Functional Diversification within the Family of B-GATA Transcription Factors through the Leucine-Leucine-Methionine Domain[1][C][W][OPEN]

Carina Behringer, Emmanouil Bastakis, Quirin L. Ranftl, Klaus F.X. Mayer, and Claus Schwechheimer*

Plant Systems Biology, Technische Universität München, 85354 Freising, Germany (C.B., E.B., Q.L.R., C.S.); and Munich Information Centre for Protein Sequences, Institute for Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany (K.F.X.M.)

The transcription of the Arabidopsis (*Arabidopsis thaliana*) GATA transcription factors GATA, NITRATE-INDUCIBLE, CARBON METABOLISM-INVOLVED (GNC) and GNC-LIKE (GNL)/CYTOKININ-RESPONSIVE GATA FACTOR1 is controlled by several growth regulatory signals including light and the phytohormones auxin, cytokinin, and gibberellin. To date, GNC and GNL have been attributed functions in the control of germination, greening, flowering time, floral development, senescence, and floral organ abscission. GNC and GNL belong to the 11-member family of B-class GATA transcription factors that are characterized to date solely by their high sequence conservation within the GATA DNA-binding domain. The degree of functional conservation among the various B-class GATA family members is not understood. Here, we identify and examine B-class GATAs from Arabidopsis, tomato (*Solanum lycopersicon*), Brachypodium (*Brachypodium distachyon*), and barley (*Hordeum vulgare*). We find that B-class GATAs from these four species can be subdivided based on their short or long N termini and the presence of the 13-amino acid C-terminal leucine-leucine-methionine (LLM) domain with the conserved motif LLM. Through overexpression analyses and by complementation of a *gnc gnl* double mutant, we provide evidence that the length of the N terminus may not allow distinguishing between the different B-class GATAs at the functional level. In turn, we find that the presence and absence of the LLM domain in the overexpressors has differential effects on hypocotyl elongation, leaf shape, and petiole length, as well as on gene expression. Thus, our analyses identify the LLM domain as an evolutionarily conserved domain that determines B-class GATA factor identity and provides a further subclassification criterion for this transcription factor family.

GATA factors are evolutionarily conserved transcriptional regulators with a type IV zinc finger domain ($C-X_2C-X_{17-20}C-X_2C$) that binds to the consensus DNA sequence WGATAR (where W is T or A and R is G or A; Reyes et al., 2004). Whereas GATA factors from nonplant species often contain multiple GATA domains with loops of variable length, all 30 previously classified GATA-like zinc finger proteins from Arabidopsis (*Arabidopsis thaliana*) contain only one zinc finger (http://www.arabidopsis.org; Reyes et al., 2004). GATA factors from Arabidopsis and rice (*Oryza sativa*) were subdivided into four classes based on their sequence similarity, the presence or absence of additional recognizable protein domains, and their exon-intron structure (Reyes et al., 2004). This classification also includes B-GATAs that share high sequence similarities between the individual family members but are devoid of any as yet identified distinguishing sequence features outside of the GATA DNA-binding domain.

To date, four of the 11 members of the Arabidopsis B-GATAs have been examined in some detail: the functionally redundant GATA, NITRATE-INDUCIBLE, CARBON METABOLISM-INVOLVED (GNC; AtGATA21), CYTOKININ-RESPONSIVE GATA FACTOR1/GNC-LIKE (CGA1/GNL; AtGATA22; hitherto GNL), HANABA TARANU (HAN; AtGATA18), and GATA23 (AtGATA23). *GNC* was originally identified as a nitrate-inducible gene with an apparent role in the control of greening (Bi et al., 2005; Hudson et al., 2011; Chiang et al., 2012). The *GNC* paralog *GNL* was also designated *CGA1* based on its transcriptional regulation by cytokinin (Naito et al., 2007). Further analyses showed that *GNC* and *GNL* are functionally redundant transcription factors whose expression is also controlled by the DELLA regulators of the GA signaling pathway, the light-labile PHYTOCHROME INTERACTING FACTORS, the AUXIN RESPONSE FACTORS ARF2 and ARF7 of the auxin signaling pathway, and the flowering regulator SUPPRESSOR-OF-constans1 (Richter et al., 2010, 2013a, 2013b). Moreover, *GNC* and *GNL* were proposed to be directly repressed by the floral development regulators APETALA3 and PISTILLATA, and it was suggested that this repression serves to prevent greening in Arabidopsis petals (Mara

and Irish, 2008). *HAN* was identified based on a mutant with a small shoot apical meristem and a reduction in the number of floral organs (Zhao et al., 2004). *HAN* overexpression results in delayed plant growth, disturbed cell division, and loss of shoot meristem activity (Zhao et al., 2004). *HAN* has an additional role in the establishment of cotyledon identity during embryogenesis, a defect that may result from the interference of HAN with the auxin transport machinery (Nawy et al., 2010; Kanei et al., 2012). Overexpression of *HAN* negatively interferes with the expression of a gene closely related to *HAN*, *HAN-LIKE2* (*HANL2*; AtGATA19), as well as with the expression of *GNC* and *GNL* (Zhang et al., 2013). This suggested that the gain of *HAN* function might be compensated by the down-regulation of potentially functionally redundant B-GATAs. Further evidence for an interplay of HAN with GNC and GNL comes from the observation that the proteins interact in the yeast (*Saccharomyces cerevisiae*) two-hybrid system (Zhang et al., 2013). Finally, *AtGATA23* is specifically expressed before the first asymmetric division in xylem pole pericycle cells, where it regulates lateral root founder cell specification and root branching patterns (De Rybel et al., 2010). Thus, these four B-class GATAs are implicated in a wide range of developmental and physiological processes. However, it remains to be seen whether these differences are determined by the differential expression of the B-GATAs or whether they additionally differ in their biochemical activities.

With regard to all phenotypes examined to date, transgenic lines overexpressing *GNC* and *GNL* have phenotypes that are opposite to those observed in *gnc* and *gnl* single or double loss-of-function mutants. Thus, the analysis of overexpressors can provide biologically relevant insights into B-GATA function in these cases that, in loss-of-function mutants, may only be discernable in specific mutant backgrounds or may be not apparent due to functional redundancies. During our previous analyses, we had noted that GNC and GNL share a conserved C-terminal domain with various other GATA factors that we designated leucine-leucine-methionine (LLM) domain based on an invariant LLM motif at its core (Richter et al., 2010). As a result of a comparative analysis of GATA factors from Arabidopsis, tomato (*Solanum lycopersicon*), Brachypodium (*Brachypodium distachyon*), and barley (*Hordeum vulgare*), we can now show that all LLM domain-containing GATAs belong to the B-GATA family and that the presence of the LLM domain is suitable criterion for the functional subclassification of B-GATAs. Furthermore, we show that LLM domain-containing B-class GATAs from different species are functionally redundant and that the LLM domain is required for some but not all biological functions of these GATAs.

RESULTS

Structural Analysis of Dicot and Monocot B-Class GATA Transcription Factors

We have previously identified and described the LLM domain as a conserved domain of the B-GATAs

GNC and GNL (Richter et al., 2010). To understand the structure and conservation of this class of GATAs and the LLM domain in plants, we searched for B-GATAs and LLM domain containing proteins in the genome databases of the dicot species Arabidopsis and tomato and the monocot species Brachypodium and barley (AGI, 2000; IBI, 2010; Mayer et al., 2012; TGC, 2012). In all four genomes, we identified B-GATAs with or without an LLM domain, but we did not identify any LLM domain containing proteins outside of the B-GATA family (Fig. 1; Supplemental Fig. S1). In each case, the LLM domain was positioned at the very C terminus of the B-GATAs, and the proteins shared substantial amino acid conservation within the LLM domain (69.9%) as well as within the GATA DNA-binding domain (72.8%; Fig. 1A; Supplemental Fig. S2). By contrast, there was only very limited sequence conservation between the N termini or in the region between the GATA domain and the LLM domain (Fig. 1A; Supplemental Fig. S3). In all species, the B-GATAs could furthermore be subdivided into long B-GATAs with an extended domain of between 74 (AtGATA19) and 230 (GNC) amino acids N-terminal to the GATA DNA-binding domain and short B-GATAs with an N-terminal domain ranging from one (SlGATA1) to 66 (SlGATA2) amino acids (Fig. 1A). Whereas the LLM domain was present in long as well as short B-GATAs, B-GATAs without an LLM domain always belonged to the family of long B-GATAs. Based on current genome annotations and gene predictions, at least two representatives for each of these three B-GATA subclasses from the four species examined were identifiable, with barley being the only exception where we could identify only one long B-GATA without an LLM domain. We concluded that the B-GATA transcription factor family has a comparable complexity in different species and can be classified into three structural categories: short or long B-GATAs with an LLM domain as well as long B-GATAs without an LLM domain. Because the GATA factors from the three non-Arabidopsis species have not yet been given trivial names, we introduced a numerical nomenclature for these B-GATAs similar to the nomenclature used in Arabidopsis (Fig. 1B). According to this classification, GNC and GNL are the only long B-GATAs with an LLM domain in Arabidopsis. In addition, the Arabidopsis genome encodes four long B-GATAs without an LLM-domain: AtGATA18 (AT3G50870/HAN), AtGATA19 (AT4G36620/HANL2), AtGATA20 (AT2G18380), and AtGATA29 (At3G20750). Five other Arabidopsis proteins are short B-GATAs with an LLM domain: AtGATA15 (AT3G06740), AtGATA16 (AT5G49300), AtGATA17 (AT3G16870), AtGATA17-LIKE (AtGATA17L; AT4G16141), an orphan GATA closely related to AtGATA17, and AtGATA23 (GATA23; AT5G26930). We noted with interest that AtGATA23, a short LLM domain B-GATA where a Cys (C) replaces the first Leu (L) of the LLM motif, was the only B-GATA with a degenerate LLM motif. Surprisingly, there were no obvious orthologs of AtGATA23 with similar features in the three non-Arabidopsis species examined here (Fig. 1B).
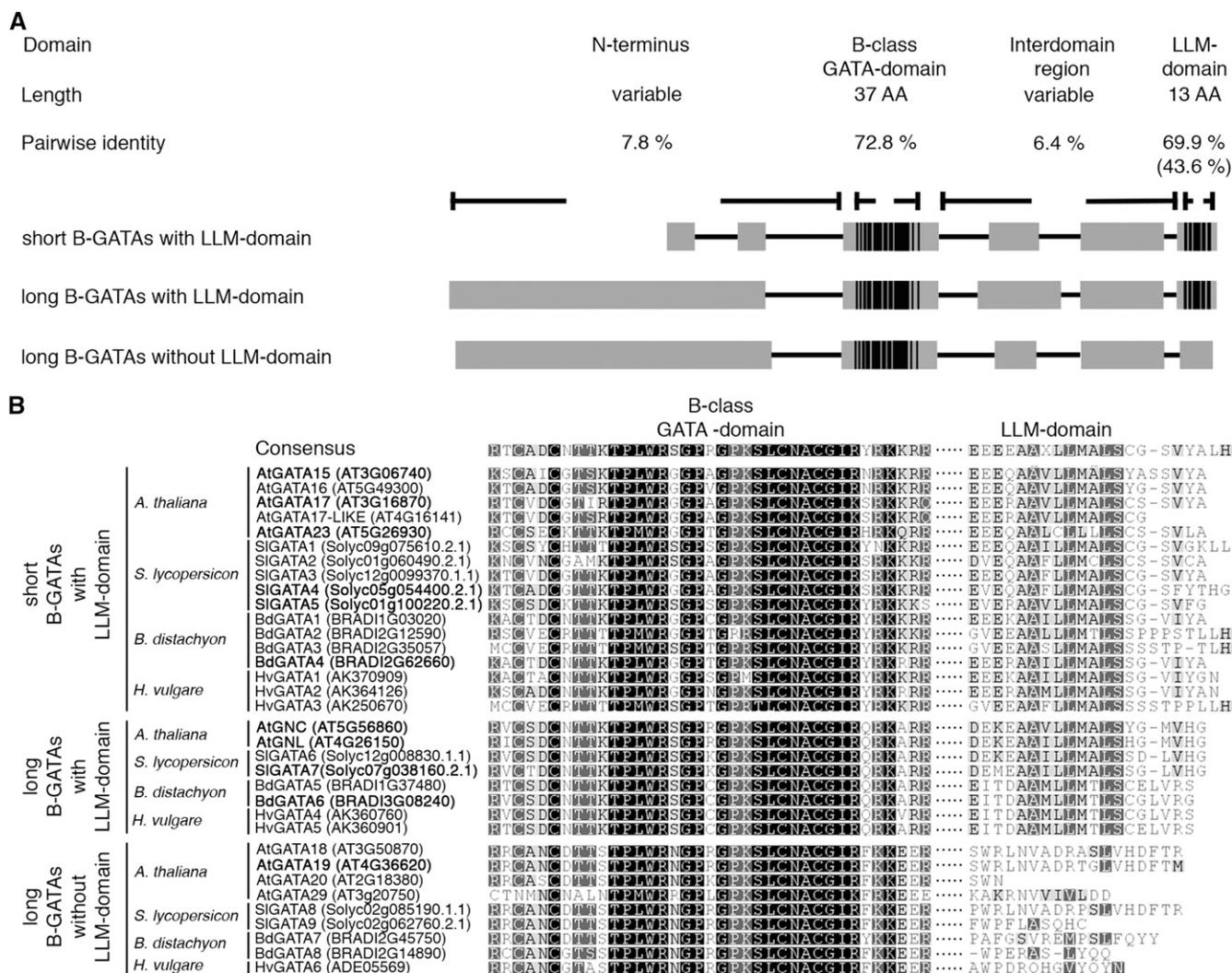
Figure 1. The B-class GATA transcription factor family can be subdivided based on structural criteria. A, Schematic representation of the three structural groups of B-GATAs that are discernable within the genomes of Arabidopsis, tomato, Brachypodium, and barley. Domain names, domain lengths (amino acids [AA]), and pairwise identities are shown. For the LLM domain, the value in the bracket indicates the overall homology within the C-terminal region among all B-GATAs, whereas the value without bracket describes the homology within the group of LLM domain-containing B-GATAs. Gray boxes indicate restricted homology; black boxes indicate protein regions of high homology; and black lines indicate protein regions of restricted homology. B, Protein sequence comparisons of the B-GATA domain and the LLM domain. Within each group of short and long B-GATAs with or without the LLM domain, the proteins are ordered according to plant species and not according to their similarities. Their phylogenetic relationship and partial and full protein sequence alignments are provided in Supplemental Figures S1 to S3. The consensus sequence shows the most frequent amino acid at the respective position; otherwise, X is shown.

