Using yeast to trace a history of expanding auxin signals

Román Ramos Báez

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Jennifer Nemhauser, Chair

Takato Imaizumi

Benjamin Kerr

Program Authorized to Offer Degree:

Biology

University of Washington

Abstract

Using yeast to trace a history of expanding auxin signals

Román Ramos Báez

Chair of the Supervisory Committee:
Jennifer Nemhauser
Department of Biology

Auxin is a hormone that plays essential roles in almost all growth and developmental processes across all plants. The way cells interpret responses to auxin is regulated by the auxin signal transduction pathway - a tangled network of proteins with overlapping functions that turn auxin signals into transcriptional responses. I set out to learn more about this signaling network: how did it come to be so expanded in flowering plants? What different roles do components play within a given plant species, or across distantly related flowers? Where do the fundamental transcriptional mechanisms that control the functions of these pathways originate from? Previously, the Nemhauser Lab has developed a synthetic auxin response circuit in yeast to facilitate studying the molecular functions of signaling components. I used this circuit to explore these questions. Firstly, I hypothesized that auxin receptors in *Arabidopsis thaliana* were diverging in their functions. I used the Visualizing Variation (ViVa) tool to process 1001 information on natural variation in these receptors across accessions to learn more about the role of these proteins across different populations. I was able to use this to create hypotheses about which proteins were most functionally important. Secondly, I hypothesized that the fundamental mechanisms of auxin signaling were conserved across distantly related flowering plants. I recapitulated the maize auxin response signaling network mixing and matching its components with those of *Arabidopsis* to see if I

could produce functional signals and explore differences in their sensitivities. I found auxin signaling networks to be highly conserved in their functions, and that this circuit can be used to rapidly characterize and compare the functions of auxin signaling components across distantly related plant species. Alex Leydon previously characterized the LisH domain of TPL as being sufficient to induce a robust and de-repressible repressive function in the auxin response circuit. In my last chapter, I hypothesized that the LisH domain's repressive function is conserved across LisH-containing proteins across eukaryotes. I was able to use the auxin response circuit and high-throughput domain libraries to test this function across different natural sequences. I found that this LisH repressive function was widely conserved, and functional in both yeast and tobacco. Understanding the roles auxin plays in plant growth is essential to our understanding of the diversity of form and function in the plant world, and to our ability to engineer plant development to better suit our needs.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

It takes a village to get through a PhD. I am so grateful for all the emotional support, mentorship, and companionship provided by communities I inhabit.

Firstly, I want to thank the local queer community. Cucci Binaca and Betty Wetter, for inviting me to perform on their stage countless times. Issa Man, who shared with me her hardest and my hardest days of the pandemic, in protest, in celebration, and in friendship. The rest of the community for reminding me to believe in abundance when there was scarcity, and for showering me with love.

Secondly, I want to thank my family. Merger House - especially Madi and Lydia - for giving me my first taste of lasting Seattle friendships and community. Pooj the cat: for being so soft and squishy. My parents, siblings, and extended family, who have every excuse to not stay in touch, and yet they do with ever-growing enthusiasm and support. Jason, with whom I share everything, and I love tremendously.

Thirdly, I want to thank my academic community. UW Biology staff of the DEC like Ben, staff like Davis, Brianna, Julia, Hayato, and Michelle. Social media science friends and Gilliam Fellows, who have been my co-conspirators, supporters, and helped me to realize I could be accepted as my full self in academia. Thank you to the Maize AuxRE group: especially Andrea and Mary for welcoming me into their corny lab, and Brit, for having me in her beautiful 1950s queer home, and spending months with me on Zoom (pre-pandemic!). Big thank you to the members of the Nemhauser Lab. Alex, who was an incredibly supportive mentor, role model, co-author, and somehow expert in all things. Amy and Eric who sciened with me most of my career. Guy, Alex K., and Sabrina, Jones, and Perfectious for trusting in my mentorship. My cohort: especially Liv who organized most of our get-togethers and paralleled me in my graduate career in many ways. Thank you to my committee, for excellent feedback, and unwavering support.

Thank you Jennifer for being an example of many of the ways in which I can be an agent of positive change within academia, and for your encouragement and support. Thank you for helping me execute big dreams for our future, in and out of the lab, and for being a nightlight.

© 2021. Published by The Company of Biologists Ltd | Development (2021) 148, dev187120. doi:10.1242/dev.187120

REVIEW

Expansion and innovation in auxin signaling: where do we grow from here?

Román Ramos Báez* and Jennifer L. Nemhauser*

ABSTRACT

The phytohormone auxin plays a role in almost all growth and developmental responses. The primary mechanism of auxin action involves the regulation of transcription via a core signaling pathway comprising proteins belonging to three classes: receptors, co-receptor/co-repressors and transcription factors. Recent studies have revealed that auxin signaling can be traced back at least as far as the transition to land. Moreover, studies in flowering plants have highlighted how expansion of the gene families encoding auxin components is tied to functional diversification. As we review here, these studies paint a picture of auxin signaling evolution as a driver of innovation.

KEY WORDS: *Marchantia polymorpha*, *Physcomitrium patens*, *Arabidopsis thaliana*, *Zea mays*, Evo-devo, Auxin response

Introduction

Auxins are a group of small molecules that regulate plant growth and development through a simple signaling pathway (Lavy and Estelle, 2016; Leyser, 2018; Woodward and Bartel, 2005). This pathway is composed of proteins that enable three functions: transcriptional activation of target genes, repression of this transcriptional activity in the absence of auxin, and perception of auxin that triggers de-repression (Fig. 1). Transcriptional activation is mediated by a group of transcription factors called A class AUXIN RESPONSE FACTORS (ARFs) that bind to AUXIN RESPONSIVE ELEMENTS (AuxREs) associated with auxin-regulated genes (Ulmasov et al., 1997). AUXIN/INDOLE-3-ACETIC ACID proteins (Aux/IAAs) then act as a bridge, connecting ARFs to members of the TOPLESS (TPL) family of co-repressors, thereby repressing auxin-responsive gene transcription when auxin levels are low (Szemenyei et al., 2008). Repression is relieved when auxin acts as a molecular glue, dramatically increasing the affinity of TRANSPORT INHIBITOR RESPONSE1 (TIR1)/AUXIN SIGNALING F-BOXES (AFBs) for Aux/IAAs, both of which directly bind to auxin. TIR1/AFBs are the substrate-recognition component in an SCF-type E3 Ubiquitin Ligase complex. Association of an Aux/IAA with a TIR1/AFB leads to its polyubiquitylation and subsequent degradation by the 26S proteasome (Kepinski and Leyser, 2005). This auxin-dependent Aux/IAA degradation leads to increased transcription of ARF-associated target genes by removing TPL-mediated repression (Kim et al., 1997; Lavy and Estelle, 2016; Leyser, 2018).

