DNA features beyond the transcription factor binding site specify target recognition by plant MYC2-related bHLH proteins

Irene López-Vidriero1,3, Marta Godoy1,3, Joaquín Grau2, María Peñuelas2, Roberto Solano2 and José M. Franco-Zorrilla1,2,3,*

1Genomics Unit, Centro Nacional de Biotecnología, CSIC, C/Darwin 3, 28049 Madrid, Spain
2Department of Plant Molecular Genetics, Centro Nacional de Biotecnología, CSIC, C/Darwin 3, 28049 Madrid, Spain
3These authors contributed equally to this article.
*Correspondence: José M. Franco-Zorrilla (jmfranco@cnb.csic.es)
https://doi.org/10.1016/j.xplc.2021.100232

ABSTRACT

Transcription factors (TFs) regulate gene expression by binding to cis-regulatory sequences in the promoters of target genes. Recent research is helping to decipher in part the cis-regulatory code in eukaryotes, including plants, but it is not yet fully understood how paralogous TFs select their targets. Here we addressed this question by studying several proteins of the basic helix-loop-helix (bHLH) family of plant TFs, all of which recognize the same DNA motif. We focused on the MYC-related group of bHLHs, that redundantly regulate the jasmonate (JA) signaling pathway, and we observed a high correspondence between DNA-binding profiles in vitro and MYC function in vivo. We demonstrated that A/T-rich modules flanking the MYC-binding motif, conserved from bryophytes to higher plants, are essential for TF recognition. We observed particular DNA-shape features associated with A/T modules, indicating that the DNA shape may contribute to MYC DNA binding. We extended this analysis to 20 additional bHLHs and observed correspondence between in vitro binding and protein function, but it could not be attributed to A/T modules as in MYCs. We conclude that different bHLHs may have their own codes for DNA binding and specific selection of targets that, at least in the case of MYCs, depend on the TF-DNA interplay.

Key words: transcription factor, bHLH, target specificity, plants

López-Vidriero I., Godoy M., Grau J., Peñuelas M., Solano R., and Franco-Zorrilla J.M. (2021). DNA features beyond the transcription factor binding site specify target recognition by plant MYC2-related bHLH proteins. Plant Comm. 2, 100232.

INTRODUCTION

Plants need to adapt their growth and development to a fluctuating environment for survival. These adaptive responses are tightly regulated, and signaling cascades most often converge at the activity of sequence-specific transcription factors (TFs). TFs are responsible for the transcriptional regulation of target genes by recognition of specific, short DNA sequences (6–12 base pairs [bp]) in their regulatory regions, referred to as TF-binding sites (TFBSs). During the last decade, a great effort has been made to decipher the transcriptional cis-regulatory code and the TFs involved in the regulation of gene expression in plants. Experimental approaches in this undertaking involve in vivo and in vitro studies (reviewed in Franco-Zorrilla and Solano, 2017). In vivo strategies, such as chromatin immunoprecipitation and sequencing (ChIP-seq), directly yield the TF target genes in planta and contribute to defining with precision the binding sites important for gene regulation. In vitro approaches also contribute to the definition of TF-bound sequences. For example, the use of universal protein binding microarrays (uPBMs) has helped in the description of hundreds of plant TFBSs (Franco-Zorrilla et al., 2014; Weirauch et al., 2014; Franco-Zorrilla and Solano, 2017). Information derived from uPBMs, although extremely helpful, does not completely reflect all the binding possibilities of the TFs in the genome, and the identification of target genes is not straightforward. DNA affinity purification and sequencing (DAP-seq) of the products overcomes the shortcomings of uPBMs, as it identifies the plant genomic fragments that interact with the protein in vitro,
and it is easily scalable to different plant species (O’Malley et al., 2016; Bartlett et al., 2017).

Jasmonates (JAs) are oxylipin hormones that regulate multiple plant processes, such as defense, growth, and development (Howe et al., 2018). The JA signaling pathway is well established in higher plants. The bioactive hormone (+)-7-iso-JA-ile (Fonseca et al., 2009) is perceived by the co-receptor complex formed by the E3-ubiquitin ligase SCF^{COI1} (Skp1-Cul1-F-box protein Coronatine-Insensitive1) and a JASMONATE-ZIM DOMAIN (JAZ) protein (Chini et al., 2007; Thines et al., 2007). In the absence of the hormone, JAZs repress several TFs through a dual mechanism, preventing interaction with MEDIATOR COMPLEX SUBUNIT25 and recruiting the co-repressor TOPLESS (Howe et al., 2018). Upon binding of the hormone, JAZ proteins are polyubiquitinated and degraded by the proteasomal complex (Chini et al., 2007; Thines et al., 2007), releasing TFs for regulation of direct targets.

Among the TFs directly regulated by this mechanism, the basic helix-loop-helix (bHLH) protein MYC2 is the master regulator of JA-mediated responses, activating defense gene expression, stomatal closure, photomorphogenic growth, and synthesis of glucosinolates, terpenoids, and other specialized metabolites (Hong et al., 2012; Schweizer et al., 2013; Gimenez-Ibanez et al., 2017; Ortigosa et al., 2020). MYC3, MYC4, and MYC5 are homolog bHLHs that act redundantly with MYC2 in the same pathway to regulate most COI1-JAZ-dependent responses and modulate the defense-growth balance (Fernández-Calvo et al., 2011; Major et al., 2017; Song et al., 2017; Ortigosa et al., 2020). Other similar bHLHs in Arabidopsis and tomato, such as bHLH03, bHLH13, and bHLH17, do not contain a transcriptional activation domain and behave as repressors by competing with activator MYCs for their TFBS (Sasaki-Sekimoto et al., 2013; Fonseca et al., 2014). Recent studies have demonstrated that the JA-signaling pathway is not exclusive to higher plants, as the non-vascular liverwort Marchantia polymorpha (Marchantia) contains all the components of this signaling pathway (COI1, JAZ, and MYC), and their functions are conserved (Monte et al., 2018, 2019; Pefuélas et al., 2019).

The DNA binding of MYCs has been widely studied, using either in vitro approaches (Dombrecht et al., 2007; Fernández-Calvo et al., 2011; Godoy et al., 2011; Fonseca et al., 2014) or ChiP-seq studies in response to JA (Wang et al., 2019; Zander et al., 2020). MYCs bind to the G-box (5’-CACGTG-3’), a DNA motif typically recognized by bHLH proteins from plants, yeast, and animals (Ledent and Vervoort, 2001). High-throughput studies have also shown the importance of two related motifs, the so-called T/G-box (5’-AAGCTG-3’) and the PIF-binding element (PBE, 5’-CATGTG-3’) (Godoy et al., 2011). Some other similar motifs are recognized by MYC2, such as the G/A and G/C elements (5’-CAGGAG-3’ and 5’-ACGGG-3’, respectively), but their contribution to the JA-mediated transcriptional pathway is lower than that of the G- and T/G-boxes or the PBE (Godoy et al., 2011).

Despite the large number of studies on the DNA sequences recognized by MYCs, we have still not identified the molecular bases that govern MYC2-related recognition of specific target genes, and by extension the bHLH family members, among a set of possible targets in the genome. In other words, it is still unclear how paralog TFs that recognize the same motif are able to discriminate their target genes. Many studies have shown that cooperativity and synergy between TFs increase specificity and restrict binding to bona fide target genes (Brljacic and Grotewold, 2017; Morgunova and Taipale, 2017). Moreover, differential methylation of cytosines in the binding site may also promote or inhibit TF recognition (O’Malley et al., 2016; Yin et al., 2017). However, in vitro approaches can be exploited to obtain all possible intrinsic binding properties that may contribute to increased specificity. In this work, we analyzed the binding patterns of MYC2, 3, 4, and bHLH17 in vitro, using bHLH-specific PBM s and DAP-seq, and we observed a correlation of these data with in vivo ChiP-seq data, indicating that these TFs are autonomous in their determination of binding specificity. Analysis of Marchantia and tomato orthologs showed that their transcriptional code for specificity is conserved from bryophytes to higher plants. We extended the analysis to 20 additional bHLHs and observed that, at least for six of them, these proteins may also have their own particular code for increasing the specificity of targets, different from that observed in MYCs.

**RESULTS**

**Refining the MYC-binding site using bHLH-specific PBM s**

To study in detail the binding properties of MYC proteins in vitro, we designed two Arabidopsis bHLH-specific Protein Binding Microarrays (bHLH-PBMs). These contained oligonucleotide probes with one of the MYC-recognized motifs, the G-box, PBE, and T/G-box, globally referred to as MYC-binding sites (MBSs), located at the center of the oligonucleotide and flanked by the sequence context from the Arabidopsis genome (Supplemental Figure 1). The two designs, ‘genome’ and ‘promoters’, differed in the number of MBSs and replicated probes (131 401 and 52 574 different probes, respectively). We used both designs to analyze Arabidopsis MYC2, 3, and 4 and the repressor bHLH17.

