



Evolution and diversification of the ACT-like domain associated with plant basic helix–loop–helix transcription factors

Yun Sun Lee^a, Shin-Han Shiu^{b,c}, and Erich Grotewold^{a,1}

Edited by Elizabeth Kellogg, Donald Danforth Plant Science Center, St. Louis, MO; received November 15, 2022; accepted March 21, 2023

Basic helix–loop–helix (bHLH) proteins are one of the largest families of transcription factor (TF) in eukaryotes, and ~30% of all flowering plants' bHLH TFs contain the aspartate kinase, chorismate mutase, and TyrA (ACT)-like domain at variable distances C-terminal from the bHLH. However, the evolutionary history and functional consequences of the bHLH/ACT-like domain association remain unknown. Here, we show that this domain association is unique to the plantae kingdom with green algae (chlorophytes) harboring a small number of bHLH genes with variable frequency of ACT-like domain's presence. bHLH-associated ACT-like domains form a monophyletic group, indicating a common origin. Indeed, phylogenetic analysis results suggest that the association of ACT-like and bHLH domains occurred early in Plantae by recruitment of an ACT-like domain in a common ancestor with widely distributed *ACT DOMAIN REPEAT* (*ACR*) genes by an ancestral bHLH gene. We determined the functional significance of this association by showing that *Chlamydomonas reinhardtii* ACT-like domains mediate homodimer formation and negatively affect DNA binding of the associated bHLH domains. We show that, while ACT-like domains have experienced faster selection than the associated bHLH domain, their rates of evolution are strongly and positively correlated, suggesting that the evolution of the ACT-like domains was constrained by the bHLH domains. This study proposes an evolutionary trajectory for the association of ACT-like and bHLH domains with the experimental characterization of the functional consequence in the regulation of plant-specific processes, highlighting the impacts of functional domain coevolution.

evolution | green algae | gene regulation | protein–protein interaction

Control of gene expression is a fundamental principle of all organisms. In eukaryotes, gene regulation is predominantly controlled by transcription factors (TFs), proteins that bind DNA in a sequence-specific fashion and which are often formed by multiple domains that participate in protein–protein and protein–DNA interactions (1, 2). Among the 60+ TF families that characterize animals or plants, the basic helix–loop–helix (bHLH) family is among the largest (3, 4). The signature bHLH domain is composed of a ~20 amino acids-long basic α -helix that recognizes one half-site of the canonical E-box (5'-CANNTG-3') or the more sequence-constrained G-box (5'-CACGTG-3') and the helix–loop–helix region that mediates homo- or heterodimer formation with other bHLH factors, indispensable for high-affinity DNA binding (5).

The first identified plant bHLH protein corresponded to maize *Lc*, a member of the *R/B* gene family of anthocyanin pigment regulators (6, 7). Subsequent plant bHLH proteins were identified by sequence similarity to R and to vertebrate bHLH domains (8–10), becoming the second largest family of plant TFs, with ~136 members in *Arabidopsis* (11). Evolutionary studies demonstrated that plant bHLHs are monophyletic, evolutionarily distinct to the rhodophytes (red algae) and metazoans (12). A significant expansion of the bHLH gene family took place after the divergence of the moss and vascular plant lineages, resulting in 26 to 32 bHLH subfamilies mostly emerging around the time that ancestral plants colonized land (12–14). The bHLH subfamilies are distinguished by differences in intron/exon structures, in the presence of conserved amino acids within the bHLH motif, and by the presence/absence of other domains (12, 13, 15).

A general characteristic of bHLH TFs is that they function by forming homo- or heterodimers through the bHLH, and by interactions with other regulatory proteins through domains other than the bHLH. In plants, this is epitomized by the interaction of maize *R/B* with a subgroup of *R2R3-MYB* regulators (characterized in maize by *C1*) through the *MYB* interaction region located at the N terminus of *R/B* (16–18). Subsequent studies demonstrated that such a bHLH/*R2R3-MYB* partnership is a characteristic of all plant bHLH factors belonging to subgroup IIIb (11, 19). The bHLH proteins in this subfamily also harbor aspartate kinase, chorismate mutase, and TyrA (ACT)-like domains C-terminal to the respective bHLH domains (20–22).

Significance

Basic helix–loop–helix (bHLH) proteins constitute one of the largest eukaryote transcription factor (TF) families. About 30% of flowering plants' bHLH TFs have aspartate kinase, chorismate mutase, and TyrA (ACT)-like domains. Here, we show that ACT-like domains associated with bHLH domains are unique to the plant kingdom and derived from a common ancestor, likely by the fusion of bHLH and ancestor *ACR* (*ACT DOMAIN REPEAT*) genes early in the evolution of the green algae. Despite the fast evolution of ACT-like domains, our results show that ACT-like and associated bHLH domains coevolved, and that the association negatively affects the DNA-binding activity of the corresponding bHLH domains. These results unveil the evolutionary history of the ACT-like/bHLH association, providing insights on potential functional consequences.

Author affiliations: ^aDepartment of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824; ^bDepartment of Plant Biology, Michigan State University, East Lansing, MI 48824; and ^cDepartment of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI 48824

Author contributions: Y.S.L. and E.G. designed research; Y.S.L. performed research; S.-H.S. contributed with ideas and analytic tools; Y.S.L. analyzed data; and Y.S.L., S.-H.S., and E.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: grotewol@msu.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2219469120/-DCSupplemental>.

Published May 1, 2023.

The ACT-like domains are characteristic of a subset of plant bHLH TFs. They are 70–80 amino acids long and have the $\beta\beta\alpha\beta\alpha$ fold (20, 22, 23), distinct from the $\beta\alpha\beta\alpha\beta$ topology of the ACT domain found in a large number of other proteins, mostly associated with primary metabolism biosynthesis enzymes (24). There are several ACT-domain variants that have similar folds (25, 26). Most relevant for this study is the presence of an ACT domain in bacterial glutamate/glutamine-sensing uridylyl-transferase (GlnD) and in plant ACT domain repeat (ACR) proteins (27–29). In GlnD, the ACT domain functions as a small-molecule sensor (27), while the function of ACT domains in ACR proteins remains unknown. Not surprisingly, given the broad distribution among bHLH TFs associated with important biological functions, mutations in the associated ACT-like domains have major functional consequences as demonstrated for maize R (20), Arabidopsis DYT1 (30), Arabidopsis SCREAM (31), and Arabidopsis LONESOME HIGHWAY(LHW) (32). For example, in DYT1, the ACT-like domain participates in nuclear localization, transcriptional activity, and interaction with other bHLH TFs implicated in anther development (23, 30). In SCREAM, the ACT-like domain contributes to providing partner selectivity among other stomatal lineage-specific bHLH factors (31). Finally, in R, homodimerization through the ACT-like domain significantly decreases the DNA-binding affinity of R for the G-box (21).

Structure similarity searches showed that ACT-like domains are not limited to bHLH subfamily III_f, but are present in ~30% of all the Arabidopsis and maize bHLH TFs (20, 22), suggesting that the association of these two domains precedes the divergence of monocot and dicot plants, estimated to have occurred around 200 Mya (33). This is supported by the recent identification of an ACT-like domain-harboring bHLH factor in the liverwort *Marchantia polymorpha* and in the *Klebsormidium nitens* (Klebsormidiophyceae, Streptophyte algae) (32, 34). However, the evolutionary history of ACT-like domains and the origin of the association with bHLH factors remain unclear, despite the important function that ACT-like domains play in the regulatory function of large family of plant bHLH TFs. It also remains unknown whether the ability of ACT-like domains to dimerize and influence the DNA-binding activity of the associated bHLH domains is conserved in green organisms other than vascular plants.

Here, we describe the analysis of the presence of ACT-like domains in bHLH proteins from over 140 eukaryotes. We show that the association of the ACT-like domain with bHLH TFs is unique to the plant kingdom and is first found in green algae (chlorophytes). ACT-like domains are characteristic of 15 bHLH subfamilies, and ACT-like domains likely descended from a common ancestor. Based on sequence and structure conservation, we propose that an ACT-like domain was acquired from a common ancestor of the plant ACR proteins early in the evolution of the Chloroplastida (green plants). A functional analysis of bHLH and ACT-like domains in the unicellular green algae *Chlamydomonas reinhardtii* indicates a striking conservation in DNA-binding and dimerization activities with land plants. Taken together, our results unveil the provenance of ACT-like domains associated with bHLH plant TFs and provide insightful structure–function information.