The protein families involved in the auxin signaling pathway have expanded dramatically alongside major innovations in plant form and function (Mutte et al., 2018), and it has been theorized that, in fact, expansion-enabled changes in paralog function drove innovation (Matthes et al., 2019). Here, we review the recent profusion of studies that are providing crucial insights into the evolutionary origins of auxin signaling, the likely original functions of auxin in shaping plant growth and morphology, and the impact of the expansion of gene families encoding auxin signaling components on novel functions. First, we delve into studies in the liverwort *Marchantia polymorpha* and the moss *Physcomitrium patens*. These models represent ancient land plant lineages with highly reduced auxin signaling repertoires, providing novel insights into protein origins and functions. As part of this molecular evolutionary approach to understanding auxin signaling, we also include insights obtained from studies of several algal species, which approximate the state of auxin signaling components in the last common ancestor before the transition to land. Second, we consider recent discoveries connecting the expansion of and diversification within each of the families encoding the major auxin signaling components – TIR1/AFBs, Aux/IAAs and ARFs – with innovations in development and physiology. Finally, we conclude with a brief perspective on unanswered questions in the field.

The auxin signaling pathway emerged with the transition to land

A fully functional auxin signaling pathway first appeared in land plants. Orthologs of TPL and TOPLESS-RELATED PROTEINS (TPRs) can be found in the earliest lineages of land plants, and they appear to have evolved from transcriptional co-repressor families found across eukaryotes (Causier et al., 2012). The evolutionary history of other core auxin components was, until now, somewhat contentious, with differing views on whether the auxin signaling pathway – as it exists in flowering plants – is truly restricted to land plants or whether a more limited version exists in algal lineages.

Insights from algal ancestors

A protein with strong similarity to TIR1/AFBs has been identified in the charophyte algae *Coleochaete irregularis* genome, but closer examination found that the region homologous to the auxin-binding pocket of this protein is highly diverged (Mutte et al., 2018). In addition, although comparative genomics show that Aux/IAAs are found only in the genomes of land plants (Mutte et al., 2018), a group of non-canonical Aux/IAAs (ncIAAs) has been found in *C. irregularis*. However, these ncIAAs contain a PB1 domain, which enables interaction with ARFs, but lack the degron domain necessary for interacting with TIR1/AFB receptors (Mutte et al., 2018). ncIAA genes are also found in *M. polymorpha*, but loss of function and transcriptome studies have been unable to connect them to an auxin response (Flores-Sandoval et al., 2018a,b; Mutte et al., 2018). The preponderance of evidence to date suggests that, although algae do show some auxin responsiveness (Ohtaka et al., 2017), this responsiveness is unlikely to be mediated by the same pathway used in land plants.

University of Washington, Department of Biology, Seattle, WA 98105-1800, USA.

*Authors for correspondence (rramosb@uw.edu; jn7@uw.edu)

R.R.B., 0000-0002-0934-5342; J.L.N., 0000-0002-8909-735X

Development (2021) 148, dev187120. doi:10.1242/dev.187120

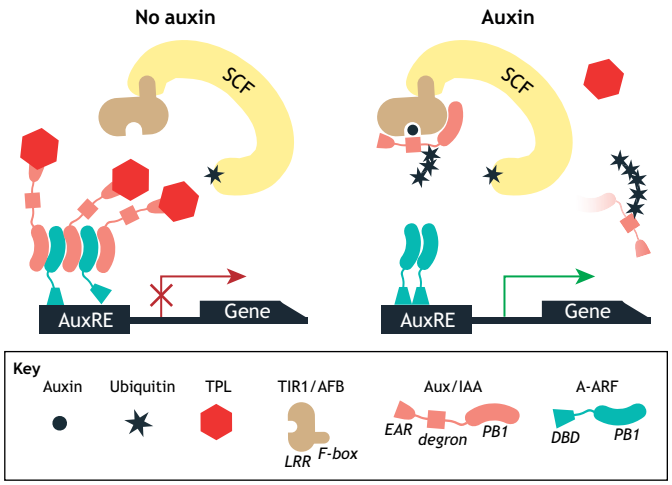


Fig. 1. A simplified representation of the canonical auxin signaling pathway. A highly simplified snapshot of the nuclear-localized, canonical auxin response pathway. In the absence of auxin, Aux/IAAs connect A-ARFs to TPL co-repressors, thereby repressing auxin-responsive gene transcription. Auxin then triggers the association of Aux/IAAs with the $SCF^{TIR1/AFB}$ complex, leading to their polyubiquitylation and subsequent degradation. The removal of Aux/IAAs (and with them the TPL) activates A-ARFs and induces gene expression.

Charophytes have also been shown to express ARFs. Phylogenetic analysis of land plant genomes divides ARFs into three distinct clades (A, B and C); however, in multiple charophyte species, there is evidence of a combined A/B clade and a C clade (Martin-Arevalillo et al., 2019). This would date the separation of the A and B groups to after the establishment of land plants (Mutte et al., 2018). This phylogenetic analysis, combined with genetic studies, implies that C class ARFs play a fundamentally different role from A or B class ARFs, and may not be involved in auxin-dependent gene regulation (Flores-Sandoval et al., 2018a,b; Mutte et al., 2018). This is further supported by phylogenetic work, which shows that the sole C class ARF found in *M. polymorpha* has a unique, apparently auxin-independent, function in growth of the gemma – an asexually produced structure aiding in vegetative propagation (Flores-Sandoval et al., 2018a,b; Kato et al., 2020; Mutte et al., 2018). A recent study argued that the C class ARFs emerged in the earliest-diverging clades of streptophyte algae, *Mesostigma viridae* and *Chlorokybus atmophyticus*, while A/B class ARFs emerge later in *Coleochaete orbicularis* (Martin-Arevalillo et al., 2019). This study also showed that C and A/B class ARFs can directly bind TPL and AuxREs in promoters, allowing them to function in an entirely auxin-independent manner (Martin-Arevalillo et al., 2019) (Fig. 2A). This direct TPL interaction suggests that the loss of direct TPL interaction, and therefore the auxin-sensitive gene repression seen in A class ARFs, was among the last pieces of auxin signaling to arise. This appears to hold true in *M. polymorpha*, where the B and C class ARFs MpARF2 and MpARF3, respectively, but not the A class ARF MpARF1, have been shown to bind TPL (Fig. 2A) (Kato et al., 2020). Together, the evidence argues for land plant auxin signaling to have evolved from neofunctionalization of a simple ancestral auxin-independent pathway in algal ancestors (Martin-Arevalillo et al., 2019).