We obtained very similar results among the four MYCs and with the two bHLH-PBM designs, consisting in a higher binding affinity to the G-box than to the other MBS motifs (Figure 1A and Supplemental Figure 2). A restrictive alignment using the top 500 probes revealed a preference for purine/pyrimidine bases flanking the G-box at the 5’ and 3’ ends, as described previously (Godoy et al., 2011), and a short stretch of A/T nucleotides flanking the G-box, about four and seven nucleotides away from the G-box. By contrast, restrictive alignment of the bottom 500 probes revealed a complementary pattern (Figure 1B and Supplemental Figure 2). We quantified the A/T content in positions flanking the G-box and observed an enrichment at positions ±5 in top probes relative to bottom probes (0.67/0.71 compared with 0.53/0.55, respectively) (Figure 1C). In addition, we could confirm that the A/T composition in these positions was associated with a higher binding intensity than in any other positions in the probe (Figure 1D). These results prompted us to speculate that the short A/T stretches, although not highly marked, might be additional determinants of MYC DNA-binding specificity.
Figure 1. Genome bHLH-PBMs provide high specificity and predict function in vivo.

(A) Signal distributions of the probes with the DNA elements indicated after incubation with the MYC proteins. Signal intensities of the probes are presented as Log2[box/Gm], where Gm represents the mutant G-box. Letters above the boxes represent groups of statistical significance at p < 0.05 (ANOVA and post hoc classification with Tukey’s test; different letters denote different groups). Data correspond to the genome design.

(B) Logos representing sequence alignment of the top 500 (left) and bottom 500 (right) probes, sorted by Log2[box/Gm], in the genome design. Short A/T-rich stretches centered at ±4–7 bp in top probes are highlighted in dark blue.

(C) [A + T] content at each position of the 500 probes bound with the highest (top) and lowest intensities (bottom). Nucleotide coordinates are numbered from the G-box toward the ends.

(D) Signal distributions (in Log2[box/Gm]) of the sets of probes containing an A/T at each of the positions of the probe. Nucleotide coordinates are as in (C). Letters below the boxes represent groups of statistical significance at p < 0.05 (ANOVA and post hoc classification with Tukey’s test; different letters denote different groups).

(legend continued on next page)
Plant Communications

In vitro DNA-binding specificity correlates with MYC target selection in vivo

We analyzed bHLH-PBM data in the context of MYC2-related functionality in order to examine a possible correlation with in vivo DNA binding. We classified oligonucleotides into deciles and analyzed the distribution of the probes in the first (D1) and the tenth (D10) deciles across different gene features. We observed that 36.7%–40% of D1 probes were in promoters, proportions markedly higher than those for the whole MBS set (30.6%) and for a random set of fragments (29.4%) (Supplemental Figure 3), with particular enrichments within 1 kb upstream of the transcription start site (TSS). As expected, proportions of D10 probes in promoters were similar to the complete MBS and almost identical among MYCs (Supplemental Figure 3).

We then examined the correspondence between MYC2-related function and DNA-binding specificity. We took advantage of an available time-course experiment in response to JA in Arabidopsis (Hickman et al., 2017) and analyzed the expression of genes associated with the D1 and D10 probes. Genes associated with D1 probes in promoters were, on average, highly expressed in response to JA, peaking at 1.5–2 h after treatment, whereas expression of the D10-associated genes did not change after hormone treatments (Figure 1E, Supplemental Figure 4). In general, we did not observe marked differences in other gene features, and all the MYC2-related TFs tested showed similar patterns (Supplemental Figure 4).

Taken together, these data point to a correspondence between MYC DNA-binding specificity in vitro and MYC function in planta. In addition, and thanks to in vitro experiments, we can conclude that the code for specificity resides in the protein-DNA interplay and does not depend, or has limited dependence, on additional cofactors.

DNA-binding properties of MYC2-related factors are conserved in plants

To further confirm the previous conclusion, we re-examined MYC2, 3, and 4 and bHLH17 DNA-binding patterns by DAP-seq. We also analyzed several MYC2-related proteins from Solanum lycopersicum (tomato), such as the tomato MYC2 ortholog (Soly08g076930, SIMYC2), the very close protein Soly08g005050 (SIMYC2 related, SIMYC2r), and the highly similar Soly10g009290 (SIMYC2r2) (Supplemental Figure 5). In addition, we studied the female MpMYCx from the non-vascular liverwort Marchantia, as recently described (Peñuelas et al., 2019).

We incubated MBP-fused proteins with DNA libraries generated with genomic DNA (gDNA) obtained from plants grown at standard conditions. We performed two independent experiments with Arabidopsis MYC2-related that showed high correlation scores and yielded similar numbers of significantly enriched peaks (q < 0.05; Supplemental Figure 6A and Supplemental Table 1). Moreover, the overlap of significant peaks between replicates ranged from 90.5% to 97% (Supplemental Figure 6B). Significant peaks differed among the four proteins analyzed, indicating that the binding activity depends on the intrinsic properties of the proteins (Supplemental Figure 6B and Supplemental Table 1). In the case of tomato and Marchantia MYC2-related, we performed only one experiment, and the numbers of significant peaks are shown in Supplemental Table 1.

To assess the quality of the results, we evaluated the enrichment of MBS in bound peaks using a naive pattern-matching method. For the Arabidopsis factors, all the motifs were very highly enriched at the centers of the peaks, the G-box being the most prominent, showing up to 1000-fold enrichment, followed by PBE (up to 150-fold enrichment) and T/G (50-, 60-fold enrichment), similar to what we observed using bHLH-specific PBMs (Figure 2A and Supplemental Figure 7). The tomato SIMYC counterparts bound the PBE more efficiently than their relatives in Arabidopsis, but overall, binding to the MBS elements was very similar among eudicot MYC2-related TFs. A similar result was obtained for MpMYCx, but binding to the T/G-box was higher than to the PBE, in contrast to eudicot MYCs (Figure 2A and Supplemental Figure 7). Therefore, although the pattern-matching method does not recover all possible binding properties of TFs, the enrichment and centrality patterns of MBS indicated that DAP-seq reflected the MYC-binding models.

Oligomerization of binding sites usually results in a higher affinity of the TF for DNA and, subsequently, in increased expression levels. Therefore, we evaluated the effect of multimeric DNA sites in protein binding, and we observed a direct relationship between the number of MBSs and the fold enrichment of the peaks, particularly in up to three repeats per 51-bp-long central fragment (Figure 2B and Supplemental Figure 8). Focusing on doublets of DNA sites, we examined whether nucleotide spacing between single motifs had any effect on protein recognition. We extracted all the possible doublets separated between 0 and 9 bp in the Arabidopsis, tomato, and Marchantia genomes, and we quantified binding by counting the number of reads that covered each doublet configuration. Preferential binding was observed for boxes spaced at three nucleotides, independent of the kind of box (G, PBE, or T/G), and similar patterns were obtained for the Arabidopsis and Marchantia MYGs tested (Figure 2C and Supplemental Figure 9). SIMYCs showed similar binding patterns to the T/G-box and PBE doublets, although this was less evident in the case of the G-box, where 7- and/or 9-bp spacings were prevalent (Figure 2C, Supplemental Figure 9).

We evaluated the correspondence between our in vitro findings and the results of recently published Arabidopsis MYC2 and tomato SIMYCY2 ChiP-seq datasets (Du et al., 2017; Wang et al., 2019), and we observed similar binding patterns among all the experiments (Figure 2C, Supplemental Figures 8 and 9). For

Table 1. Moreover, the overlap of significant peaks between replicates ranged from 90.5% to 97% (Supplemental Figure 6B). Significant peaks differed among the four proteins analyzed, indicating that the binding activity depends on the intrinsic properties of the proteins (Supplemental Figure 6B and Supplemental Table 1). In the case of tomato and Marchantia MYC2-related, we performed only one experiment, and the numbers of significant peaks are shown in

4 Plant Communications 2, 100232, November 8 2021 © 2021 The Author(s).
DNA-binding specificity of plant bHLH transcription factors