Results

ACT-Like Domains Are Uniquely Associated with a Subset of Plant bHLH Proteins. To establish the distribution of ACT-like domains and determine whether they are always associated with bHLH proteins, we investigated 140 species which encompass

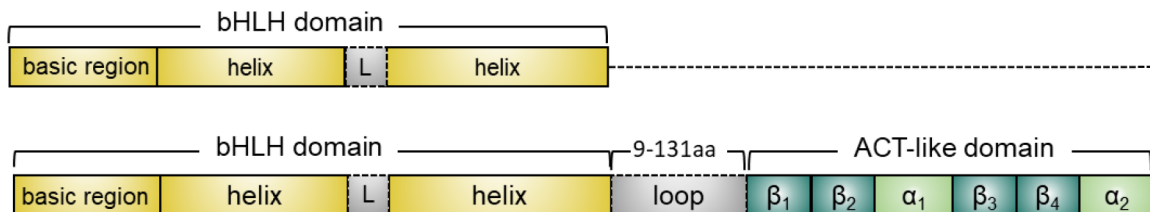
major taxa of eukaryotes, of which 49 are photosynthetic organisms, including 37 chlorophytes, five streptophyte algae, and seven embryophytes (land plants) (Dataset S1) and identified the presence of the ACT-like domains (SI Appendix, Materials and Methods). The resulting list included 354 bHLH proteins containing ACT-like domains (hereby referred as bHLH^{ACT+}, while those bHLH proteins without ACT domain are referred as bHLH^{ACT−}). Without exception, the ACT-like domain was located C-terminal to the bHLH, with a linker region between the two domains of variable length, but which in no case exceeded 131 amino acids (Fig. 1A). Every single bHLH^{ACT+} identified belonged to Plantae and included chlorophyte, streptophyte algae, and embryophyte species (Fig. 1B and Dataset S1).

To assess whether ACT-like domains are associated with plant TFs other than bHLHs and given that the ACT-like domains are difficult to identify owing to their low sequence identities (37), we employed complementary approaches based on AlphaFold prediction and Pfam annotation (SI Appendix, Fig. S1A; see the detailed methods in SI Appendix, Materials and Methods). In the first approach, AlphaFold (38, 39) searches using the Foldseek interface (40) with the predicted tertiary structure of the ACT-like domain from maize R (22) as a query retrieved all the proteins with an ACT-like domain-related tertiary structure. We found that ACT-like domains are only present in bacterial GlnDs, plant ACRs, and plant bHLH TFs (SI Appendix, Fig. S1B and Dataset S2). In the second approach, because most chlorophyte bHLH^{ACT+} proteins contain one of the Pfam ACT domains (PF01842), we retrieved protein sequences using the Pfam IDs corresponding to the ACT domain and identified TFs among the retrieved sequences. The Pfam database v35.0 search produced 1,366 ACT-like domain-associated TFs, all belonging to the bHLH family (SI Appendix, Fig. S1A and B and Dataset S2), supporting the notion that the association of bHLH and ACT-like domains is a phenomenon unique to the Plantae.

ACT-Like and bHLH Domain Association Precedes Colonization of Land by Plants. To determine the distribution of ACT-like domains before the radiation of vascular plants, where ACT-like domains are well described (20, 22), we investigated the number of bHLH genes and how many of them are bHLH^{ACT+} in the 49 Plantae species. From the 37 chlorophyte species investigated, we identified bHLH genes in all of them, and bHLH^{ACT+} in 29, encompassing several chlorophyte lineages (highlighted green in SI Appendix, Fig. S2 and Dataset S1). The eight taxa with no bHLH^{ACT+} are not monophyletic and may be a result from incomplete genome sequencing or provide examples of gene/domain loss. The genomes of all seven analyzed embryophytes and of 4/5 streptophyte algae investigated encode for bHLH^{ACT+} proteins (Dataset S1). While in the land plants bHLH^{ACT+} represent roughly 30% of the total bHLH proteins present (20, 22), the fraction of bHLH TFs harboring ACT-like domains is larger in the chlorophytes, averaging 52% of the identified bHLH factors harboring ACT-like domains (highlighted in dark green in SI Appendix, Fig. S2 and Dataset S1). However, no bHLH^{ACT+} was identified in the single glaucophyte species with sequence available, or in the six rhodophytes analyzed (Dataset S1). The results indicate that the association of the ACT-like and bHLH domains began close to 1.5 billion years ago after the Plantae diverged from other related eukaryotic lineages (Fig. 1B and Dataset S1) (41).

To assess whether the association of ACT-like and bHLH domains has single or multiple origins in the chlorophytes, a maximum likelihood (ML) tree with the bHLH domains derived from 135 bHLH proteins from the 37 chlorophytes was constructed

A



B

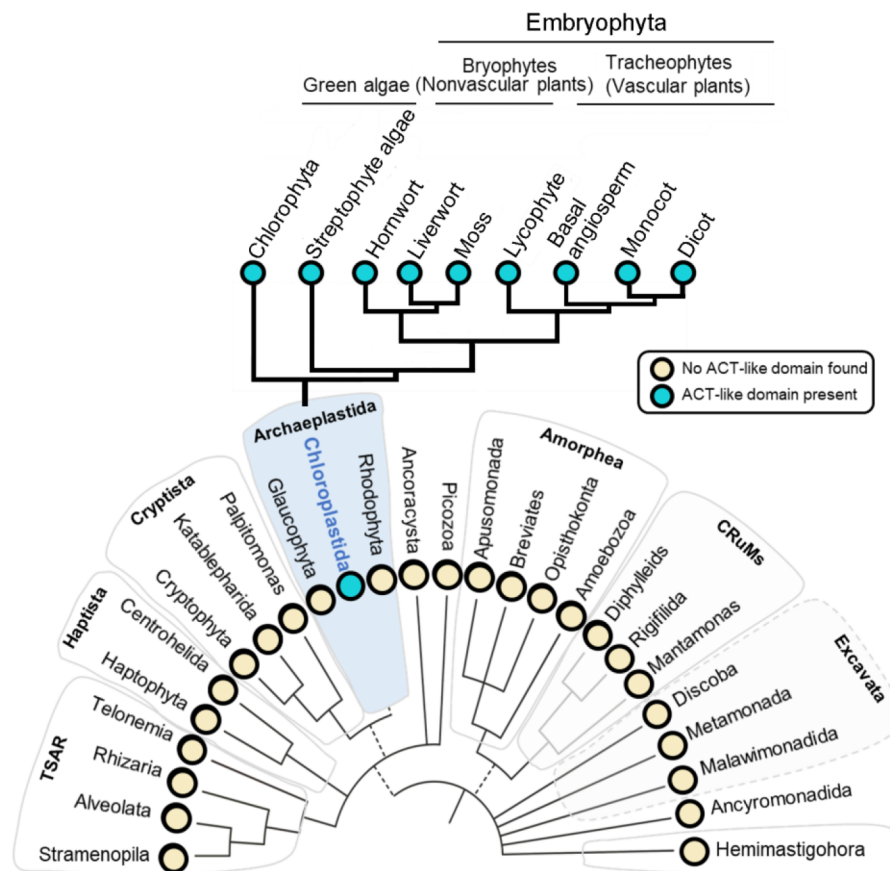


Fig. 1. Association of ACT-like and bHLH domains. (A) Schematic representation of bHLH and ACT-like domains. α and β represent the α -helices and β -sheets, respectively. The L within the gray box denotes the loop region. (B) Presence of bHLH-associated ACT-like domains in various species. Green and yellow circles indicate bHLH proteins with or without the ACT-like domains, respectively. The eukaryote phylogenetic tree was modified from the literature (35, 36). Species information and data source for the bHLH sequences used in this study are presented in [Dataset S1](#), and Pfam analysis result is in [SI Appendix, Fig. S1B](#) and [Dataset S2](#). AlphaFold search result is in [SI Appendix, Fig. S1B](#) and [Dataset S2](#). Numbers of the ACT-like domains in each species are described in [Dataset S1](#).

([SI Appendix, Fig. S3](#)). We found a monophyletic clade comprised of the 62 bHLH^{ACT+} proteins (indicated by the red arrow in [SI Appendix, Fig. S3](#)), of which 16 proteins contain incomplete ACT-like domain lacking β_1 among the $\beta\alpha\beta\beta\alpha$ topology ([SI Appendix, Fig. S4](#)). This result supports the interpretation that a single event contributed to the ACT-like and bHLH association in chlorophytes. The sequence difference between chlorophyte bHLH^{ACT+} and bHLH^{ACT-} proteins is evident in a conserved motif in the second α -helix of the bHLH domain and a long loop region ([SI Appendix, Fig. S5](#)).

Establishment of bHLH^{ACT+} Subfamilies after Land Colonization.

