism leading to colorless floral organs.

The prominent greening defect of LLM-domain B-GATA mutants has been explained by defects of the GATAs in the activation of chlorophyll biosynthesis genes, chloroplast sigma factors, and chloroplast development (8, 120). Transcriptomics analyses in combination with ChIP-qPCR suggested a prominent role for GATAs in the regulation of many genes in the chlorophyll and the heme biosynthesis pathways (8). Genetic interaction analysis revealed that *AtGNL* overexpression was sufficient to suppress the greening defect of the Mg^{2+}-chelatase subunit gene mutants *genome uncoupled5* (*gun5*) and *chli* (8) (**Figure 2*d***). Overexpression of 3,8-divinyl

protochlorophyllide a 8-vinyl reductase (*DVR*) suppressed the *gnc gnl* greening defect, indicating that *DVR* levels are critically low in this background (8). An interesting physiological ramification of the LLM-domain B-GATA-dependent regulation of the chlorophyll pathway is the apparent diversion of metabolites from the heme branch of the tetrapyrrole pathway and consequently reduced phytochromobilin synthesis, resulting in decreased phytochrome function (8). The latter can explain the occurrence of a longer hypocotyl in *AtGNL* overexpression lines, which should be due to a partial light insensitivity of the overexpressor seedlings (8).

Chloroplast development in the root, repressed by auxin signaling in intact wild-type plants, is induced in a CK-dependent manner when the shoot is removed (51). This CK response is mediated by the type-B ARABIDOPSIS RESPONSE REGULATORS ARR11 and ARR12 that promote the expression of *AtGNL* (51). In turn, *AtGNL* expression in the root is negatively regulated by auxin, such that positive and negative regulation of *AtGNL* may be causal for the differential greening of roots in intact and decapitated plants (51).

In the control of greening, AtGNC and AtGNL act together with the *Arabidopsis* GOLDEN-LIKE transcription factors GLK1 and GLK2. At least with regard to the regulation of a number of genes from the tetrapyrrole biosynthesis pathway, as well as two important *SIGMA* factor genes, the transcription defects detected in the *gnc gnl* and *glk1 glk2* loss-of-function mutants are affected in a largely additive manner in the *gnc gnl glk1 glk2* quadruple mutant, indicating that the two types of transcription regulators act additively and thus, most likely, independently of each other (8, 120).

LLM-domain B-GATAs can also induce stomata formation, most strongly in hypocotyls but also in *Arabidopsis* seedling cotyledons (49) (**Figure 2*d***). The effects of the LLM-domain B-GATAs on stomata formation are light dependent but can be induced by red, far-red, and blue light treatments and also in the dark in a quadruple mutant deficient in the function of the PHYTOCHROME-INTERACTING FACTORS *PIF1*, *PIF3*, *PIF4*, and *PIF5* (49).

Two independent studies have approached the genome-wide identification of AtGNC and AtGNL downstream genes at the genome level (114, 120). In a first study, AtGNC and AtGNL were categorized as transcriptional activators and found to have 1,475 and 638 target genes, respectively (114). Interestingly, the overlap between these two gene sets was comparatively small, leading to the surprising conclusion that these two GATAs may not share the functional redundancy suggested by the ample genetic evidence (114). In the second study and using protein-binding microarrays, several biologically relevant genes were identified as AtGNC repression targets, namely *PIF*, BR biosynthesis, and BR signaling genes, as well as genes for the development of stomatal regulators (120). The latter findings are interesting since they are in line with the known biological functions and regulation of the GATAs but await genetic validation.

2.2.3. Upstream regulators of LLM-domain B-GATAs. In contrast to HAN-domain B-GATAs, where essentially nothing is known about upstream regulatory pathways, multiple phytohormones have been implicated in the regulation of *LLM-domain B-GATA* gene expression (**Figure 2*d***). Common to all *Arabidopsis* LLM-domain B-GATAs is regulation by CK, initially reported for *AtGNC* and *AtGNL* (**Figure 2*d***) (84). However, the strong light-dependent regulation reported for *AtGNL* is not shared by any of the other family members (72, 84). The expression of *AtGNC* and *AtGNL* is also induced by GA and mediated by PIFs that are repressed by the GA-labile DELLA proteins (90) (**Figure 2*d***).

AtGNC and *AtGNL* are important regulators downstream of the GA pathway since their loss-of-function mutations partially suppress the severe GA-deficiency phenotype of a *ga1* mutant, e.g., its late-flowering phenotype (90). This observation has motivated investigations into the role of the LLM-domain B-GATAs in flowering-time control (**Figure 2*d***). In long-day conditions, loss-of-function mutants and overexpression lines flower earlier and later, respectively, than the

wild type (84, 90). AtGNC and AtGNL directly repress the transcription of *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*) (**Figure 2*d***). Inversely, SOC1 represses the transcription of *AtGNC* and *AtGNL* to control greening and cold tolerance, two physiological responses, besides flowering, mediated by SOC1 (89) (**Figure 2*d***). *AtGNC* and *AtGNL* overexpressors share a number of phenotypes with mutants of the *AUXIN RESPONSE FACTOR2* (*ARF2*). The genetic suppression of *arf2* greening and late-flowering-time phenotypes in the *arf2 gnc gnl* mutant indicated that *AtGNC* and *AtGNL* may be direct targets of ARF2, which was then confirmed by ChIP-qPCR experiments (91) (**Figure 2*d***). Similarly, a regulation of *AtGNC* and *AtGNL* downstream from ARF7 and its negative auxin/indole-3-acetic acid (AUX/IAA) regulator SOLITARY ROOT (SLR) could be demonstrated (91) (**Figure 2*d***). LLM-domain B-GATAs are thus subject to regulation by a number of hormonal inputs and are well connected in the plant signaling transduction landscape.

2.3. Class C GATA Factors

Arabidopsis class C GATA factors (C-GATAs) contain 20 amino acids in the zinc finger loop and are characterized by the presence of a TIFY and a CCT motif, located on the N terminus of the GATA domain (**Figure 1**). The TIFY motif is named after the conserved amino acid sequence, threonine-isoleucine-phenylalanine-tyrosine, found in TIFY-motif-containing proteins (106). The TIFY-corresponding sequence in GATA factors is, however, comparatively more divergent. The approximately 45-amino-acid CCT domain gets its name from three important plant regulatory proteins, the flowering-time regulators CONSTANS and CONSTANS-LIKE, as well as the circadian clock regulator TIMING OF CAB EXPRESSION1 (TOC1) (**Figure 2*d***). The function of the CCT motif is unknown, but many proteins with a CCT motif have a role in integrating day length and rhythmicity. In this regard, it is striking that a grass-specific clade of the C-GATAs lacks the CCT motif but contains a FAR-RED IMPAIRED RESPONSE1 (FAR1)/FAR-RED ELONGATED HYPOCOTYL3 (FHY3) protein domain, which also signals environmental light cues to the plant (59, 109).

Biological information about C-GATAs is scarce. In *A. thaliana*, overexpression of *ZINC FINGER PROTEIN EXPRESSED IN THE INFLORESCENCE MERISTEM* (*AtZIM*) results in seedlings with elongated hypocotyls and petioles (75, 96, 106). A mutant of the *Brassica napus ZML1* ortholog, *BnA5.ZML*, has reduced self-incompatibility, suggesting that the gene plays a role in this process (22). C-GATAs thus require deeper investigations, particularly since they have an important role in integrating light information.

2.4. Class D GATA Factors

Class D GATA factors (D-GATAs) have an 18-amino-acid zinc finger loop, and their N-terminally located GATA domain is flanked, in the case of the two *Arabidopsis* family members, by an ADDITIONAL SEX COMBS homology (ASXH; PF13919) domain (**Figure 1**). Characteristic of the α-helical ASXH domain is an LLXXL motif, which is present in diverse transcription factors, coactivators, and repressors, where it supposedly mediates interactions among them (82). The rice D-GATA does not have a recognizable ASXH domain but shares another motif with the *Arabidopsis* D-GATAs (**Figure 1**). There is no information about the biological or biochemical role of the GATAs from this small GATA class, but it is noteworthy that one of the two *Arabidopsis* family members has a degenerated zinc finger domain (**Figure 1**). Further, after modeling of *Arabidopsis* or rice D-GATAs, none of the models predicted a typical zinc finger consisting of an α-helix and an antiparallel β-sheet for these proteins (**Figure 1**). D-GATAs may thus have evolved a new structure and possibly a new function.