From liverworts and mosses to angiosperms

Functional genetic studies on emerging bryophyte models have been a welcome complement to the long history of work in angiosperms. In *M. polymorpha*, there is a single TIR1/AFB receptor, a single Aux/IAA, a single TPL co-repressor and three ARFs (Flores-Sandoval et al., 2015). In *P. patens* there are four TIR1/AFBs, three Aux/IAAs, two TPL proteins (Paponov et al., 2009) and 16 ARFs (Lavy et al., 2016). However, despite the highly reduced size of these gene families, auxin signaling appears to function similarly in these plants as it does in angiosperms (Flores-Sandoval et al., 2015, 2016; Kato et al., 2015, 2017, 2018). In *Arabidopsis*, for example, there are six TIR1/AFBs, 29 Aux/IAAs, 23ARFs and four TPL relatives. The reduced auxin signaling network in bryophytes, along with their position as one of the earliest diverging groups of land plants (Harris et al., 2020; Morris et al., 2018), make them useful for inferring how the functions of this network have evolved.

In *M. polymorpha*, all the functional domains of MpTIR1, MpIAA, the MpARFs and MpTPL are intact (Flores-Sandoval et al., 2016; Kato et al., 2020). Translational fusions of MpTPL to either MpARF1 or the PB1 domain of MpIAA1 repress the auxin response and lead to severe stunting of growth (Flores-Sandoval et al., 2015, 2016). These experiments prove that MpTPL can repress auxin-responsive genes, and that MpARFs are able to dimerize with MpARF1 and MpIAAs (Flores-Sandoval et al., 2015, 2016). Although MpTIR1 function in auxin-dependent degradation of MpIAAs has not been measured directly, it has been shown that auxin-dependent growth and development require MpIAA, MpTPL and the A class ARF MpARF1 (Flores-Sandoval et al., 2015, 2016). A more recent study further detailed the specific role of MpARF1 in regulating the timing and position of cell divisions early in the development of gemma primordia, determining their axis of growth and ultimately their three-dimensional structure (Kato et al., 2017). This developmental function is highly reminiscent of the role of A class ARFs in controlling asymmetric cell divisions during early embryogenesis in the angiosperm *Arabidopsis thaliana* (Prigge et al., 2020). MpARF1 knockout plants are still able to develop structures necessary to survive (Kato et al., 2017). This initially seems surprising as MpARF1 is the only activator ARF in *M. polymorpha*, and MpARF2 and MpARF3 have been shown to be unable to complement ARF1 (Kato et al., 2020). However, it is likely that auxin-independent expression of these genes is mediated by other transcription factors (Kato et al., 2017). Together, these studies show that both the molecular functions and the developmental roles of auxin signaling circuit components are highly conserved across land plants.

Studies have employed the simplicity of the *P. patens* auxin signaling pathway to reveal functions of its components that were difficult to probe in angiosperms. For example, work in *P. patens* has shown that no auxin response is possible in full Aux/IAA knockouts, suggesting that Aux/IAAs are required for auxin response regulation (Lavy et al., 2016). These Aux/IAA null strains also show that repression is mediated by a combination of Aux/IAAs and B class repressor ARFs (Lavy et al., 2016). The B class ARF-mediated repression is weaker than that conferred by Aux/IAAs and does not require association with TPL, leading to the conclusion that it results from direct competition with the A class ARFs for DNA-binding sites (Lavy et al., 2016). Another study highlighted the complexities of Aux/IAA-induced repression using a *P. patens* strain in which PpIAA1a is the only functional Aux/IAA (Tao and Estelle, 2018). Importantly, PpIAA1a on its own is sufficient to enable wild-type auxin responses. When the interaction with TPL is removed, PpIAA1a is still able to repress the auxin response (Tao and Estelle, 2018). Moreover, further mutagenesis revealed that an Aux/IAA monomer retains function (Tao and Estelle, 2018). These suggest an additional dimension to Aux/IAA repression. In addition to multiple Aux/IAAs oligomerizing with