Short and Long LLM Domain B-GATAs Are Functionally Redundant

Because the length of the N terminus is one distinguishing feature of the individual members of the B-GATA family, we examined the contribution of the short and long N termini to B-GATA function. Previous analyses of the long LLM domain containing B-GATAs GNC and GNL had shown that their loss-of-function mutants have greening phenotypes opposite to those observed in the overexpression lines GNC (GNCox) and GNL (GNLox; Bi et al., 2005; Richter et al., 2010; Hudson

et al., 2011). While *gnc* single and *gnc gnl* double loss-of-function mutants are light green and have decreased chlorophyll levels, GNCox and GNLox seedlings are dark green, accumulate chlorophyll, and have epinastic unexpanded cotyledons (Fig. 2A; Richter et al., 2010). In addition, the GNCox and GNLox overexpressors have other strong phenotypes such as a delay in germination and a hypersensitivity to the GA biosynthesis inhibitor paclobutrazol (PAC) as well as a strong delay in flowering (Richter et al., 2010, 2013a). To understand to what extent short and long LLM domain-containing B-GATAs are functionally redundant, we compared overexpression

lines of the short B-GATAs *AtGATA15* (AtGATA15ox) and *AtGATA17* (AtGATA17ox) with the long B-GATA overexpressors GNCox and GNLox. Phenotypic analyses revealed that plants overexpressing all four GATAs have very similar phenotypes (Fig. 2, A and B). For each of the four transgenes, we identified weak and strong overexpressors, and in each case, the strong accumulation of the transgenic protein correlated with an enhancement of the different phenotypes, including increased chlorophyll accumulation, enhanced epinasty of the cotyledons, and severely delayed flowering (Fig. 2, B and C). Additionally, we noted that the strong overexpressors
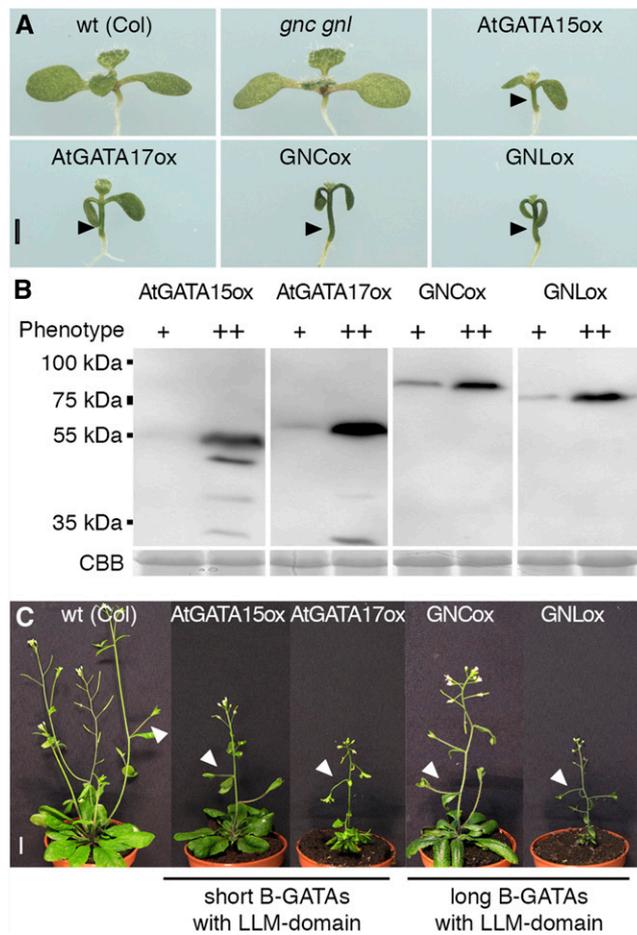
Figure 2. Short and long LLM domain B-GATAs are functionally redundant. A, Representative photographs of 9-d-old light-grown Arabidopsis seedlings of the wild type, the *gnc gnl* mutant, and strong overexpression lines of the four Arabidopsis LLM domain-containing B-class GATAs. Please note the particularly strong chlorophyll accumulation at the base of the seedling hypocotyls in the overexpressors as indicated by the arrowheads. Bar = 1 mm. B, Immunoblots with an anti-HA antibody to detect transgene protein expression in the overexpression lines with a weak (+) and a strong (++) phenotype, respectively. The seedlings shown in A correspond to seedlings with a strong phenotype. CBB, Coomassie Brilliant Blue (protein-loading control). C, Representative photographs of 6-week-old Arabidopsis plants as specified in the figure. The arrowheads point at the angles between the primary inflorescence and a lateral inflorescence, which is increased in the B-GATA overexpressors. Bar = 1 cm. wt, Wild type; Col, Columbia.

had an increased angle between the primary inflorescence and lateral inflorescences, a phenotype of these overexpressors that had gone unnoticed during our previous analyses of *GNC* and *GNL* overexpressors (Fig. 2, A–C).

When germinated on the GA biosynthesis inhibitor PAC, seeds of strong overexpressors showed a significant delay in germination when compared with the wild type (Fig. 3A). In addition, light-grown overexpressor seedlings had an elongated hypocotyl compared with the wild type, another phenotype that had gone unnoticed during our previous analyses of the GNCox and GNLox overexpressors (Fig. 3B). Finally, chlorophyll measurements showed that the overexpressor seedlings accumulated high levels of chlorophyll when compared with the wild type (Fig. 3C). Multiple independent transgenic lines were analyzed for each of the overexpression constructs, and the results obtained with representative strong lines are shown in each case. Based on the analysis of multiple lines, we judge that the quantitative differences between the individual overexpression lines reflect differences in transgene expression or may inversely also be the result of transgene silencing. In summary, we concluded that the overexpression of long and short LLM domain-containing B-GATAs gives rise to highly similar if not identical growth phenotypes.

To understand whether the phenotypic similarities between the various overexpression lines correlate with similarities at the gene expression level, we performed a comparative microarray analysis with seedlings overexpressing the short LLM domain B-GATA AtGATA17 (AtGATA17ox) and the long LLM domain B-GATAs GNC (GNCox) and GNL (GNLox). Here, we found a substantial overlap in the identity of genes that were differentially regulated in the overexpression lines of the three GATAs when compared with the ecotype Columbia wild type (Fig. 4), e.g. 62% (1,492 genes) of the 2,399 genes that were down-regulated in AtGATA17ox were also down-regulated in GNCox or GNLox when compared with the wild type. Inversely, 42% (1,041 genes) of the 2,499 genes that were up-regulated in AtGATA17ox were also up-regulated in GNCox or GNLox (Fig. 4). In this respect, the overexpression line AtGATA17ox was as similar to GNCox or GNLox as GNCox and GNLox were to each other. Because all our previous analyses suggested that GNC and GNL are functionally redundant, we concluded that the short LLM domain B-GATAs are functionally redundant with the long LLM domain B-GATAs, at least with regard to the phenotypes examined here. This conclusion found its support in our subsequent finding that *AtGATA17* as well as *GNC*, when expressed under the control of a *GNC* promoter fragment (P_{GNC}), were able to suppress the greening defect of the *gnc gnl* double mutant (Fig. 5).

The Function of LLM Domain B-GATAs Is Conserved across Species

To test whether the LLM domain B-GATAs from the other non-Arabidopsis species are functionally redundant
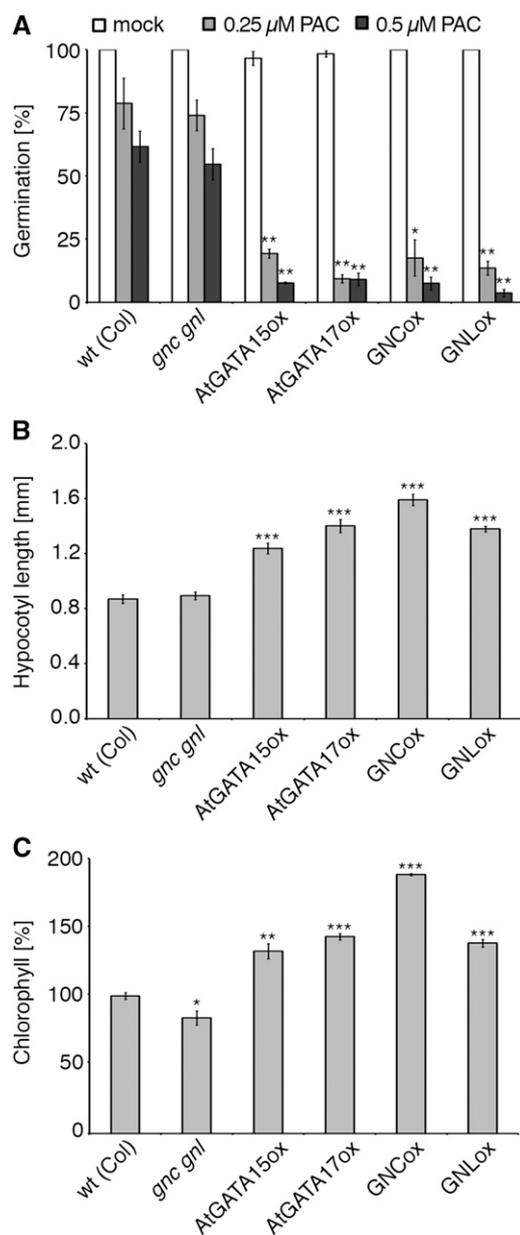
Figure 3. Physiological analysis of overexpression lines of short and long LLM domain B-GATAs. A, Relative germination rate of seeds from the wild type, *gnc gnl* mutant, and B-GATA overexpression lines grown on one-half-strength MS medium and on medium supplemented with the GA biosynthesis inhibitor PAC after 4 d of stratification and 3 d of growth at 21°C as determined based on endosperm rupture ($n \geq 25$). B, Quantitative analysis of hypocotyl elongation in 8-d-old light-grown seedlings ($n \geq 20$). C, Quantitative analysis of chlorophyll accumulation in 9-d-old light-grown seedlings ($n \geq 10$). Averages and sEs are provided in each case. Student's *t* tests were performed compared with the wild type (*$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$). wt, Wild type; Col, Columbia.

with their Arabidopsis counterparts, we overexpressed selected short and long LLM domain-containing B-GATAs from tomato and Brachypodium in Arabidopsis, namely the short *SlGATA4*, *SlGATA5*, and *BdGATA4* as well as the long *SlGATA7* and *BdGATA6*. In each case, B-GATA overexpression resulted in phenotypes identical to those

observed after overexpression of the Arabidopsis LLM domain B-GATAs (Supplemental Fig. S4, A and B). Transgenic seedlings and plants were dark green, accumulated high levels of chlorophyll, had epinastic cotyledons, were sensitive to the inhibition of germination by PAC treatment, and had elongated hypocotyls when grown in the light (Supplemental Fig. S4, C–E). We also performed a genome-wide gene expression analysis with seedlings overexpressing *BdGATA6* as a representative non-Arabidopsis long LLM domain B-GATA (Supplemental Fig. S4F). Again, the overlap among the down- or up-regulated genes between the BdGATA6ox line and GNCox or GNLox was as strong as it was between the GNCox and GNLox, indicating that overexpression of these three B-GATAs results in comparable molecular phenotypes in Arabidopsis. Compared with GNCox and GNLox, we found that 62% (457) of the 735 genes that were down-regulated in BdGATA6ox were also down-regulated in at least one, GNCox or GNLox (Supplemental Fig. S4F). Similarly, 53% (663) of the 1,191 genes up-regulated in BdGATA6ox were also up-regulated in GNCox or GNLox (Supplemental Fig. S4F). Furthermore, we could show that the expression of *BdGATA6* under control of P_{GNC} complemented the greening defect of the *gnc gnl* mutant (Fig. 5). Based on these analyses, we concluded that short and long LLM domain B-GATAs from tomato and Brachypodium are functionally redundant with their Arabidopsis counterparts.

B-GATAs Are Unstable Proteins

Many regulatory proteins are subject to constitutive or induced degradation by the ubiquitin-proteasome system (Vierstra, 2012). We therefore tested whether B-GATAs are also regulated by ubiquitin-dependent protein degradation. To this end, we incubated B-GATA-overexpressing seedlings with the protein synthesis inhibitor cycloheximide (CHX) and examined the abundance of the transgenic B-GATAs over time in immunoblots using an anti-hemagglutinin (HA) antibody directed against the C-terminal YELLOW FLUORESCENT PROTEIN (YFP)-HA-tag of the overexpressed proteins. In these analyses, we reproducibly detected a significant decrease in the abundance of the LLM domain-containing B-GATAs after CHX treatment that became already obvious after 15 to 30 min (Fig. 6; Supplemental Fig. S5). Further experiments then showed that the 26S proteasome inhibitor MG132 [for *N*-(benzyloxycarbonyl)leucinylleucinylleucinalZ-Leu-Leu-Leu-al] could block this decrease in protein abundance and that MG132 treatment alone resulted in the accumulation of the proteins (Fig. 6). Based on these data, we concluded that the B-class GATAs are subject to proteasomal turnover in light-grown seedlings.