3. EVOLUTIONARY CONSERVATION OF GATA FACTORS IN PHOTOSYNTHETIC EUKARYOTES

Pfam: a database of protein families and domains that are defined with hidden Markov models (HMMs)

Phylogeny: a diagram, typically in tree format, that represents predicted evolutionary relationships

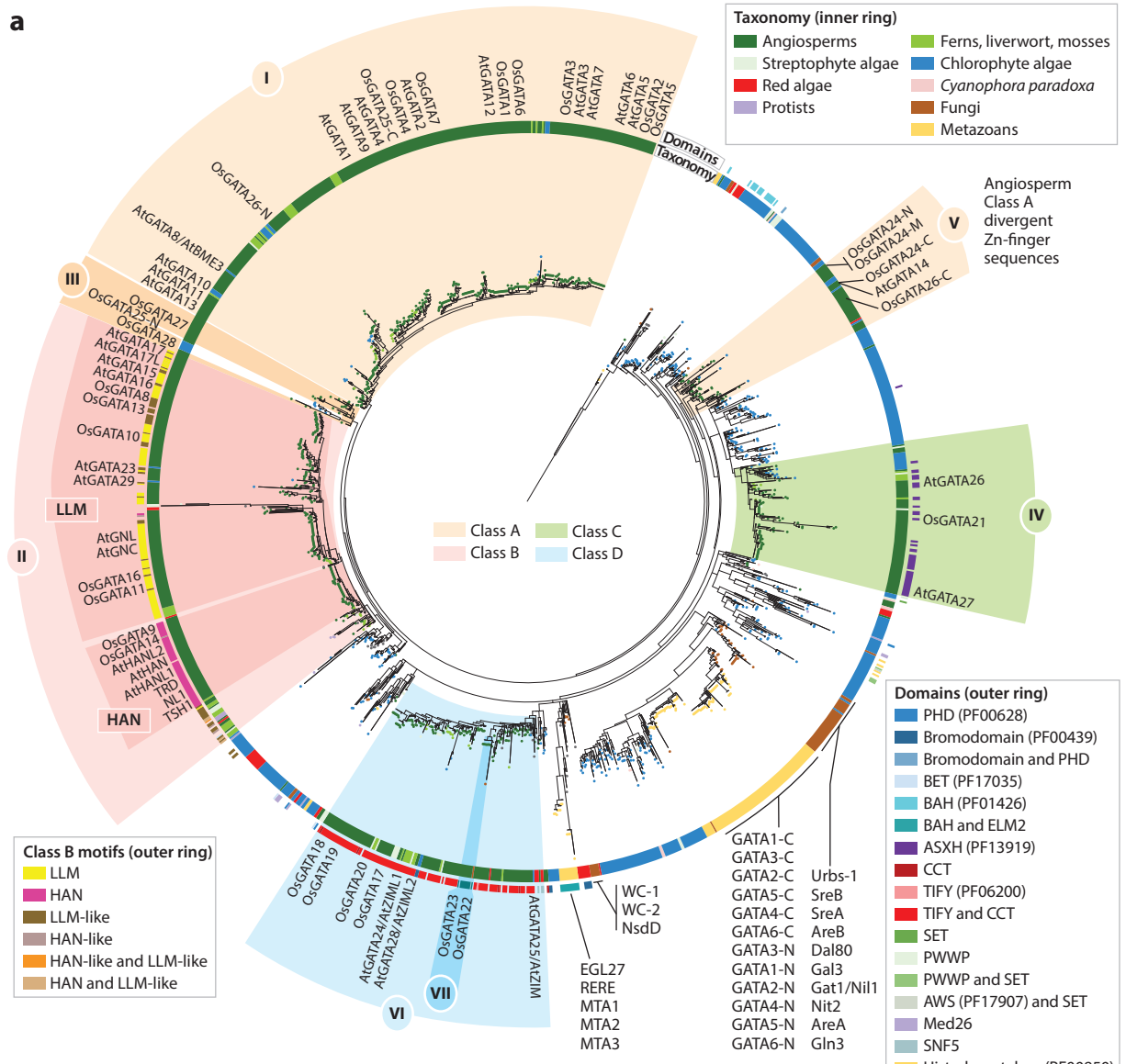
Modular evolution: the combination or recombination of existing units, such as individual protein domains, to form new proteins with new functionality

The availability of numerous land plant and algal genomes has enabled a broad analysis of the GATA family throughout the photosynthetic lineages (105) (**Figure 3**). Searching for proteins containing the GATA zinc finger domain as defined by the Pfam database (PF00320) results in the identification of numerous GATA-type zinc fingers encoded in the genomes of vascular and non-vascular plants, streptophyte algae, chlorophyte algae, red algae, the glaucophyte alga *Cyanophora paradoxa*, and three algae with complex plastids (70) (**Figure 3**). The presence of comparatively large, expanded GATA gene families is not unique to land plants. The green alga *Chlamydomonas reinhardtii*, for example, contains 20 uncharacterized GATA-type zinc finger proteins. On the other hand, the red algae and the only sequenced glaucophyte alga have few GATA zinc finger–containing proteins. For instance, only one GATA factor is found in the genome of the seaweed *Pyropia yezoensis*, while six GATA factors are detected in the genome of the unicellular red alga *Cyanidioschyzon merolae*. As noted previously, GATA factors are absent from diatoms, and only a few sequenced algae from the protist lineages contain GATA factors (87).

Large-scale phylogenetic analyses confirm that the four classes of land plant GATA-type zinc fingers have ancient origins (88) (**Figure 3**). Clades representing each class share a common ancestor with distinct green algal homologs and, in some cases, at least one red algal homolog (**Figure 3**). This pattern suggests that separate ancestral proteins for each GATA class likely existed in a common ancestor of the green and red lineages, with multiple gene duplications occurring at different times and in different lineages during evolution (**Figure 3**). Each land plant genome analyzed, including genomes from ferns and mosses, contains homologs from each class. Each class also contains at least one streptophyte algal homolog, but only the genome of the streptophyte alga *Klebsormidium nitens* contains homologs from all classes, which could be because the other streptophyte algal genomes are incomplete. The class A–D GATA factors from streptophyte algae often resemble land plant GATA factors, with motifs and domains originally described for the *Arabidopsis* GATAs (**Figures 1** and **3**). By contrast, most of the chlorophyte and red algal zinc finger sequences are not easily defined to one class or the other based on phylogeny. Some algal clades have low branch support, some algal sequences branch early in the tree, and other algal proteins are more similar to human and fungal GATA domains than to those of land plants (**Figure 3**). In the latter case, it is possible that red and green algae have retained GATA factors from the last common eukaryotic ancestor, which were lost during land plant evolution. However, GATA factors appear to evolve via modular evolution, involving domain fusion and shuffling. As a result, apart from the zinc finger, there is low to zero sequence similarity between subfamilies, and resolving deep phylogenetic relationships based on the zinc finger sequence is difficult.

3.1. Evolution of Class A GATA Factors

As in *Arabidopsis* and rice, class A is the largest of the four GATA classes in most land plant genomes (**Figure 3**). Based on relatedness to A-GATAs from the ferns, at least three gene duplications likely occurred early in the evolution of vascular plants (**Figure 3**). When using the zinc finger sequence identified with PF00320 to build the phylogenetic tree, the previously class A–classified zinc fingers of AtGATA14, OsGATA24, OsGATA27, and OsGATA28, as well as the C-terminal zinc finger from OsGATA26 and orthologous zinc finger sequences, are not found in the class A clade (**Figure 3**). This is in part due to the absence of a lysine at amino acid position 15 after the first CXXC in the zinc finger, which is conserved in all other A-GATAs in the tree. Although these proteins contain divergent zinc finger sequences, the presence of the



(Caption appears on following page)

Figure 3 (*Figure appears on preceding page*)

Phylogenetic tree of GATA-type zinc finger sequences from representative land plant, algal, metazoan, and fungal proteins. (*a*) The zinc finger region defined by PF00320 was extracted from representative proteins, aligned with NCBI's COBALT alignment tool (77), and used to build a phylogenetic tree with W-IQ-TREE (102). The consensus tree was visualized in iTOL (56), and branches with bootstrap values less than 51% were deleted. The tree can be viewed and explored at **https://tinyurl.com/2p8ja587**. Clades representing previously identified classes, subfamilies, and subclasses are marked. Taxonomic classifications for each leaf (*inner ring*, labeled as taxonomy in the corresponding legend) and the presence of conserved domains (*outer ring*, labeled as domains in the corresponding legend) are colored according to the appropriate legend. For Class B, clades containing the HAN and LLM subfamilies are indicated with darker shading, and motifs identified for each protein are shown as different colors on the outer ring and colored according to the Class B motifs legend. LLM refers to the canonical motif AAXLLMXLSXG, while LLM-like motifs have one or more amino acid substitutions. HAN refers to the canonical motif VDCTLSL, while HAN-like motifs have one or more amino acid substitutions. Branches representing *Arabidopsis* and rice proteins as well as proteins mentioned in the text are labeled. Subfamily classifications from Gupta et al. (32) are indicated with Roman numerals. Darker shading is used to delineate the presence of subfamilies within a single class. In Reyes et al. (88), AtGATA14 was ascribed to subfamily I, but in this tree, AtGATA14 shows more affinity with the rice subfamily V. (*b*) Diagrams of representative proteins with conserved domains identified with RPS-BLAST against the NCBI conserved-domain database (61) or identified by MEME (for HAN and LLM-like). Blocks representing the same domains are colored the same. The name of each domain is given above the first block shown in panel *b*. See **Supplemental Table 1** for a list of all sequences used to generate the phylogenetic tree and accompanying sequence-specific information used to annotate the tree. Abbreviations: ASXH, Additional sex combs homology; AWS, Associated with SET; Med26, Mediator of RNA polymerase II transcription subunit 26; BAH, Bromo-adjacent homology; BET, Bromodomain and extra-terminal domain; FAR1, Far-red impaired response1; MEME, Multiple expectation maximizations for motif elicitation; NCBI, National Center for Biotechnology Information; PHD, plant homeodomain; PWWP, proline-tryptophan-tryptophan-proline; SET, Su(var)3-9, Enhancer of zeste, and Trithorax; SNF5, Sucrose nonfermenting 5; WC, White collar; Zn, zinc.