**Results in vitro correlate with MYC2-related functionality in vivo**

As a first analysis of the biological significance of DNA-binding profiles *in vitro*, we scored the distribution of bound peaks in different gene features. We observed that 51.0%–56.9% of *Arabidopsis* peaks corresponded to promoter regions up to 3 kb upstream of the TSS, whereas these proportions in the complete MBS set or in a random set of fragments were 30.6% and 29.1%, respectively (Supplemental Figure 10A). The distributions were markedly enriched within 1-kb promoters and downstream of the transcriptional termination site (TTS) (Supplemental Figure 10A). Peaks downstream of the transcriptional termination site were also enriched relative to the control distributions (Supplemental Figure 10A).
This contrasts with our observations in bHLH-PBMs, suggesting that DAP-seq reflects the binding specificities of MYCs more accurately than bHLH-PBM. Similar tendencies were observed for SIMYCs and MpMYCx (Supplemental Figure 10A), although the enrichment of peaks in upstream and downstream regions was lower than in Arabidopsis. Remarkably, tomato and Marchantia peaks were particularly enriched in distal promoters and were almost as high as in proximal promoters in Marchantia (Supplemental Figure 10A), suggesting that regulatory regions may be larger in these species than in Arabidopsis, in line with the greater length of intergenic regions in these species. The distribution of peaks in gene features observed in the Arabidopsis and tomato ChIP-seq experiments showed the same tendency but were slightly higher in promoters: 68.2% of the peaks in Arabidopsis and 44% of the peaks in tomato corresponded to upstream regions (Supplemental Figure 10A). Taken together, these results indicate that plant MYC2 and related factors bind very efficiently to regulatory regions, as the proportions of bound fragments in Arabidopsis DAP-seq experiments that corresponded to promoters were 1.67- to 1.86-fold higher than the proportions of the complete MBS set. Similarly, enrichments in promoter fragments for SIMYCs were 1.71- to 2.42-fold higher than the proportion of the complete MBS set, and this value was 1.53-fold for Marchantia MpMYCx. Furthermore, we observed large overlaps in the target gene lists derived from the DAP- and ChIP-seq experiments (Supplemental Figure 10B), indicating parallelism between the different approaches. Therefore, in vitro experiments performed with DAP-seq approaches are good indicators of plant MYC2 and related activity in planta.

To place these results into a biological context, we analyzed the binding of MYCs to genes regulated by JA in the three species and to direct targets of MYC2 in Arabidopsis. We analyzed the binding profiles of MYC2-related proteins in gene clusters regulated by JA (Hickman et al., 2017) and observed the strongest DNA-binding signal in promoters, very close to the TSS, in clusters of genes highly upregulated in response to JA, particularly between 1 and 3 h after treatment, identical to that observed in ChIP-MYC2 (Supplemental Figure 11). Moreover, 1094 and 918 genes out of 1705 (64.2% and 53.8%) in upregulated gene clusters were found to be direct targets of MYC2 and MYC3, respectively (Zander et al., 2020), proportions much higher than those expected by chance (18.6% and 12.6%, respectively; p = 0; chi-square test). By contrast, downregulated genes showed low binding intensities in vitro, slight over-representation of MYC2 targets, and the same proportion of MYC3 targets as that expected by chance (26.3% and 12.6% for MYC2 and MYC3, respectively; p (MYC2) = 2.5e-17; p (MYC3) = 0.688; chi-square test) (Supplemental Figure 11). Graphical representation of MYC binding to some well-known targets revealed overlap in highly enriched peaks from both in vitro and in vivo experiments, and peak summits matched with at least one MBS (Figure 3A). Similar results were obtained in tomato and Marchantia experiments, where the highest signals were observed close to the TSS and just downstream of the TTS (Supplemental Figure 12 and 13). This pattern was particular to genes induced by JA and/or wounding and downregulated in the tomato RNAi-SIMYC2 line (Du et al., 2017) and in the loss-of-function Mpmycx and Mpmycx mutants in Marchantia (Peñuelas et al., 2019) (Supplemental Figures 12 and 13). As in Arabidopsis, SIMYC2, SIMYC2-related, and MpMYCx bound to regulatory regions of some presumed MYC targets, and peak summits matched at least one MBS (Figure 3B and 3C). It is worth noting that not all the MBSs matched with a highly enriched peak, suggesting that the presence of the MBS itself is not sufficient for recognition by MYC2-related proteins. Together, these results indicate that in vitro DNA binding by DAP-seq perfectly reproduces binding in vivo and suggest that recognition of biologically relevant motifs requires additional DNA features beyond a mere MBS.

Identification of highly specific MYC-binding sequences

We were interested in identifying particular sequence features in DNA motifs recognized by MYC2-related TFs, and we therefore performed de novo discovery of motifs with MEME. We observed an enrichment of A/T nucleotides centered at positions ±5 relative to the MBS, including Arabidopsis and tomato ChIP-seq data, suggesting that these additional sequence modules are important for MYC2-related recognition and that they may be implicated in MYC-related functions (Figure 3D; Supplemental Table 2). Scanning the central sequences from DAP- and ChIP-seq experiments revealed large proportions of peaks without a canonical MBS (G-, T/G-box, or PBE) (Supplemental Figure 14). We then performed an additional de novo discovery of motifs and obtained a sequence resembling a non-perfect MBS flanked by identical A/T-rich motifs centered at ±5, supporting the notion that these regions may be important for MYC2-related recognition, even in the absence of a perfect MBS (Figure 3E).

We explored the possibility of exploiting DNA-binding data in vitro to increase the predictive capacity for target genes. Accordingly, in addition to position weight matrices (PWMs) obtained with MEME, we generated additional binding models for a more accurate definition of TFBSs. We used TF flexible models (TFFMs), a method based on hidden Markov models, to capture interdependencies of successive nucleotides (Mathelier et al., 2016). We did not observe particular nucleotide interdependencies beyond the G-box (Supplemental Figure 15), suggesting that the MEME-derived PWMs already reflected the most informative compositions. We then used a method that incorporates information on the tridimensional structure of DNA, also referred to as DNA shape, to predict TFBSs. We followed previously described algorithms (Mathelier et al., 2016) to create a pipeline in which DNA sequences in potential TFBSs, predicted either with PWM or TFFM, and their corresponding background sequences are enriched with DNA shape information and used to train a supervised machine learning (ML) algorithm. The ML classifier is then applied to the testing sequences and obtains the probability of the sequence being a foreground sequence. Foreground and background probabilities are used to compute receiver operating characteristic (ROC) area under the curve (AUC) scores to assess the binding model, and these scores can be used for a stricter prediction of TF targets and comparison between models (Figure 3F). Consistent with previous observations, the AUC values for PWM and TFFM MYC2 models were almost identical, but DNA shape features increased the predictive power (Figure 3G), suggesting a role of
DNA-binding specificity of plant bHLH transcription factors

Figure 3. Highly specific MYC DNA binding in vitro.

(A) Screen captures of Integrated Genome Viewer visualizations of DAP- and ChIP-seq binding profiles of Arabidopsis MYCs in some predicted targets. Arrowheads indicate the positions of different MBSs: red, G-box; green, PBE; yellow, T/G-box. Scale bars represent 1 kb.

(B and C) Same as in (A) but data are from tomato and (C) from Marchantia.

(D) Sequence logos for Arabidopsis, tomato, and Marchantia MYCs obtained from MEME analysis. A/T modules centered at position ±5 relative to the MBS are boxed in red.

(E) Same as in (D) but peak sequences did not include a canonical MBS.

(F) Diagram of the ML pipeline to generate new binding models for better target prediction.

(G) ROC and AUC values of the four binding models generated with the Arabidopsis, tomato, and Marchantia MYC2 proteins. A theoretical diagonal with an AUC value of 0.5 would indicate that the models have no predictive potential (true-positive rate = true-negative rate), whereas AUC = 1 would represent a perfect predictive model.

(H) Proportion of Arabidopsis MYC2 targets identified by ChIP-seq and predicted from the PWM and PWM + DNA shape models with binding score ≥0.98.

(I) Same as in (H) but with tomato SlMYC2 targets.

(J) Expression of the genes corresponding to binding score ≥0.98 identified with any of the four models (PWM, TFFM, PWM + shape, TFF + shape). Gene expression data are from the wild-type (Tak1) and the mpmycx (mycx) mutant in response to OPDA and wounding (W).

DNA structure during target selection. It is worth noting that PWM/TFFM models for MYC4, and particularly bHLH17, captured the vast majority of the binding information (AUC values 0.96 and 0.99, respectively; Supplemental Figure 16). The apparent discrepancy between these proteins and MYC2/3 may be due to the different number of significant peaks obtained from each experiment and the fact that the training set from MYC4/bHLH17 recovered the most relevant features for DNA binding, rather than to a different role of DNA shape in the TFBS definition. Very similar patterns of DNA-shape-augmented models were observed in the Solanum and Marchantia orthologs (Figure 3G and Supplemental Figure 16), indicating that the involvement of DNA shape is conserved in plants.
In order to associate the improved DNA-binding models to the function of MYC2, we scored the number of MYC2 target genes among the lists of predicted targets in the PWM and PWM + shape studies. We observed an increase in MYC2 target numbers when the DNA-shape variable was included, indicating that this model actually increases the predictive capacity for target genes (Figure 3H). By contrast, the proportions of MYC2 target genes were very similar in PWM and PWM + shape predicted targets, coincident with a smaller difference in their rates of true-positives (Figure 3G and 3I). A direct relationship between enhanced models and function was also observed in MpMYCx. In this case, TFFM and shape-enhanced models increased the predictive potential for MpMYCx targets, revealed by their higher expression in response to wounding or OPDA, and their more drastic decrease in the Mpmycx loss-of-function mutant. Interestingly, the increased potential for prediction of target genes in TFFM and shape-enhanced models perfectly correlated with their corresponding AUC values in model assessment (Figure 3G and 3J).