The LLM Domain of GNC and GNL Regulates Plant Growth

We next examined the relevance of the LLM domain for B-class GATA function. To this end, we generated
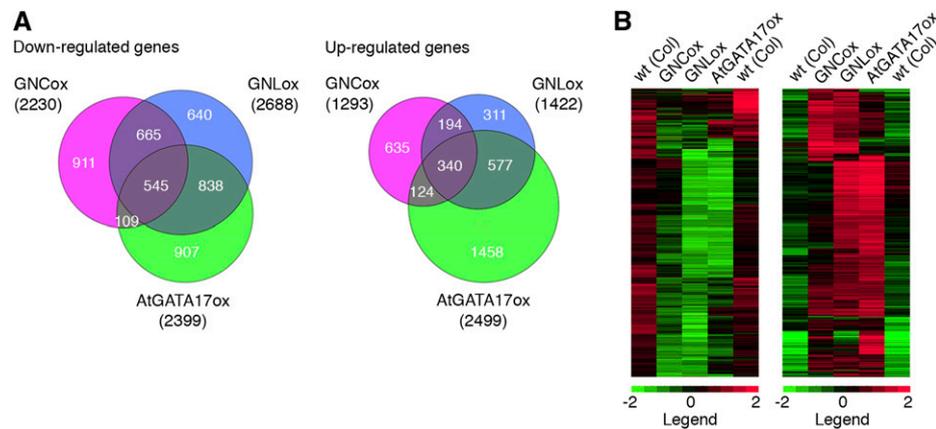
Figure 4. Short and long LLM domain B-GATAs have similar gene expression profiles. A, Venn diagrams of the genes down- and up-regulated in overexpression lines of GNC, GNL, and AtGATA17. Indicated are the numbers of differentially expressed genes in the respective genotypes and their intersections. The total number of differentially regulated genes is indicated for each genotype in brackets. B, Heat map of the genes that are up- or down-regulated in at least two overexpression lines out of GNCox, GNLox, and AtGATA17ox when compared with the wild-type (wt, Columbia [Col]) control. The experiment was performed in two rounds, and two separate wild-type control samples were therefore analyzed. A list of the differentially regulated genes used for these comparisons is provided in Supplemental Table S1.

transgenic plants overexpressing B-GATAs with a deletion or mutation of the LLM domain. In GNCΔLLMox, we deleted the entire 13-amino acid LLM domain together with the residual C-terminal residues (Fig. 7A). In GNC_LLM/AAAox and GNL_LLM/AAAox, we replaced the LLM motif by three Ala residues (AAA; Fig. 7A). Interestingly, transgenic lines expressing these mutant B-GATA variants phenotypically resembled lines expressing the wild-type constructs with regard to the greening and germination phenotypes as well as the altered lateral inflorescence angle phenotype and the delays in flowering and senescence (Fig. 7, B–D and F; Supplemental Figs. S6–S8). By contrast, however, overexpressors with a mutation or deletion of the LLM motif did not display the elongated hypocotyl phenotype that was observed in the wild-type overexpressors (Fig. 7E), indicating that the LLM domain may be specifically required for the regulation of hypocotyl elongation but may be dispensable for the control of germination, greening, flowering, plant architecture, or senescence. In turn, we noted changes in leaf morphology in the LLM domain-mutated overexpressors that were not apparent in the wild-type overexpressors. Specifically, the leaves of the LLM domain-mutated overexpressors were rounder than leaves of the wild type or of wild-type overexpressors, and their petioles were strongly shortened (Fig. 8). Furthermore, we found that a substantial number of genes that were differentially regulated in both GNCox and GNLox wild-type overexpressors was not differentially expressed in GNC_LLM/AAAox and GNL_LLM/AAAox and that the overall number of differentially expressed genes was strongly reduced in the transgenic lines expressing these LLM domain-mutated variants (Fig. 9). We thus concluded that there are substantial differences between the wild type and the mutant overexpressors at the gene expression level. Because the

overexpression lines of the mutant variants contained rather more than less transgenic protein than overexpression lines of the wild-type proteins, we judge, at the same time, that these gene expression differences cannot be explained by differences in transgene protein abundance (Fig. 7C). Differences between the wild-type and mutant GNC proteins also became apparent when we examined the ability of the mutant GNC_LLM/AAA to complement the *gnc gnl* mutant phenotype. Here, we observed that the LLM-mutated construct was unable to complement the greening defects of *gnc gnl* double mutants, indicating that GNC requires a functional LLM domain when the protein is expressed under control of an endogenous promoter fragment that is sufficient to control the expression of the wild-type protein (Fig. 10). Thus, the LLM domain is required for full B-GATA function of the LLM domain-containing GNC and GNL proteins. Furthermore, there are phenotypic differences between the LLM domain-mutated and wild-type transgenic lines with regard to hypocotyl elongation and leaf formation phenotypes of the overexpressors and the chlorophyll accumulation phenotype when the LLM domain-mutated transgenes are expressed from a *GNC* promoter fragment. This apparent relevance of the LLM domain for B-GATA function is also substantiated by the gene expression differences that we observed in the microarray analyses.

In our efforts to understand the role of the LLM domain, we also examined the possibility that the LLM domain mutations may affect the proteasomal degradation of the B-GATAs. However, we found no evidence for a role of the LLM domain in protein degradation when we compared the degradation of the overexpressed GNC protein with that of GNC_LLM/AAA or GNCΔLLM (Supplemental Fig. S9). Because it had previously been reported that GNC and GNL can interact with
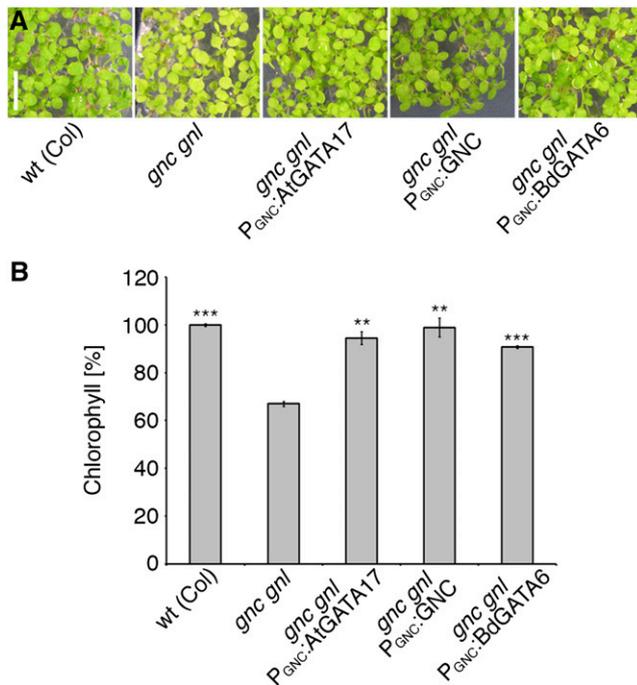
Figure 5. Complementation of the *gnc gnl* chlorophyll phenotype by AtGATA17 and BdGATA6. A, Representative photographs of 9-d-old seedlings of the genotypes as indicated in the figure. B, Quantitative analysis of chlorophyll accumulation in 9-d-old light-grown seedlings ($n \geq 10$). Averages and SES are indicated in each case. Student's *t* tests were performed compared with the *gnc gnl* mutant (**$P \leq 0.01$; ***$P \leq 0.001$). wt, Wild type; Col, Columbia.

the B-GATA HAN in the yeast two-hybrid system, we also examined whether GNC and GNL can interact in the yeast system. However, GNC and GNL did not interact in these experiments, arguing against a role of the LLM domain as a homo- or heterodimerization domain for these B-GATAs (Supplemental Fig. S10).

AtGATA23 and AtGATA19 Are Distinct from LLM Domain-Containing GATAs

We next turned our interest to AtGATA23 and AtGATA19. We chose AtGATA23 because it is an LLM domain-containing B-GATA with a degenerated LLM domain and AtGATA19 as a representative long B-GATA without an LLM domain (Figs. 1 and 7A; Supplemental Figs. S1–S3). We analyzed these B-GATAs also because both AtGATA23 and AtGATA19 (HANL2) had previously been implicated in specific biological processes. AtGATA23 is an early marker gene for lateral root initiation, and AtGATA18/HAN together with AtGATA19/HANL2 has a described role in flower development (Zhao et al., 2004; De Rybel et al., 2010; Zhang et al., 2013). While our attempts failed to generate overexpression lines for the better studied HAN, we were successful in generating overexpression lines for AtGATA19/HANL2.

Importantly, neither AtGATA23ox nor AtGATA19ox lines displayed the full range of phenotypes that was

typical for the overexpressors of the LLM domain-containing B-GATAs (Fig. 7). While we failed to identify an AtGATA23ox line that expresses the protein at levels comparable to those of the other B-GATA overexpressors, AtGATA23ox still contained significantly more chlorophyll than the wild type (Fig. 7, B and F). By contrast, however, there was no chlorophyll accumulation visible at the base of the hypocotyl. Furthermore and in contrast to the wild type as well as the other LLM domain B-GATA overexpressors, AtGATA23ox was insensitive to the inhibition of germination by PAC (Fig. 7D), but similarly to the other LLM domain B-GATA overexpressors, AtGATA23ox had an elongated hypocotyl when compared with the wild type (Fig. 7E). Hypocotyl elongation in the AtGATA23 overexpressors was however minor when compared with the GNC or GNL overexpressors (Fig. 7E). Thus, AtGATA23ox lines display a differential spectrum of overexpression phenotypes than the other B-GATA overexpression lines, although our possibilities to properly compare the AtGATA23ox lines are limited due to the reduced transgene expression of AtGATA23ox.