Supplemental Material >

highly conserved C-terminal extension of the class A zinc finger, $EYRPX_3PX_6HX_3H$, confirms that the proteins are phylogenetically related to A-GATAs (**Figure 1**).

3.2. Evolution of Class B GATA Factors

Class B is the second largest GATA class in most land plant genomes, where it can be subdivided into LLM- and HAN-domain-containing subclasses based on phylogeny and sequence motif conservation (**Figures 1–3**). The HAN subclass contains a VDCTLSL signature motif near the N terminus (**Figure 1**) with a couple of variants, such as MNCTLSL in one of four *Physcomitrium* (*Physcomitrella*) *patens* paralogs or IDCTLSL in one of two paralogs in *Hordeum vulgare* (**Supplemental Table 1**). Most proteins in the LLM subclass from vascular plants contain the AAXLLMXLSXG signature motif near the C terminus with degenerate sequences found in AtGATA23, AtGATA29, and closely related proteins (9) (**Figure 1**). Variants of the canonical motif, such as GAXLLMXLFXG from the streptophyte alga *Chara braunii*, are also present in both algae and land plants (**Supplemental Table 1**).

Subfunctionalization: the process by which, after gene duplication, each duplicated gene retains a separate function of the ancestral gene, such that both copies are needed to preserve the ancestral gene functions

By analyzing B-GATAs from vascular plants, nonvascular plants, and green algae, instances of neofunctionalization and possible subfunctionalization events can be traced that likely occurred very early during the evolution of streptophytes (**Figure 3*b***). B-GATAs from the ferns *Salvinia cucullata* and *Azolla filiculoides*; the spikemoss *Selaginella moellendorffii*; the streptophyte algae *C. braunii* and *K. nitens*; and the chlorophyte algae *Bathycoccus prasinos*, *Ostreococcus lucimarinus*, *Micromonas pusilla*, and *Micromonas* sp. RCC299 contain LLM-like motifs (**Figure 3**; **Supplemental Table 1**). The GATA factors from the mosses *P. patens*, *S. cucullata*, and *S. moellendorffii* and the liverwort *Marchantia polymorpha* contain both the LLM and HAN motifs, while GATA factors from *S. moellendorffii* and *A. filiculoides* only contain the HAN motif. In the case of *A. filiculoides*, the HAN motif has a serine-to-histidine substitution. The presence of the LLM motif in green algae suggests that it may have arisen first in a green alga ancestor followed by the early evolution of the HAN motif in land plants (**Figure 3**). In this context, the LLM-domain B-GATAs from land plants, such as AtGNC and AtGNL, have been attributed with mainly physiological functions

that are already present in algae, notably in chlorophyll biosynthesis and chloroplast division, while HAN-domain B-GATAs may have acquired functions as regulators of morphology and differentiation in land plants (8, 10, 11, 15, 21, 47, 73, 84, 95, 111, 116, 119, 120). It is interesting to speculate that the homologs with HAN and LLM domains found in *P. patens*, *S. cucullata*, *S. moellendorffii*, and *M. polymorpha* may resemble an ancestral state in land plant B-GATAs before the subfunctionalization event that led to separate LLM- and HAN-type class B-GATA lineages.

3.3. Evolution of Class C GATA Factors

The majority of C-GATAs as found in land plants and streptophyte algae contain both the TIFY (PF06200) and CCT (PF06203) domains (**Figure 3**). OsGATA22 and OsGATA23 are the founding members of a small grass-specific subfamily within class C, which, instead of a TIFY-CCT-GATA architecture, contain a FAR1/FHY3 (PF03101) domain located C-terminally of the zinc finger (88) (**Figure 1**). The transcription factor FHY3 and its paralog FAR1 regulate light-induced *CIRCADIAN CLOCK ASSOCIATED1* (*CCA1*) gene expression in *Arabidopsis* (59). The presence of the FHY3/FAR1 domain in these GATAs may also suggest a role in the integration of environmental light signals. The zinc finger sequences from plant C-GATAs share a clade with animal and fungal zinc finger sequences (**Figure 3**). There are also distinct clades representing alga and plant proteins [such as *Aquilegia coerulea* (Aqcoe5G389300) and *Kalanchoe fedtschenkoi* (Kaladp0024s0676)], which show affinity to the C-class clade. These proteins often contain bromodomains or bromodomain extraterminal (BET) domains (**Figure 3**). Bromodomain-related sequences mediate protein–protein interactions in proteins involved in transcription and chromatin remodeling (94). Chromatin remodeling could thus be a biochemical function associated with C-GATAs.

3.4. Evolution of Class D GATA Factors

Class D is the smallest family of GATA factors found in land plant genomes, suggesting selective pressure to maintain these proteins at a relatively low copy number (**Figure 3**). When multiple class D homologs are present in the same genome, the paralogs tend to be structurally different. For instance, in *Arabidopsis*, there are two class D GATA factors (D-GATAs), AtGATA26 and AtGATA27, but AtGATA26 is missing the first two cysteines of the zinc finger sequence. In *M. truncatula*, there are six D-GATAs, and all but one are missing the first two cysteines of the zinc finger sequence. Class D is also the only family that has a large number of closely related sequences from the chlorophytes, several of which contain a detectable ASXH domain that is found in most, but not all, D-GATAs. The ASXH mediates protein–protein interactions among transcription regulators (82). Human ASXL interacts with the tumor suppressor protein and ubiquitin carboxy-terminal hydrolase (UCH) BRCA1-associated protein 1 (BAP1) through the ASXH domain, and the interaction is required for BAP1 functionality (3, 80). In humans, ASXH domains are not associated with GATA factors, but their occurrence in plant GATAs is an indication that UCH hydrolases may act together with plant D-GATAs. In *A. thaliana*, UCH1, UCH2, and UCH3 are closely related to BAP1, which could function together with ASXH-domain-containing GATAs (36, 115).

3.5. GATA Factors Contain Domains Commonly Found in Chromatin Readers

Algal proteins containing GATA-type zinc fingers often contain regions similar to characterized histone readers, including the bromodomain, the plant homeodomain (PHD), the bromo-adjacent homology (BAH) domain, the proline-tryptophan-tryptophan-proline (PWWP) domain, and the

Tudor domain (94) (**Figure 3**). The BAH domain is also found in the GATA-domain-containing proteins EGL27 from *Caenorhabditis elegans* and MTA1 from human, which can function as transcriptional corepressor and coactivator and is part of a histone deacetylase multiprotein complex (69, 74, 98). In line with a direct role for GATAs in histone modification, some green algal GATA factors contain a class I histone deacetylase domain (PF00850) (**Figure 3**). The bromodomain, which recognizes acetylated lysine residues on the N-terminal tails of histones, was commonly found to be associated with GATA domains in green and red algae, sometimes in combination with other domains, such as the Su(var)3-9, Enhancer of zeste, and Trithorax (SET) domain, a domain with methyltransferase activity, and the Tudor-like Agenet domain, a domain of unknown function (67, 68). Taken together with the presence of the ASXH domain in D-GATAs from land plants and two green algae, members of these GATA families have conserved roles or interactions with chromatin readers and remodelers.

Porphyrin: a chemical compound composed of four modified pyrrole subunits

4. EVOLUTIONARY CONSERVATION OF GATA FACTOR FUNCTIONS ACROSS KINGDOMS

Although the zinc finger domain and the structural organization differ between GATAs from the three kingdoms, the evolutionary conservation of GATAs invites the hypothesis that there may be conserved, ancient functions under GATA control. In the following sections, we highlight a few cases where such conserved biological functions can be found. In many cases, the underlying similarities are vague, and the knowledge is often still very fragmentary, but our comparative overview may allow the formulation of a new hypothesis for future research and lead to a better understanding of GATA evolution.

4.1. GATAs in the Regulation of Metal-Binding Complexes

One commonality between GATAs from animals, plants, and yeasts is their role in the synthesis of metal-binding complexes (**Figure 4*a***). Due to their capacity for electron transport, metal-binding complexes are essential mediators of biological redox processes, such as nitrogen fixation, photosynthesis, and mitochondrial respiration (35). Porphyrins are heterocyclic organic tetrapyrroles with different side chain substituents that often form metalloporphyrin complexes with metal ions such as iron(Fe)(II/III), magnesium(Mg)(II), copper(II), and zinc(II) (13, 100). Porphyrins appear colored because they absorb light in the visible range of the spectrum. Chlorophylls, the Mg(II) complexes of protoporphyrin IX, and hemes, the Fe(II) complexes of protoporphyrin IX, are two important groups of metalloporphyrins. The formation of 5-aminolevulinate (5ALA) is divergent between animals, fungi, and plants, but the enzymatic reactions from 5ALA to protoporphyrin IX, the precursor of chlorophyll and heme, are identical, even though the reactions take place in different intracellular compartments between the kingdoms (50).