Development (2021) 148, dev187120. doi:10.1242/dev.187120

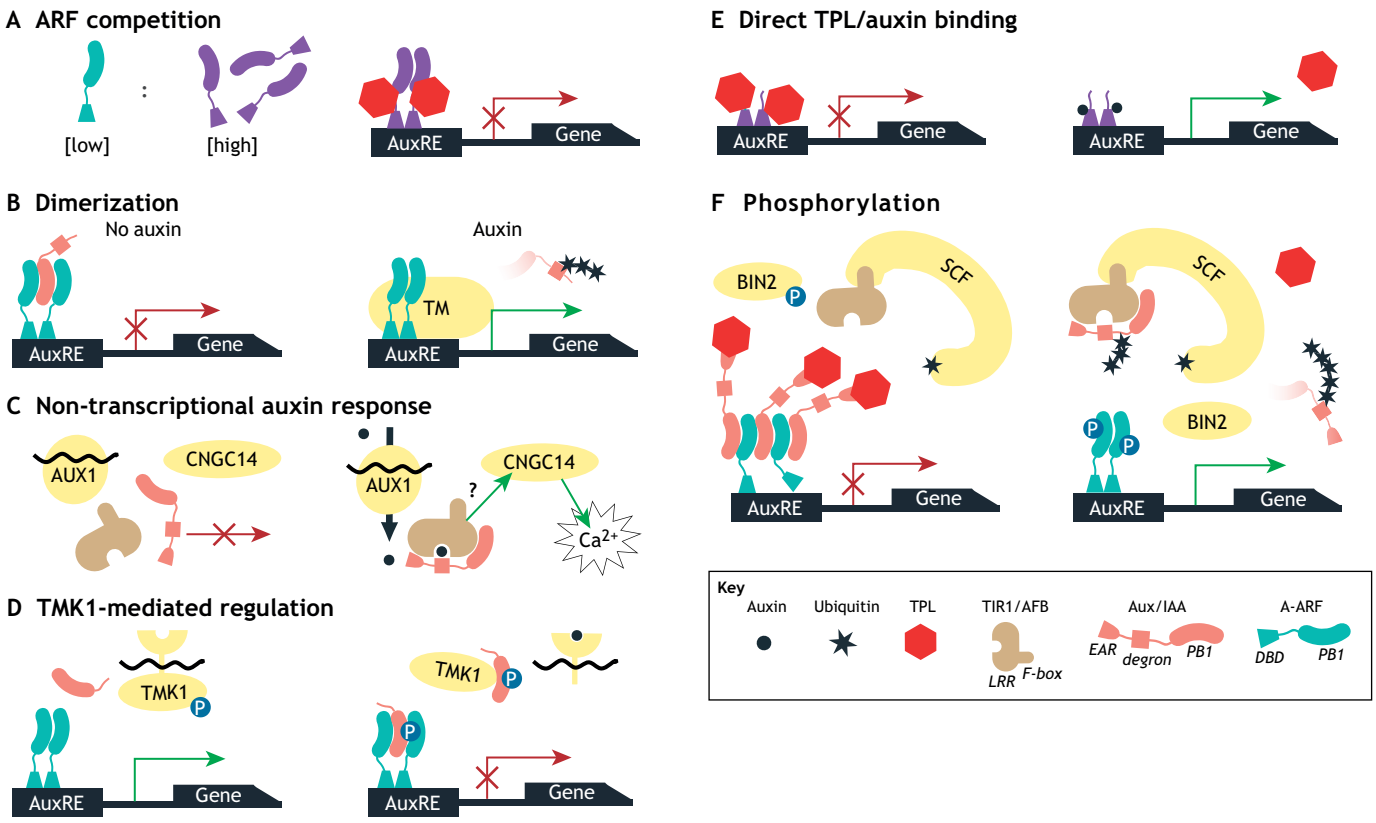


Fig. 2. A plethora of ways to regulate and signal through the simple auxin signaling pathway. (A) At high concentrations in the nucleus, A/B, B and C ARFs (purple) outcompete a low concentration of A-ARFs (blue) in binding AuxREs. This replaces the auxin-dependent repression of A-ARF with the auxin-independent repression of B and C ARFs, where these ARFs recruit TPL to AuxREs and repress auxin-inducible genes. (B) ARF dimerization is needed for full activation, acting at least in part by leading to increased affinity of class A ARFs for DNA. However, even without TPL, Aux/IAAs can contribute to repression of class A ARFs by disrupting their dimerization. (C) A rapid, non-transcriptional auxin response can also occur. This takes place outside the nucleus, where AUX1 transports auxin into the cytoplasm, triggering cytoplasmic TIR1 and likely Aux/IAAs to activate CNGC14 through an unknown pathway. This in turn allows calcium (Ca^{2+}) into the cell, rapidly depolarizing it. (D) TMK1 is a plasma membrane-localized, trans-membrane receptor-like kinase. When auxin is present, the kinase domain of TMK1 is cleaved and phosphorylates non-canonical IAAs (ncIAAs) that lack a degron. This phosphorylation then strengthens repression of some auxin-induced genes. (E) ETTIN/ARF3 (ETT, purple) is a B-class ARF that has no PB1 domain but is able to directly bind TPL. Auxin perception disrupts the interaction between ETT and TPL, thereby activating gene expression. (F) BIN2 phosphorylates (P) A class ARFs, preventing Aux/IAAs from binding to them, and strengthening their bond to AuxREs. BIN2 also facilitates Aux/IAA degradation in an auxin-independent manner, de-repressing auxin-inducible genes.

activator ARF pairs and recruiting TPL, single Aux/IAAs may disrupt A-ARF pairs by wedging themselves between ARF PB1 domains (Fig. 2B). In *A. thaliana*, A-ARF dimerization at the PB1 domain contributes to DNA binding (Pierre-Jerome et al., 2016). It is worth noting that quantifying this effect for *A. thaliana* proteins requires reconstructing the entire auxin pathway in a heterologous context, another argument in favor of making full use of bryophyte models.

Studies of bryophyte auxin signaling networks have therefore proven a useful tool, in both clarifying the mechanisms of nuclear auxin signaling, as well as for ascertaining the functions of individual components. Considering that the auxin nuclear response first appears early in land plant evolution, it is surprising that bryophyte auxin signaling is so similar to that in flowering plants at the molecular, cellular and organismal levels. This suggests that auxin signaling proteins have retained their fundamental functions throughout most of their evolutionary history.

Evolving complexity in signaling components enables diversity in plant form

In parallel to the evolution of morphological and physiological complexity in land plants, the gene families encoding auxin signaling components have undergone sizable expansions (Mutte et al., 2018). Several recent studies bolster the argument that this expansion contributed to evolutionary innovations (Lavy and Estelle, 2016; Matthes et al., 2019; Mutte et al., 2018; Wang and Estelle, 2014; Israeli et al., 2020). The 1001 Genome Project (Weigel and Mott, 2009), with its ever-growing number of sequenced *A. thaliana* accessions, has been a tremendous boom to plant evolutionary studies. This community resource greatly facilitates the linking of morphological or physiological variation to differences in molecular expression and/or function caused by changes in genome sequence. In one study, a natural hypermorphic variant of AtTIR1 was identified (Wright et al., 2017). This variant alters root architecture when expressed in a commonly used lab accession, and the roots of most accessions carrying this variant were found to be more sensitive to exogenous auxin. A follow-up to this work was the development of a webtool called Visualizing Variation (ViVa), which aims to make use of the 1001 Genomes data a routine and readily accessible part of every molecular biologist's toolkit (Hamm et al., 2019). When ViVa was applied to auxin pathway components, several interesting patterns emerged, including the observation that recently duplicated sister proteins in the large *Arabidopsis* Aux/IAA family frequently have very different rates of sequence diversification, suggesting that

Development (2021) 148, dev187120. doi:10.1242/dev.187120

one sister was becoming neo- or sub-functionalized, while the other was retaining a more ancestral function (Hamm et al., 2019). This is consistent with the detailed analysis of the diversification in sequence, expression pattern and function of one such sister pair: *IAA6* and *IAA19* (Winkler et al., 2017). Together, these and other approaches are beginning to shed light on how expanding auxin signaling networks are evolving, and how this expansion might have driven diversification in the dynamics of auxin signaling pathways and plant form.

Evolution of the rapid auxin response

One evolutionary innovation observed in *Arabidopsis* roots enables auxin to induce very rapid growth responses that are too fast to require *de novo* protein production but still require *AtTIR1* (Shih et al., 2015). The key to pinpointing this response, and connecting it to AtTIR1, was the development of two new technologies. The first allowed researchers to measure rapid changes in auxin response in combination with a means to rapidly apply and remove exogenous auxin. Work with this system provided conclusive evidence that the first auxin-regulated growth responses precede transcriptional effects (Fendrych et al., 2018). By combining this set-up with a second technology – an engineered *Arabidopsis* TIR1 (ccvTIR1) that responds exclusively to an engineered auxin molecule (cvxIAA) (Uchida et al., 2018) – researchers could illustrate unambiguously that auxin sensing by AtTIR1 is sufficient for triggering the rapid response (Fendrych et al., 2018). Additional work in root hairs connects this rapid response to AUXIN RESISTANT 1 (AUX1)-mediated auxin uptake followed by CYCLIC NUCLEOTIDE-GATED CHANNEL 14 (CNGC14)-mediated Ca^{2+} depolarization of cells (Dindas et al., 2018). MITOGEN-ACTIVATED PROTEIN KINASE (MAPK) cascade-mediated regulation of auxin-induced cell expansion via RHO-LIKE GTPASES FROM PLANTS (ROPs) (Enders et al., 2017) may also be connected to this mode of auxin signaling. Given work in *P. patens* showing that Aux/IAAs are required for an auxin response (Lavy et al., 2016), the non-transcriptional auxin response likely requires both TIR1/AFBs and Aux/IAAs, but may not need ARFs or TPL (Fig. 2C). However, this hypothesis only holds true if the rapid auxin response exists in *P. patens*, which remains to be determined.