Previous studies suggested that major radiation of bHLH subfamilies arose before the origin of the mosses (12–14), yet

the evolutionary relationship of the ACT-like and bHLH domains is unexplored. To establish the evolutionary history of bHLH^{ACT+} proteins, we constructed an ML phylogenetic tree with sequences corresponding to the bHLH domain of 797 proteins derived from 49 plant species (Fig. 2 and [Datasets S3–S5](#)). The phylogenetic analysis recovered 27 bHLH subfamilies (Ia through XV), supported by Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) values greater than 80 and largely aligned with the 30 subfamilies into which land plant bHLH proteins were previously classified using four chlorophytes and six streptophytes (streptophyte algae and embryophytes) (Fig. 2 and [Datasets S4](#) and [S5](#)) (12, 14). Notably, the chlorophyte bHLH^{ACT+} (highlighted yellow in Fig. 2) forms an independent group (SH-aLRT = 99%), distinct from the bHLH^{ACT+} subfamilies

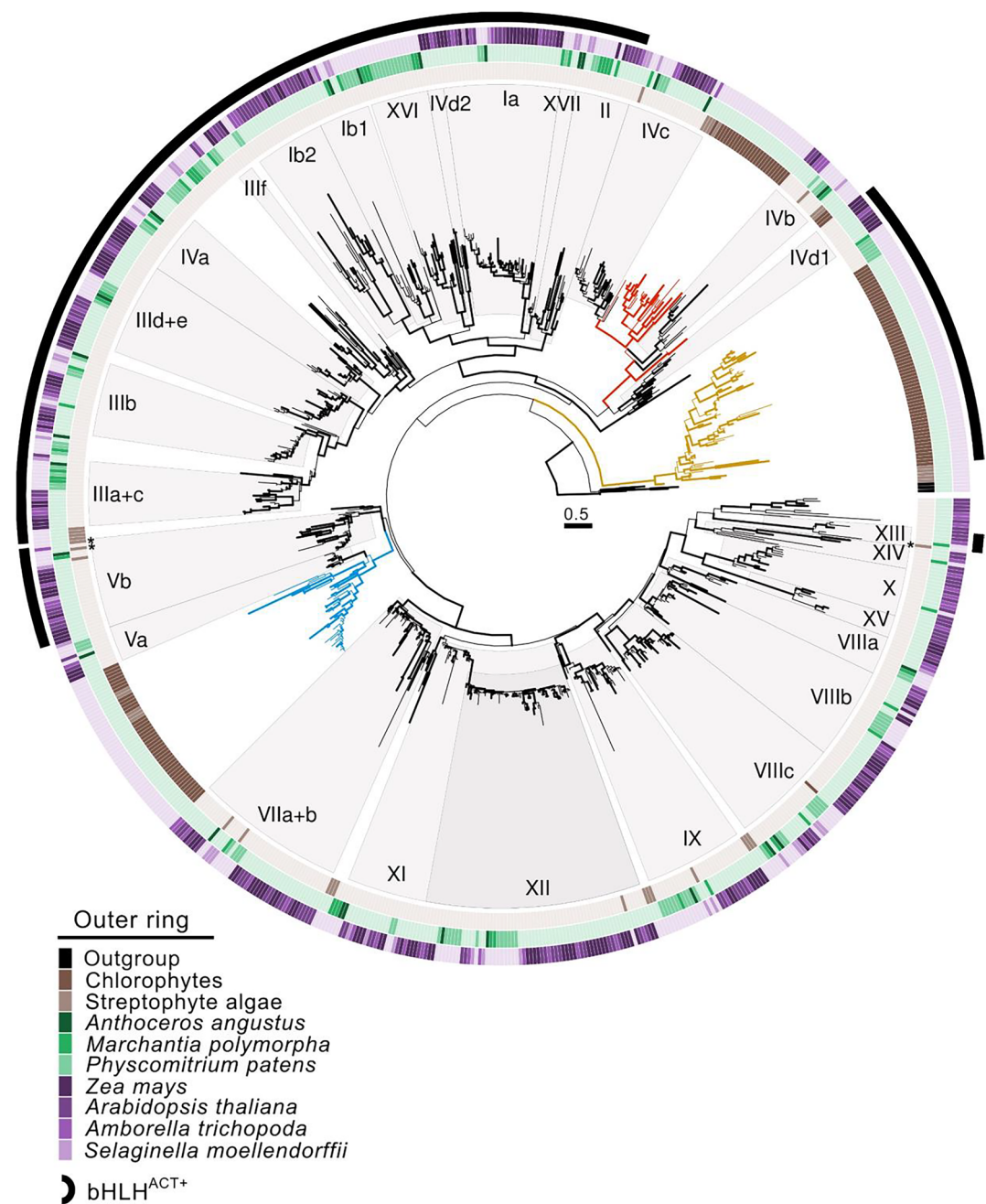


Fig. 2. bHLH domains associated with ACT-like domains form distinct clades in their phylogeny. ML phylogenetic tree was reconstructed with bHLH domains of 797 bHLH proteins retrieved from 37 chlorophytes, five streptophyte algae, and seven embryophytes, with an outgroup of three bHLH proteins from haptophytes. Colored rectangles on outer rings denote the species from which the bHLH proteins derived. Arcs of a circle in black indicate the bHLH proteins containing ACT-like domain. Shades of gray in the tree indicate clades corresponding to the described 30 bHLH subfamilies. Thick lines indicate branches with SH-aLRT support >80. Branches in yellow color highlight the chlorophyte bHLH^{ACT+}. Branches in red and blue colors indicate two respective chlorophyte bHLH^{ACT-} clades. Asterisks indicates the streptophyte algae bHLH^{ACT+} proteins belonging to subfamilies Vb and XIII. The phylogeny was calculated using RAXML-NG, 1,000 iterations, and Jones-Taylor-Thornton model with a gamma distribution (JTT+G). Full details of the phylogenetic tree are provided in [Dataset S5](#).

that consist of the streptophyte algae and the embryophytes. In the streptophyte algae and the embryophytes, 13 out of 15 bHLH^{ACT+} subfamilies and 12 out of 15 bHLH^{ACT-} subfamilies largely divided into two clades, with supporting values of 80% and 97%, respectively (Fig. 2 and [Dataset S5](#)). Chlorophyte bHLH^{ACT-} group (highlighted red in Fig. 2) and subfamilies IVb and IVc have the characteristic leucine-zipper motif in their C terminus ([SI Appendix, Fig. S6](#) and [Dataset S6](#)) and form a clade nested within multiple subclades of bHLH^{ACT+} genes from both chlorophyte and land plants, indicative of a domain loss event in the bHLH^{ACT-} clade, prior to the divergence of the chlorophyte and the land plant lineages. However, there are clearly limitations

on the evolutionary relationships that can be drawn from the short bHLH region reflected, for example, in that the position of subfamily Vb cannot be precisely established in the phylogeny (SH-aLRT = 62%). The bHLH^{ACT-} clade also includes bHLH^{ACT+} subfamily XIII. In subfamily XIII, ACT-like domains are present in vascular plants, but not in the liverworts ([Dataset S4](#)) (32). This is different from other bHLH^{ACT+} subfamilies, where all proteins harbor the ACT-like domains, regardless of which species they were derived from, except a few proteins that contain a stop codon after the C terminus of the bHLH domains ([Dataset S4](#)). The position of subfamily XIII based on the bHLH alignment suggests the

independent gain of an ACT-like domain after the diversification of this subfamily. However, we cannot rule out the possibility that the phylogeny of subfamily XIII is misestimated due to limitations imposed by the short length of the alignment.

Our phylogenetic tree reconstruction also gives insights into when the major radiation of the bHLH^{ACT+} subfamilies occurred. In streptophyte algae, two bHLH^{ACT+} proteins from *K. nitens* and *Chara braunii* cluster with subfamily Vb and one bHLH^{ACT+} from *K. nitens* in XIII (indicated by asterisks within the Vb and XIII clades in Fig. 2), while most streptophyte algae bHLH^{ACT-} proteins belong to subfamilies that are interspersed with embryophyte bHLH^{ACT-} (IVb, IVc, VIIa+b, IX, and XII; Fig. 2 and [Datasets S4 and S5](#)).

Taken together, our results indicate that the establishment of the bHLH^{ACT+} subfamilies took place after the embryophyte lineage diverged from the streptophyte algae lineage, likely after land colonization. In contrast, the bHLH^{ACT-} subfamilies were established prior to the divergence of the embryophyte and streptophyte algae lineages, but after their divergence from the chlorophyte, likely coinciding with the transition to multicellularity. Collectively, our findings indicate that the bHLH^{ACT+} proteins were established following the appearance and diversification of the bHLH^{ACT-} proteins.

Potential Origin of bHLH-Associated ACT-Like Domain. To identify the provenance of the ACT-like domain that characterizes a significant fraction of plant bHLH proteins, the ancestral sequence from all the ACT-like domains used in this study was inferred and subjected to DELTA-BLAST and PSI-BLAST in NCBI ([SI Appendix, Materials and Methods](#)). These searches led to the identification of the bacterial GlnD and plant ACR sequences as potential relatives of ACT-like sequences. As predicted by AlphaFold (38, 39), GlnDs proteins have two adjacent ACT-like domains ([SI Appendix, Fig. S7A](#)). The group I and II ACRs contain one ACT domain flanked by two or three ACT-like domains, respectively, while group III ACRs only have two ACT-like domains ([SI Appendix, Fig. S7 A and B](#)), and all the ACRs share high sequence similarity in β_2 - α_1 loop region with the GlnDs (28, 29).