Heme is an indispensable constituent of hemoglobin, the metalloprotein that enables oxygen transport via the erythrocytes of most vertebrates, but it has a variety of other functions in animals, e.g., as a cofactor in enzyme catalysis and electron transfer. Human HsGATA-1 is the master regulator of heme biosynthesis and erythropoiesis in bone marrow and liver (12, 26, 54) (**Figure 4*a***). HsGATA-1 activates heme biosynthesis by activating 5-aminolevulinic acid synthase-2 (ALAS2) through binding of *ALAS2* intronic *cis* elements. ALAS2 is a rate-limiting factor for heme biosynthesis as it catalyzes the first step of heme biosynthesis, the reaction of glycine and succinyl-CoA forming 5-aminolevulinate (81).

Chlorophylls are the major pigments involved in photosynthesis in plants. The abundant chlorophyll a and b are structurally very similar. Analogous to the role of animal GATAs in heme

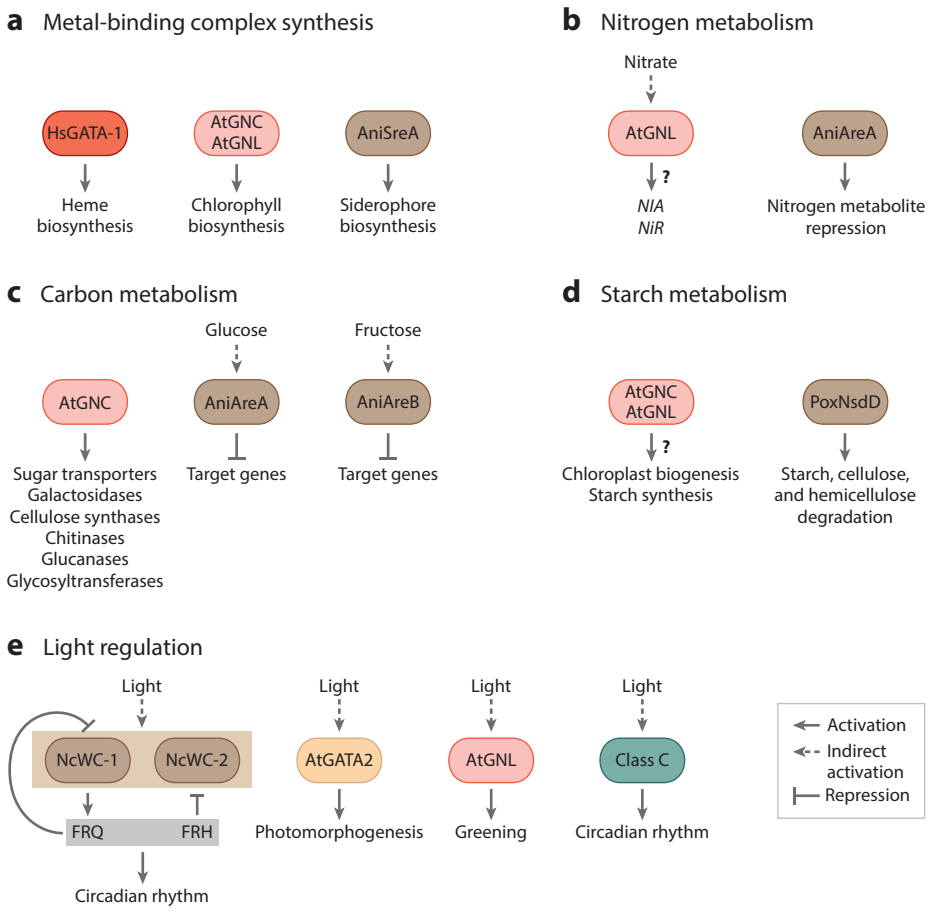


Figure 4

Biological functions of animal, plant, and fungal GATA factors. (*a–e*) Schematic representation of prominent roles of GATA factors in plant growth and development. Dashed arrows indicate upstream signals that do not act directly on GATA factors; solid arrows indicate that GATA factors may act directly on the regulation of genes from the respective pathways. Abbreviations: Ani, *Aspergillus nidulans*; At, *Arabidopsis thaliana*; Hs, *Homo sapiens*; Ncr, *Neurospora crassa*; Pox, *Penicillium oxalicum*.

biosynthesis, B-GATAs promote chlorophyll biosynthesis in *A. thaliana* (8, 84, 90) (**Figure 4*a***). Transcriptomic analyses revealed the regulation of almost all genes of the chlorophyll biosynthesis pathway in *AtGNC* and *AtGNL* mutants or overexpressors, suggesting that these B-GATAs are central regulators of the pathway (8). AtGNC and AtGNL directly regulate genes encoding enzymes of the chlorophyll biosynthesis pathway, like the Mg-chelatase complex and *DVR* (8). At the same time, they regulate the nucleus-encoded *SIGMA* factor genes, which are general transcription regulators in chloroplasts, and act upstream of and in concert with the *GOLDEN2-LIKE* transcription factors, another class of important chlorophyll biosynthesis regulators (8, 120).

Fungal GATAs have a role in the synthesis of metal-binding complexes through their function in the biosynthesis of siderophores, cyclic or linear metal-binding complexes (**Figure 4*a***). Siderophores are structurally and chemically distinct from porphyrins but similar in their capacity

for Fe(III) binding and uptake from the environment (39). Fungi mainly use hydroxamate-type siderophores, which are derived from the nonproteinogenic amino acid ornithine (103). GATA factors from a number of different fungal species regulate siderophore biosynthesis, e.g., Urbs-1 from the basidiomycete *Ustilago maydis* (107), SreP from the ascomycete *Penicillium chrysogenum* (33), and SreA from the ascomycete *Aspergillus nidulans* (34). *A. nidulans* SreA represses siderophore biosynthesis through its negative regulation of L-ornithine-N^5-oxygenase, which catalyzes the first step in the biosynthesis of hydroxamate siderophores (34, 55). Similar observations were made in *U. maydis* and in *Blastomyces dermatitidis* where the GATA factor SreB represses genes involved in the biosynthesis and uptake of siderophores, as well as transport of ornithine from the mitochondria into the cytosol (29, 55). In summary, the synthesis of metal-binding complexes is a common function of GATAs from all three eukaryotic kingdoms.

4.2. GATAs in Nitrogen Metabolism

The regulation of nitrogen metabolism is a common theme between GATAs from fungi and plants (**Figure 4*b***). Early discoveries related to plant GATA functions in nitrate metabolism were motivated, and may potentially have been biased, by the discovery that several fungal GATAs regulate nitrogen responses. The *A. nidulans* GATA AreA functions in nitrogen metabolite repression, which controls the transcription of genes involved in nitrogen uptake and catabolism (5, 18). During the transition from nitrogen-starving to nitrogen-rich conditions, nitrogen metabolite repression prevents the synthesis of enzymes and permeases for the utilization of inferior nitrogen sources like proline, allantoin, and γ-aminobutyric acid (GABA) when superior nitrogen sources like glutamine and asparagine are present (14, 30, 40). AreA acts directly upstream of nitrogen catabolic enzymes and ammonium permeases (14, 71). The *A. nidulans* GATA AreB is also involved in nitrogen metabolite repression by repressing the formamidase *fmdS* under nitrogen-limiting conditions as well as the arginine catabolism genes arginase aminotransferase (*agaA*) and ornithine aminotransferase (*otaA*) in the presence of ammonium (18). Similarly, 4 of the 10 GATA factors, Gln3, Nil1, Nil2, and Dal80, regulate nitrogen metabolite repression in *S. cerevisiae* (30). In this context, Gln3 and Nil1 activate, while Nil2 and Dal80 repress Gln3 and Nil1 targets (19, 40, 64). The absence of obvious additional domains and the low sequence similarity between the different fungal GATAs outside the zinc finger region may suggest that the GATA domain itself has a role in these regulatory processes.

Arabidopsis B-GATAs were initially identified based on their induction by nitrate (11, 93) (**Figure 4*b***). *Arabidopsis* seedlings transferred from medium containing glutamine as a nitrogen source to medium containing nitrate increase *AtGNC* expression twofold after 2 h (11). Additionally, GATA motifs are present in the promoters of nitrate reductase (*NIA*), nitrite reductase (*NiR*), and glutamine synthetase genes. Footprinting experiments suggested that the *Neurospora crassa* GATA NIT2 can bind to these GATA elements in the promoter of spinach *NiR* (85). Furthermore, NIT2 can bind promoter fragments of tomato *NIA* in vitro (45). Although the expression of the *NIA* and *NiR* genes was not altered in the *gnc* mutant (11), the link between plant GATAs and nitrogen metabolism deserves investigation at a deeper level.