In contrast to the chlorophyll accumulation phenotype of the other B-GATA overexpression lines, strong AtGATA19 overexpression lines had reduced rather than increased chlorophyll levels also when compared with the wild type and even the *gnc gnl* mutants (Fig. 7, B and F). Furthermore, AtGATA19ox was similarly sensitive to the germination inhibitory effects of PAC treatment, as were GNCox and GNLox lines (Fig. 7D), but, unlike GNCox and GNLox, had no apparent effect on hypocotyl elongation (Fig. 7E). In summary, we concluded that the overexpression of AtGATA19 and AtGATA23 gives rise to a phenotypic spectrum that is distinct from that observed after overexpression of the LLM domain-containing B-GATAs. In summary, these observations support the idea that the presence or absence of the LLM domain contributes to the functional identity of the B-GATA transcription factors.
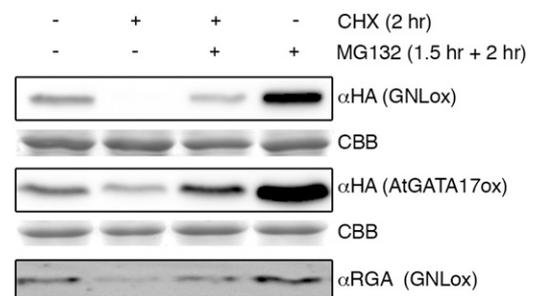
Figure 6. B-class GATAs are short-lived proteins. Anti-HA immunoblots of total protein extracts prepared from 9-d-old seedlings from GNLox and AtGATA17ox overexpression lines that had been pretreated for 1.5 h with the 26S proteasome inhibitor MG132 (100 μM) and then treated for an additional 2 h with the protein synthesis inhibitor CHX (50 μM). The anti-RGA immunoblot detects the abundance of the endogenous RGA protein, a DELLA protein, and known proteasomal degradation target in the GNLox samples (Willige et al., 2007). CBB, Coomassie Brilliant Blue (protein-loading control).
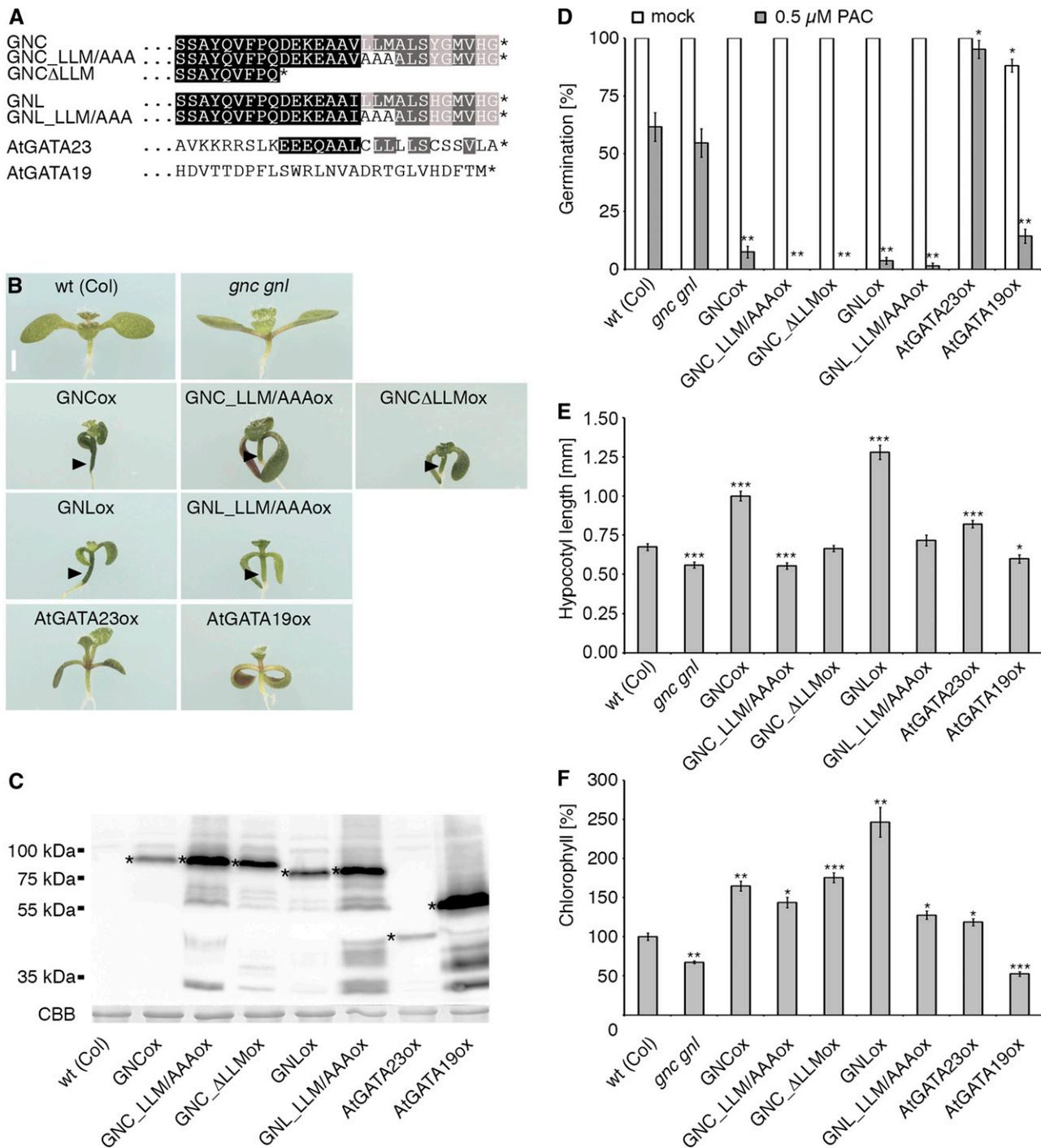
Figure 7. The LLM domain contributes to B-GATA function. A, Protein alignment of the C termini of the specified B-GATAs and the LLM domain mutant variants as used in this part of the study. B, Representative photographs of 9-d-old light-grown Arabidopsis seedlings of the wild type and strong overexpression lines carrying the overexpression transgenes specified in the figure. Note the particularly strong chlorophyll accumulation at the base of the seedling hypocotyls in the GNCox and GNLox overexpressors that is not visible in the AtGATA23ox and AtGATA19ox lines as indicated by the arrowheads. Bar = 1 mm. C, Immunoblot with an anti-HA antibody to detect transgene protein expression in the overexpression lines that are shown in B. CBB, Coomassie Brilliant Blue (protein-loading control). The respective transgenic proteins are marked by asterisks. D, Relative germination rate of seeds from the wild type, *gnc gnl* mutant, and B-GATA overexpression lines grown on one-half-strength MS medium and on medium supplemented with the GA biosynthesis inhibitor PAC after 4 d of stratification and 3 d of growth at 21°C as determined based on endosperm rupture. The control samples are identical to those shown in Figure 3A ($n \geq 25$). E, Quantitative analysis of hypocotyl elongation in 8-d-old light-grown seedlings ($n \geq 20$). F, Quantitative analysis of chlorophyll accumulation in 9-d-old light-grown seedlings ($n \geq 10$). Averages and SES are shown in each case. Student's *t* tests were performed compared with the wild type (*$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$). wt, Wild type; Col, Columbia.
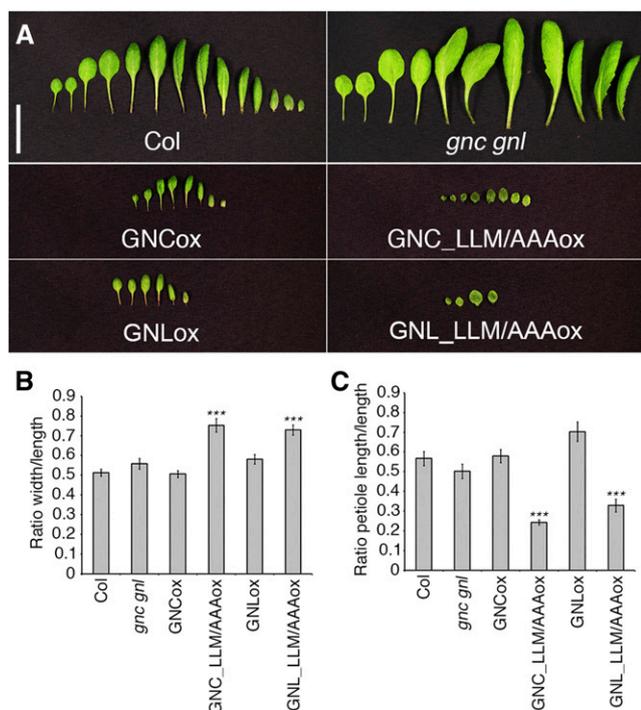
Figure 8. The LLM domain determines leaf shape and petiole length of GNC/GNL overexpressors. A, Representative photographs of dissected leaves from 21-d-old Arabidopsis plants of the specified genotypes. Bars = 2 cm. B and C, Ratios between maximum leaf width and leaf length (B) or between petiole length and leaf length (C; *n* > 10). Averages and sEs are provided in each case. Student's *t* tests were performed compared with the wild type (***$P < 0.001$). Col, Columbia. [See online article for color version of this figure.]

DISCUSSION

We have functionally analyzed representative members of the plant family of B-GATA transcription factors. B-GATAs had previously been defined based on the overall sequence conservation between the individual family members in Arabidopsis and rice and the absence of any other recognizable protein sequence features, which distinguishes B-GATAs from the other GATA families (Reyes et al., 2004). Based on this previous classification, the Arabidopsis B-GATA family from Arabidopsis comprised 10 members. Here, we identified AtGATA17L as an additional B-GATA closely related to AtGATA17. As a result of the comparative sequence analysis of B-GATAs from Arabidopsis, tomato, Brachypodium, and barley, we could further subdivide the B-GATAs into short and long B-GATAs with an LLM domain but also long B-GATAs without an LLM domain (Fig. 1). Our subsequent analyses aimed at identifying the degree of functional redundancy between representative members of the subfamilies within the B-GATA family to understand the contribution of their N termini and the LLM domain to B-GATA protein function.

Previous analyses had shown that the overexpression of *GNC* and *GNL* results in a number of strong phenotypes that we now used for comparisons with other B-GATA family members (Richter et al., 2010, 2013a, 2013b; Hudson et al., 2011; Chiang et al., 2012). Importantly, several of these phenotypes are opposite to those observed in *gnc gnl* loss-of-function mutants (Bi et al., 2005; Richter et al., 2010, 2013a, 2013b; Hudson et al., 2011; Chiang et al., 2012). Respectively, these include increased and decreased chlorophyll biosynthesis, increased and decreased germination efficiency, and delayed and accelerated flowering in the overexpressors and loss-of-function mutants. Because our previous studies had shown that most of the phenotypes of the *gnc gnl* loss-of-function mutants are, with the exception of the greening phenotype (Bi et al., 2005), only prominent in specific genetic backgrounds such as the GA-deficient *ga1* mutant or the auxin-signaling impaired *arf2* mutant (Richter et al., 2010, 2013a, 2013b), we opted to use the phenotype of B-GATA overexpressors as a primary criterion to assess the functional redundancy between the individual B-GATAs tested. As a result of our phenotypic comparisons of overexpressors of short and long B-GATAs from Arabidopsis, tomato, and Brachypodium with *GNC* and *GNL* overexpressors, we could show that the length of the N terminus does not allow distinguishing between these B-GATAs with regard to the phenotypes examined in our study (Figs. 2–4; Supplemental Fig. S4). Along the same lines, we found that the overexpression of the short LLM domain-containing *AtGATA17* or the long Brachypodium LLM domain *BdGATA6* induces similar gene expression changes as the overexpression of *GNC* and *GNL* (Fig. 4; Supplemental Fig. S4). Finally, we could provide proof for the functional redundancy of AtGATA17 and BdGATA6 with GNC and GNL by demonstrating that both B-GATAs suppress the greening defect of the *gnc gnl* double mutant when expressed from the *GNC* promoter fragment (Fig. 5). In summary, our findings suggest that short and long LLM domain-containing B-GATAs are functionally redundant.

We further addressed the relevance of the LLM domain for B-GATA function and overexpressed *GNC* and *GNL* with mutation or deletions of the LLM motif in Arabidopsis. Interestingly, overexpressors of these mutant variants did not display the hypocotyl elongation phenotype that we had identified in the wild-type overexpressors but shared with the wild-type overexpressors the greening phenotype as well as the germination defect when seeds were grown on PAC (Fig. 7). When comparing the gene expression profiles of the wild type and mutant *GNC* and *GNL* overexpressors, we found that the overall number of differentially expressed genes was strongly reduced in the case of the GNC_LLM/AAA or GNL_LLM/AAA overexpressors (Fig. 9). Furthermore, the majority of genes that we had identified as differentially regulated in both the *GNC* and *GNL* wild-type overexpressors were not differentially expressed in the LLM domain-mutated overexpressors. Based on these criteria, we judge that the LLM domain is essential for full B-GATA function and that it is responsible for a majority of the gene expression changes observed in the overexpression lines of the wild-type
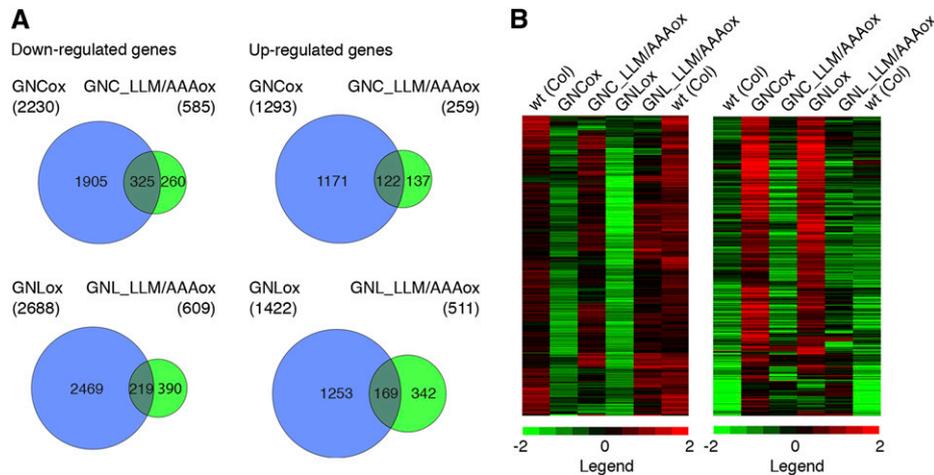
Figure 9. LLM domain-specific expression profile analysis. A, Venn diagrams of the genes down- and up-regulated in overexpression lines of *GNC*, *GNL*, and their *GNC_LLM/AAA* and *GNL_LLM/AAA* mutant counterparts. Provided are the total numbers of differentially regulated genes for each genotype (in brackets) and in their intersections. B, Heat map showing the differential regulation of the genes that are up- or down-regulated in GNCox and GNLox when compared to the wild-type (wt, Columbia [Col]) control and their expression in the GNC_LLM/AAAox and GNL_LLM/AAAox lines as indicated in the figure. The experiment was performed in two rounds, and two separate wild-type control samples were analyzed and are therefore also shown. The list of the differentially regulated genes used for this analysis is provided in Supplemental Table S1.

B-GATAs *GNC* and *GNL*. At the phenotypic level, these gene expression changes led to detectable alterations in hypocotyl elongation but also to the appearance of a novel leaf formation phenotype specific for the LLM domain-mutated version (Fig. 8). Interestingly, the LLM/AAA variants retained the ability to induce chlorophyll accumulation when overexpressed, while the GNC_LLM/AAA variant was unable to rescue the chlorophyll accumulation defect of the *gnc gnl* mutant when expressed from a *GNC* promoter fragment. We thus suggest that the effect of *GNC* and *GNL* expression on chlorophyll biosynthesis may be controlled in an LLM domain-dependent as well as in an LLM domain-independent manner and that this differential activity may depend on protein dosage or the respective gene expression domains.