To understand the evolutionary relationships of ACT-like domains, phylogenetic trees were reconstructed with sequences of the ACT-like domains of 67 bacterial GlnDs, 18 chlorophyte ACRs, and 279 bHLH^{ACT+} proteins (Fig. 3 and [Datasets S7–S9](#)). The group III ACRs closely clustered with GlnDs with a supporting value of 73%/0.80 (bootstrap in ML tree/posterior probability in Bayesian tree; [Datasets S8 and S9](#)), consistently with previous studies (29). Noteworthy is that the bHLH-associated ACT-like domains form an independent clade, apart from the GlnDs and ACRs (bootstrap = 75%; Bayesian posterior probability = 0.61; [Datasets S8 and S9](#)). In the Bayesian tree, the bHLH-associated ACT-like domains grouped with the group I and II ACRs with high posterior probability (0.80), suggesting that the ACT-like domain from Plantae species likely arose from a single origin from its common ancestor with group I and II ACR. (Fig. 3 and [Datasets S8 and S9](#)). Furthermore, subfamily Vb clusters into a monophyletic clade, not grouped with other subfamilies, but is rather close to chlorophyte ACT-like domains (gray shade in Fig. 3 and [Datasets S8 and S9](#)), suggesting that subfamily Vb might have diverged earlier than other bHLH^{ACT+} subfamilies from their common ancestor. In accordance with this, the ancestor ACT-like domains inferred from the sequences of all ACT-like domains used in this study possess the conserved putative ligand-binding site represented by the DRPGLL motif in the β_2 - α_1 loop region as present in the GlnDs and ACRs ([SI Appendix, Fig. S7B](#)). A very similar motif is also found in chlorophyte ACT-like domains (DR[K/L]GLL) and

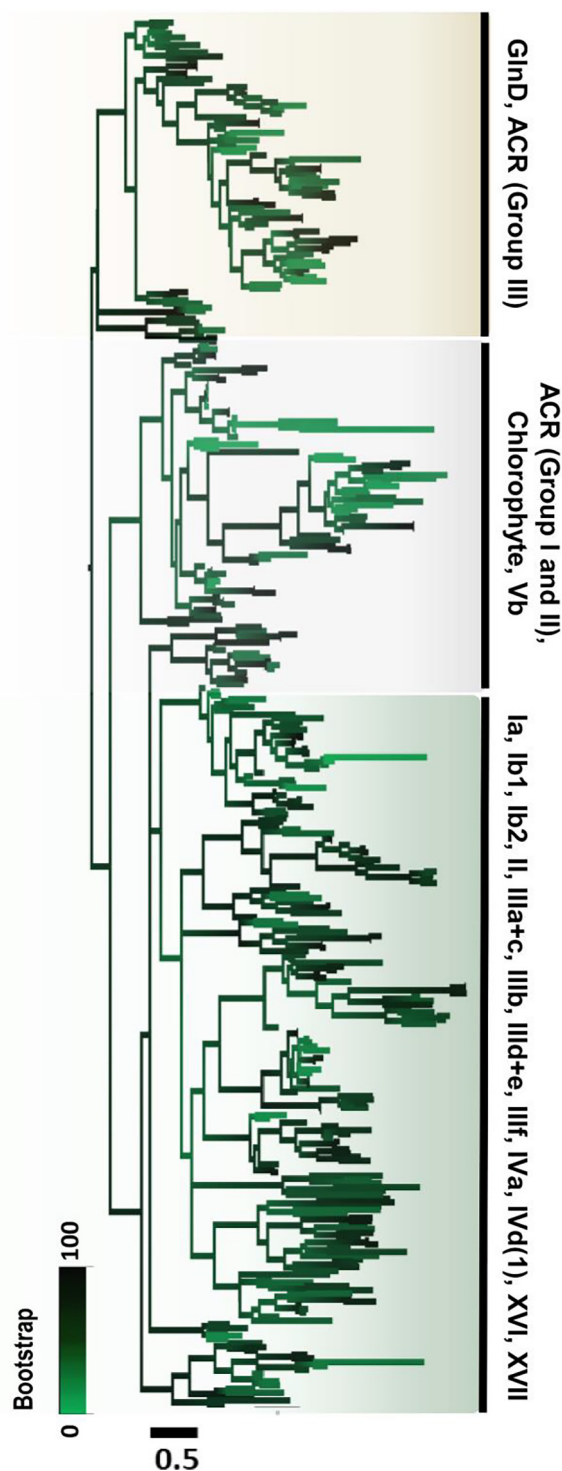


Fig. 3. bHLH-associated ACT-like domains form an independent clade distinct from ACRs and GlnDs. ML phylogeny was inferred from ACT-like domains of 279 bHLH proteins in 48 chlorophytes, two streptophyte algae, and 229 embryophytes; from 18 ACRs in chlorophytes, and 67 GlnDs in bacteria, using a JTT model implemented in RAXML-NG with 2,000 bootstraps. Shades in brown, gray, and green indicate monophyletic groups that consist of ACT-like domains of ACRs (Group III) and GlnDs, ACRs (Group I and II), chlorophytes and subfamily Vb, and 12 subfamilies (Ia, Ib1, Ib2, II, IIIa+c, IIIb, III+d+e, IIIf, IVa, IVd(1), XVI, and XVII). Blackish lines indicate branches with bootstrap support >70%. A detailed phylogenetic tree is presented in [Dataset S8](#).

subfamily Vb (DRPDLL), but not in other bHLH^{ACT+} subfamilies ([SI Appendix, Fig. S7B](#)) (42). Collectively, these results suggest that the bHLH-associated ACT-like domain descended from a common ancestor with ACR groups I and II.

Different Evolutionary Rates of the bHLH and Associated ACT-Like Domains Are Constrained by Correlated Substitutions.

Our phylogenetic analyses suggest that the ACT-like domains associated with the bHLH domains derived from a common ancestor and evolved by duplication. This notion implies that the ACT-like and bHLH domains coevolved; consequently, the two domains evolved with correlated substitution rates, despite higher amino acid sequence diversity of the ACT-like domain than that of the bHLH domains in the streptophytes (streptophyte algae and land plants) (Fig. 4A). Unlike the bHLH domains, MEME analyses showed that the streptophyte algae and embryophyte ACT-like domains have less conserved motifs compared to those of the chlorophytes (SI Appendix, Fig. S5).

To determine the substitution rates of the ACT-like and bHLH domains, the nonsynonymous-to-synonymous substitution rate ratio (ω) of the two domains was estimated with Brassicaceae sequences from 13 bHLH^{ACT+} subfamilies (Ia through Vb and XIII, those that have Arabidopsis orthologs) using PAML v4.9 (44) (SI Appendix, Tables S1 and S2 and Dataset S10). Under the

M_0 model (model = 0, fix_omega = 0) that allows the branches to have a single ω rate (45, 46), ACT-like domains showed significantly higher ω values ($0.081 < \omega < 0.51$) than those of the bHLH domains ($0.0072 < \omega < 0.27$) ($P = 2.4 \times 10^{-4}$), indicating much faster evolution rates than those of the associated bHLH-like domains, yet both under strong purifying selection (Fig. 4B). Despite the significant difference in ω values, the analyses showed good correlation between the ω values of the two domains (permutation test P -value = $2E-04$; $R^2 = 0.73$ from the observed data versus $R^2 = 0.29$ from 95% quantile in the randomized pairwise comparison) (Fig. 4C and SI Appendix, Fig. S8), suggesting that either the faster evolution of the ACT-like domains was constrained by the bHLH domains, or that other factors influence the evolutionary constraint of the two domains. In addition, the tanglegram displaying a pair of the phylogenetic trees of the ACT-like and bHLH domains with edges linked those derived from the same proteins demonstrates that the ACT-like domains among the same bHLH subfamilies cluster together (Fig. 4D and Dataset S11).