4.3. GATAs in Carbon Metabolism

Fungal and plant GATAs play a role in carbon metabolism (**Figure 4*c***). Through the negative transcriptional regulation of *agaA* and *otaA* by the GATAs AreA and AreB, *A. nidulans* can utilize arginine as a source of nitrogen and carbon under nitrogen-repressing conditions (63). The effects of AreA and AreB depend on their primary carbon source. While AreA acts as a transcriptional

repressor in the presence of glucose, AreB represses its target genes in the presence of fructose. The activity of AreA is regulated by several posttranscriptional and posttranslational mechanisms. One of these is the physical interaction of the negative transcriptional regulator NmrA with AreA (53). Similarly, *N. crassa* Nmr1 interacts with the AreA ortholog NIT2 (112).

The *Arabidopsis* B-GATA AtGNC gets its name from its proposed role in carbon metabolism (11) (**Figure 4*c***). The *gnc* mutant is hypersensitive to glucose, and many genes misregulated in the *gnc* mutant are connected to carbon metabolism, such as genes encoding sugar transporters and genes encoding galactosidases (disaccharide metabolism), cellulose synthases (polysaccharide metabolism), chitinases, glucanases, and glycosyltransferases (glycosyl transfer) (11).

Starch plays an important role in both plants and fungi (**Figure 4*d***). Fungi, with their parasitic or necrotrophic behavior, degrade starch from plants, while plants use starch to store glucose generated through photosynthesis. In both kingdoms, starch accumulation and degradation, respectively, are regulated by GATA factors. The GATA NsdD from *Penicillium oxalicum* regulates major genes involved in starch, cellulose, and hemicellulose degradation, namely the glucoamylase gene *POX01356/AMY15A*, the α-amylase gene *POX09352/AMY13A*, and their transcriptional activator POX03890/AMYR (38) (**Figure 4*d***). The *ΔPoxNsdD* mutant shows a strong decrease in starch-degrading enzyme activity. Concomitantly, *Arabidopsis* AtGNC and AtGNL influence the number of chloroplasts, which are the organelles for starch synthesis, as well as the amount of starch itself in proportion to their transcript levels. The same effect on starch biosynthesis was observed in rice and poplar (1, 42, 43). Taken together, these findings confirm, or at least suggest, a conserved role of GATAs in carbon metabolism.

4.4. GATAs in Light Response

Light controls specific developmental programs, photosynthesis, circadian rhythms, and tropic growth responses in fungi and plants. In fungi, blue light signaling depends on the White Collar photoreceptor GATA factors WC-1 and WC-2, originally described in *N. crassa* (7) (**Figure 4*e***). The light-sensory WC-1 interacts with WC-2 in the White Collar Complex (WCC) where WC-2 confers the signal to downstream genes through binding of the promoters of its targets, such as Frequency (FRQ), the major regulator of the circadian clock (27, 108). Within the WCC, FRQ acts in a negative feedback loop together with its partner FRQ-INTERACTING RNA HELICASE (FRH) to inhibit WCC and to regulate the *N. crassa* circadian clock (27) (**Figure 4*e***). *N. crassa white collar* loss-of-function mutants are unable to produce carotenoid photoprotectants, e.g., by regulating the enzyme phytoene dehydrogenase (*al-1*) (27). The interplay between WC-1 and WC-2 appears highly conserved in fungi since fungal genomes always encode both GATA factors (23, 44).

Plant GATAs were originally studied because GATA elements were enriched in the promoters of light- and circadian clock–responsive genes (4, 101). Several *A. thaliana* GATA gene family members are light or dark regulated (65). *AtGATA2* is a key light-signaling transcription factor that mediates photomorphogenesis (62) (**Figures 2*b*** and **4*e***). Further, *AtGNC* and *AtGNL* take part in chloroplast biogenesis and chlorophyll synthesis, two light-dependent processes (8, 15) (**Figure 4*e***). Their light regulation is, among others, repressed by the phytochrome-regulated PIFs (90) (**Figure 1*b***). Additionally, *AtGNL* is strongly light inducible (72, 84). Another interesting link between plant GATAs and the fungal WCC is the occurrence of CCT domains in several C-GATAs (**Figure 1**). CCT domains are commonly found in proteins regulating information from the light environment, and they are present in TOC1, a major regulator of plant circadian rhythms (99).

5. CONCLUSIONS

Research on plant GATA factors over the past 10 to 15 years has greatly advanced the understanding of this conserved transcription factor family. Phylogenomic analysis reveals that the four GATA classes found in angiosperms are already present in land plants. The occurrence of LLM-domain B-GATAs in algal species may suggest that these GATAs, which control photosynthesis and chloroplast biogenesis in angiosperms, could control similar functions in algae. The second motif typical for LLM-domain B-GATAs, the HAN domain, is only found in land plants where HAN-domain B-GATAs control growth and differentiation processes typical for multicellular organisms. C-GATAs are outstanding because of the small size of this family in any tested organism and because all species analyzed only contain one protein with an intact GATA domain. Since GATAs are evolutionarily conserved, it can be speculated that they have also conserved aspects of their biological function, e.g., in the synthesis of metal-binding complexes, in nitrogen or carbon metabolism, or in light-responsive growth. In contrast to animal GATAs, where GATAs are known to function together with interacting proteins, the biochemical mode of function of plant GATAs still needs to be unraveled. Similarly, future research has to concentrate on the identification of direct GATA target genes at the genome-wide level.

SUMMARY POINTS

1. GATA factors are evolutionarily conserved transcription factors present in animals, plants, and fungi.
2. Plant GATAs can be subdivided into four classes and seven structural subfamilies that have their origins in algae.
3. Angiosperm B-class GATA factors (B-GATAs) contain HAN and LLM domains but only LLM-domain-containing GATAs can be found in algae.
4. Early land plants likely contained B-GATAs with both HAN and LLM domains.
5. LLM-domain B-GATAs control chlorophyll biosynthesis and chloroplast division, and they may also have this ancestral function in algae.
6. The HAN domain may have been acquired for the control of differentiation processes in land plants.
7. The class D clade contains an assortment of paralogs with and without conserved GATA-type zinc fingers and the ASXH domain.
8. GATAs may have maintained ancestral biological functions in the biosynthesis of metal-binding complexes and nitrogen and carbon metabolism as well as in light regulation.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Work on GATA transcription factors in the laboratory of C.S. is funded by the Deutsche Forschungsgemeinschaft (grants SCHW751/9-1, SCHW751/9-2, and SCHW751/15-1). C.E.B.-H. is supported by the US Department of Energy, Office of Science, Office of Biological

and Environmental Research, as part of the Quantitative Plant Science Initiative at Brookhaven National Laboratory and award DE-SC0018301.

LITERATURE CITED

1. An Y, Han X, Tang S, Xia X, Yin W. 2014. Poplar GATA transcription factor *PdGNC* is capable of regulating chloroplast ultrastructure, photosynthesis, and vegetative growth in *Arabidopsis* under varying nitrogen levels. *Plant Cell Tissue Organ Cult.* 119:313–27
2. An Y, Zhou Y, Han X, Shen C, Wang S, et al. 2020. The GATA transcription factor GNC plays an important role in photosynthesis and growth in poplar. *J. Exp. Bot.* 71:1969–84
3. **Aravind L, Iyer LM. 2012. The HARE-HTH and associated domains. *Cell Cycle* 11:119–31**
4. Argüello-Astorga G, Herrera-Estrella L. 1998. Evolution of light-regulated plant promoters. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:525–55
5. Arst HN, Cove DJ. 1973. Nitrogen metabolite repression in *Aspergillus nidulans*. *Mol. Gen. Genet.* 126:111–41
6. Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2:28–36
7. Ballario P, Vittorioso P, Magrelli A, Talora C, Cabibbo A, Macino G. 1996. White collar-1, a central regulator of blue light responses in *Neurospora*, is a zinc finger protein. *EMBO J.* 15:1650–57
8. **Bastakis E, Hedtke B, Klermund C, Grimm B, Schwechheimer C. 2018. LLM-domain B-GATA transcription factors play multifaceted roles in controlling greening in Arabidopsis. *Plant Cell* 30:582–99**
9. **Behringer C, Bastakis E, Ranftl QL, Mayer KF, Schwechheimer C. 2014. Functional diversification within the family of B-GATA transcription factors through the leucine-leucine-methionine domain. *Plant Physiol.* 166:293–305**
10. Behringer C, Schwechheimer C. 2015. B-GATA transcription factors—insights into their structure, regulation, and role in plant development. *Front. Plant Sci.* 6:90
11. Bi YM, Zhang Y, Signorelli T, Zhao R, Zhu T, Rothstein S. 2005. Genetic analysis of Arabidopsis GATA transcription factor gene family reveals a nitrate-inducible member important for chlorophyll synthesis and glucose sensitivity. *Plant J.* 44:680–92
12. Bresnick EH, Katsumura KR, Lee HY, Johnson KD, Perkins AS. 2012. Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. *Nucleic Acids Res.* 40:5819–31
13. Buchler JW. 1975. Static coordination chemistry of metalloporphyrins. In *Porphyrins and Metalloporphyrins: A New Edition Based on the Original Volume by J. E. Falk*, ed. KM Smith, pp. 157–231. Amsterdam: Elsevier
14. Caddick MX. 1994. Nitrogen metabolite repression. *Prog. Ind. Microbiol.* 29:323–53
15. **Chiang YH, Zubo YO, Tapken W, Kim HJ, Lavanway AM, et al. 2012. Functional characterization of the GATA transcription factors GNC and CGA1 reveals their key role in chloroplast development, growth, and division in Arabidopsis. *Plant Physiol.* 160:332–48**
16. **Chlon TM, Crispino JD. 2012. Combinatorial regulation of tissue specification by GATA and FOG factors. *Development* 139:3905–16**
17. Choi D, Choi J, Kang B, Lee S, Cho YH, et al. 2014. iNID: an analytical framework for identifying network models for interplays among developmental signaling in *Arabidopsis*. *Mol. Plant* 7:792–813
18. Chudzicka-Ormaniec P, Macios M, Koper M, Weedall GD, Caddick MX, et al. 2019. The role of the GATA transcription factor AreB in regulation of nitrogen and carbon metabolism in *Aspergillus nidulans*. *FEMS Microbiol. Lett.* 366:fnz066
19. Cooper TG. 2002. Transmitting the signal of excess nitrogen in *Saccharomyces cerevisiae* from the Tor proteins to the GATA factors: connecting the dots. *FEMS Microbiol. Rev.* 26:223–38
20. De Rybel B, Vassileva V, Parizot B, Demeulenaere M, Grunewald W, et al. 2010. A novel aux/IAA28 signaling cascade activates GATA23-dependent specification of lateral root founder cell identity. *Curr. Biol.* 20:1697–706