**Involvement of MBS-flanking sequences in TF recognition**

Given that the DNA-binding domain of MYC2 and related TFs does not interact with nucleotides outside the MBS (Zhang et al., 2015; Lian et al., 2017) and that DNA shape is relevant to target recognition, it is plausible that distantly flanking nucleotides may help to confer a particular conformation to DNA for better recognition by plant MYCs. To investigate this possibility, we classified genomic fragments centered at the MBS as bound or unbound if they were or were not included among the peaks enriched in DNA-binding experiments. Sequence alignment revealed a preference for A/T stretches at flanking positions, and this was more evident when increasing the affinity, represented in the first decile (D1) of peaks (Figure 4A). We obtained calculations for three DNA shape features involving intra-base-pair interactions (minor groove width [MGW], electrostatic potential of the minor groove [EP], and propeller twist [ProT]) and two inter-base pair features (helix twist [HelT] and Roll). We observed clear differences in intra-base-pair features between bound and unbound motifs that reached their maxima at the A/T modules centered at ±5. Moreover, shape differences correlated with the binding affinity of the proteins, as they were higher in D1 sequences (Figure 4B; Supplemental Figure 17 and 18). All the Arabidopsis MYC2-related proteins showed very similar preferences, suggesting that DNA shape requirements are intrinsic to the DNA-binding activity of MYCs. Consistently, binding of MYC2 and SIMYC2 derived from ChiP-seq experiments reflected similar DNA shape requirements (Supplemental Figure 17 and 18), pointing to a prominent role of A/T modules during target recognition in vivo. Finally, alignment of sequences centered at a non-perfect MBS also reflected a clear preference for flanking A/T stretches that correlated with binding affinity and with the highest deviations in DNA shape values (Figure 4C and 4D).

To evaluate the biological relevance of flanking A/T modules, we performed DNA-binding experiments of MYC3 to different genomic fragments that differed at their MBS-flanking positions. Binding experiments consisted of focused DAP-seq experiments with a pool of amplicons containing any of the MBSs and their corresponding point mutations, in which A/T nucleotides at positions ±4, 5, or 6 were replaced by G/C (Figure 4E). We constructed a DNA library with the mix of amplicons and performed several DAP experiments in parallel. Sequencing of the samples revealed higher binding to the G-box than to the PBE and T/G in all experiments (Supplemental Table 3). More interestingly, MYC3 bound very poorly to point mutations that affected the ±4, 5, and 6 flanking positions. This was particularly true for the two amplicons that involved the G-box and the one that contained the T/G-box, as point mutations reduced binding 19.5-, 11.2-, and 8.1-fold, respectively (Figure 4E; Supplemental Table 3). In the case of the PBE-containing fragment, binding of the protein was not very efficient, although recognition of the mutant fragment was reduced 1.8-fold. Together with the samples above, we included a collection of amplicons that contained G/C nucleotides at positions ±4, 5, or 6, as well as their corresponding mutant versions in which these nucleotides were replaced by A/T (Figure 4F). Binding to the mutant versions in the context of the two G-boxes and the T/G-element increased 11.3-, 6.5-, and 3.3-fold, respectively, in relation to their corresponding wild-type versions. Also, in this case, binding to the PBE fragment was not optimal, but recovery of the mutant fragment increased 1.7-fold (Figure 4F; Supplemental Table 3). We evaluated whether mutations in A/T modules affected intra-base-pair shape features. Point mutations correlated with drastic changes in these parameters, and these tended to be opposite in loss-of-affinity and gain-of-affinity mutations (Supplemental Figure 19).

Taken together, these data indicate that a particular nucleotide composition of distantly flanking modules, which probably confers a particular shape to DNA, is necessary for the correct binding of MYC3, and probably for MYC2-related proteins, and that this characteristic is likely to be essential for MYC function in vivo.

**Biological implications of increasing binding specificity to DNA**

The existence of additional factors beyond the MBS suggests an increased specificity in target recognition and a decreased risk of cross-interaction among transcriptional cascades. If this were true, we should be able to predict MYC2-related targets from DAP-seq experiments. We clustered Arabidopsis genes according to their MYC2/3/4 and bHLH17 DAP and MYC2-ChIP intensities in promoters, and we searched for enriched gene ontology (GO) terms in the four clusters with the highest signal 1 kb upstream of the TSS (Figure 5A). The most enriched GO terms in the clusters with the highest binding close to the TSS were related to JA signaling and defense (Figure 5A). This correlation is conserved, at least in higher plants, as identical patterns and GO enrichment were also observed in tomato DAP- and ChiP-seq experiments (Figure 5B). Therefore, enhanced binding specificity provided by additional flanking sequences increases target selectivity, to the point that we were able to predict MYC targets from in vitro experiments.

**Variability in target recognition specificity within the bHLH family**

We analyzed whether other unrelated bHLHs share similar constraints to confer specificity or whether each bHLH subgroup employs a different strategy for target recognition. We selected
bHLHs from several subfamilies and analyzed their binding patterns by DAP-seq; we included in this analysis some bHLHs that have been characterized previously by DAP-seq (O’Malley et al., 2016) (Supplemental Figure 5). We analyzed 15 bHLH proteins, obtaining a variable number of significant peaks, and we discarded proteins bHLH93, AKS2, GL3, MUTE, PYE, and UNE12, which had a low number of significant peaks and lacked an enriched G-box (Supplemental Table 1). We also discarded bHLH74, bHLH77, bHLH104, BIM3, and PIF7 because of their poor yields, despite their having been shown to bind the G-box (O’Malley et al., 2016).

We searched for MBSs in peaks and observed the greatest enrichments for the G-box, ranging from 200- to >1000-fold,
whereas the relative affinities for PBE and T/G-boxes were variable (Supplemental Figure 20). For example, PIF4 and PIF5 showed slight differences in PBE binding, but they did not recognize the T/G-element, similar to previous reports (Hornitschek et al., 2012). Interestingly, bHLH157 peaks were not enriched in the G-box and PBE but were enriched in the T/G-box (Supplemental Figure 20). Taken together, these results suggest that different bHLHs have distinct DNA-binding properties to the core MBS.

We also analyzed the synergistic effect of MBS doublets and observed differences among bHLHs (Supplemental Figure 21A). For example, PIF4 and PIF5 showed slight differences in PBE binding, but they did not recognize the T/G-element, similar to previous reports (Hornitschek et al., 2012). Interestingly, bHLH157 peaks were not enriched in the G-box and PBE but were enriched in the T/G-box (Supplemental Figure 20). Taken together, these results suggest that different bHLHs have distinct DNA-binding properties to the core MBS.

We assessed new binding models to define more accurately the binding properties of the bHLHs. TFFMs revealed little or no nucleotide interdependencies beyond the MBS and the adjacent downstream position (Supplemental Figure 24). We incorporated DNA shape information and applied the ML classifier to generate and evaluate new binding models. As before, TFFM and PWM scores were almost identical in all the experiments, but we observed some variation in DNA shape models, ranging from no (PIFs and bHLH31) or very little contribution (BIM2) of DNA shape.
shape to a moderate effect in bHLH66 and particularly in bHLH18 (Figure 6B). We also observed little or no difference in intra-base-pair shape features between bound and unbound fragments in flanking positions, except for bHLH18 (Figure 6C and Supplemental Figure 25). In the latter, the differences were not restricted to specific positions as in MYC2-related proteins, but they were homogeneous along the flanks. Interestingly, shape features correlated with [A + T] content, as this parameter did not differ or barely differed between bound and unbound fragments in PIFs, bHLH31, bHLH66, and BIM2, whereas it was markedly lower and homogeneously distributed in bHLH18-bound fragments, in contrast to MYCs (Figure 6D; Supplemental Figure 25).

Predictive capacity of bHLH functionality

We first scored the distribution of significant peaks across different gene features. Most of the proteins bound preferentially to upstream regions of the genes, in particular to proximal promoters (<1 kb). In addition, bHLH92, -99, and -115 also recognized more motifs downstream of the TTS than the distribution of MBSs in this region (Supplemental Figure 26). An exception was observed for bHLH18, whose bound fragments were located preferentially in gene bodies, probably because of their preference for a G + C-rich context. Taken together, these data support the idea that DNA-binding data obtained from in vitro experiments may reflect the actual functionality of bHLHs in vivo.