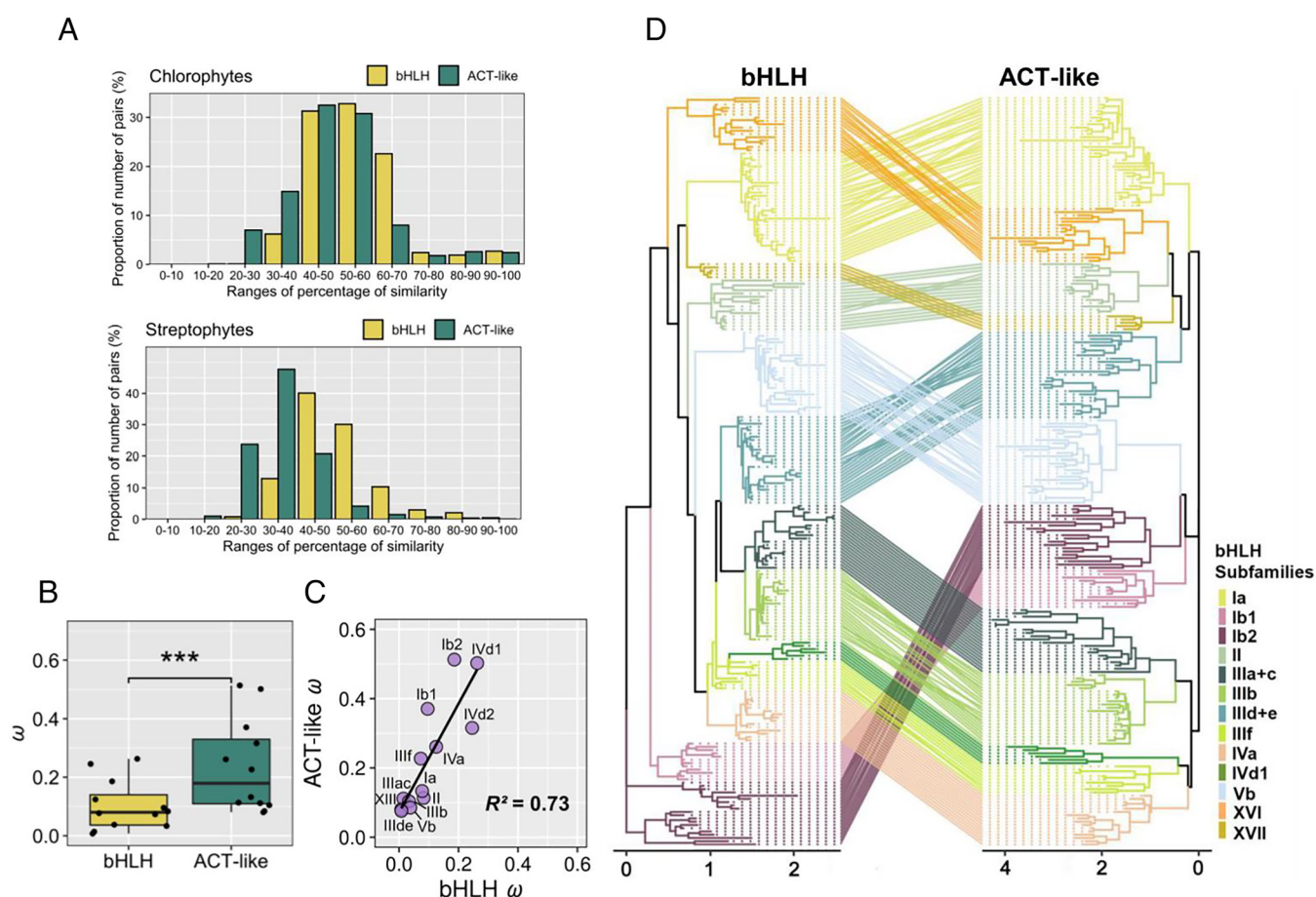


Fig. 4. Coevolution of ACT-like and bHLH domains. (A) Proportion of amino acid sequence similarity of ACT-like and bHLH domains in chlorophytes and streptophytes. The pairwise comparison analysis was performed with ACT-like and bHLH domains of 54 chlorophytes and 231 streptophytes. The proportion of the number of sequence pairs within the same range of similarity (number of pairs within same ranges of similarity/total number of pairs of bHLH domains and ACT-like domains $\times 100$) is displayed. (B) Nonsynonymous (dN)-to-synonymous substitutions (dS) rates (dN/dS; ω) in bHLH and ACT-like domains. The box plot shows the range of the global ω values of the bHLH and ACT-like domains in bHLH orthologs in Brassicaceae. The ω values of individual subfamilies are indicated as dots, and the data are provided in SI Appendix, Table S2. The horizontal lines in the boxes represent median values, box heights interquartile range, and the vertical lines data point within 1.5 \times interquartile range. Asterisks denote statistically significant difference ($P < 0.001$) in Wilcoxon signed-rank test. The species and genes used in this analysis are provided in SI Appendix, Table S1 and Dataset S10. (C) Correlated evolutionary rates between ACT-like domain and bHLH domains. Global ω rates were estimated from nucleotide sequences encoding the ACT-like and bHLH domains of bHLH orthologs in Brassicaceae. The line represents the linear regression, and the coefficient correlation value (R^2) is indicated in the graph. The ω values, species, and genes used in this analysis are provided in SI Appendix, Tables S1 and S2 and Dataset S10. (D) Tanglegram of bHLH and ACT-like domains in 231 streptophytes. ML phylogenetic trees of bHLH domains and ACT-like domains in streptophytes were individually reconstructed by RAXML-NG (43), with 1,000 bootstraps, based on "JTT+G" model, and plotted face to face with links between same proteins in the two trees. Line and tip colors indicate the various bHLH subfamilies. The x-axis represents branch lengths. Detailed phylogenetic trees with tip labels and bootstrap values are provided in Dataset S11.

Biochemical Functional Analyses of bHLH Proteins from a Chlorophyte. Previous studies reported that ACT-like domains can dimerize (20, 22, 30, 31), and that some modulate the DNA-binding and/or dimerization activity of the associated bHLH (21). To characterize possible functional consequence of the coevolution of the ACT-like and bHLH in chlorophyte bHLH TFs, three bHLH^{ACT+} proteins, four bHLH proteins containing incomplete ACT-like domains lacking the first β sheet (SI Appendix, Fig. S4), and one bHLH^{ACT-} protein

from *C. reinhardtii* were biochemically characterized (Fig. 5 and SI Appendix, Fig. S9). Briefly, we synthesized, cloned, and expressed for each protein the regions corresponding to the bHLH domain, ACT-like domain or bHLH domain and C-terminal regions (containing the ACT-like domain in bHLH^{ACT+}; Fig. 5A) in *Escherichia coli* as N-terminal SUMO-containing histidine-tagged (N₆His-SUMO-), histidine-tagged (N₆His-) and glutathione *S*-transferase (GST-) fusions that facilitated affinity purification by Ni-NTA and glutathione