3. An analysis of the shuffling of protein domains in different species and kingdoms in combination with DNA-binding domains.

8. This detailed genetic and molecular analysis identifies several genes related to chlorophyll biosynthesis and chloroplast transcription as direct targets of AtGNC and AtGNL.

9. Through LLM domain deletion and mutation, this study proves the relevance of the LLM domain for the full functionality of LLM domain–containing B-GATAs.

15. A study on the role of AtGNC and AtGNL in chloroplast development and division.

16. FOG is a GATA-interacting factor of human and *Drosophila* GATA; no bona fide protein interactors of plant GATAs have been identified to date.

21. **Ding L, Yan S, Jiang L, Zhao W, Ning K, et al. 2015. *HANABA TARANU* (*HAN*) bridges meristem and organ primordia boundaries through *PINHEAD*, *JAGGED*, *BLADE-ON-PETIOLE2* and *CYTOKININ OXIDASE 3* during flower development in *Arabidopsis*. *PLOS Genet.* 11:e1005479**

21. This study describes the interplay of HAN GATA factors with other developmental regulators during flower development.

22. **Duan Z, Zhang Y, Tu J, Shen J, Yi B, et al. 2020. The *Brassica napus* GATA transcription factor BnA5.ZML1 is a stigma compatibility factor. *J. Integr. Plant Biol.* 62:1112–31**

22. The only report to date ascribing a biological role to C-GATAs.

23. Dunlap JC, Loros JJ. 2004. The *Neurospora* circadian system. *J. Biol. Rhythms* 19:414–24
24. Endo H, Yamaguchi M, Tamura T, Nakano Y, Nishikubo N, et al. 2015. Multiple classes of transcription factors regulate the expression of *VASCULAR-RELATED NAC-DOMAIN*7, a master switch of xylem vessel differentiation. *Plant Cell Physiol.* 56:242–54
25. Friml J, Vieten A, Sauer M, Weijers D, Schwarz H, et al. 2003. Efflux-dependent auxin gradients establish the apical–basal axis of *Arabidopsis*. *Nature* 426:147–53
26. Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH. 1996. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *PNAS* 93:12355–58
27. Fuller KK, Dunlap JC, Loros JJ. 2016. Fungal light sensing at the bench and beyond. *Adv. Genet.* 96:1–51
28. Gaudinier A, Rodriguez-Medina J, Zhang L, Olson A, Liseron-Monfils C, et al. 2018. Transcriptional regulation of nitrogen-associated metabolism and growth. *Nature* 563:259–64
29. Gauthier GM, Sullivan TD, Gallardo SS, Brandhorst TT, Vanden Wymelenberg AJ, et al. 2010. *SREB*, a GATA transcription factor that directs disparate fates in *Blastomyces dermatitidis* including morphogenesis and siderophore biosynthesis. *PLOS Pathog.* 6:e1000846
30. Georis I, Feller A, Vierendeels F, Dubois E. 2009. The yeast GATA factor Gat1 occupies a central position in nitrogen catabolite repression-sensitive gene activation. *Mol. Cell. Biol.* 29:3803–15
31. Gillis WQ, Bowerman BA, Schneider SQ. 2008. The evolution of protostome GATA factors: Molecular phylogenetics, synteny, and intron/exon structure reveal orthologous relationships. *BMC Evol. Biol.* 8:112
32. Gupta P, Nutan KK, Singla-Pareek SL, Pareek A. 2017. Abiotic stresses cause differential regulation of alternative splice forms of GATA transcription factor in rice. *Front. Plant Sci.* 8:1944
33. Haas H, Angermayr K, Stöffler G. 1997. Molecular analysis of a *Penicillium chrysogenum* GATA factor encoding gene (*sreP*) exhibiting significant homology to the *Ustilago maydis urbs1* gene. *Gene* 184:33–37
34. Haas H, Zadra I, Stöffler G, Angermayr K. 1999. The *Aspergillus nidulans* GATA factor SREA is involved in regulation of siderophore biosynthesis and control of iron uptake. *J. Biol. Chem.* 274:4613–19
35. Harel A, Bromberg Y, Falkowski PG, Bhattacharya D. 2014. Evolutionary history of redox metal-binding domains across the tree of life. *PNAS* 111:7042–47
36. Hayama R, Yang P, Valverde F, Mizoguchi T, Furutani-Hayama I, et al. 2019. Ubiquitin carboxyl-terminal hydrolases are required for period maintenance of the circadian clock at high temperature in Arabidopsis. *Sci. Rep.* 9:17030
37. He P, Wang X, Zhang X, Jiang Y, Tian W, et al. 2018. Short and narrow flag leaf1, a GATA zinc finger domain-containing protein, regulates flag leaf size in rice (*Oryza sativa*). *BMC Plant Biol.* 18:273
38. He QP, Zhao S, Wang JX, Li CX, Yan YS, et al. 2018. Transcription factor NsdD regulates the expression of genes involved in plant biomass-degrading enzymes, conidiation, and pigment biosynthesis in *Penicillium oxalicum*. *Appl. Environ. Microbiol.* 84:e01039-18
39. Hider RC, Kong X. 2010. Chemistry and biology of siderophores. *Nat. Prod. Rep.* 27:637–57
40. Hofman-Bang J. 1999. Nitrogen catabolite repression in *Saccharomyces cerevisiae*. *Mol. Biotechnol.* 12:35–73
41. Horiguchi G, Kim GT, Tsukaya H. 2005. The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. *Plant J.* 43:68–78
42. Hudson D, Guevara DR, Hand AJ, Xu Z, Hao L, et al. 2013. Rice cytokinin GATA transcription factor1 regulates chloroplast development and plant architecture. *Plant Physiol.* 162:132–44
43. Hudson D, Guevara DR, Yaish MW, Hannam C, Long N, et al. 2011. *GNC* and *CGA1* modulate chlorophyll biosynthesis and glutamate synthase (*GLU1/Fd-GOGAT*) expression in *Arabidopsis*. *PLOS ONE* 6:e26765
44. Idnurm A, Verma S, Corrochano LM. 2010. A glimpse into the basis of vision in the kingdom *Mycota*. *Fungal Genet. Biol.* 47:881–92
45. Jarai G, Truong H-N, Daniel-Vedele F, Marzluf GA. 1992. NIT2, the nitrogen regulatory protein of *Neurospora crassa*, binds upstream of *nia*, the tomato nitrate reductase gene, in vitro. *Curr. Genet.* 21:37–41