A recent genetic tour-de-force combining all *TIR1/AFB* loss of function mutants in *A. thaliana* revealed that *AtTIR1* and its closest paralog *AtAFB1* have been sub-functionalized to elicit these rapid, non-transcriptional auxin responses (Prigge et al., 2020). *AtAFB1* and *AtTIR1* arose from a duplication at the base of Brassicales – a group of flowering plants including *A. thaliana* (Prigge et al., 2020). Although *AtAFB1* is highly conserved in *A. thaliana* (Hamm et al., 2019), it is divergent across plant species, evolving three times more rapidly than *AtTIR1* (Prigge et al., 2020). It also has a substitution in a residue crucial for full function in auxin-dependent Aux/IAA degradation (Yu et al., 2015), leading to the hypothesis that it has undergone non-functionalization. However, when the function of all TIR1/AFBs was measured during rapid bending in response to a change in gravity vector, AtAFB1 function was seen to be essential for normal growth kinetics involving the rapid response (Prigge et al., 2020). Whether this special role for AtTIR1 and AtAFB1 among the receptors represents an early bifurcation in function among the receptor family, or instead reflects a division of labor seen only in some species, will be exciting to investigate in future work.

Evolution of Aux/IAA degradation rates

Expansion of the gene families encoding auxin components raises the question of whether there are unique roles for other members, as seen for AtAFB1 and AtTIR1. The heterologous expression of auxin pathway components in yeast has enabled direct testing of many components in isolation from one another (Pierre-Jerome et al., 2014). One insight gleaned from this approach is that the identity of both TIR1/AFB and Aux/IAA components can have profound impacts on auxin-induced degradation rates (Havens et al., 2012). In addition, experiments using the yeast system to study auxin signaling components from *Zea mays* (maize) found fairly limited differences in degradation rates between ZmAux/IAAs when expressed with a maize receptor, whereas a broad range of rates could still be detected when the assays included a receptor from *A. thaliana* (Ramos Báez et al., 2020).

Aux/IAAs act as co-receptors with TIR1/AFBs, directly binding auxin along their degron domain. Although mutations within the degron can dramatically stabilize Aux/IAAs, their degradation rate is also impacted by sequences outside this region. Degron domains that enable fast or slow degradation in *A. thaliana* confer different behaviors in *P. patens* (Tao and Estelle, 2018). Swapping the domains of two *A. thaliana* Aux/IAAs showed that most domains play a role in the auxin-induced affinity of the Aux/IAAs to the AtTIR1 receptor (Niemeyer et al., 2020). The same study demonstrated that AtTIR1 has specific residues outside the auxin-binding region that greatly facilitate Aux/IAA interaction (Niemeyer et al., 2020). Importantly, changing the rate of auxin-induced degradation was able to alter the pace of lateral root development in transgenic *A. thaliana* plants (Guseman et al., 2015), suggesting that this property may be optimized during evolution. Future work that includes more species, and especially more receptors, could shed light on how important auxin sensitivity/degradation rate is in different contexts.

Evolution of distinct auxin signaling circuits

Auxin can produce a multitude of distinct transcriptional responses depending on which cells or tissues are involved. One way to explain such context-specific auxin functions would be if each ARF had a distinct DNA-binding preference, as ARFs vary in their expression patterns (Truskina et al., 2021). Several recent studies have provided evidence that this is at least partially true when comparing DNA-binding preferences of A and B class ARFs (Boer et al., 2014; Galli et al., 2018; Freire-Rios et al., 2020). However, a recent study in *P. patens* revealed that both classes of ARFs have the same binding preferences (Lavy et al., 2016; Kato et al., 2020). Moreover, in all species examined to date, ARFs within a single class share binding preferences. A large-scale analysis of ZmARF binding sites (Galli et al., 2018) and a synthetic study in yeast that used clade A ARFs from both *A. thaliana* and maize (Lanctot et al., 2020) point to nearly identical binding site preferences within the class A ARFs. When comparing binding specificities of A, B and C class ARFs, as well as those of the charophyte algae *C. atmophyticus* ARF, C class ARFs are seen to bind to a broader array of promoters (Martin-Arevalillo et al., 2019). These differences in DNA specificity could be due to observed differences in residues within the DNA-binding domain (Kato et al., 2020), as well as to differences in ARF interactors. Recent *in vivo* analysis and transcriptome analysis has also shown that differences in DNA architecture, such as AuxRE repeat organization, affect the affinity of AtARF5 (class A) and AtARF1 (class B) for different promoters (Freire-Rios et al., 2020). Because significant differences in DNA specificity have not been shown in A class ARFs, it remains challenging to resolve how distinct transcriptional modules are encoded.

One emerging model poses that the ratios of ARFs and other signaling components provide specificity. In *Marchantia*, MpARF1

Development (2021) 148, dev187120. doi:10.1242/dev.187120

(class A) and MpARF2 (class B) have similar DNA affinity and binding preferences (Kato et al., 2020). In addition, it has been observed that inducing the overexpression of MpARF1 alone maintains auxin responsiveness but inhibits growth, while overexpressing MpARF2 leads to auxin insensitivity (Kato et al., 2020). Overexpression of both ARFs together restores auxin sensitivity. Thus, the relative stoichiometry and not absolute levels of the different ARF classes is proposed to be the key parameter for determining the amplitude, and perhaps duration, of auxin-induced transcriptional effects (Fig. 2A) (Kato et al., 2020). The same study was also able to quantify the levels of the two ARFs across gemma cells, revealing differences in the ratio of protein accumulation in different cells (Kato et al., 2020). Moreover, the expression of A class ARFs in *Arabidopsis* was recently shown to be regulated by a network of repressors, with the interaction networks varying for different ARFs (Truskina et al., 2021). This provides a mechanism by which different A class ARFs might vary in their accumulation across different plant tissues, leading to context-specific auxin responses.

The ratio of ARFs and Aux/IAAs also contributes to signaling dynamics. A comprehensive molecular genetic study of A class ARFs in *Solanum lycopersicum* (tomato) has linked subfunctionalization within the family to different aspects of leaf shape determination (Israeli et al., 2019), revealing that a specific expression ratio of SlAux/IAAs and SlARFs in the developing leaf primordia is required to ensure the appropriate formation of discrete leaflets. Specifically, SlAux/IAAs act to suppress growth between leaflets, whereas SlARFs stimulate expansion. Changing the ratios of these proteins results in leaves that range from simple and fully expanded to extremely reduced and needle-like (Israeli et al., 2019). Interestingly, gene redundancy and subsequent robustness are also showcased in this work. In double mutants for an SlAux/IAA and an SlARF, the average number of leaflets per leaf is similar to that observed in wild-type plants, although the variation in leaflet number between leaves increases (Israeli et al., 2019, 2020).