To assess the capacity for in vitro experiments to predict TF function, we clustered the genes in relation to protein-binding intensity in promoters and searched for significant GO terms related to the function of the proteins. In the case of PIFs, the strongest binding signal near the TSS correlated with an increased number of PIF targets identified by ChIP-seq and lower gene expression in the quadruple pifq mutant (Pfeiffer et al., 2014) (Supplemental Figure 27A and 27B). Moreover, these clusters were enriched in GO terms related to light signaling, photosynthesis, and, to a lesser extent, response to far-red light (Supplemental Figure 27C and 27D). Gene clustering of the bHLH34-binding signal showed the relevance of proximal promoters, only around the TSS, in the expression of genes involved in iron deficiency and transport (Supplemental Figure 27E and 27F). Clustering of bHLH99-binding data revealed its involvement in defense-related functions and JA- and salicylic acid-mediated pathways (Supplemental Figure 27G). In a second group of proteins, we obtained GO terms related to the proteins’ functions, but these were not among the most enriched terms. This was the case with bHLH92, which is involved in abiotic stress signaling (Jiang et al., 2009) (Supplemental Figure 27H), and bHLH18 and
bHLH115, which participate in iron deficiency-related pathways (Gao et al., 2019) (Supplemental Figure 27I and 27J). A third group included bHLH66, -31, and BIM2; we could not detect significant GO terms related directly to their functions, but instead, we observed a correlation between binding patterns and gene expression related to their function. For example, bHLH66 is implicated in the differentiation of root hair cells (Bruex et al., 2012), and there was particular enrichment of core hair cell genes in clusters corresponding to binding close to the TSS (Supplemental Figure 28A). In addition, we observed an overlap of bHLH66-binding peaks with transcriptionally active regions in root hair cells identified by assays for transposase-accessible chromatin with sequencing (ATAC-seq) (Maher et al., 2018) (Supplemental Figure 28B). bHLH31, also called BIG PETAL, is expressed in petals (Varaud et al., 2011), and there was a consistently higher proportion of petal-expressed genes, with the highest binding around 0.5 kb upstream of the TSS (Supplemental Figure 28C). In the case of BIM2, known to participate in brassinosteroid signaling together with BES1, we observed an increase in BIM2 binding in the proximal promoters close to the TSS in a group of genes downregulated in the bri1-301 mutant, which is affected in the brassinosteroid receptor BRASSINOESTEROID-INSENSITIVE1 (Supplemental Figure 28D).

In summary, we found an enrichment of bound peaks in regulatory regions for 13 out of 14 bHLHs and observed direct relationships between the functions of the proteins and their DNA-binding patterns. To find particular patterns associated with binding affinity, we clustered the proteins in relation to their relative binding to all the G-boxes in the Arabidopsis genome (Figure 7). We observed that binding to all possible G-boxes was not uniform among the proteins, revealing particular specificities in the DNA-protein interplay in vitro that do not depend on external cofactors. On the other hand, highly homologous bHLHs from the same subgroup share very similar binding patterns, as is the case with MYC2-related or PIFs, in agreement with the fact that they share some common targets in vivo, and their binding preferences are already detected in vitro (Figure 7). Interestingly, G-boxes in cluster 8 were bound by all the proteins, but there were some clusters that could be considered TF or group specific, such as cluster 4 for BIM2, cluster 3 for MYCs, and, to some extent, cluster 2 for PIFs (Figure 7).

**DISCUSSION**

A major unanswered question in transcriptional regulation is how TFs select their targets without affecting the expression of other genes that may contain the same binding sequence. Here we demonstrate that MYC2 and related TFs, a group of proteins that redundantly regulate the JA-signaling pathway, specifically recognize their targets thanks to an extension at both sides of their cognate element, the MBS. These extensions, formed by short A/T-rich modules, are necessary for correct binding of the TF and may cooperate with imperfect G-boxes in the recognition of targets. Furthermore, DAP-seq experiments revealed the existence of specific features or fingerprints for bHLHs that allow for the prediction of gene targets from in vitro binding data. Our results indicate that, at least for a group of bHLHs, in vitro determination of binding specificities provides information close to that of ChIP-seq experiments in a faster way, thereby overcoming the drawbacks of this technique.

During the last decade, a great effort has been made to understand the molecular basis of protein-DNA recognition using in vitro approaches, some of which have involved plant bHLHs (reviewed in Franco-Zorrilla and Solano, 2017). These studies used uPBMs, which do not completely cover all the possible sequence features that influence protein recognition. One step beyond is provided by custom TF family-specific PBMs. For example, custom PBMs for Transcription Activator-Like Effector (TALE) proteins increased the prediction of TALE-binding sites while decreasing potential off-target sites (Rogers et al., 2015). More similar to our approach, Gordan et al. (2013) designed a custom PBM for two yeast bHLHs that contained ChIP-seq-derived bound and unbound sequences in their genomic contexts. The two bHLH-PBM designs used in this work were able to reflect, at least in part, the biological functions of MYC2-related factors, and they pointed to the existence of additional nucleotides important for protein recognition (Figure 1). DAP-seq experiments have been decisive for defining the additional DNA features necessary for binding of MYC2-related bHLHs to DNA, based on several pieces of evidence. First, we obtained almost identical patterns for three redundant activators...
involved in the JA-signaling pathway and for one repressor that recognizes the same binding sites (Fernández-Calvo et al., 2011; Fonseca et al., 2014). Second, we observed conservation of identical sequence features across distant lineages, thanks to the analysis of three tomato MYC2 orthologs and MpMYC2x from Marchantia, a non-vascular plant that conserves the same JA-signaling pathway components as higher plants, including MYC2-related functions (Monte et al., 2018, 2019; Peñuelas et al., 2019) (Figure 2, Supplemental Figures 7–13). Third, we took advantage of available data on MYC2 binding in vivo in Arabidopsis and tomato, and we observed identical binding patterns, even the existence of non-perfect G-boxes surrounded by A/T-rich stretches (Figure 3). Fourth, we demonstrated experimentally that the A/T modules are necessary for correct MYC3 recognition, providing a basis for artificially fine tuning TF-DNA interactions by simply modifying the surrounding context of the core MBS (Figure 4). Finally, DAP-seq analysis of representative bHLHs indicated that MYC2 and related associated features are unique, raising the possibility of finding novel features in other bHLHs (Figure 7). Some other works have used DAP-seq data to define in more detail the binding features of DNA, including differences in bHLH/bZIP recognition of the G-box (Ezer et al., 2017) or sequence-spacing preferences on auxin response factors (ARFs) (O’Malley et al., 2016; Galli et al., 2018; Stigliani et al., 2019). However, these studies focused on studying core motifs or, in the case of ARFs, the spacing between basic modules, known to be decisive for ARF recognition (Boer et al., 2014).

The role of flanking A/T stretches remains a central question in the recognition of targets by MYC2 and related TFs. These motifs are not expected to directly interact with the protein (Zhang et al., 2015; Lian et al., 2017), but they could confer a particular shape to DNA that would increase MYC2 binding specificity and/or stability. In fact, A/T modules overlap with the largest differences in intra-base-pair shape features in all the MYC2-related proteins tested, and they can be reversed by changing the nucleotide composition (Figure 4). Moreover, local disruptions in DNA shape have not been observed in the set of bHLHs analyzed in this study, suggesting that A/T stretches are specific to plant MYCs. A previous report (Figueroa and Browse, 2012) pointed to a TTTT stretch immediately 3′ to the G-box as responsible for providing JA specificity, but we demonstrate that this requirement is more distal to the G-box. This may not be exclusive to plant MYC2-related, as the presence of three- to five-A/T base-pair flanking core sequences recognized by human TFs is relatively common (Jolma et al., 2013), and (dA:dT) tracts immediately flanking the TFBS have a role in the DNA binding of yeast Gcn4 and Gal4 TFs (Levo et al., 2015).

Recent research has highlighted the involvement of DNA shape in modulating TFBS recognition, and these features have been shown to improve the prediction of binding sites (Gordan et al., 2013; Abe et al., 2015; Dror et al., 2015; Mathelier et al., 2016; Lai et al., 2019). It was proposed that DNA-protein recognition results from an interplay between base readout (direct interaction) and shape readout (indirect interaction) (Slattery et al., 2014), and the relative contributions of both mechanisms vary across protein families and, as we show here, within one particular family. A/T stretches flanking the core sequence have been associated with a narrowing of the minor groove, facilitating the interaction with an adjacent major groove (Rohs et al., 2009; Jolma et al., 2013; Levo et al., 2015). However, this may not be the case with MYC2 and related proteins, given that these TFs directly interact with the major groove of DNA (Zhang et al., 2015; Lian et al., 2017) and that the A/T stretches do not immediately flank the MBS but are centered at position ±5. A plausible alternative is that A/T stretches may favor DNA bending, contributing to the generation of a preferred DNA structure and thereby facilitating interaction with MYC2s.