46. Jiang K, Yung V, Chiba T, Feldman LJ. 2018. Longitudinal patterning in roots: A *GATA2*–auxin interaction underlies and maintains the root transition domain. *Planta* 247:831–43
47. Kanei M, Horiguchi G, Tsukaya H. 2012. Stable establishment of cotyledon identity during embryogenesis in *Arabidopsis* by *ANGUSTIFOLIA3* and *HANABA TARANU*. *Development* 139:2436–46
48. Kim M, Xi H, Park J. 2021. Genome-wide comparative analyses of GATA transcription factors among 19 *Arabidopsis* ecotype genomes: intraspecific characteristics of GATA transcription factors. *PLOS ONE* 16:e0252181
49. Klermund C, Ranftl QL, Diener J, Bastakis E, Richter R, Schwechheimer C. 2016. LLM-domain B-GATA transcription factors promote stomatal development downstream of light signaling pathways in *Arabidopsis thaliana* hypocotyls. *Plant Cell* 28:646–60
50. Kloehn J, Harding CR, Soldati-Favre D. 2021. Supply and demand—heme synthesis, salvage and utilization by Apicomplexa. *FEBS J.* 288:382–404
51. Kobayashi K, Ohnishi A, Sasaki D, Fujii S, Iwase A, et al. 2017. Shoot removal induces chloroplast development in roots via cytokinin signaling. *Plant Physiol.* 173:2340–55
52. Kollmer I, Werner T, Schmulling T. 2011. Ectopic expression of different cytokinin-regulated transcription factor genes of *Arabidopsis thaliana* alters plant growth and development. *J. Plant Physiol.* 168:1320–27
53. Lamb HK, Ren J, Park A, Johnson C, Leslie K, et al. 2004. Modulation of the ligand binding properties of the transcription repressor NmrA by GATA-containing DNA and site-directed mutagenesis. *Protein Sci.* 13:3127–38
54. Lentjes MHFM, Niessen HEC, Akiyama Y, de Bruïne AP, Melotte V, van Engeland M. 2016. The emerging role of GATA transcription factors in development and disease. *Expert Rev. Mol. Med.* 18:e3
55. Leong SA, Winkelmann G. 1998. Molecular biology of iron transport in fungi. *Met. Ions Biol. Syst.* 35:147–86
56. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49:W293–96
57. Li Y, Mullin M, Zhang Y, Drews F, Welch LR, Showalter AM. 2020. Identification of cis-regulatory sequences controlling pollen-specific expression of hydroxyproline-rich glycoprotein genes in *Arabidopsis thaliana*. *Plants* 9:1751
58. Liu PP, Koizuka N, Martin RC, Nonogaki H. 2005. The *BME3* (*Blue Micropylar End 3*) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination. *Plant J.* 44:960–71
59. Liu Y, Ma M, Li G, Yuan L, Xie Y, et al. 2020. Transcription factors FHY3 and FAR1 regulate light-induced *CIRCADIAN CLOCK ASSOCIATED1* gene expression in Arabidopsis. *Plant Cell* 32:1464–78
60. Lowry JA, Atchley WR. 2000. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *J. Mol. Evol.* 50:103–15
61. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, et al. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48:D265–68
62. Luo X-M, Lin W-H, Zhu S, Zhu JY, Sun Y, et al. 2010. Integration of light- and brassinosteroid-signaling pathways by a GATA transcription factor in *Arabidopsis*. *Dev. Cell* 19:872–83
63. Macios M, Caddick MX, Weglenski P, Scazzocchio C, Dzikowska A. 2012. The GATA factors AREA and AREB together with the co-repressor NMRA, negatively regulate arginine catabolism in *Aspergillus nidulans* in response to nitrogen and carbon source. *Fungal Genet. Biol.* 49:189–98
64. Magasanik B. 2005. The transduction of the nitrogen regulation signal in *Saccharomyces cerevisiae*. *PNAS* 102:16537–38
65. Manfield IW, Devlin PF, Jen C-H, Westhead DR, Gilmartin PM. 2007. Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the Arabidopsis GATA gene family. *Plant Physiol.* 143:941–58
66. Mara CD, Irish VF. 2008. Two GATA transcription factors are downstream effectors of floral homeotic gene action in Arabidopsis. *Plant Physiol.* 147:707–18
67. Marmorstein R. 2003. Structure of SET domain proteins: a new twist on histone methylation. *Trends Biochem. Sci.* 28:59–62

68. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. 2003. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci.* 28:69–74
69. Millard CJ, Fairall L, Schwabe JWR. 2014. Towards an understanding of the structure and function of MTA1. *Cancer Metastasis Rev.* 33:857–67
70. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, et al. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49:D412–19
71. Monahan BJ, Askin MC, Hynes MJ, Davis MA. 2006. Differential expression of *Aspergillus nidulans* ammonium permease genes is regulated by GATA transcription factor AreA. *Eukaryot. Cell* 5:226–37
72. Naito T, Kiba T, Koizumi N, Yamashino T, Mizuno T. 2007. Characterization of a unique GATA family gene that responds to both light and cytokinin in *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.* 71:1557–60

73. A careful developmental analysis of HAN-domain B-GATA function during embryogenesis.

73. Nawy T, Bayer M, Mravec J, Friml J, Birnbaum KD, Lukowitz W. 2010. The GATA factor *HANABA TARANU* is required to position the proembryo boundary in the early *Arabidopsis* embryo. *Dev. Cell* 19:103–13

74. Nicolson GL, Nawa A, Toh Y, Taniguchi S, Nishimori K, Moustafa A. 2003. Tumor metastasis-associated human *MTA1* gene and its MTA1 protein product: role in epithelial cancer cell invasion, proliferation and nuclear regulation. *Clin. Exp. Metastas.* 20:19–24
75. Nishii A, Takemura M, Fujita H, Shikata M, Yokota A, Kohchi T. 2000. Characterization of a novel gene encoding a putative single zinc-finger protein, *ZIM*, expressed during the reproductive phase in *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.* 64:1402–9
76. Osuna D, Usadel B, Morcuende R, Gibon Y, Bläsing OE, et al. 2007. Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. *Plant J.* 49:463–91
77. Papadopoulos JS, Agarwala R. 2007. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23:1073–79
78. Park J, Kim H, Kim S, Kong S, Park J, et al. 2006. A comparative genome-wide analysis of GATA transcription factors in fungi. *Genom. Inform.* 4:147–160
79. Patient RK, McGhee JD. 2002. The GATA family (vertebrates and invertebrates). *Curr. Opin. Genet. Dev.* 12:416–22
80. Peng H, Prokop J, Karar J, Park K, Cao L, et al. 2018. Familial and somatic *BAP1* mutations inactivate ASXL1/2-mediated allosteric regulation of BAP1 deubiquitinase by targeting multiple independent domains. *Cancer Res.* 78:1200–13
81. Phillips JD. 2019. Heme biosynthesis and the porphyrias. *Mol. Genet. Metab.* 128:164–77
82. Plevin MJ, Mills MM, Ikura M. 2005. The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem. Sci.* 30:66–69
83. Preston JC, Powers B, Kostyun JL, Driscoll H, Zhang F, Zhong J. 2019. Implications of region-specific gene expression for development of the partially fused petunia corolla. *Plant J.* 100:158–75

84. A description of a complex mutant with defects in most Arabidopsis LLM-domain B-GATA genes.

84. Ranftl QL, Bastakis E, Klermund C, Schwechheimer C. 2016. LLM-Domain containing B-GATA factors control different aspects of cytokinin-regulated development in *Arabidopsis thaliana*. *Plant Physiol.* 170:2295–311

85. Rastogi R, Bate NJ, Sivasankar S, Rothstein SJ. 1997. Footprinting of the spinach nitrite reductase gene promoter reveals the preservation of nitrate regulatory elements between fungi and higher plants. *Plant Mol. Biol.* 34:465–76
86. Ravindran P, Verma V, Stamm P, Kumar PP. 2017. A novel RGL2-DOF6 complex contributes to primary seed dormancy in *Arabidopsis thaliana* by regulating a GATA transcription factor. *Mol. Plant* 10:1307–20
87. Rayko E, Maumus F, Maheswari U, Jabbari K, Bowler C. 2010. Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol.* 188:52–66
88. Reyes JC, Muro-Pastor MI, Florencio FJ. 2004. The GATA family of transcription factors in Arabidopsis and rice. *Plant Physiol.* 134:1718–32