Our observation that the LLM domain contributes to the identity of this B-GATA protein family was also substantiated in the studies of AtGATA19, a long Arabidopsis B-GATA without an LLM domain. AtGATA19 is also known as HANL2, a protein closely related to the floral development regulator HAN. Because we failed in repeated attempts to generate overexpressors of the better characterized *HAN*, we examined *AtGATA19/HANL2* overexpression plants. Unlike the overexpressors of the LLM domain-containing GATAs, *AtGATA19* overexpression did not induce chlorophyll overexpression and also did not promote hypocotyl elongation (Fig. 7). Thus, AtGATA19ox seedlings had a phenotypic spectrum that allowed distinguishing them from any of the other overexpression lines examined here. This observation does not, however, allow excluding that B-GATAs without and with an LLM domain act in concert in the regulation of developmental or physiological responses. In fact, a recent study of *han* mutants in combination with

gnc and *gnl* mutations indicated that the phenotype of the *han* mutant is enhanced in the absence of *GNC* and *GNL* (Zhang et al., 2013). Furthermore, it was found that HAN
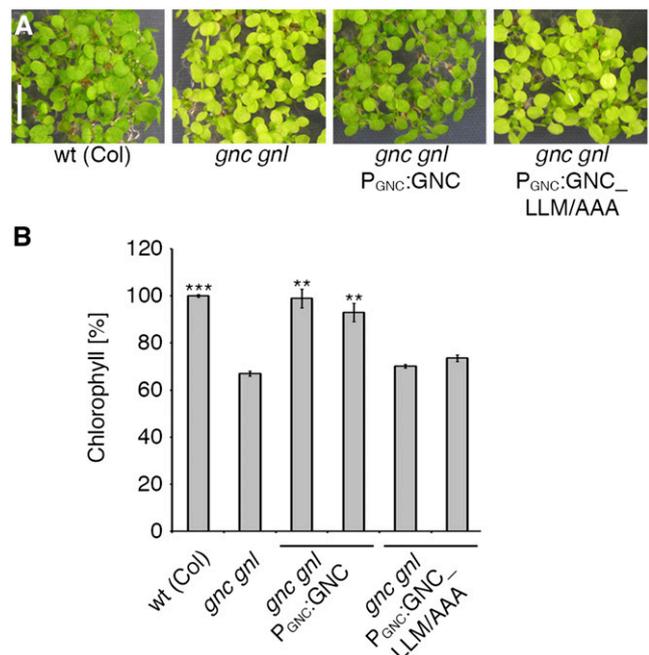
Figure 10. Complementation of the *gnc gnl* chlorophyll phenotype is dependent on the LLM domain. A, Representative photographs of 9-d-old seedlings of the genotypes as indicated in the figure. B, Quantitative analysis of chlorophyll accumulation in 9-d-old light-grown seedlings. Two independent transgenic lines were analyzed ($n \geq 10$). Averages and SES are shown. Student's *t* tests were performed compared with the *gnc gnl* mutant (**$P \leq 0.01$; ***$P \leq 0.001$). wt, Wild type; Col, Columbia.

interacts with GNC and GNL in the yeast two-hybrid system and that the expression of *GNC* and *GNL* was altered in the *han* background, suggesting that the loss of HAN may be partially suppressed through an up-regulation of these two B-GATAs (Zhang et al., 2013). In contrast to the findings and conclusions from this earlier study, our findings with HANL2 now suggest that this B-GATA has a function that is distinct from that of GNC and GNL, which of course does not exclude that GNC or GNL functionally interacts with HAN at the genetic or the protein level.

More difficult is the interpretation of the results that we obtained with AtGATA23. Among the four species examined here, AtGATA23 from Arabidopsis is a short B-GATA and the only protein with a degenerate LLM domain. At the phenotypic level, the overexpression of *AtGATA23* induces a slight increase in chlorophyll levels, a slight elongation of the hypocotyl, and, in contrast to the other overexpressors and the wild type, a PAC-insensitive germination phenotype (Fig. 7). Thus, with regard to these phenotypes, AtGATA23 is functionally distinct from the other LLM domain-containing B-GATAs. However, the fact that the strongest AtGATA23ox lines did not express the protein as strongly as the other available B-GATA overexpressors makes it ultimately difficult to interpret these phenotypes compared with those of the other B-GATA overexpressors. Importantly, AtGATA23 clusters in our phylogenetic analysis with the long B-GATAs BdGATA2, BdGATA3, and HvGATA3 that have a perfectly conserved LLM motif and LLM domain (Fig. 1; Supplemental Figs. S1–S3). In turn, our search for AtGATA23 orthologs with a degenerate LLM domain resulted in the identification of such AtGATA23 orthologs only in the Brassicaceae species *Arabidopsis lyrata* (XP_002872245), *Capsella rubella* (XP_006289523), and *Eutrema salsugineum* (XP_006394919) but not from non-Brassicaceae species. This finding, seen in the context of our other results, could indicate that AtGATA23 may be derived from bona fide LLM domain-containing B-GATAs and acquired a specific biochemical function during the evolution of the Brassicaceae.

We have also attempted to explain the various phenotypes and phenotypic differences at the molecular level through analysis of the available set of gene expression data, e.g. by examining the differential regulation of known key regulators of hypocotyl elongation and greening (Supplemental Table S1). However, we have failed as yet to identify good candidates for genes whose differential regulation would allow us to rationalize the observed physiological phenotypes at the molecular level. Thus, the identification of specific direct downstream targets of the B-GATAs as well as the identification of LLM domain-specific targets awaits further analyses.

The LLM domain is a protein domain of unknown function. Sequence comparisons of the LLM domains from the four plant species allowed us to define the sequence EEEEAAXLLMALS as the consensus of this 13-amino acid domain (Fig. 1B). In Arabidopsis and outside of the B-GATA protein family, a protein sequence related to the LLM domain is only present in

AT3G29140 as EEEqAAVLLMqLS (amino acids diverging from the consensus are shown in lowercase letters). AT3G29140 is a protein of unknown function without any recognizable protein domains that would allow us to draw any interpretable conclusions about the function of the protein or its LLM domain sequence. In turn, structural predictions using the EMBOSS algorithm (http://emboss.sourceforge.net) indicate that the LLM domain forms an α-helix, a protein structure suitable for interactions with other protein domains. Thus, the LLM domain is likely a protein-protein interaction domain. In this context, our experimental findings allow excluding the possibility that the LLM domain regulates the turnover of the protein (Supplemental Fig. S9). In fact, the observation that the loss of the LLM domain does not impair B-GATA degradation is also in line with our observations that the loss of the LLM domain results in a reduced, rather than an increased, protein accumulation in the overexpression lines. Secondly, we were unable to demonstrate intra- and intermolecular interactions between the LLM domain-containing B-GATAs in the yeast two-hybrid system, suggesting that the LLM domain may rather engage in interactions with other protein partners, possibly other transcription factors. The latter hypothesis is supported by our observation that LLM domain-mutated B-GATAs display a strong decrease in the number of differentially regulated genes and by the fact that interactions of GATAs with other transcription regulatory proteins have been described for mammalian GATA transcription factors (Mackay et al., 1998; Fox et al., 1999; Ross et al., 2012). Thus, in view of the apparent strong importance of the LLM domain for gene expression regulation, we would like to propose that the LLM domain is required for the interaction of B-GATAs with other gene expression regulatory proteins. Future research will have to reveal the identity of such regulators.

MATERIAL AND METHODS

Biological Material

All experiments were performed in Arabidopsis (*Arabidopsis thaliana*) ecotype Columbia wild-type background or *gnc* (SALK_001778) *gnl* (SALK_003995) double mutant backgrounds as previously described (Richter et al., 2010). All B-GATA complementary DNA (cDNA) clones were amplified by PCR from reverse-transcribed RNA prepared from Arabidopsis, tomato (*Solanum lycopersicon* 'Moneymaker'), and Brachypodium (*Brachypodium distachyon*) seedlings.

Protein Alignment and Phylogeny

B-GATA and LLM domain-containing proteins from tomato and Brachypodium were identified based on their homology to Arabidopsis GNC and GNL protein sequences by searching the annotated genome sequences of tomato (http://solgenomics.net/tools/blast/index.pl) and Brachypodium (http://plants.ensembl.org/Brachypodium_distachyon/). Barley (*Hordeum vulgare*) B-GATAs were identified in BLAST searches of the B-class GATAs identified from Brachypodium (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The protein alignment of the B-GATA factors was generated using the full-length protein sequences and the CLUSTALW2 algorithm at the EMBL-EBI Web site (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Pairwise identities were calculated using the Geneious software package for all sequences or, where applicable, for all B-GATAs or only the LLM domain-containing GATAs. The phylogenetic tree was generated based on an

alignment of trimmed B-GATAs using the entire B-GATA DNA-binding domain and the C termini with or without the LLM domain. The phylogenetic tree was constructed with MEGA5.05 (http://www.megasoftware.net) using the neighbor-joining method and the bootstrap method with 1,000 bootstrap replications as well as the Jones-Taylor-Thornton model with gaps/missing data treatment set to pairwise deletion.

Molecular Cloning

All B-GATA cDNA constructs were prepared using Gateway technology (Life Technologies) and the primers listed in Supplemental Table S2. cDNAs were PCR amplified from the respective species, inserted into pDONR201 or pDONR207 (Life Technologies), and, from there, inserted into the overexpression vector pEarleyGate101 for the fusion of a C-terminal YFP-HA tag to the respective protein (Earley et al., 2006). The GNCΔLLMox deletion variant was obtained by PCR-based deletion with the primers GNC-FW and GNC_ΔLLM-RV from the GNCox construct as a template. GNC_LLM/AAAox was prepared using GNCox as a template by overlap PCR with the primers GNC-FW and GNC-RV in combination with GNC_LLM/AAA-RV and GNC_LLM/AAA-FW, respectively. After purification, these PCR fragments were combined and used as template for a fusion PCR with the primers GNC-FW and GNC-RV before insertion into pDONR201 and ultimately pEarleyGate101. The same procedure was performed with a GNL-specific primer set to obtain the construct GNL_LLM/AAAox.

For P_GNC-driven GNC expression construct P_GNC:GNC, a genomic DNA fragment from PCR amplification with the primers GNCpro-FW-1 and GNC-RV-1 was obtained by inserting a genomic GNC fragment into the Gateway system-compatible cloning vector pEarleyGate 301. For P_GNC:GATA17 and P_GNC:BdGATA6, the genomic sequences for GATA17 and BdGATA6 were cloned into pJET1.2 as PCR fragments amplified from genomic DNA of the respective species using the primers GATA17-FW-2/GATA17-RV and BdGATA6-FW-2/BdGATA6-RV with an EcoRI restriction site directly upstream of the ATG start codon. Subsequently, a 2.3-kb P_GNC was amplified with the primers GNCpro-FW-2 and RV-2, cloned into pJET1.2, and, from there, subcloned as an Xho1/EcoRI fragment into the pJET1.2 vectors containing the individual B-GATA gene fragments. Finally, the linearized P_GNC-GATA gene fragments were inserted into pEarlyGate301 using Gateway technology. To obtain the LLM/AAA amino acid exchange variant of GNC, P_GNC:GNC_LLM/AAA, a mutation PCR with a 5'-phosphorylated primer was performed. All primer sequences are listed in Supplemental Table S2.

All transfer DNA constructs were directly transformed into the wild type or gnc gnl mutants using the floral dip transformation method (Clough and Bent, 1998). Multiple transgenic lines were isolated for each transgene. Strongly expressing transgenic lines were identified based on anti-HA immunoblots and analyzed phenotypically. The results obtained with one representative transgenic line are shown.

Yeast (Saccharomyces cerevisiae) two-hybrid constructs were obtained by cloning PCR-amplified and digested GNC and GNL cDNAs into the EcoR1/Sal1 sites of the vectors pGBKT7 (DNA-binding domain [DBD]) and pGADT7 (activation domain [AD]; Life Technologies). All relevant primer sequences are provided in Supplemental Table S2. The empty vector control constructs and the respective DBD and AD constructs were transformed into yeast strain Y190 and selected on drop-out growth media and 3-aminotriazole as previously described (Schwechheimer, 2002). Protein expression of the fusion proteins was verified by immunoblotting.

Physiological Assays

All plants were cultivated on sterile one-half-strength Murashige and Skoog (MS) medium without sugar under continuous white light (120 μmol m^{-2} s^{-1}). The germination rate was determined after 72 h by counting seeds with successful endosperm rupture. For hypocotyl length measurements, seedlings were grown on vertically oriented plates for 8 d. Plates were then scanned, and hypocotyl length was measured using the National Institutes of Health ImageJ software. Chlorophyll measurements were performed with 9-d-old seedlings, and chlorophyll content was determined as previously described and normalized to the chlorophyll content of the wild-type seedlings grown in the same conditions (Inskeep and Bloom, 1985). The experiments were repeated three times with comparable outcomes, and the result of one representative experiment is shown. Adult plants were grown on soil under continuous white light (120 μmol m^{-2} s^{-1}) and photographed at the times indicated in the figure legends.

Microarray Analyses

Microarray analyses were performed based on total RNA extracted with the NucleoSpin RNA Plant Kit (Macherey-Nagel) from 10-d-old seedlings grown in continuous white light (120 μmol m^{-2} s^{-1}). One hundred fifty nanograms of total RNA was labeled with Cy3 using the Low-Input Quick Amp Labeling Protocol (Agilent Technologies). Three biological replicate samples were prepared for each genotype, and Arabidopsis arrays (V4, design identification 21169; Agilent Technologies) were hybridized at 65°C for 17 h in rotating hybridization chambers (Agilent Technologies). Subsequently, the arrays were washed and scanned using an Agilent Microarray Scanner. Total RNA and probe quality were controlled with a Bioanalyzer 2100 (Agilent Technologies). Raw data were extracted using the Feature Extraction software v.10.7.3.1 (Agilent Technologies). Raw data files were imported into GeneSpring GX (v.12) and normalized by choosing the scale-to-median and baseline-to-median algorithms. Data were then subjected to an ANOVA analysis ($P < 0.05$) with an S-N-K posthoc test and filtered for genes with 2-fold expression differences to the respective wild-type controls (Supplemental Table S1). In two cases, one of the three replicate arrays (one wild type and one GNLox) had to be excluded from further analysis because these arrays were identified as outliers in the principal component analysis. Gene clustering was performed with GeneCluster3 using Hierarchical Clustering (correlation uncentered and centroid linkage options) and visualized using Java TreeView. Microarray data were deposited as GSE55442 to the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/).