There are also reports of unexpected avenues of diversification of auxin pathway components. One recent study showed that AtIAA32 and AtIAA34 – both highly diverged Aux/IAAs with minimal, if any, response to auxin – interact with TRANSMEMBRANE KINASE 1 (TMK1) to activate auxin-responsive growth in the apical hook (Cao et al., 2019) (Fig. 2D). A second study focused on AtARF3/ETTIN (ETT), which directly binds to auxin (Kuhn et al., 2020) (Fig. 2E), thereby bypassing the ubiquitylation of Aux/IAAs and directly disrupting TPL binding to relieve repression on target genes. This appears to be a textbook example of neofunctionalization. A third study connects the kinase BRASSINOSTEROID-INSENSITIVE 2 (BIN2) to auxin-independent regulation of auxin-inducible genes. BIN2 was found to phosphorylate A class ARFs in *Arabidopsis* AtARF7 and AtARF19 with multiple auxin response-promoting results, including a facilitation of their interaction with AuxREs, a weakening of their interaction with Aux/IAAs and a strengthening of the Aux/IAA interaction with the SCF^{TIR1} complex (Fig. 2F) (Cho et al., 2014). Phosphorylation by BIN2 also results in loss of DNA binding and repression of activities of AtARF2, a class B ARF (Cho et al., 2014). It is exciting to ponder whether more of this type of radically unexpected variation will be uncovered as auxin responses are studied in more detail in more species.

Conclusions and future perspectives

We still have much to learn about the evolutionary origins of auxin signaling, especially with regard to how individual genes are co-opted to serve new functions and how crosstalk is minimized between auxin modules. Outside of the model that A and B class ARFs bind to and compete for the same AuxREs, there remain many mysteries about B and C class ARF function. Learning more about the roles of A/B class ARF in algae might help us to better understand what makes A and B class ARFs fundamentally different from C class ARFs and provide clues into the origins of ARF classes. Intriguingly, some flowering plant C class ARFs interact with specific Aux/IAAs (Piya et al., 2014; Vernoux et al., 2011). Have some of these ARFs become neofunctionalized to perform an auxin signaling function in other plants? In addition, there are tantalizing unanswered questions about the rapid auxin response. Is it conserved in all land plants? If, as the work in bryophytes suggests, all auxin signaling requires Aux/IAAs, are Aux/IAAs necessary for the non-transcriptional auxin response? What is the relationship between events at the plasma membrane and those in the nucleus, and is that relationship different in different cell types or plant lineages? More studies in *Marchantia* and *Physcomitrium*, as well as studies in a broader array of angiosperms, would help to answer these questions.

Despite its deceptive simplicity, the nuclear auxin signaling pathway continues to reveal complexity in functions. Studies in tomato, *Arabidopsis* and maize have shown that many of the fundamental functions of auxin signaling components are highly conserved across land plants. They have also revealed a number of new signaling mechanisms. Included in these mechanisms are: differences in protein ratios (Israeli et al., 2019); non/sub-functionalization of components (Israeli et al., 2019); binding competition between ARFs as well as with other signaling proteins (Kato et al., 2018); and the conservation and extent of non-canonical functions, such as direct auxin interactions and direct TPL interactions (Causier et al., 2012; Simonini et al., 2016). These novel functions translate into diversity in plant architecture, and highlight subtle control of the auxin response beyond a simple on/off growth signal.

Although genetics has enabled great advances in identifying the function of auxin pathway components, many single mutants are lethal, highly pleiotropic or have no phenotype at all. Moving forward, it would be particularly useful to observe phenotypes in specific cell types or at crucial time points in development (Decaestecker et al., 2019). Although studies have shown similarities in circuit function between angiosperms and flowering plants, there are many other plant families that remain understudied. CRISPR/Cas9-based genome editing makes it feasible to build multi-knockout lines in Aux/IAAs and ARFs in diverse species. In combination with new technologies that allow for easier comparative analysis of natural or engineered variation (Ramos Báez et al., 2020), generating genetic resources in a phylogenetically informed collection of plants should ultimately make it possible to connect changes in specific amino acid sequences in auxin signaling components to key innovations in development. It will also be fascinating to learn whether evolution has re-parameterized auxin signaling with compensatory changes across the network when confronted with novel hyper- or hypomorphic variants.

Finally, work on the auxin pathway could be used as a scaffold to better understand other plant signaling pathways, especially those with similar structures. Jasmonic acid (JA) is of particular interest, as the evolutionary history of auxin and JA perception are closely linked (Blázquez et al., 2020; Wang et al., 2015). What was the function of the shared ancestral signaling pathway in early land plants? Genomics studies of charophyte algae and bryophytes, alongside ancestral sequence reconstruction of critical nodes combined with functional analysis in bryophytes and angiosperms, could help us reconstruct

this pivotal moment in plant evolution. By understanding how plants came to survive a completely new existence on land, we may find clues for engineering plants that are able to meet the challenges of tomorrow.

Acknowledgements
We thank the members of the Nemhauser Lab for many helpful discussions and critical reviews of the manuscript.

Competing interests
The authors declare no competing or financial interests.

Funding
Work in our group on evolution of auxin response is supported by the National Science Foundation (IOS-1546873), the National Institutes of Health (R01-GM107084) and a Howard Hughes Medical Institute Faculty Scholar Award. R.R.B. is additionally supported by a Howard Hughes Medical Institute Gilliam Fellowship (GT11355). Deposited in PMC for release after 12 months.

References

Blázquez, M. A., Nelson, D. C. and Weijers, D. (2020). Evolution of plant hormone response pathways. *Annu. Rev. Plant Biol.* **71**, 327-353. doi:10.1146/annurev-arplant-050718-100309

Boer, D. R., Freire-Rios, A., van den Berg, W. A. M., Saaki, T., Manfield, I. W., Kepinski, S., López-Vidrieo, 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. doi:10.1016/j.cell.2013.12.027

Cao, M., Chen, R., Li, P., Yu, Y., Zheng, R., Ge, D., Zheng, W., Wang, X., Gu, Y., Gelová, Z. et al. (2019). TMK1-mediated auxin signalling regulates differential growth of the apical hook. *Nature* **568**, 240-243. doi:10.1038/s41586-019-1069-7

Causier, B., Lloyd, J., Stevens, L. and Davies, B. (2012). TOPLESS co-repressor interactions and their evolutionary conservation in plants. *Plant Signal. Behav.* **7**, 325. doi:10.4161/psb.19283

Cho, H., Ryu, H., Rho, S., Hill, K., Smith, S., Audenaert, D., Park, J., Han, S., Beeckman, T., Bennett, M. J. et al. (2014). A secreted peptide acts on BIN2-mediated phosphorylation of ARFs to potentiate auxin response during lateral root development. *Nat. Cell Biol.* **16**, 66-76. doi:10.1038/ncb2893