From a practical point of view, the existence of additional features beyond a core sequence increases the accuracy of target prediction. In addition to a direct comparison of in vitro (bHLH-PBMs and DAP-seq) and in vivo (transcriptomic or ChIP-seq) data, we followed a top-down strategy based on gene clustering in relation to their TF-binding signal and the analysis of GO terms associated with each cluster. This approach successfully predicted targets of Arabidopsis and tomato MYCs but notMpMYC2x. The lack of prediction in Marchantia may be explained by the very recent and incomplete annotation of its genome (Montgomery et al., 2020) and/or by a different distribution of target sites along the promoters, as indicated by an increase in significant DAP peaks in the 1- to 3-kb promoter interval. In fact, Marchantia promoters and intergenic regions in general are surprisingly large compared with those in eudicots and may influence the distribution of binding sites.

The top-down approach also predicted putative targets of bHLH34, one of the upstream regulatory proteins involved in iron deficiency (Gao et al., 2019) and of PIF5/4, two well-known bHLHs involved in growth under dark conditions (Horvitschek et al., 2012). Target prediction of bHLH99 revealed its participation in biotic stress responses, as expected given its involvement in limiting the spread of necrotrophic fungal pathogens (Dobón et al., 2015). In three other proteins, bHLH66, bHLH31, and BIM2, we could not detect significant GO terms associated with particular targets; however, we observed a correlation between a higher binding to proximal promoters and transcriptomic data associated with the proteins (Yin et al., 2005; Varaud et al., 2011; Bruex et al., 2012). These results, together with a higher-than-expected distribution of DAP peaks in regulatory regions, suggest the existence of protein-specific characteristics for the correct selection of target genes. DNA-shape-enriched binding models and distribution of shape features along the core flanking sequences indicated little or no effect of DNA shape on protein recognition (Figure 6, Supplemental Figure 24 and 25), with the exception of bHLH18. In this case, however, preferentially bound sequences occur in a G/C-rich context and provoke opposite DNA-shape configurations to MYC2-related factors. Interestingly, in this experiment, we detected one of the highest frequencies of non-canonical elements in bound peaks (Supplemental Figure 22). Although we cannot explain how the shape of DNA influences bHLH18 binding, we speculate that it may facilitate protein anchorage to DNA and allow interaction with the non-perfect G-box in bound fragments.

An additional mode of bHLH target gene selection may be the identification of secondary motifs with slightly lower affinities than those of primary motifs. This mechanism may be operating...
Plant Communications

in bHLH38, bHLH34, and BIM2 (Supplemental Figure 23). Secondary motifs have been proposed previously, although their real relevance has been controversial (Badis et al., 2009; Zhao and Stormo, 2011; Franco-Zorrilla et al., 2014), probably owing to the technology used to identify the motifs. Nevertheless, it would be interesting to analyze the role of secondary, non-canonical sequences in bHLH binding in future work.

In this work, we demonstrate the involvement of A/T-rich flanking modules in MYC binding and propose a role of DNA shape in bHLH18 recognition. In addition to the limited effect of particular configurations of motif doublets, we have not identified additional features that contribute to protein-DNA interactions. However, differential binding to G-boxes and clustering classification revealed that individual proteins or groups of TFs each have a particular code that determines binding specificity, and this method makes it possible to identify characteristic fingerprints for each protein (Figure 7). Moreover, paralog proteins that share most of the gene targets in vivo have very similar fingerprints, as is the case with MYC2 and related TFs, or even the two PIFs studied (Figure 7). The cooperative activity of TFs during target selection has been widely demonstrated. Interestingly, it has been described that side-by-side tetramerization of MYC2 increases binding affinity to distant G-boxes spaced >100 bp by facilitating DNA looping (Lian et al., 2017). Furthermore, chromatin accessibility, defined by nucleosome number and position, is a determining factor for TF target recognition. Our results indicate that plant MYC-binding patterns to targets in vitro and in vivo are almost identical, and they may depend on particular fingerprints more than on the presence of external cofactors or the accessibility of chromatin.

This mechanism provides novel insights into the study of transcriptional regulation from several points of view. On the one hand, as has been mentioned throughout the text, it allows for more specific and rapid recognition of gene targets, a feature that may be of particular relevance to MYC2 and related TFs for their roles in early defense responses and in maintaining a defense/growth balance (Ortígosa et al., 2020). On the other hand, stricter requirements for target recognition may also be used by pioneer TFs to bind to DNA assembled into nucleosomes and unlock the chromatin (Pajoro et al., 2014; Sayou et al., 2016; Jin et al., 2021). The study of the intrinsic properties of bHLH TFs that allow target recognition, even in in vitro approaches without co-adjuvant proteins, will be of interest in the future. Furthermore, our observations emphasize the suitability of in vitro techniques for faster and easier prediction of gene targets compared with current in vivo methodologies.

METHODS

Detailed methods are available in the supplemental information.

bHLH-PBMs

Details on the probes included in promoter and genome bHLH-PBM designs are in the supplemental information. Synthesis of double-stranded DNA, protein binding, detection of protein-DNA complexes, and scanning of PBMs were performed as for the uPBMs (Franco-Zorrilla et al., 2014). Probe intensities referred to the mean intensity of the probes containing a mutant G-box. Oligonucleotide deciles for each experiment were defined by sorting the probes in relation to their signal intensity (in Log2 [G-box/Gmutant]) from the highest to the lowest. Then, the top 10% of the probes with the highest signal was defined as the first decile (D1), whereas the bottom 10% corresponded to the tenth decile (D10).

bHLH constructs and protein expression

cDNAs for bHLHs were obtained from the TRANSPLANTA collection (Coego et al., 2014) and mobilized to pDEST-H1 to obtain fusions to maltose-binding protein (MBP). Constructs for MYC2, 3, 4, bHLH17, PIF4, and PIF5 in pDEST-H1 were described previously (Fernández-Calvo et al., 2011; Godoy et al., 2011; Hornitschek et al., 2012; Fonseca et al., 2014). Tomato MYC genes were amplified from gDNA (oligonucleotide sequences in Supplemental Table 4) and cloned in pDEST-H1. MBP-fused Marchantia MpMYCx was obtained previously (Peñuelas et al., 2019). Recombinant proteins were expressed in Escherichia coli as described previously (Godoy et al., 2011; Franco-Zorrilla et al., 2014); cells were pelleted and kept frozen at −80°C until use.

gDNA isolation and DAP-seq

gDNA was isolated from whole plants using the CTAB method (Doyle and Doyle, 1987) and quantified with a Qubit 4 fluorometer. DAP-seq experiments were carried out as described previously (Bartlett et al., 2017), but cleared soluble protein extracts obtained from induced E. coli cultures were incubated with gDNA libraries. Input samples were obtained by direct amplification of gDNA libraries.

Focused DAP-seq

Amplicons were obtained by PCR amplification using gDNA as a template. Site-directed mutagenesis was performed by sequential PCR using complementary oligonucleotides (sequences in Supplemental Table 4) containing the mutations. Amplicons were pooled to construct a TruSeq DNA library, which was incubated with MBP-MYC3 cleared protein extracts.

Sequencing analysis

 Reads were mapped with Bowtie2 (Langmead and Salzberg, 2012) using default parameters to the latest genome versions of Arabidopsis, tomato, and Marchantia. BAM files were used for peak calling with GEM (Guo et al., 2012), using input samples as controls, and extracted 51-bp sequences were used for de novo discovery with MEME (Bailey et al., 2009).

ChiP-seq and RNA-seq analysis

Raw reads from Arabidopsis and tomato MYC2 ChiP-seq were downloaded from the Genome Sequence Archive of the Beijing Institute of Genomics Data Center (GSA BIG) and reanalyzed as above to obtain datasets comparable with those from the DAP-seq experiments.

Arabidopsis RNA sequencing (RNA-seq) data were obtained from Hickman et al. (2017), and tomato RNA-seq raw reads were downloaded from GSA BIG, filtered for low-quality reads, and mapped to the reference genome with HISAT2 (Kim et al., 2015). Differentially expressed genes were identified with DESeq2 (Love et al., 2014).