89. Richter R, Bastakis E, Schwechheimer C. 2013. Cross-repressive interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the control of greening, cold tolerance, and flowering time in Arabidopsis. *Plant Physiol.* 162:1992–2004
90. Richter R, Behringer C, Müller IK, Schwechheimer C. 2010. The GATA-type transcription factors GNC and GNL/CGA1 repress gibberellin signaling downstream from DELLA proteins and PHYTOCHROME-INTERACTING FACTORS. *Genes Dev.* 24:2093–104
91. Richter R, Behringer C, Zourelidou M, Schwechheimer C. 2013. Convergence of auxin and gibberellin signaling on the regulation of the GATA transcription factors *GNC* and *GNL* in *Arabidopsis thaliana*. *PNAS* 110:13192–97
92. Scazzocchio C. 2000. The fungal GATA factors. *Curr. Opin. Microbiol.* 3:126–31
93. Scheible W-R, Morcuende R, Czechowski T, Fritz C, Osuna D, et al. 2004. Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. *Plant Physiol.* 136:2483–99
94. Scheid R, Chen J, Zhong X. 2021. Biological role and mechanism of chromatin readers in plants. *Curr. Opin. Plant Biol.* 61:102008
95. Schiessl K, Muiño JM, Sablowski R. 2014. Arabidopsis JAGGED links floral organ patterning to tissue growth by repressing Kip-related cell cycle inhibitors. *PNAS* 111:2830–35
96. Shikata M, Matsuda Y, Ando K, Nishii A, Takemura M, et al. 2004. Characterization of *Arabidopsis ZIM*, a member of a novel plant-specific GATA factor gene family. *J. Exp. Bot.* 55:631–39
97. Shin JM, Chung K, Sakamoto S, Kojima S, Yeh C-M, et al. 2017. The chimeric repressor for the GATA4 transcription factor improves tolerance to nitrogen deficiency in *Arabidopsis*. *Plant Biotechnol.* 34:151–58
98. Solari F, Bateman A, Ahringer J. 1999. The *Caenorhabditis elegans* genes *egl-27* and *egr-1* are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. *Development* 126:2483–94
99. Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, et al. 2000. Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. *Science* 289:768–71
100. Tahoun M, Gee CT, McCoy VE, Sander PM, Müller CE. 2021. Chemistry of porphyrins in fossil plants and animals. *RSC Adv.* 11:7552–63
101. Terzaghi WB, Cashmore AR. 1995. Light-regulated transcription. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 46:445–74
102. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44:W232–35
103. Tripathi K, Kumar N, Singh M, Singh RK. 2020. Fungal siderophore: biosynthesis, transport, regulation, and potential applications. In *Rhizosphere Microbes*, ed. SK Sharma, UB Singh, PK Sahu, HV Singh, PK Sharma, pp. 387–408. Singapore: Springer
104. Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, et al. 1997. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* 90:109–19
105. Unamba CIN, Nag A, Sharma RK. 2015. Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. *Front. Plant Sci.* 6:1074
106. Vanholme B, Grunewald W, Bateman A, Kohchi T, Gheysen G. 2007. The tify family previously known as ZIM. *Trends Plant Sci.* 12:239–44
107. Voisard C, Wang J, McEvoy JL, Xu P, Leong SA. 1993. *urbsl*, a gene regulating siderophore biosynthesis in *Ustilago maydis*, encodes a protein similar to the erythroid transcription factor GATA-1. *Mol. Cell Biol.* 13:7091–100
108. Wang B, Zhou X, Loros JJ, Dunlap JC. 2015. Alternative use of DNA binding domains by the *Neurospora* White Collar complex dictates circadian regulation and light responses. *Mol. Cell Biol.* 36:781–93
109. Wang H, Wang H. 2015. Multifaceted roles of *FHY3* and *FAR1* in light signaling and beyond. *Trends Plant Sci.* 20:453–61
110. Wang L, Yin H, Qian Q, Yang J, Huang C, et al. 2009. *NECK LEAF 1*, a GATA type transcription factor, modulates organogenesis by regulating the expression of multiple regulatory genes during reproductive development in rice. *Cell Res.* 19:598–611

111. Identification of a novel HAN factor from grass species with a conserved role in the suppression of bract leaf development.

111. Whipple CJ, Hall DH, DeBlasio S, Taguchi-Shiobara F, Schmidt RJ, Jackson DP. 2010. A conserved mechanism of bract suppression in the grass family. *Plant Cell* 22:565–78

112. Xiao X, Fu YH, Marzluf GA. 1995. The negative-acting NMR regulatory protein of *Neurospora crassa* binds to and inhibits the DNA-binding activity of the positive-acting nitrogen regulatory protein NIT2. *Biochemistry* 34:8861–68
113. Xu Y, Wang H, Lu Z, Wen L, Gu Z, et al. 2021. Developmental analysis of the GATA factor *HANABA TARANU* mutants in *Medicago truncatula* reveals their roles in nodule formation. *Front. Plant Sci.* 12:616776
114. Xu Z, Casaretto JA, Bi Y-M, Rothstein SJ. 2017. Genome-wide binding analysis of AtGNC and AtCGA1 demonstrates their cross-regulation and common and specific functions. *Plant Direct* 1:e00016
115. Yang P, Smalle J, Lee S, Yan N, Emborg TJ, Vierstra RD. 2007. Ubiquitin C-terminal hydrolases 1 and 2 affect shoot architecture in Arabidopsis. *Plant J.* 51:441–57
116. Zhang X, Zhou Y, Ding L, Wu Z, Liu R, Meyerowitz EM. 2013. Transcription repressor HANABA TARANU controls flower development by integrating the actions of multiple hormones, floral organ specification genes, and GATA3 family genes in *Arabidopsis*. *Plant Cell* 25:83–101
117. Zhang Y-J, Zhang Y, Zhang L-L, Huang H-Y, Yang B-J, et al. 2018. *OsGATA7* modulates brassinosteroids-mediated growth regulation and influences architecture and grain shape. *Plant Biotechnol. J.* 16:1261–64
118. Zhao T, Wu T, Pei T, Wang Z, Yang H, et al. 2021. Overexpression of *SlGATA17* promotes drought tolerance in transgenic tomato plants by enhancing activation of the phenylpropanoid biosynthetic pathway. *Front. Plant Sci.* 12:634888
119. Zhao Y, Medrano L, Ohashi K, Fletcher JC, Yu H, et al. 2004. HANABA TARANU is a GATA transcription factor that regulates shoot apical meristem and flower development in Arabidopsis. *Plant Cell* 16:2586–600
120. Zubo YO, Blakley IC, Franco-Zorrilla JM, Yamburenko MV, Solano R, et al. 2018. Coordination of chloroplast development through the action of the GNC and GLK transcription factor families. *Plant Physiol.* 178:130–47

Annual Review of Plant Biology

Volume 73, 2022

Contents

Errata

An online log of corrections to *Annual Review of Plant Biology* articles may be found at http://www.annualreviews.org/errata/arplant

Related Articles

From the ***Annual Review of Biochemistry***, Volume 90 (2021)

Preparing Better Samples for Cryo–Electron Microscropy: Biochemical Challenges Do Not End with Isolation and Purification
Robert M. Glaeser

From the ***Annual Review of Biophysics***, Volume 50 (2021)

Structure of Phycobilisomes
Sen-Fang Sui

From the ***Annual Review of Cell and Developmental Biology***, Volume 37 (2021)

Dynamic Nutrient Signaling Networks in Plants
Lei Li, Kun-hsiang Liu, and Jen Sheen

From the ***Annual Review of Food Science and Technology***, Volume 12 (2021)

Cellular Agriculture: Opportunities and Challenges
Regine Eibl, Yannick Senn, Géraldine Gubser, Valentin Jossen, Christian van den Bos, and Dieter Eibl

Soy Protein: Molecular Structure Revisited and Recent Advances in Processing Technologies
Xiaonan Sui, Tianyi Zhang, and Lianzhou Jiang

Feruloylated Arabinoxylan and Oligosaccharides: Chemistry, Nutritional Functions, and Options for Enzymatic Modification
Shang Lin, Jane W. Agger, Casper Wilkens, and Anne S. Meyer

Hormesis Mediates Acquired Resilience: Using Plant-Derived Chemicals to Enhance Health
Edward J. Calabrese

Recent Advances in Food Emulsions and Engineering Foodstuffs Using Plant-Based Nanocelluloses
Long Bai, Siqi Huan, Ya Zhu, Guang Chu, D. Julian McClements, and Orlando J. Rojas

Carotenoids: Considerations for Their Use in Functional Foods, Nutraceuticals, Nutricosmetics, Supplements, Botanicals, and Novel Foods in the Context of Sustainability, Circular Economy, and Climate Change
Antonio J. Meléndez-Martínez, Volker Böhm, Grethe Iren Andersen Borge, M. Pilar Cano, Martina Fikselová, Ruta Gruskiene, Vera Lavelli, Monica Rosa Loizzo, Anamarija I. Mandić, Paula Mapelli Brahm, Aleksandra Č. Mišan, Adela M. Pintea, Jolanta Sereikaitė, Liliana Vargas-Murga, Sanja S. Vlaisavljević, Jelena J. Vulić, and Nora M. O'Brien

Glucosinolates in *Brassica* Vegetables: Characterization and Factors That Influence Distribution, Content, and Intake
Xianli Wu, Hui Huang, Holly Childs, Yanbei Wu, Liangli Yu, and Pamela R. Pehrsson

Molecular Origins of Polymorphism in Cocoa Butter
Saeed M. Ghazani and Alejandro G. Marangoni

From the ***Annual Review of Genetics***, Volume 55 (2021)

Plant Cell Identity in the Era of Single-Cell Transcriptomics
Kook Hui Ryu, Yan Zhu, and John Schiefelbein

Green Algal Models for Multicellularity
James Umen and Matthew D. Herron

Genetics of Shoot Meristem and Shoot Regeneration
Leor Eshed Williams

From the ***Annual Review of Phytopathology***, Volume 59 (2021)

Damage-Associated Molecular Patterns (DAMPs) in Plant Innate Immunity: Applying the Danger Model and Evolutionary Perspectives
Kiwamu Tanaka and Martin Heil

Characterization of Effector–Target Interactions in Necrotrophic Pathosystems Reveals Trends and Variation in Host Manipulation
Timothy L. Friesen and Justin D. Faris

Emerging Roles of Posttranslational Modifications in Plant-Pathogenic Fungi and Bacteria
Wende Liu, Lindsay Triplett, and Xiao-Lin Chen

Models of Plant Resistance Deployment
Loup Rimbaud, Frédéric Fabre, Julien Papaïx, Benoît Moury, Christian Lannou, Luke G. Barrett, and Peter H. Thrall

Interkingdom Signaling Interference: The Effect of Plant-Derived Small Molecules on Quorum Sensing in Plant-Pathogenic Bacteria
Janak Raj Joshi, Netaly Khazanov, Amy Charkowski, Adi Faigenboim, Hanoch Senderowitz, and Iris Yedidia