Decaestecker, W., Buono, R. A., Pfeiffer, M. L., Vangheluwe, N., Jourquin, J., Karimi, M., Van Isterdael, G., Beeckman, T., Nowack, M. K. and Jacobs, T. B. (2019). CRISPR-TSKO: a technique for efficient mutagenesis in specific cell types, tissues, or organs in arabidopsis. *Plant Cell* **31**, 2868-2887. doi:10.1105/tpc.19.00454

Dindas, J., Scherzer, S., Roelfsema, M. R. G., von Meyer, K., Müller, H. M., Al-Rasheid, K. A. S., Palme, K., Dietrich, P., Becker, D., Bennett, M. J. et al. (2018). AUX1-mediated root hair auxin influx governs $SCF^{TIR1/AFB}$ -type Ca^{2+} signaling. *Nat. Commun.* **9**, 1174. doi:10.1038/s41467-018-03582-5

Enders, T. A., Frick, E. M. and Strader, L. C. (2017). An Arabidopsis kinase cascade influences auxin-responsive cell expansion. *Plant J.* **92**, 68-81. doi:10.1111/tpj.13635

Fendrych, M., Akhmanova, M., Merrin, J., Glanc, M., Hagihara, S., Takahashi, K., Uchida, N., Torii, K. U. and Friml, J. (2018). Rapid and reversible root growth inhibition by TIR1 auxin signalling. *Nat. Plants* **4**, 453-459. doi:10.1038/s41477-018-0190-1

Flores-Sandoval, E., Eklund, D. M. and Bowman, J. L. (2015). A simple auxin transcriptional response system regulates multiple morphogenetic processes in the liverwort marchantia polymorpha. *PLoS Genet.* **11**, e1005207. doi:10.1371/journal.pgen.1005207

Flores-Sandoval, E., Eklund, D. M. and Bowman, J. L. (2016). Correction: a simple auxin transcriptional response system regulates multiple morphogenetic processes in the liverwort Marchantia polymorpha. *PLoS Genet.* **12**, e1005900. doi:10.1371/journal.pgen.1005900

Flores-Sandoval, E., Eklund, D. M., Hong, S.-F., Alvarez, J. P., Fisher, T. J., Lampugnani, E. R., Golz, J. F., Vázquez-Lobo, A., Dierschke, T., Lin, S.-S. et al. (2018a). Class C ARFs evolved before the origin of land plants and antagonize differentiation and developmental transitions in Marchantia polymorpha. *New Phytol.* **218**, 1612-1630. doi:10.1111/nph.15090

Flores-Sandoval, E., Romani, F. and Bowman, J. L. (2018b). Co-expression and transcriptome analysis of marchantia polymorpha transcription factors supports class C ARFs as independent actors of an ancient auxin regulatory module. *Front. Plant Sci.* **9**, 1345. doi:10.3389/fpls.2018.01345

Freire-Rios, A., Tanaka, K., Crespo, I., van der Wijk, E., Sizentsova, Y., Levitsky, V., Lindhoud, S., Fontana, M., Hohlbein, J., Boer, D. R. et al. (2020). Architecture of DNA elements mediating ARF transcription factor binding and auxin-responsive gene expression in Arabidopsis. *Proc. Natl Acad. Sci. USA* **117**, 24557-24566. doi:10.1073/pnas.2009554117

Galli, M., Khakhar, A., Lu, Z., Chen, Z., Sen, S., Joshi, T., Nemhauser, J. L., Schmitz, R. J. and Gallavotti, A. (2018). The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. *Nat. Commun.* **9**, 4526. doi:10.1038/s41467-018-06977-6

Guseman, J. M., Hellmuth, A., Lanctot, A., Feldman, T. P., Moss, B. L., Klavins, E., Calderon Villalobos, L. I. A. and Nemhauser, J. L. (2015). Auxin-induced degradation dynamics set the pace for lateral root development. *Development* **142**, 905-909. doi:10.1242/dev.117234

Hamm, M. O., Moss, B. L., Leydon, A. R., Gala, H. P., Lanctot, A., Ramos, R., Klaeser, H., Lemmex, A. C., Zahler, M. L., Nemhauser, J. L. et al. (2019). Accelerating structure-function mapping using the ViVa webtool to mine natural variation. *Plant Direct* **3**, e00147. doi:10.1002/pld3.147

Harris, B. J., Harrison, C. J., Hetherington, A. M. and Williams, T. A. (2020). Phylogenomic evidence for the monophyly of bryophytes and the reductive evolution of stomata. *Curr. Biol.* **30**, 2001-2012.e2. doi:10.1016/j.cub.2020.03.048

Havens, K. A., Guseman, J. M., Jang, S. S., Pierre-Jerome, E., Bolten, N., Klavins, E. and Nemhauser, J. L. (2012). A synthetic approach reveals extensive tunability of auxin signaling. *Plant Physiol.* **160**, 135-142. doi:10.1104/pp.112.202184

Israeli, A., Capua, Y., Shwartz, I., Tal, L., Meir, Z., Levy, M., Bar, M., Efroni, I. and Ori, N. (2019). Multiple auxin-response regulators enable stability and variability in leaf development. *Curr. Biol.* **29**, 1746-1759.e5. doi:10.1016/j.cub.2019.04.047

Israeli, A., Reed, J. W. and Ori, N. (2020). Genetic dissection of the auxin response network. *Nat. Plants* **6**, 1082-1090. doi:10.1038/s41477-020-0739-7

Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J. L., Nishihama, R. and Kohchi, T. (2015). Auxin-mediated transcriptional system with a minimal set of components is critical for morphogenesis through the life cycle in marchantia polymorpha. *PLoS Genet.* **11**, e1005084. doi:10.1371/journal.pgen.1005084

Kato, H., Kouno, M., Takeda, M., Suzuki, H., Ishizaki, K., Nishihama, R. and Kohchi, T. (2017). The roles of the sole activator-type auxin response factor in pattern formation of marchantia polymorpha. *Plant Cell Physiol.* **58**, 1642-1651. doi:10.1093/pcp/pcx095

Kato, H., Nishihama, R., Weijers, D. and Kohchi, T. (2018). Evolution of nuclear auxin signaling: lessons from genetic studies with basal land plants. *J. Exp. Bot.* **69**, 291-301. doi:10.1093/jxb/erx267

Kato, H., Mutte, S. K., Suzuki, H., Crespo, I., Das, S., Radoeva, T., Fontana, M., Yoshitake, Y., Hainiwa, E., van den Berg, W. et al. (2020). Design principles of a minimal auxin response system. *Nat. Plants* **6**, 473-482. doi:10.1038/s41477-020-0662-y

Kepinski, S. and Leyser, O. (2005). The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature* **435**, 446-451. doi:10.1038/nature03542

Kim, J., Harter, K. and Theologis, A. (1997). Protein-protein interactions among the Aux/IAA proteins. *Proc. Natl. Acad. Sci. USA* **94**, 11786-11791. doi:10.1073/pnas.94.22.11786