Additional analysis

Heatmaps and K-means clustering of DAP-seq intensity data were obtained from BIGWIG files with plotHeatmap in deepTools (Ramírez et al., 2016). Distribution of DAP peaks across gene features was obtained with the R package ChIPseeker (Yu et al., 2015). DNA shape features were calculated with the R package DNAshaper (Chiu et al., 2016). Read coverage of single and doubled MBS was extracted from BAM files with bedtools coverage (Quinlan, 2014), and normalized ratios against inputs were calculated with custom bash scripts. Binding models and scores for PWMs, TFFMs, and DNA-shape-enriched models were generated as in Mathelier et al. (2016).
DNA-binding specificity of plant bHLH transcription factors

ACCESSION NUMBERS
Raw intensity files for bHLH-PBMs and raw reads for DAP-seq experiments are available in GEO (https://www.ncbi.nlm.nih.gov/geo/) under the Accession Series GSE155321 and Subseries GSE155317, GSE155318, and GSE155320.

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at Plant Communications Online.

FUNDING
Research in J.M.F.-Z.’s laboratory was funded by the Spanish Ministry of Science and Innovation grant BIO2017-86651-P (MINECO/FEDER). Research in R.S.’s laboratory was funded by the Spanish Ministry of Science and Innovation grant BIO2016-77216-R (MINECO/FEDER).

AUTHOR CONTRIBUTIONS
J.M.F.-Z. and R.S. designed the research. I.L.-V., M.G., J.G., R.S., and J.M.F.-Z. analyzed and interpreted data. J.M.F.-Z. wrote the manuscript with input from the authors.

ACKNOWLEDGMENTS
We thank Javier Paz-Ares for critical reading of the manuscript and José A. Jarillo for providing bHLH clones used in this work. We also thank Christian Fankhauser for providing the PIF4- and PIF5-MBP constructs. The authors declare no competing interests.

Received: April 12, 2021
Revised: July 9, 2021
Accepted: August 10, 2021
Published: August 12, 2021

REFERENCES
Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, H.J., Rohs, R., and Mann, R.S. (2015). Deconvolving the recognition of DNA shape from sequence. Cell 161:307–318.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. Science 324:1720–1723.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. 37:W202–W208.

Bartlett, A., O’Malley, R.C., Huang, S.C., Gallavotti, A., and Ecker, J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat. Protoc. 12:1659–1672.

Boer, D.R., Freire-Rios, A., van den Berg, W.A.M., Saaki, T., Manfield, I.W., Kepinski, S., López-Vidriero, I., Franco-Zorrilla, J.M., De Vries, S.C., Solano, R., et al. (2014). Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. Cell 156:577–589.

Brkljacic, J., and Grotewold, E. (2017). Combinatorial control of plant gene expression. Biochim. Biophys. Acta 1860:31–40.

Bruex, A., Keinkanyam, R.M., Wieckowski, Y., Kang, Y.H., Bernhardt, C., Xia, Y., Zheng, X., Wang, J.Y., Lee, M.M., Benfey, P., et al. (2012). A gene regulatory network for root epidermis cell differentiation in Arabidopsis. Plant Cell Physiol. 53:330–343.

Chini, A., Fonseca, S., Fernandez, G., Adie, B., Chico, J.M., Lorenzo, O., Garcia-Casado, G., Lopez-Vidriero, I., Lozano, F.M., Ponce, M.R., et al. (2007). The JAZ family of repressors is the missing link in jasmonate signalling. Nature 448:666–U4.

Chiu, T.P., Comoglio, F., Zhou, T., Yang, L., Paro, R., and Rois, R. (2016). DNAshaper: an R/Bioconductor package for DNA shape prediction and feature encoding. Bioinformatics 32:1211–1213.

Coego, A., Brizuela, E., Castillejo, P., Ruiz, S., Koncz, C., del Pozo, J.C., Piñeiro, M., Jarillo, J.A., Paz-Ares, J., León, J., et al. (2014). The TRANSPLENTA collection of Arabidopsis lines: a resource for functional analysis of transcription factors based on their conditional overexpression. Plant J 77:944–953.

Dobón, A., Canet, J.V., García-Andrade, J., Angulo, C., Neumetzler, L., Persson, S., and Vera, P. (2015). Novel disease susceptibility factors for fungal necrotrophic pathogens in Arabidopsis. PLoS Pathog. 11:e1004800.

Dombrecht, B., Gang, P.X., Sprague, S.J., Kirkegaard, J.A., Ross, J.J., Reid, J.B., Pitt, G.P., Sewelam, N., Schenk, P.M., Manners, J.M., et al. (2007). MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. Plant Cell 19:2225–2245.

Doyle, J., and Doyle, J. (1987). A rapid DNA isolation procedure from small quantities of fresh leaf tissue. Phytochem. Bull. 19:11–15.

Dror, I., Golan, T., Levy, C., Rois, R., and Mandel-Gutfreund, Y. (2015). A widespread role of the motif environment in transcription factor binding across diverse protein families. Genome Res. 25:1268–1280.

Du, M., Zhao, J., Tzeng, D.T.W., Liu, Y., Deng, L., Yang, T., Zhai, Q., Wu, F., Huang, Z., Zhou, M., et al. (2017). MYC2 orchestrates a hierarchical transcriptional cascade that regulates jasmonate-mediated plant immunity in tomato. Plant Cell 29:1883–1906.

Ezer, D., Shepherd, S.J.K., Brestovskiy, A., Dickinson, P., Cortijo, S., Charoensawan, V., Box, M.S., Biswas, S., Jaeger, K.E., and Wigge, P.A. (2017). The G-box transcriptional regulatory code in Arabidopsis. Plant Physiol. 175:628–640.

Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J.-M.M., Gimenez-Ibanez, S., Gerinck, J., Eckhout, D., Schweizer, F., Godoy, M., Franco-Zorrilla, J.M., et al. (2011). The Arabidopsis bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. Plant Cell 23:701–715.

Figueroa, P., and Browse, J. (2012). The Arabidopsis JAZ2 promoter contains a G-box and thymidine-rich module that are necessary and sufficient for jasmonate-dependent activation by MYC transcription factors and repression by JAZ proteins. Plant Cell Physiol. 53:330–343.

Fonseca, S., Chini, A., Hamberg, M., Adie, B., Porzel, A., Kramell, R., Miersch, O., Westernack, C., and Solano, R. (2009). (+)-7-isoojasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. Nat. Chem. Biol. 5:344–350.

Fonseca, S., Fernández-Calvo, P., Fernández-Gimenez, G.M., Díez-Díaz, M., Chico, J.-M.M., Gimenez-Ibanez, S., Gerinck, J., Eckhout, D., Schweizer, F., Godoy, M., Franco-Zorrilla, J.M., et al. (2011). The Arabidopsis bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. Plant Cell 23:701–715.

Gao, F., Robe, K., Gaymard, F., Izquierdo, E., and Dubos, C. (2019). The transcriptional control of iron homeostasis in plants: a tale of bHLH transcription factors? Front. Plant Sci. 10:6.
DNA-binding specificity of plant bHLH transcription factors

Gimenez-Ibanez, S., Boter, M., Ortigosa, A., García-Casado, G., Chini, A., Lewsey, M.G., Ecker, J.R., Ntoukakis, V., and Solano, R. (2017). JA22 controls stomata dynamics during bacterial invasion. New Phytol. 213:1378–1392.

Godoy, M., Franco-Zorrilla, J.M., Pérez-Pérez, J., Oliveros, J.C., Lorenzo, Ó., and Solano, R. (2011). Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. Plant J. 66:700–711.

Gordan, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., Bulyk, M.L., Gordan, R., Shen, N., Dror, I., et al. (2013). Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. Cell Rep. 3:1093–1104.

Guo, Y., Mahony, S., and Gifford, D.K. (2012). High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. PLoS Comput. Biol. 8:e1002638.

Hickman, R., Van Verk, M.C., Van Dijken, A.J.H., Mendes, M.P., Vroegop-Vos, I.A., Caarls, L., Steenbergen, M., Van der Nagel, I., Wesselink, G.J., Jirönkin, A., et al. (2017). Architecture and dynamics of the jasmonic acid gene regulatory network. Plant Cell 29:2086–2105.

Hong, G.J., Xue, X.Y., Mao, Y.B., Wang, L.J., and Chen, X.Y. (2012). Arabidopsis MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. Plant Cell 24:2635–2648.

Hornitschek, P., Kohnen, M.V., Lorrain, S., Rougemont, J., Ljung, K., López-Vidriero, I., Franco-Zorrilla, J.M., Solano, R., Trevisan, M., Pradervand, S., et al. (2012). Phytochrome interacting factors 4 and 5 control seedling growth in changing light conditions by directly controlling auxin signaling. Plant J. 71:399–711.

Howe, G.A., Major, I.T., and Koo, A.J. (2018). Modularity in jasmonate signaling for multistress resilience. Annu. Rev. Plant Biol. 69:387–415.

Jiang, Y., Yang, B., and Deyholos, M.K. (2009). Functional characterization of the Arabidopsis bHLH92 transcription factor in abiotic stress. Mol. Genet. Genomics 282:503–516.