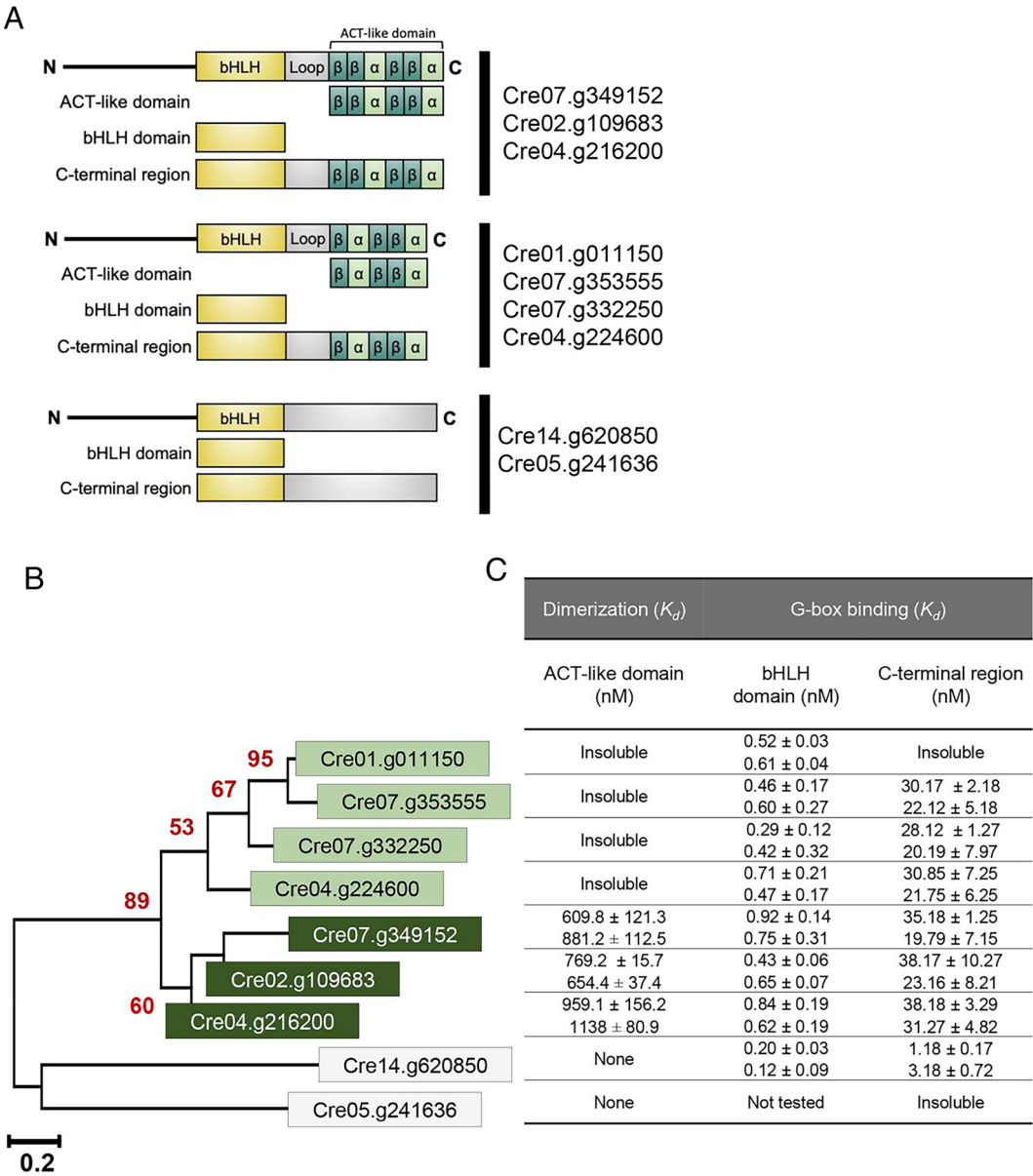


Fig. 5. Dimerization and influence of *C. reinhardtii* ACT-like domains on the DNA-binding activity of bHLH domains. (A) Schematic representation of the domain architecture of seven *C. reinhardtii* bHLH^{ACT+} and one bHLH^{ACT-} proteins showing the fragments expressed in *E. coli* as N-terminal fusions to N₆His-SUMO or N₆His and GST and affinity-purified. Yellow rectangles depict the bHLH domain, with gray indicating the loop regions, green the β -sheets, and yellowish green the α -helices on the ACT-like domain. (B) Phylogenetic relationships of the bHLH domains of the nine *C. reinhardtii* bHLH factors biochemically analyzed. The ML phylogenetic tree was constructed with the bHLH domains from the nine *C. reinhardtii* species using Le and Gascuel model with gamma distribution (LG+G) and 1,000 Felsenstein bootstraps in MEGA. Bootstrap values >50 are shown as percentages at the respective branch points. The dark green shades highlight proteins harboring ACT-like domains, and light green shades indicate those with ACT-like domain lacking the β_1 sheet. Gray shades indicate proteins without ACT-like domains. (C) Dissociation constants (K_d) for dimerization of ACT-like domains and G-box binding of bHLH domains and C-terminal regions (both bHLH and ACT-like domains). K_d values were determined by saturation binding assays using ALPHA. The ACT-like domains lacking the first β -sheet formed inclusion bodies when expressed in *E. coli* (Cre01.g011150, Cre07.g353555, Cre07.g332250, and Cre04.g224600). For the DNA-binding assays, various concentrations (0 to 4 μ M) of the GST-tagged proteins (Cre07.g349152, Cre02.g109683, and Cre04.g216200) were incubated with 500 nM of the same protein fused to N₆His-SUMO and subjected to fluorescence signal measurement. The G-box-binding strengths were determined by competition binding assays with N₆His-SUMO-tagged bHLH domains or His₆-SUMO-tagged C-terminal regions (SI Appendix, Materials and Methods). K_d values were calculated by one-site fit model in GraphPad Prism v6.0. For each interaction, K_d values (mean \pm SD) of two biological repeats are shown, and the SD reflects the variation between three technical replicates.

affinity chromatography, respectively (*SI Appendix, Fig. S10 and Materials and Methods*). We analyzed ACT homodimerization affinity by saturation binding assays using amplified luminescent proximity homogeneous assay (ALPHA) using the N₆His- and GST-tagged proteins and DNA-binding affinity using the N₆His-SUMO-tagged protein and a biotin-labeled double-stranded DNA oligonucleotide containing a G-box (*SI Appendix, Materials and Methods*).

The dimerization assays showed that the ACT-like domains from the three bHLH^{ACT+} proteins (Cre07.g349152, Cre02.g109683, and Cre04.g216200) have similar apparent dissociation constants (K_d) values (from 711.8 to 1,048.6 nM; Fig. 5C and *SI Appendix, Fig. S11 A and B*) as we previously determined for the Arabidopsis GL3 bHLH^{ACT+} TF (~1,540 nM) (22). When we quantitatively analyzed DNA-binding K_d for the eight N₆His-SUMO-tagged bHLH domains, we determined that they all bind the G-box with strong affinity (0.16 to 0.85 nM, Fig. 5C and *SI Appendix, Fig. S11 C and D*) and did not bind a mutated G-box double-stranded oligonucleotide probe (*SI Appendix, Fig. S12A*). However, when we tested for DNA binding for the N₆His-SUMO conjugated C-terminal regions containing both the ACT-like and bHLH domains, we determined that they exhibited significantly reduced DNA binding (almost 50 times lower) against the wild-type G-box probe (Fig. 5C and *SI Appendix, Fig. S11 E and F*), similar as was previously shown for R (21), with no binding to a mutated G-box probe (*SI Appendix, Fig. S12B*). The reduction in the DNA-binding activities was observed for all the eight proteins regardless of the presence of the first β -sheet in the ACT-like domains, suggesting that the presence of this β -sheet is not essential for this activity of ACT-like domains. To test for the possibility that any sequence C-terminal to the bHLH would negatively affect DNA binding, we assayed the DNA-binding activity of the bHLH plus C-terminal region of the bHLH^{ACT-} Cre14.g620850. Compared to the DNA binding of the respective bHLH, a modest decrease was observed, but certainly not as dramatic as for the bHLH^{ACT+} proteins (Fig. 5C), suggesting a potential difference in bHLH^{ACT-} and bHLH^{ACT+} bHLH domains, the affinity for the G-box of Cre14.g620850 (bHLH^{ACT-}) was 2 to 3 times higher than that for the bHLH domains associated with ACT-like motifs (compare the K_d for the bHLH domains in Fig. 5C). Taken together, these results indicate that chlorophyte ACT-like domains retain some of the functional characteristics attributed to vascular plant ACT-like domains.

Discussion

ACT-like domains are associated with a significant fraction of the large plant bHLH TFs and have been shown to be important for regulatory activity. While structurally conserved, ACT-like domains diverged significantly at the amino acid sequence level, complicating their global identification using solely protein sequence alignments. Using structure-based homology searches, we show that the bHLH and ACT-like domain association is first present in Plantae common ancestor. Similar to what was determined for land plant bHLH^{ACT+} proteins (20–22), ACT-like domains present in the chlorophyte *C. reinhardtii* dimerize and negatively modulate the DNA-binding activity of the associated bHLH motifs (Fig. 5), indicating that the regulatory functions on bHLH domains are likely ancestral activities of ACT-like domains. In addition, based on sequence and structural similarities, we propose a model in which the ACT-like domain associated with plant bHLH TFs derived from the ACT-like domain present in a common ancestor with ACR

proteins (Fig. 6), which in turn is likely to have originated from the ACT domain present in widely distributed GlnD proteins.

Our results suggest a monophyletic origin of the ACT-like domain associated with plant bHLH TFs, instead of multiple domain-shuffling events (12). Previous studies suggested that the origin of the bHLH domain precedes the split between the Plantae and opisthokont (including fungi and animals) lineages (12), and the bHLH family underwent independent amplification in both the land plants and the vertebrates (11). In accordance with these notions, we found prevalent presence of the ACT-like domains associated with the bHLH domains in Plantae, and could not find such evidence in opisthokont (*Dataset S1*), indicating a plant-specific origin of the ACT-like domains.

Our results reveal variable distribution of ACT-like domains among bHLH TFs in the green algae. Compared to vascular plants, green algae show a rather variable number of bHLH TFs (ranging from one to nine), and the fraction of them harboring ACT-like domains (ranging from 0 to 100%) is also variable (*SI Appendix, Fig. S2*). While the absolute numbers must be taken with some caution as some of the genomes are probably not fully sequenced/annotated yet, it is remarkable that six green algae have just one bHLH TF, which may or may not harbor an ACT-like domain. These results suggest that the radiation of bHLH^{ACT+} genes in green algae was associated with frequent amplification/loss of genes in different lineages.

Possible functions of the ACT-like domains in chlorophytes were investigated in *C. reinhardtii* for which there is a well-annotated genome (47). Previous studies in maize and Arabidopsis showed that ACT-like domains can provide additional dimerization opportunities for bHLH^{ACT+} TFs (20, 22), and that ACT-like domains negatively affect the DNA-binding activity of the associated bHLH motif (21). Our result shows that this is indeed the case for *C. reinhardtii* bHLH^{ACT+} (Fig. 5). *C. reinhardtii* ACT-like domains dimerize with an affinity comparable to those of vascular plants (22), and the ACT-like domain significantly diminishes the DNA-binding activity of the associated bHLH domain toward canonical E-box DNA motifs. In addition, out of the nine bHLH TFs present in *C. reinhardtii*, four bHLH^{ACT+} genes with ACT-like domains with the $\beta\alpha\beta\beta\alpha$ fold, instead of the more common $\beta\beta\alpha\beta\alpha$ (Fig. 5), retain the ability to inhibit DNA binding by the adjacent bHLH domain (Fig. 5 and *SI Appendix, Fig. S11*). Our results suggest that the $\beta\beta\alpha\beta\alpha$ fold represents the ancestral ACT-like domain and that $\beta\alpha\beta\beta\alpha$ variants derive from it, but establishing this will require additional analyses and experimentation. The mechanism by which the ACT-like domain has an inhibitory effect on the DNA-binding activity of the bHLH remains unknown and so are the regulatory consequences of this modulation, but the small number of bHLH genes in *C. reinhardtii* furnishes a singular opportunity to investigate this. From an evolutionary perspective, the apparently conserved homodimerization activity of ACT-like domains may have allowed bHLH^{ACT+} genes to be more stably maintained after gene duplication events, as homodimerization will be less affected than heterodimerization by the gene dosage imbalance resulting from the duplication (48). ACT-mediated heterodimer formation observed in some instances may have evolved after neo- or sub-functionalization of the duplicates has been stabilized (30, 31).