Immunoblots

To assay for transgene protein expression, total protein plant extracts were prepared from 9-d-old light-grown seedlings in an extraction buffer containing 50 mM Tris-HCl, pH 7.0, 150 mM NaCl, 0.5% (v/v) Triton-X100, 1 mM phenylmethylsulfonyl fluoride, and 1× Protease Inhibitor Cocktail (Roche). Forty-five micrograms of total protein extract was separated on a 10% (v/v) SDS-PAGE gel. For immunoblotting, the following antibodies were used: anti-HA-horseradish peroxydase (HRP) (1:1,000; Roche), anti- REPRESSOR-of-ga1 (RGA; 1:1,000; Willige et al., 2007), and anti-rabbit-HRP (1:2,000; Sigma-Aldrich). For protein stability assays, seedlings were transferred to one-half-strength MS liquid medium for up to 1.5 h in medium containing the protein synthesis inhibitor CHX (50 μM). For treatments with the 26S proteasome inhibitor MG132 (100 μM), seedlings were preincubated with MG132 or a mock solution for 1.5 h before addition of CHX, as specified in the figure legends. The verification of protein expression of DBD and AD fusion proteins expressed in yeast was performed as previously described with the antibodies anti-GAL4 DBD (Santa Cruz Biotechnology) and anti-HA-HRP (AD; Sigma-Aldrich; Katsiarimpa et al., 2011).

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers AtGNC (AT5G56860), AtGNL (AT4G26150), AtGATA15 (AT3G06740), AtGATA16 (AT5G49300), AtGATA17 (AT3G16870), AtGATA17-LIKE (AT4G16141), AtGATA18 (AT3G50870), AtGATA19 (AT4G36620), AtGATA20 (AT2G18380), AtGATA23 (AT5G26930), AtGATA29 (AT3G20750), SlGATA1 (Solyc09g075610.2.1), SlGATA2 (Solyc01g060490.2.1), SlGATA3 (Solyc12g099370.1.1), SlGATA4 (Solyc05g054400.2.1), SlGATA5 (Solyc01g100220.2.1), SlGATA6 (Solyc12g008830.1.1), SlGATA7 (Solyc07g038160.2.1), SlGATA8 (Solyc02g085190.1.1), SlGATA9 (Solyc02g062760.2.1), BdGATA1 (BRADI1G03020), BdGATA2 (BRADI2G12590), BdGATA3 (BRADI2G35057), BdGATA4 (BRADI2G62660), BdGATA5 (BRADI1G37480), BdGATA6 (BRADI3G08240), BdGATA7 (BRADI2G45750), BdGATA8 (BRADI2G14890), HvGATA1 (AK370909), HvGATA2 (AK364126), HvGATA3 (AK250674), HvGATA4 (AK360760), HvGATA5 (AK360901), HvGATA6 (ADE05569), RGA (AT2G01570).

Supplemental Data

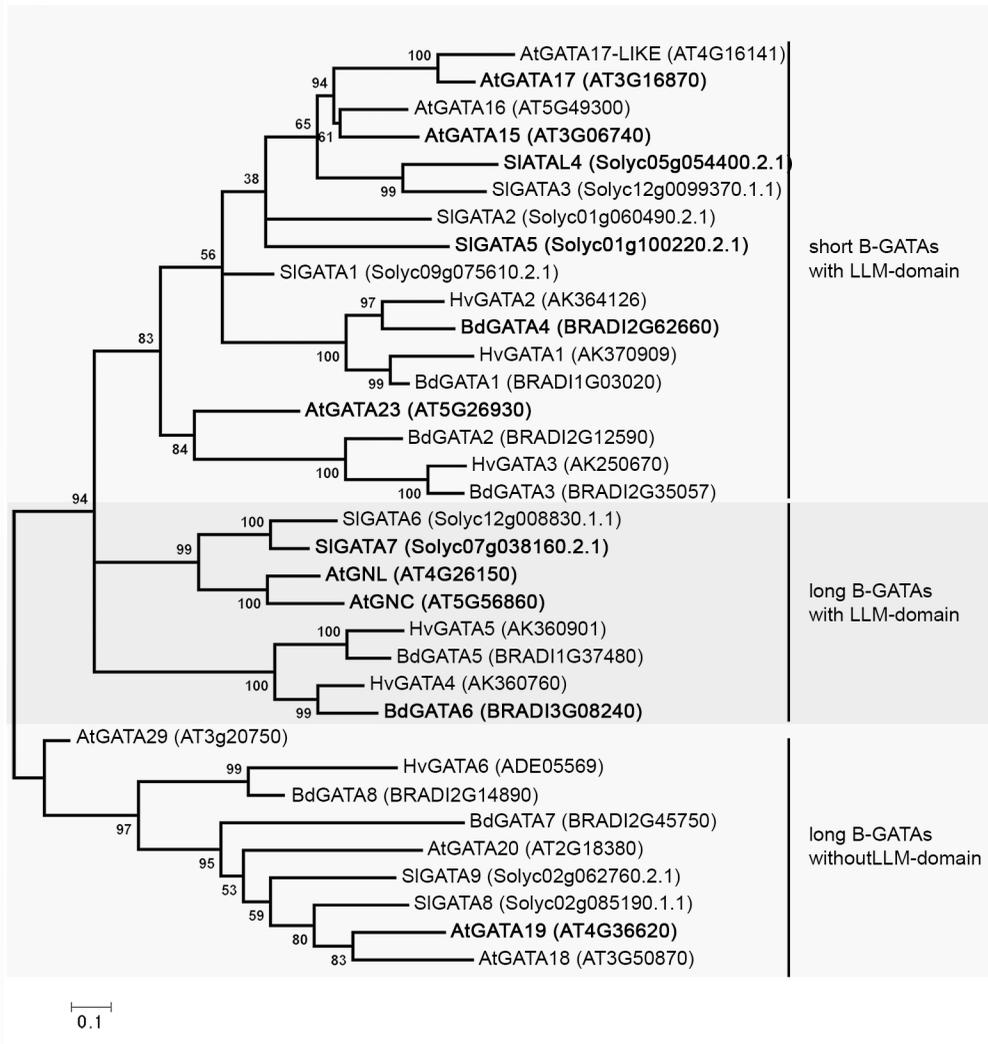The following materials are available in the online version of this article.

Supplemental Figure S1. Phylogenetic tree of the B-class GATAs from Arabidopsis, tomato, Brachypodium, and barley.

Supplemental Figure S2. Full-length protein sequence alignment of the B-class GATAs.

Supplemental Figure S3. Protein sequence alignment of the trimmed B-class GATA sequences as used for the phylogenetic analysis.
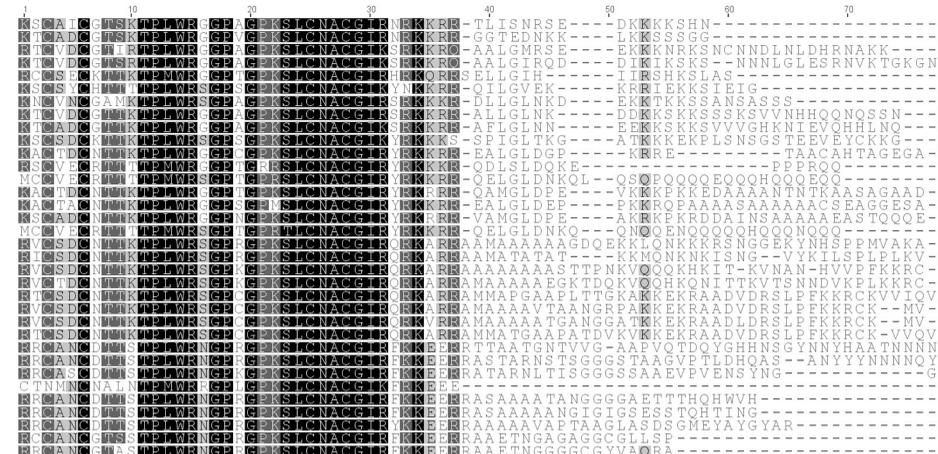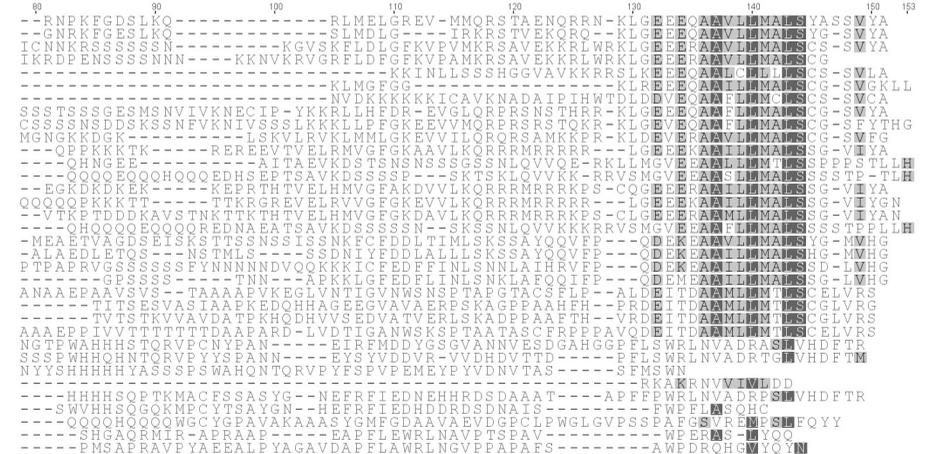
Supplemental Figure S4. B-GATAs from tomato and Brachypodium are functionally redundant with B-GATAs from Arabidopsis.

Supplemental Figure S5. B-class GATAs are short-lived proteins.

Supplemental Figure S6. The LLM domain does not determine lateral inflorescence angles in the GNC/GNL overexpressors.

Supplemental Figure S7. Flowering time delays of GNC/GNL overexpressors are independent from the LLM domain.

Supplemental Figure S8. The delayed senescence phenotype of the GNC/GNL overexpression lines is not dependent on the presence of the LLM-domain.

Supplemental Figure S9. The LLM domain is dispensable for proteasomal degradation of B-GATAs.

Supplemental Figure S10. GNC and GNL do not interact in the yeast two-hybrid system.

Supplemental Table S1. Tables of differentially expressed genes as used for the analyses presented in Figures 4 and 8 and Supplemental Figure S4.

Supplemental Table S2. List of primers used in this study.

ACKNOWLEDGMENTS

LITERATURE CITED

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

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:** 680–692

Chiang YH, Zubo YO, Tapken W, Kim HJ, Lavanway AM, Howard L, Pilon M, Kieber JJ, Schaller GE (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:** 332–348

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J **16:** 735–743

De Rybel B, Vassileva V, Parizot B, Demeulenaere M, Grunewald W, Audenaert D, Van Campenhout J, Overvoorde P, Jansen L, Vanneste S, et al (2010) A novel aux/IAA28 signaling cascade activates GATA23-dependent specification of lateral root founder cell identity. Curr Biol **20:** 1697–1706

Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, Pikaard CS (2006) Gateway-compatible vectors for plant functional genomics and proteomics. Plant J **45:** 616–629

Fox AH, Liew C, Holmes M, Kowalski K, Mackay J, Crossley M (1999) Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. EMBO J **18:** 2812–2822

Hudson D, Guevara D, Yaish MW, Hannam C, Long N, Clarke JD, Bi YM, Rothstein SJ (2011) GNC and CGA1 modulate chlorophyll biosynthesis and glutamate synthase (GLU1/Fd-GOGAT) expression in Arabidopsis. PLoS ONE **6:** e26765

Inskeep WP, Bloom PR (1985) Extinction coefficients of chlorophyll *a* and *b* in *N,N*-dimethylformamide and 80% acetone. Plant Physiol **77:** 483–485

Mayer KF, Waugh R, Brown JW, Schulman A, Langridge P, Platzer M, Fincher GB, Muehlbauer GJ, Sato K, Close TJ, et al (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature **491:** 711–716

International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature **463:** 763–768

Kanei M, Horiguchi G, Tsukaya H (2012) Stable establishment of cotyledon identity during embryogenesis in Arabidopsis by ANGUSTIFOLIA3 and HANABA TARANU. Development **139:** 2436–2446

Katsiarimpa A, Anzenberger F, Schlager N, Neubert S, Hauser MT, Schwechheimer C, Isono E (2011) The *Arabidopsis* deubiquitinating enzyme AMSH3 interacts with ESCRT-III subunits and regulates their localization. Plant Cell **23:** 3026–3040

Mackay JP, Kowalski K, Fox AH, Czolij R, King GF, Crossley M (1998) Involvement of the N-finger in the self-association of GATA-1. J Biol Chem **273:** 30560–30567

Mara CD, Irish VF (2008) Two GATA transcription factors are downstream effectors of floral homeotic gene action in Arabidopsis. Plant Physiol **147:** 707–718

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:** 1557–1560

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:** 103–113

Reyes JC, Muro-Pastor MI, Florencio FJ (2004) The GATA family of transcription factors in Arabidopsis and rice. Plant Physiol **134:** 1718–1732

Richter R, Bastakis E, Schwechheimer C (2013a) 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

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:** 2093–2104

Richter R, Behringer C, Zourelidou M, Schwechheimer C (2013b) Convergence of auxin and gibberellin signaling on the regulation of the GATA transcription factors GNC and GNL in *Arabidopsis thaliana*. Proc Natl Acad Sci USA **110:** 13192–13197

Ross J, Mavoungou L, Bresnick EH, Milot E (2012) GATA-1 utilizes Ikaros and polycomb repressive complex 2 to suppress Hes1 and to promote erythropoiesis. Mol Cell Biol **32:** 3624–3638

Schwechheimer C (2002) Studying protein-protein interactions with the yeast two-hybrid system. *In* P Gilmartin, C Bowler, eds, Molecular Plant Biology: A Practical Approach, Band 2, Vol 2. Oxford University Press, Oxford

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

Vierstra RD (2012) The expanding universe of ubiquitin and ubiquitin-like modifiers. Plant Physiol **160:** 2–14

Willige BC, Ghosh S, Nill C, Zourelidou M, Dohmann EM, Maier A, Schwechheimer C (2007) The DELLA domain of GA INSENSITIVE mediates the interaction with the GA INSENSITIVE DWARF1A gibberellin receptor of *Arabidopsis*. Plant Cell **19:** 1209–1220

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

Zhao Y, Medrano L, Ohashi K, Fletcher JC, Yu H, Sakai H, Meyerowitz EM (2004) HANABA TARANU is a GATA transcription factor that regulates shoot apical meristem and flower development in *Arabidopsis*. Plant Cell **16:** 2586–2600

Figure S1: Phylogenetic tree of the B-class GATAs from Arabidopsis, tomato, Brachypodium and barley. Unrooted phylogenetic tree of the B-class GATAs reveals a common evolutionary origin of short and long B-GATAs with and without an LLM-domain from the four different species. The phylogenetic tree was drawn to scale and generated based on an alignment of trimmed B-GATAs using the entire B-GATA DNA-binding domain and the C-termini with or without the LLM-domain. Bootstrap values are indicated by each node. Bar, 0.2 amino acid substitutions per site. The underlying alignment is provided in Supplemental Fig. S2.