Jin, R., Klasfeld, S., Zhu, Y., Fernandez Garcia, M., Xiao, J., Han, S.K., Konkol, A., and Wagner, D. (2021). LEAFY is a pioneer transcription factor and licenses cell reprogramming to floral fate. Nat. Commun. 12:626.

Jolma, A., Yan, J., Whittington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgounova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell 152:327–339.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast splice-aware aligner with low memory requirements. Nat. Methods 12:357–360.

Lai, X., Stigliani, A., Vachon, G., Carles, C., Smaczniak, C., Zubieta, C., Kaufmann, K., and Parcy, F. (2019). Building transcription factor binding site models to understand gene regulation in plants. Mol. Plant 12:743–763.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Ledent, V., and Vervoort, M. (2001). The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. Genome Res. 11:754–770.

Levo, M., Zalcikar, E., Sharon, E., Machado, A.C.D., Kalma, Y., Lotam-Pompan, M., Weinberger, A., Yakhini, Z., Rohs, R., and Segal, E. (2015). Unraveling determinants of transcription factor binding outside the core binding site. Genome Res. 25:1018–1029.

Lian, T., Xu, Y.P., Li, L.F., and Su, X.D. (2017). Crystal structure of tetrameric Arabidopsis MYC2 reveals the mechanism of enhanced interaction with DNA. Cell Rep. 19:1334–1342.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

Maher, K.A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D.A., Zumstein, K., Woodhouse, M., Bubb, K., Dorrity, M.W., et al. (2018). Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. Plant Cell 30:15–36.

Major, I.T., Yoshida, Y., Campos, M.L., Kapali, G., Xin, X.-F., Sugimoto, K., de Oliveira Ferreira, D., He, S.Y., and Howe, G.A. (2017). Regulation of growth-defense balance by the JASMONATE ZIM-DOMAIN (JAZ)-MYC transcriptional module. New Phytol. 215:1533–1547.

Mathelier, A., Xiong, B., Chiu, T.P., Yang, L., Rohs, R., and Wasserman, W.W. (2016). DNA shape features improve transcription factor binding site predictions in vivo. Cell Syst. 3:278–286.e4.

Monte, I., Ishida, S., Zamarreño, A.M., Hamberg, M., Franco-Zorrilla, J.M., García-Casado, G., Gouhier-Darimont, C., Reymond, P., Takahashi, K., García-Mina, J.M., et al. (2018). Ligand-receptor co-evolution shaped the jasmonate pathway in land plants. Nat. Chem. Biol. 14:480–488.

Monte, I., Franco-Zorrilla, J.M., García-Casado, G., Zamarreño, A.M., García-Mina, J.M., Nishihama, R., Kohchi, T., and Solano, R. (2019). A single JAZ repressor controls the jasmonate pathway in Marchantia polymorpha. Mol. Plant 12:185–198.

Montgomery, S.A., Tanizawa, Y., Galik, B., Wang, N., Ito, T., Mochizuki, T., Akimcheva, S., Bowman, J.L., Cognat, V., Maréchal-Drouard, L., et al. (2020). Chromatin organization in early land plants reveals an ancestral association between H3K27me3 transposons, and constitutive heterochromatin. Curr. Biol. 30:573–588.e7.

Morgunova, E., and Taijale, J. (2017). Structural perspective of cooperative transcription factor binding. Curr. Opin. Struct. Biol. 47:1–8.

O’Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and epistrome features shape the regulatory DNA landscape. Cell 165:1280–1292.

Ortigosa, A., Fonseca, S., Franco-Zorrilla, J.M., Fernández-Calvo, P., Zander, M., Lewsey, M.G., García-Casado, G., Fernández-Barbero, G., Ecker, J.R., and Solano, R. (2020). The JAs-pathway MYC transcription factors regulate photomorphogenic responses by targeting HY5 gene expression. Plant J. 102:138–152.

Pajaro, A., Madrigal, P., Muino, J.M., Matus, J.T., Jin, J., Mechcia, M.A., Debernardi, J.M., Palatnik, F.J., Balazadeh, S., Arif, M., et al. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biol. 15:R41.

Peñuelas, M., Monte, I., Schweizer, F., Vallat, A., Reymond, P., Garcia-Casado, G., Franco-Zorrilla, J.M., and Solano, R. (2019). Jasmonate-related MYC transcription factors are functionally conserved in Marchantia polymorpha. Plant Cell 31:2491–2509.

Pfeiffer, A., Shi, H., Tepperman, J.M., Zhang, Y., and Quail, P.H. (2014). Combinatorial complexity in a transcriptionally centered signaling hub in Arabidopsis. Mol. Plant 7:1598–1618.

Quinlan, A.R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinformatics 47:1–34.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44:W160–W165.
DNA-binding specificity of plant bHLH transcription factors

Rogers, J.M., Barrera, L.A., Reyon, D., Sander, J.D., Kellis, M., Keith Joung, J., and Bulyk, M.L. (2015). Context influences on TALE-DNA binding revealed by quantitative profiling. Nat. Commun. 6:7440.

Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009). The role of DNA shape in protein-DNA recognition. Nature 461:1248–1253.

Sasaki-Sekimoto, Y., Jikumaru, Y., Obayashi, T., Saito, H., Masuda, S., Kamiya, Y., Ohta, H., and Shirasu, K. (2013). Basic helix-loop-helix transcription factors JASMONATE-ASSOCIATED MYC2-LIKE1 (JAM1), JAM2, and JAM3 are negative regulators of jasmonate responses in Arabidopsis. Plant Physiol. 163:291–304.

Sayou, C., Nanao, M.H., Jamin, M., Posé, D., Thévenon, E., Grégoire, L., Tichtinsky, G., Denay, G., Ott, F., Llobet, M.P., et al. (2016). A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. Nat. Commun. 7:11222.

Schweizer, F., Fernandez-Calvo, P., Zander, M., Diez-Diaz, M., Fonseca, S., Glauser, G., Lewsey, M.G., Ecker, J.R., Solano, R., and Reymond, P. (2013). Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. Plant Cell 25:3117–3132.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordan, R., Rohs, R., Gordan, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. Trends Biochem. Sci. 39:381–399.

Song, S., Huang, H., Wang, J., Liu, B., Qi, T., and Xie, D. (2017). MYC5 is involved in jasmonate-regulated plant growth, leaf senescence and defense responses. Plant Cell Physiol. 58:1752–1763.

Stigliani, A., Martin-Arevalillo, R., Lucas, J., Bessy, A., Vinos-Poyo, T., Mironova, V., Vernoux, T., Dumas, R., and Parcy, F. (2019). Capturing auxin response factors syntax using DNA binding models. Mol. Plant 12:822–832.

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., Nomura, K., He, S.Y., Howe, G.A., and Browse, J. (2007). JAZ repressor proteins are targets of the SCFCO11 complex during jasmonate signalling. Nature 448:661–665.

Varaud, E., Brioudes, F., Szécsi, J., Leroux, J., Brown, S., Perrot-Rechenmann, C., and Bendahmane, M. (2011). AUXIN RESPONSE FACTOR8 regulates Arabidopsis petal growth by interacting with the bHLH transcription factor BIGPETALp. Plant Cell 23:973–983.

Wang, H., Li, S., Li, Y., Xu, Y., Wang, Y., Zhang, R., Sun, W., Chen, Q., Wang, X., Li, C., et al. (2019). MED25 connects enhancer-promoter looping and MYC2-dependent activation of jasmonate signalling. Nat. Plants 5:616–625.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell 158:1431–1443.

Yin, Y., Vafeados, D., Tao, Y., Yoshida, S., Asami, T., and Chory, J. (2005). A new class of transcription factors mediates brassinosteroid-regulated gene expression in Arabidopsis. Cell 120:249–259.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356:eaaj2239.

Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIP seeker: An R/Biocenter package for ChIP peak annotation, comparison and visualization. Bioinformatics 31:2382–2383.

Zander, M., Lewsey, M.G., Clark, N.M., Yin, L., Bartlett, A., Saldiera Guzmán, J.P., Hann, E., Langford, A.E., Jow, B., Wise, A., et al. (2020). Integrated multi-omics framework of the plant response to jasmonic acid. Nat. Plants 6:290–302.

Zhang, F., Yao, J., Ke, J., Zhang, L., Lam, V.Q., Xin, X.-F., Zhou, X.E., Chen, J., Brunzelle, J., Griffin, P.R., et al. (2015). Structural basis of JAZ repression of MYC transcription factors in jasmonate signalling. Nature 525:269–273.

Zhao, Y., and Stormo, G.D. (2011). Quantitative analysis demonstrates most transcription factors require only simple models of specificity. Nat. Biotechnol. 29:480–483.