The bHLH family of TFs is characterized by the presence of other domains (besides the ACT-like) that, when present, could provide additional phylogenetic information (9). However, not all bHLH subfamilies possess these additional domains or conserved sequences outside the bHLH. Thus, our phylogenetic reconstructions rely solely on the conservation of the bHLH (8). The challenges associated with drawing conclusions regarding

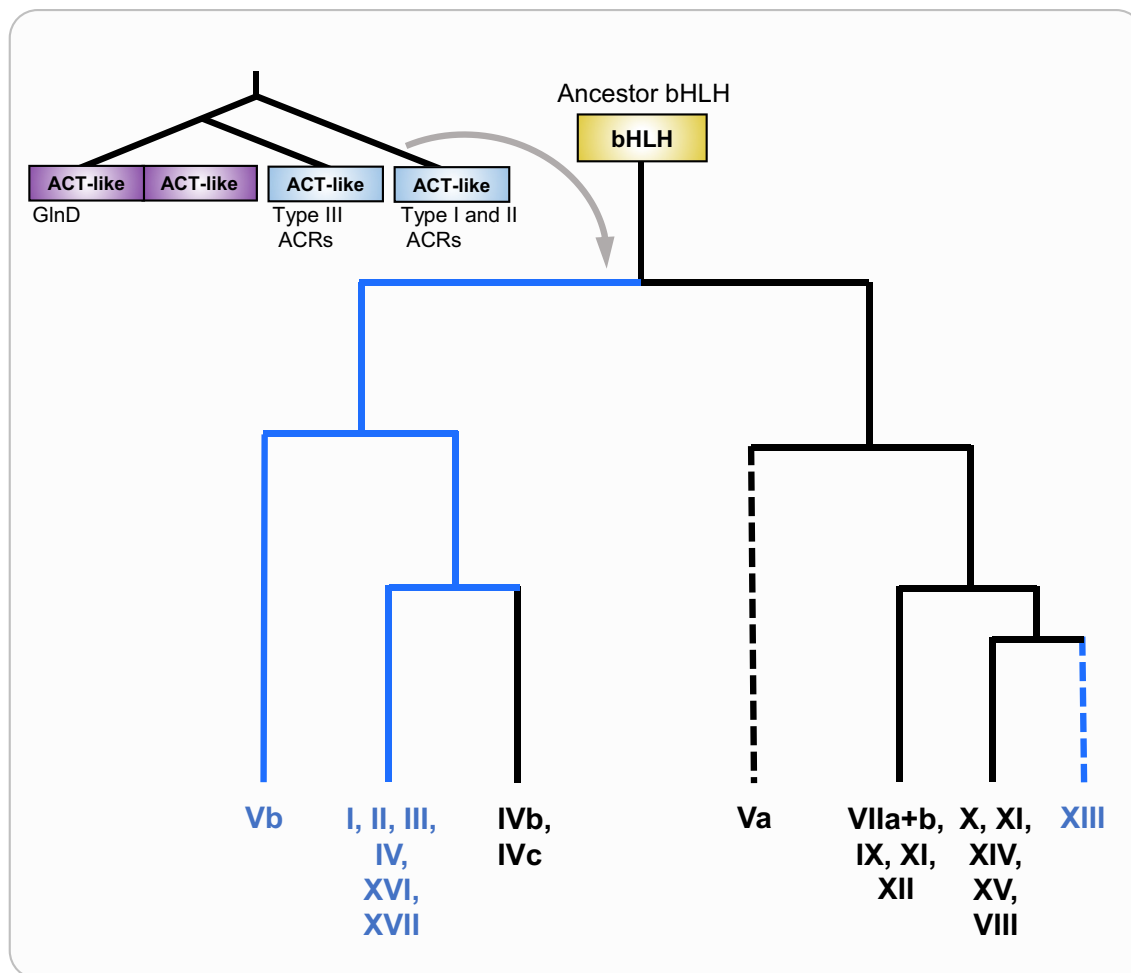


Fig. 6. Proposed evolutionary model of ACT-like domain in plant bHLH TFs. Possible parsimonious evolutionary model of ACT-like domain in bHLH TFs is proposed based on the relationships inferred from the phylogenies presented in this study. Ancestor bHLH protein acquired the ACT-like domain from type I or II ACRs. Then, subfamilies XIII and Vb diverged early before the division of streptophyte algae and embryophytes, while other bHLH^{ACT+} subfamilies radiated in the evolution of embryophytes. Subfamilies IVb and IVc might descend from a common ancestor with most bHLH^{ACT+} subfamilies but might have lost the ACT-like domain later. Most bHLH^{ACT-} subfamilies, except for IVb and IVc, evolved along different lineages. The arrow in gray represents the association event between the ACT-like and bHLH domains. The blue line indicates the lineages of the bHLH^{ACT+}. Dashed lines indicate branches whose positions are not clearly resolved, based on the available data.

major events in the evolution of the family are illustrated by members of subfamilies Vb and XIII. In *Arabidopsis*, subfamily XIII is rather small with four members, including *LHW* and three *LHW-LIKE* (*LL1-3*) genes (32, 49). *LHW* participates in the regulation of vascular cell proliferation together with *TARGET OF MONOPTEROS* (*TMO5*), another bHLH TF which belongs to subfamily Vb. Both *LHW* and *TMO5* have ACT-like domains, and the two proteins interact through the bHLH domain (50), although *LHW* can also form homodimers (49). Subfamily XIII *LHW* orthologs were identified in all major plant lineages; however, bryophyte *LHW* orthologs lack the ACT-like domain, and domain-swapping experiments demonstrated that this domain was essential for vascular cell proliferation (32). The position of subfamily XIII in the bHLH phylogeny (Fig. 2) and the finding that liverworts subfamily XIII lacks ACT-like domains while vascular plants have it suggest the independent acquisition of an ACT-like domain to result in subfamily XIII in a common ancestor of the streptophyte algae and the embryophytes, and subsequently lost it in the liverworts lineage (51). In support of this, we identified at least one subfamily XIII bHLH gene (GAQ82294) harboring an ACT-like domain in *K. nitens* (Klebsormidiophyceae; streptophyte algae) (*SI Appendix*, Fig. S13).

The emergence of subfamily Vb is difficult to determine based on the phylogeny of the bHLH domain because it is not strongly supported (SH-aLRT = 62%) (Fig. 2). However, the presence of Vb in the streptophyte algae (Fig. 2) suggests a very ancient origin basal to other bHLH^{ACT+} genes (Fig. 6). The ACT-like domains of subfamilies Vb and XIII are more similar to each other than expected from the bHLH phylogeny (*Dataset S12*). When taken together with the finding that the bHLHs of *LHW* (XIII) and *TMO5* (Vb) interact (32), it is possible that selection acted on subfamily XIII bHLH, resulting in the divergence that characterizes this subfamily, to preserve interaction with Vb members. If so, then the position of subfamily XIII could branch out from subfamily Vb (Fig. 6).

As is the case for many other eukaryotic TFs, bHLH factors often have a number of other domains that participate in the formation of unique regulatory complexes. How such ability to recruit additional partners has evolved is of fundamental importance, and the association of bHLH with ACT-like domains provides an attractive paradigm for how this might happen, and the biological consequences associated. TFs harboring bHLH domains have so far only been found in eukaryotes, and our studies indicate that the association of the bHLH and ACT-like domain is unique to the plants. It is intriguing however that both archaeobacteria and eubacteria harbor a group of TFs that belong to the Feast/Famine

Regulatory Proteins in which the N-terminal DNA-binding helix–turn–helix domain is separated by a loop region from a C-terminal Regulation of Amino-acid Metabolism (RAM) domain (51, 52). The RAM domain has a very similar topology ($\beta\alpha\beta\beta\alpha\beta$) to ACT domains, but their functions and small-molecule binding activities have rendered them a separate name (51, 53). Thus, evolution might have toyed with similar combinations of structural domains at least twice independently.

Materials and Methods

Protein Sequence Analyses. A total of 140 eukaryote bHLH TF sequences were identified by Pfam ID (PF00010). The $\beta\alpha\beta\beta\alpha\beta$ secondary structures and tertiary protein structures of the ACT-like domains were predicted using PSIPRED (54) and AlphaFold (38, 39) via the ColabFold interface (55). The presence of ACT-like domains in various organisms was investigated with PSI-BLAST, Conserved Domain Database search in NCBI (<https://blast.ncbi.nlm.nih.gov>) (56), and Pfam v34.0 database search (<http://pfam.xfam.org/>) (57). Detailed methods for the identification of the ACT-like domains are presented in *SI Appendix, Materials and Methods*.

Phylogenetic Tree Analyses and Ancestral Sequence Reconstruction. ML trees of the ACT-like and bHLH domains derived from Chloroplastida were constructed by best-fit substitution models chosen by ProtTest v3.4.2 (58) with 1,000 or 2,000 iterations using the Transfer Bootstrap method (59), or using

the SH-aLRT implemented in RAxML-NG v1.1.0 (43). The selected models and number of iterations are provided for each figure and in *SI Appendix, Materials and Methods*. Bayesian trees were estimated using MrBayes v3.2.7 (60) by summarizing 25,000 trees generated from two independent runs that converged after 7.5 million generations with the SD of split frequencies <0.02. The resulting extended majority consensus tree was used. Detailed methods are provided in *SI Appendix, Materials and Methods*.

DNA Binding and Protein–Protein Interaction Assays. The homodimerization activities of the ACT-like domains and the G-box DNA-binding activities of the bHLH and ACT-like domains, and of the C-terminal regions containing both the ACT-like and bHLH domains, were determined in vitro by ALPHA. Detailed information is described in *SI Appendix, Materials and Methods*.