Figure S10_Behringer et al.
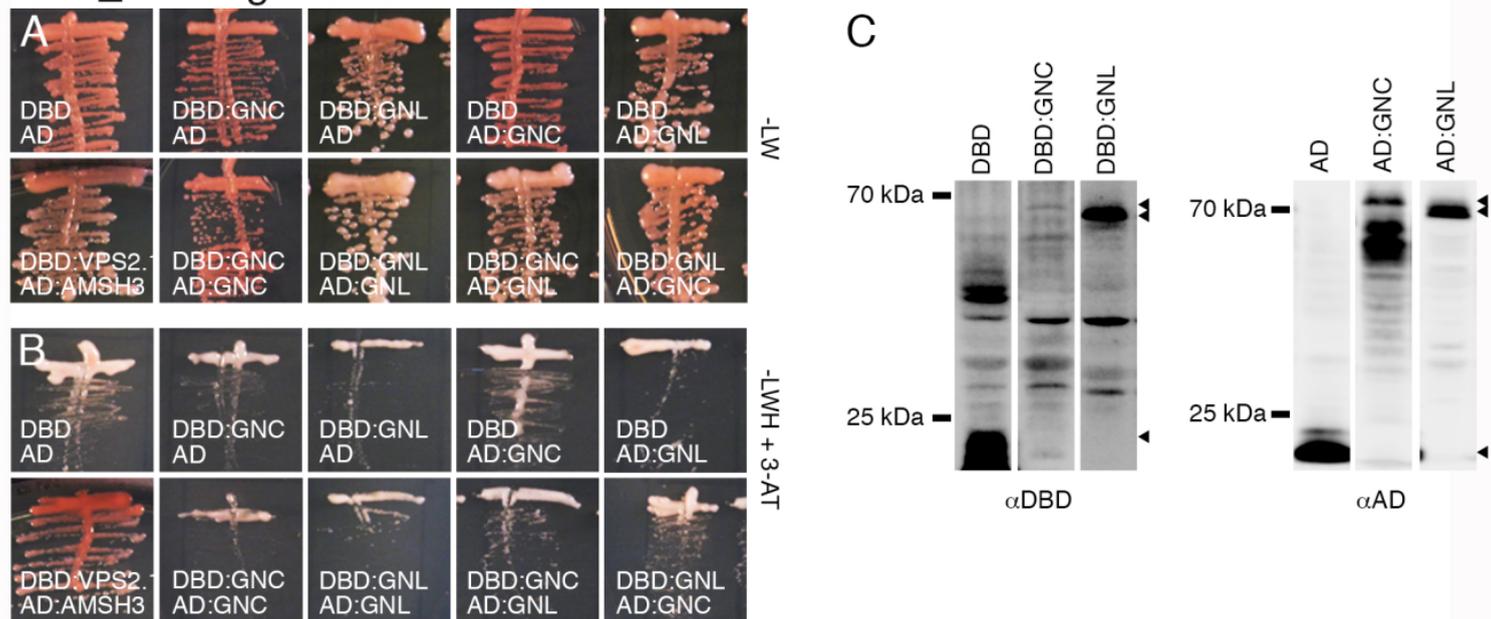
Figure S10: GNC and GNL do not interact in the yeast two-hybrid system. A. and **B.** Growth of yeast transformed with the bait (DBD) and prey (AD) plasmids as indicated in the figure on leucine (L)- and tryptophane (W)-deficient medium (A, growth and transformation control) or L, W, histidine (H)-deficient medium supplemented with 2.5 mM 3-aminotriazole (B, interaction experiment). The presence of reddish yeast colonies is indicative for an interaction as can be seen in the positive control experiment with the yeast expressing the interacting proteins DBD:VPS2.1 and AD:AMSH3 (Katsiarimpa et al., 2011). **C.** Immunoblots with αDBD (anti-GAL4DB) and αHA (anti-AD) antibodies to verify expression of the DBD and AD fusion proteins, respectively. The arrowheads point at the bands corresponding to the protein domains alone or their fusion protein products

Figure S2_Behringer et al.
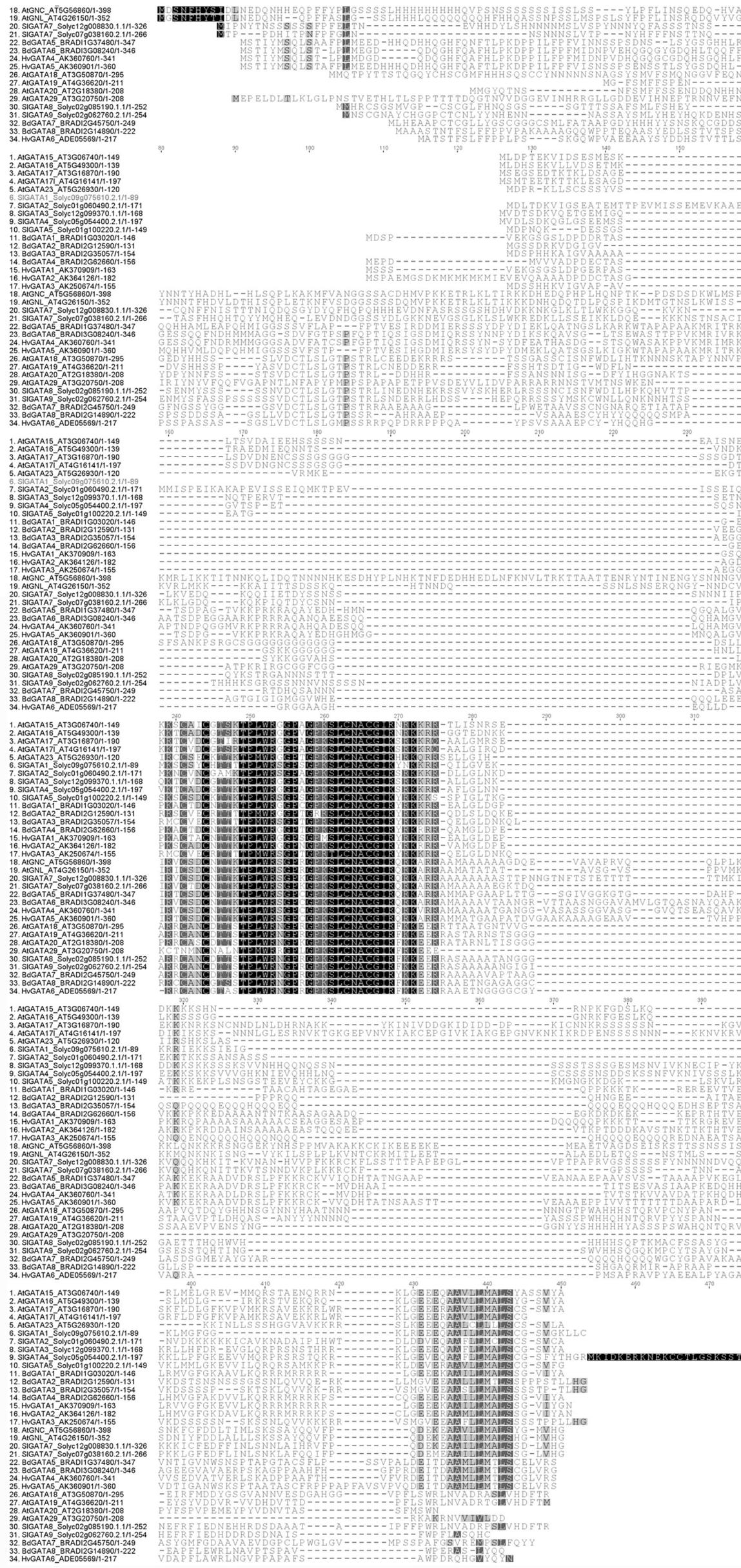
Figure S2: Protein sequence alignment of the trimmed B-class GATA sequences as used for the phylogenetic analysis. The protein alignment of the B-GATA DNA-binding domain and C-termini was generated using the full-length protein sequences and the CLUSTALW2 algorithm at the EMBL-EBI website (www.ebi.ac.uk/Tools/msa/clustalw2/) using default settings

Figure S3_Behringer et al.

Figure S3: Full-length protein sequence alignment of the B-class GATAs. The protein alignment of the B-GATA factors was generated using the full-length protein sequences and the CLUSTALW2 algorithm at the EMBL-EBI web-site (www.ebi.ac.uk/Tools/msa/clustalw2/) using default settings.

Figure S4_Behringer et al.
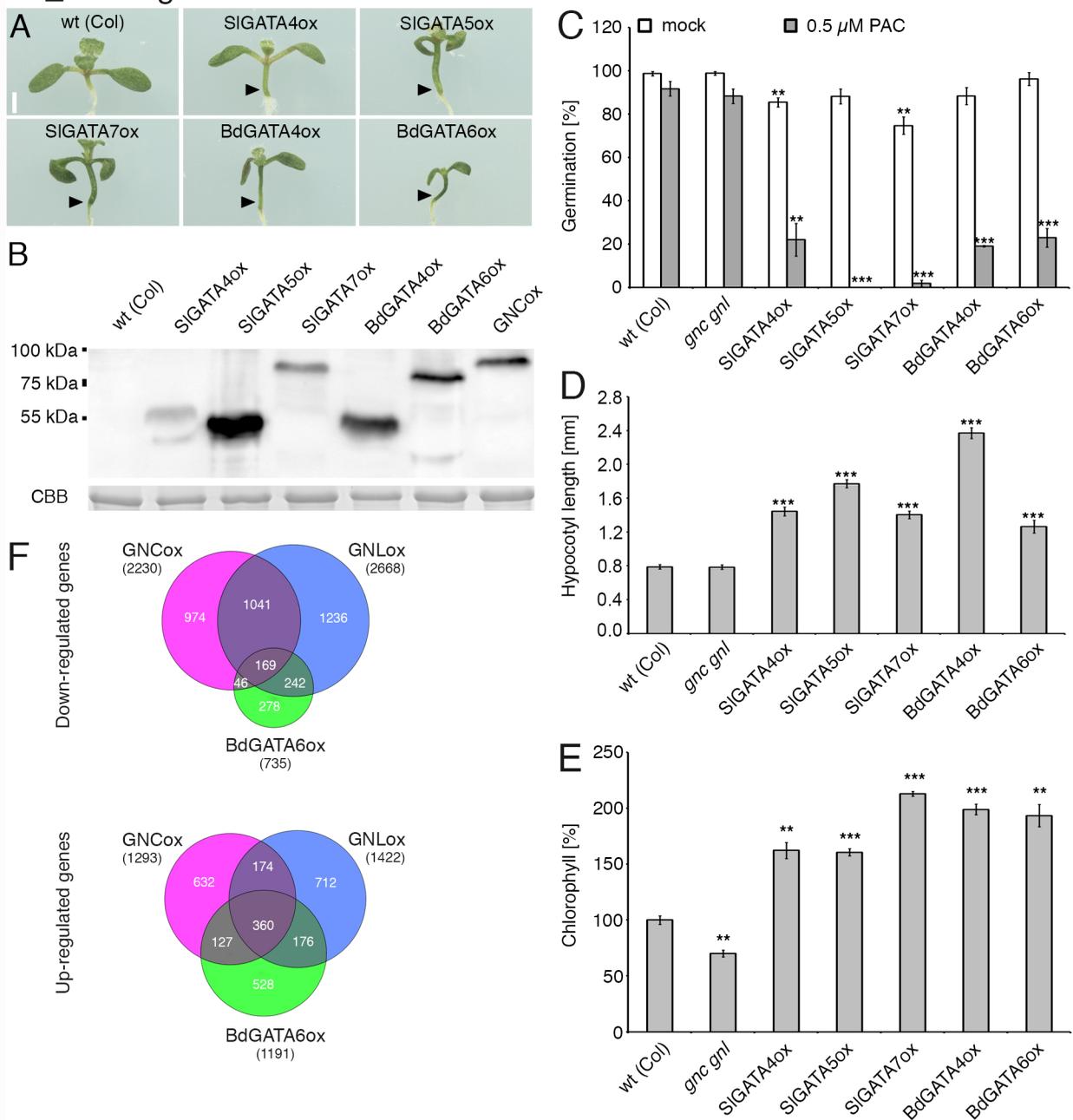


Figure S4: B-GATAs from tomato and Brachypodium are functionally redundant with B-GATAs from Arabidopsis. A. Representative photographs of 9-days-old light-grown *Arabidopsis thaliana* seedlings of the wild type and strong overexpression lines of short and long LLM-domain containing B-class GATAs from tomato and Brachypodium. Note the particularly strong chlorophyll accumulation at the base of the seedling hypocotyls in the overexpressors as indicated by the arrowheads. Scale bar = 1 mm. **B.** Immunoblot with an anti-HA antibody to detect transgene protein expression in the overexpression lines as shown in (A). CBB, Coomassie Brilliant Blue, protein loading control. **C.** Relative germination rate of seeds from wild type, *gnc gnl* mutant and B-GATA overexpression lines grown on 1/2 MS medium and on medium supplemented with the GA biosynthesis inhibitor paclobutrazol (PAC) after 4 days of stratification and 3 days of growth at 21 °C as determined based on endosperm rupture. n > 25. **D.** Quantitative analysis of hypocotyl elongation in 8-days-old light-grown seedlings. n > 20. **E.** Quantitative analysis of chlorophyll accumulation in 9-days-old light-grown seedlings. n > 10. Averages and standard errors are indicated in each case. Student's t-tests were performed in comparison to the wild type: * ,$p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. **F.** Venn-diagrams of the genes down- and up-regulated in GNCox, GNLox and BdGATA6ox overexpression lines. Numbers indicate the number of differentially expressed genes in the respective genotypes and their intersections. The total number of differentially regulated genes is indicated in brackets for each genotype. A list of the differentially regulated genes is provided in Supplemental Table S2.