Detailed methods for the identification of the ACT-like domains, the protein sequence analyses, the reconstruction of the phylogenetic trees, the estimation of the evolutionary rates, the plasmid cloning, the recombinant protein purification, and the ALPHA are available in *SI Appendix, Materials and Methods*.

Data, Materials, and Software Availability. All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank Eric Mukundi Maina for extensive assistance with computational analyses. We appreciate the bioinformatics assistance of Fabio Gomez-Cano and the technical assistance provided by Ryan Krueger. This work was supported by grants from the NSF MCB-1822343 and IOS-1733633 to E.G.; IOS-2107215 and MCB-2210431 to S.-H.S.

1. S. A. Lambert *et al.*, The human transcription factors. *Cell* **172**, 650–665 (2018).
2. J. L. Riechmann, O. J. Ratcliffe, A genomic perspective on plant transcription factors. *Curr. Opin. Plant Biol.* **3**, 423–434 (2000).
3. V. Ledent, M. Vervoort, The basic helix-loop-helix protein family: Comparative genomics and phylogenetic analysis. *Genome Res.* **11**, 754–770 (2001).
4. J. L. Riechmann *et al.*, Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* **290**, 2105–2110 (2000).
5. M. E. Massari, C. Murre, Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. *Mol. Cell Biol.* **20**, 429–440 (2000).
6. S. R. Ludwig, L. F. Habera, S. L. Dellaporta, S. Wessler, Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7092–7096 (1989).
7. S. R. Ludwig, S. R. Wessler, Maize R gene family: Tissue-specific helix-loop-helix proteins. *Cell (Cambridge)* **62**, 849–851 (1990).
8. G. Toledo-Ortiz, E. Huq, P. H. Quail, The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* **15**, 1749–1770 (2003).
9. M. A. Heim *et al.*, The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* **20**, 735–747 (2003).
10. X. Li *et al.*, Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiol.* **141**, 1167–1184 (2006).
11. A. Feller, K. Machemer, E. L. Braun, E. Grotewold, Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* **66**, 94–116 (2011).
12. N. Pires, L. Dolan, Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol. Biol. Evol.* **27**, 862–874 (2010).
13. L. Carretero-Paulet *et al.*, Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiol.* **153**, 1398–1412 (2010).
14. B. Catarino, A. J. Hetherington, D. M. Emms, S. Kelly, L. Dolan, The stepwise increase in the number of transcription factor families in the Precambrian predated the diversification of plants on land. *Mol. Biol. Evol.* **33**, 2815–2819 (2016).
15. W. R. Atchley, W. M. Fitch, A natural classification of the basic helix-loop-helix class of transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5172–5176 (1997).
16. S. A. Goff, K. C. Cone, V. L. Chandler, Functional analysis of the transcriptional activator encoded by the maize B gene: Evidence for a direct functional interaction between two classes of regulatory proteins. *Genes Dev.* **6**, 864–875 (1992).
17. E. Grotewold *et al.*, Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13579–13584 (2000).
18. I. M. Zimmermann, M. A. Heim, B. Weisshaar, J. F. Uhrig, Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like bHLH proteins. *Plant J.* **40**, 22–34 (2004).
19. N. A. Ramsay, B. J. Glover, MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* **10**, 63–70 (2005).
20. A. Feller, J. M. Hernandez, E. Grotewold, An ACT-like domain participates in the dimerization of several plant basic-helix-loop-helix transcription factors. *J. Biol. Chem.* **281**, 28964–28974 (2006).
21. Q. Kong *et al.*, Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2091–2097 (2012).
22. Y. S. Lee, A. Herrera-Tequila, J. Silwal, J. H. Geiger, E. Grotewold, A hydrophobic residue stabilizes dimers of regulatory ACT-like domains in plant basic helix-loop-helix transcription factors. *J. Biol. Chem.* **296**, 100708 (2021).
23. A. Feller, L. Yuan, E. Grotewold, The BIF domain in plant bHLH proteins is an ACT-like domain. *Plant Cell* **29**, 1800–1802 (2017).
24. G. A. Grant, The ACT domain: A small molecule binding domain and its role as a common regulatory element. *J. Biol. Chem.* **281**, 33825–33829 (2006).
25. D. M. Chipman, B. Shaanan, The ACT domain family. *Curr. Opin. Struct. Biol.* **11**, 694–700 (2001).
26. E. J. Lang, P. J. Cross, G. Mittelstädt, G. B. Jameson, E. J. Parker, Allosteric ACTION: The varied ACT domains regulating enzymes of amino-acid metabolism. *Curr. Opin. Struct. Biol.* **29**, 102–111 (2014).
27. Y. Zhang, E. L. Pohlmann, J. Serate, M. C. Conrad, G. P. Roberts, Mutagenesis and functional characterization of the four domains of GlnD, a bifunctional nitrogen sensor protein. *J. Bacteriol.* **192**, 2711–2721 (2010).
28. M.-H. Hsieh, H. M. Goodman, Molecular characterization of a novel gene family encoding ACT domain repeat proteins in Arabidopsis. *Plant Physiol.* **130**, 1797–1806 (2002).
29. T.-Y. Sung, T.-Y. Chung, C.-P. Hsu, M.-H. Hsieh, The ACR11 encodes a novel type of chloroplastic ACT domain repeat protein that is coordinately expressed with Gln2 in Arabidopsis. *BMC Plant Biol.* **11**, 118 (2011).
30. J. Cui *et al.*, Feedback regulation of DYT1 by interactions with downstream bHLH factors promotes DYT1 nuclear localization and anther development. *Plant Cell* **28**, 1078–1093 (2016).
31. H. Seo *et al.*, Intragenic suppressors unravel the role of the SCREAM ACT-like domain for bHLH partner selectivity in stomatal development. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117774119 (2022).
32. K.-J. Lu *et al.*, Evolution of vascular plants through redeployment of ancient developmental regulators. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 733–740 (2020).
33. K. H. Wolfe, M. Gouy, Y.-W. Yang, P. M. Sharp, W.-H. Li, Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6201–6205 (1989).
34. H. Arai, K. Yanagiura, Y. Toyama, K. Morohashi, Genome-wide analysis of MpBHLH12, a IIIa basic helix-loop-helix transcription factor of Marchantia polymorpha. *J. Plant Res.* **132**, 197–209 (2019).
35. F. Burki, A. J. Roger, M. W. Brown, A. G. B. Simpson, The new tree of eukaryotes. *Trends Ecol. Evol.* **35**, 43–55 (2020).
36. S. A. Rensing, How plants conquered land. *Cell* **181**, 964–966 (2020).
37. R. Lewin, When does homology mean something else? *Science* **237**, 1570–1570 (1987).
38. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
39. M. Varadi *et al.*, AlphaFold Protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
40. M. van Kempen *et al.*, Foldseek: Fast and accurate protein structure search. *bioRxiv* [Preprint] (2022). <https://doi.org/10.1101/2022.02.07.479398> (Accessed 28 March 2023).
41. H. S. Yoon, J. D. Hackett, C. Ciniglia, G. Pinto, D. Bhattacharya, A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* **21**, 809–818 (2004).
42. H.-S. Liao, Y.-H. Chung, C. Chardin, M.-H. Hsieh, The lineage and diversity of putative amino acid sensor ACR proteins in plants. *Amino Acids* **52**, 649–666 (2020).
43. A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
44. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
45. N. Goldman, Z. Yang, A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).
46. Z. Yang, R. Nielsen, Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **46**, 409–418 (1998).
47. S. S. Merchant *et al.*, The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* **318**, 245–250 (2007).
48. D. Lang, S. A. Rensing, “The evolution of transcriptional regulation in the viridiplantae and its correlation with morphological complexity” in *EVOLUTIONARY TRANSITIONS Multicellular Life: Principles and mechanisms*, I. Ruiz-Trillo, A. M. Nedelcu, Eds. (SpringerLink, 2015), pp. 301–333.

49. K. Ohashi-Ito, D. C. Bergmann, Regulation of the Arabidopsis root vascular initial population by LONESOME HIGHWAY. *Development* **134**, 2959–2968 (2007).
50. B. De Rybel *et al.*, A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in Arabidopsis. *Developmental Cell* **24**, 426–437 (2013).
51. C. A. Ziegler, P. L. Freddolino, The leucine-responsive regulatory proteins/feast-famine regulatory proteins: an ancient and complex class of transcriptional regulators in bacteria and archaea. *Critical Rev. Biochem. Mol. Biol.* **56**, 373–400 (2021).
52. P. M. Leonard *et al.*, Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. *EMBO J.* **20**, 990–997 (2001).
53. T. J. Ettema, A. B. Brinkman, T. H. Tani, J. B. Rafferty, J. Van Der Oost, A novel ligand-binding domain involved in regulation of amino acid metabolism in prokaryotes. *J. Biol. Chem.* **277**, 37464–37468 (2002).
54. L. J. McGuffin, K. Bryson, D. T. Jones, The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
55. M. Mirdita *et al.*, ColabFold-Making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2021).
56. S. Lu *et al.*, CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265–D268 (2020).
57. J. Mistry *et al.*, Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2020).
58. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
59. F. Lemoine *et al.*, Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
60. F. Ronquist *et al.*, MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biol.* **61**, 539–542 (2012).