Figure S5_Behringer et al.
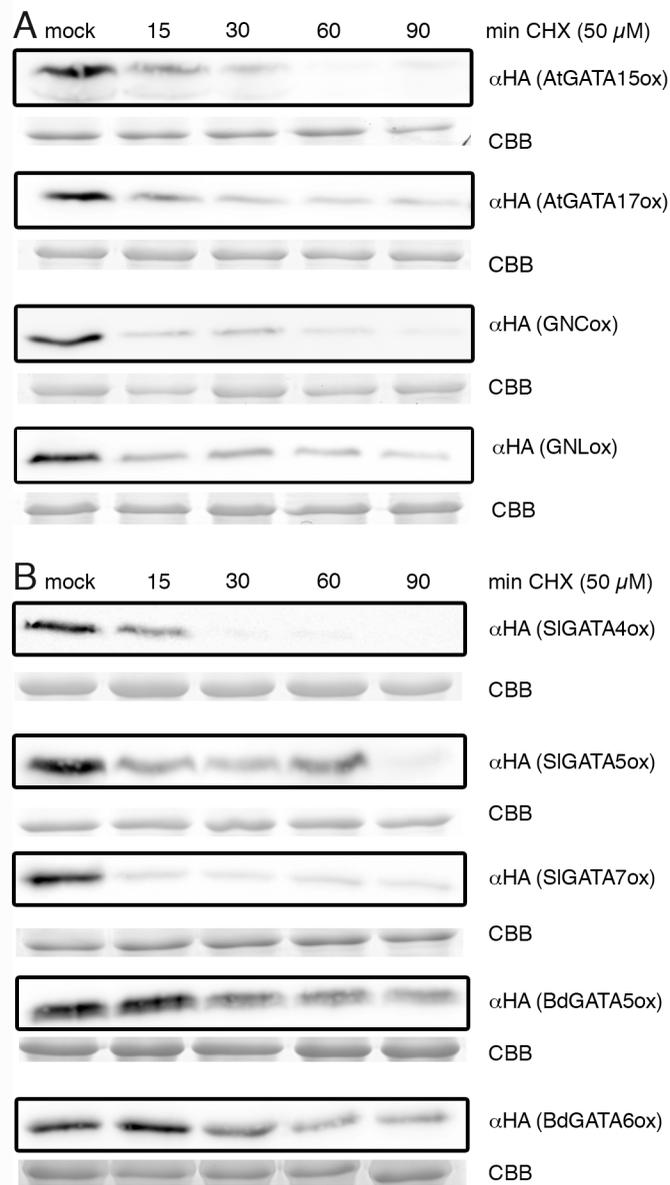
Figure S5: B-class GATAs are short-lived proteins. Anti-HA immunoblots of total protein extracts from 9-days-old seedlings of all LLM-domain containing B-GATAs examined in this study. Seedlings had been treated with the protein biosynthesis inhibitor cycloheximide (CHX) for the periods of time specified in the figure. CBB, Coomassie-Brilliant Blue, protein loading control.

Figure S6_Behringer et al.
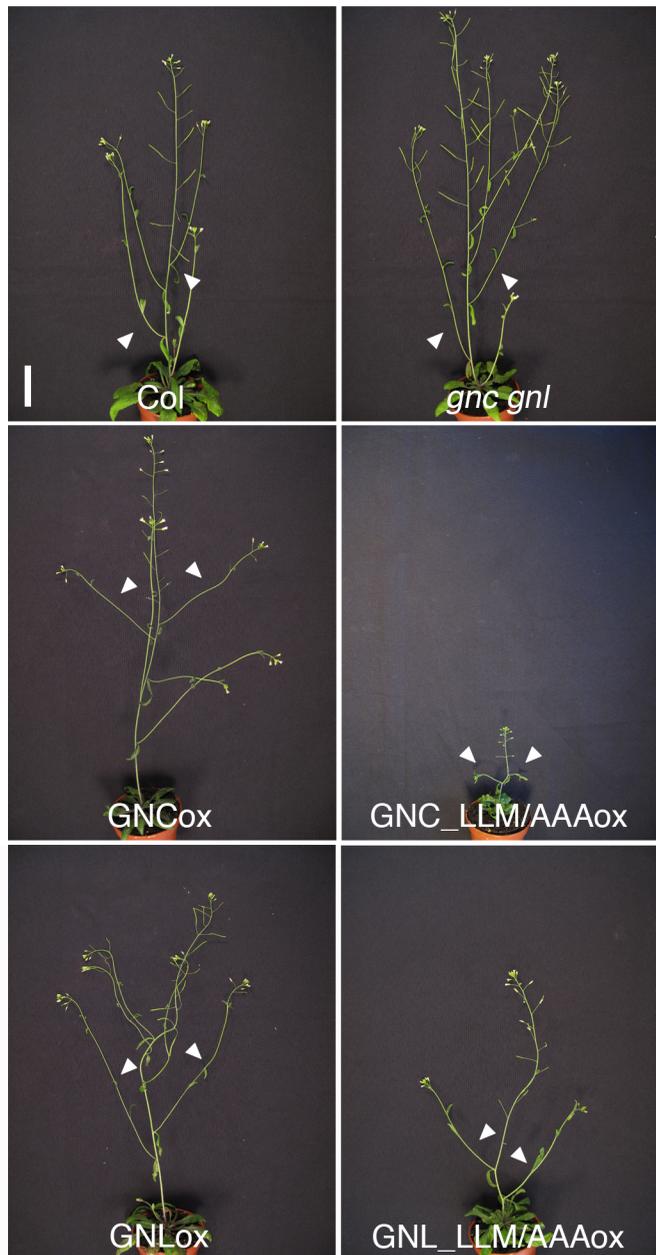
Figure S6: The LLM-domain does not determine lateral inflorescence angels in the GNC/GNL overexpressors. Representative photographs of dissected leaves from 32 days-old Arabidopsis plants of the specified genotypes. In the case of GNC_LLMAAAox, a 39 days-old plant is shown. The arrowheads point at the angles between the primary inflorescence and a lateral inflorescence, which is increased in the B-GATA overexpressors. Scale bar = 2.5 cm.

Figure S7_Behringer et al.

Figure S7: Flowering time delays of GNC/GNL overexpressors are independent from the LLM-domain: Representative photographs of 4-weeks-old *Arabidopsis thaliana* seedlings of the wild type, the *gnc gnl* mutant and strong overexpression lines of GNC and GNL overexpressors with a wildtype or a mutated LLM-domain. Scale bar = 2 cm.

Figure S8_Behringer et al.
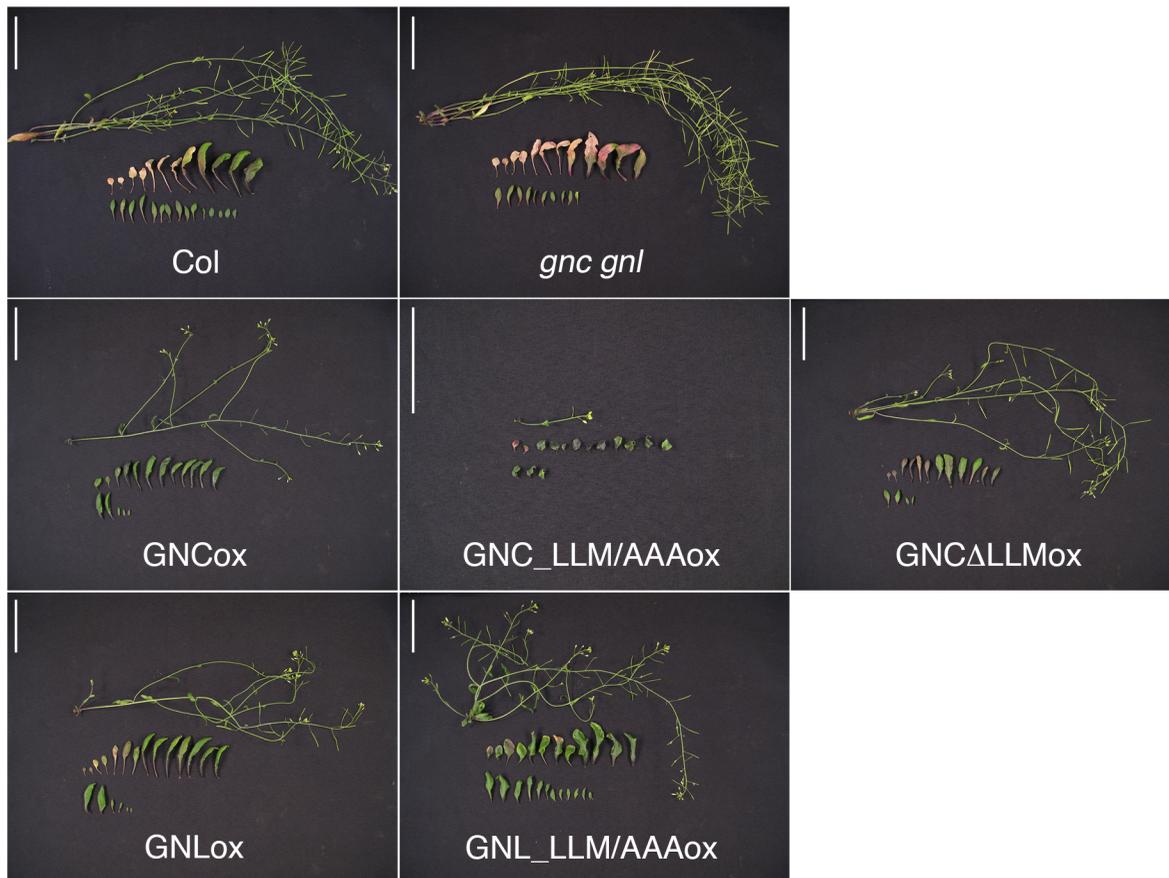
Figure S8: The delayed senescence phenotype of GNC/GNL overexpression lines is not dependent on the presence of the LLM-domain. Representative photographs of dissected 39 days-old Arabidopsis plants of the specified genotypes. Please note the absence of senescing leaves from the GNC and GNL overexpression lines. Scale bars = 5 cm.

Figure S9_Behringer et al.
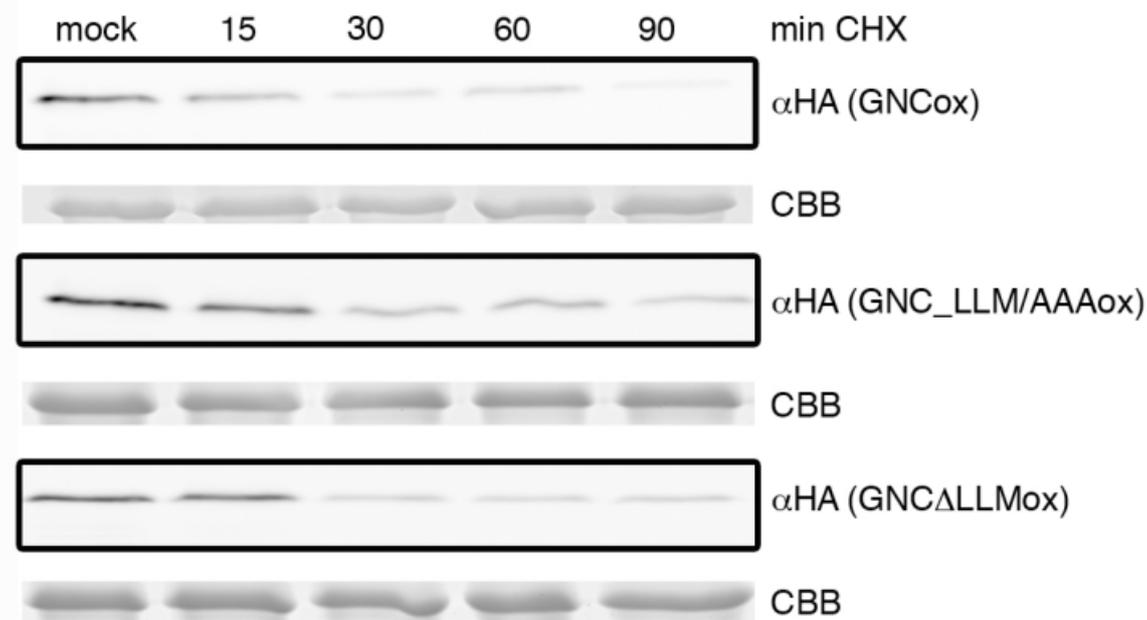
Figure S9: The LLM-domain is dispensable for proteasomal degradation of B-GATAs. Anti-HA immunoblots performed with total protein extracts from 9-days-old GNCox seedlings and seedlings of the same age expressing its two LLM-domain mutant variants. Seedlings had been treated with CHX for the periods of time specified in the figure. CBB, Coomassie-Brilliant Blue, protein loading control.