Kuhn, A., Ramans Harborough, S., McLaughlin, H. M., Natarajan, B., Verstraeten, I., Friml, J., Kepinski, S. and Østergaard, L. (2020). Direct ETTIN-auxin interaction controls chromatin states in gynoecium development. *eLife* **9**, e51787. doi:10.7554/eLife.51787

Lanctot, A., Taylor-Teeples, M., Oki, E. A. and Nemhauser, J. L. (2020). Specificity in auxin responses is not explained by the promoter preferences of activator ARFs. *Plant Physiol.* **182**, 1533-1536. doi:10.1104/pp.19.01474

Lavy, M. and Estelle, M. (2016). Mechanisms of auxin signaling. *Development* **143**, 3226-3229. doi:10.1242/dev.131870

Lavy, M., Prigge, M. J., Tao, S., Shain, S., Kuo, A., Kirchsteiger, K. and Estelle, M. (2016). Constitutive auxin response in Physcomitrella reveals complex interactions between Aux/IAA and ARF proteins. *eLife* **5**, e13325. doi:10.7554/eLife.13325

Leyser, O. (2018). Auxin Signaling. *Plant Physiol.* **176**, 465-479. doi:10.1104/pp.17.00765

Martin-Arevalillo, R., Thévenon, E., Jégu, F., Vinos-Poyo, T., Vernoux, T., Parcy, F. and Dumas, R. (2019). Evolution of the auxin response factors from charophyte ancestors. *PLoS Genet.*. **15**, e1008400. doi:10.1371/journal.pgen.1008400

Matthes, M. S., Best, N. B., Robil, J. M., Malcomber, S., Gallavotti, A. and McSteen, P. (2019). Auxin EvoDevo: conservation and diversification of genes regulating auxin biosynthesis, transport, and signaling. *Mol. Plant* **12**, 298-320. doi:10.1016/j.molp.2018.12.012

Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C. H., Yang, Z., Schneider, H. and Donoghue, P. C. J. (2018). The timescale of early land plant evolution. *Proc. Natl. Acad. Sci. USA* **115**, E2274-E2283. doi:10.1073/pnas.1719588115

Mutte, S. K., Kato, H., Rothfels, C., Melkonian, M., Wong, G. K.-S. and Weijers, D. (2018). Origin and evolution of the nuclear auxin response system. *eLife* **7**, e33399. doi:10.7554/eLife.33399

Niemeyer, M., Moreno Castillo, E., Ihling, C. H., Iacobucci, C., Wilde, V., Hellmuth, A., Hoehenwarter, W., Samodelov, S. L., Zurbriggen, M. D., Kastritis, P. L. et al. (2020). Flexibility of intrinsically disordered degrons in AUX/IAA proteins reinforces auxin co-receptor assemblies. *Nat. Commun.* **11**, 2277. doi:10.1038/s41467-020-16147-2

Ohtaka, K., Hori, K., Kanno, Y., Seo, M. and Ohta, H. (2017). Primitive auxin response without TIR1 and Aux/IAA in the charophyte alga klebsormidium nitens. *Plant Physiol.* **174**, 1621-1632. doi:10.1104/pp.17.00274

Paponov, I. A., Teale, W., Lang, D., Paponov, M., Reski, R., Rensing, S. A. and Palme, K. (2009). The evolution of nuclear auxin signalling. *BMC Evol. Biol.* **9**, 126. doi:10.1186/1471-2148-9-126

Pierre-Jerome, E., Jang, S. S., Havens, K. A., Nemhauser, J. L. and Klavins, E. (2014). Recapitulation of the forward nuclear auxin response pathway in yeast. *Proc. Natl. Acad. Sci. USA* **111**, 9407-9412. doi:10.1073/pnas.1324147111

Pierre-Jerome, E., Moss, B. L., Lanctot, A., Hageman, A. and Nemhauser, J. L. (2016). Functional analysis of molecular interactions in synthetic auxin response circuits. *Proc. Natl Acad. Sci. USA* **113**, 11354-11359. doi:10.1073/pnas.1604379113

Piya, S., Shrestha, S. K., Binder, B., Stewart, C. N., Jr. and Hewezi, T. (2014). Protein-protein interaction and gene co-expression maps of ARFs and Aux/IAAs in Arabidopsis. *Front. Plant Sci.* **5**, 744. doi:10.3389/fpls.2014.00744

Prigge, M. J., Platre, M., Kadakia, N., Zhang, Y., Greenham, K., Szutu, W., Pandey, B. K., Bhosale, R. A., Bennett, M. J., Busch, W. et al. (2020). Genetic analysis of the Arabidopsis TIR1/AFB auxin receptors reveals both overlapping and specialized functions. *eLife* **9**, e54740. doi:10.7554/eLife.54740

Ramos Báez, R., Buckley, Y., Yu, H., Chen, Z., Gallavotti, A., Nemhauser, J. L. and Moss, B. L. (2020). A synthetic approach allows rapid characterization of the maize nuclear auxin response circuit. *Plant Physiol.* **182**, 1713-1722. doi:10.1104/pp.19.01475

Shih, H.-W., DePew, C. L., Miller, N. D. and Monshausen, G. B. (2015). The cyclic nucleotide-gated channel CNGC14 regulates root gravitropism in arabidopsis thaliana. *Curr. Biol.* **25**, 3119-3125. doi:10.1016/j.cub.2015.10.025

Simonini, S., Deb, J., Moubayidin, L., Stephenson, P., Valluru, M., Freire-Rios, A., Sorefan, K., Weijers, D., Friml, J. and Østergaard, L. (2016). A noncanonical auxin-sensing mechanism is required for organ morphogenesis in Arabidopsis. *Genes Dev.* **30**, 2286-2296. doi:10.1101/gad.285361.116

Szemenyei, H., Hannon, M. and Long, J. A. (2008). TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis. *Science* **319**, 1384-1386. doi:10.1126/science.1151461

Tao, S. and Estelle, M. (2018). Mutational studies of the Aux/IAA proteins in Physcomitrella reveal novel insights into their function. *New Phytol.* **218**, 1534-1542. doi:10.1111/nph.15039

Truskina, J., Han, J., Chrysanthou, E., Galvan-Ampudia, C. S., Lainé, S., Brunoud, G., Macé, J., Bellows, S., Legrand, J., Bågman, A.-M. et al. (2021). A network of transcriptional repressors modulates auxin responses. *Nature* **589**, E7. doi:10.1038/s41586-020-03066-x

Uchida, N., Takahashi, K., Iwasaki, R., Yamada, R., Yoshimura, M., Endo, T. A., Kimura, S., Zhang, H., Nomoto, M., Tada, Y. et al. (2018). Chemical hijacking of auxin signaling with an engineered auxin–TIR1 pair. *Nat. Chem. Biol.* **14**, 299-305. doi:10.1038/nchembio.2555

Ulmasov, T., Hagen, G. and Guilfoyle, T. J. (1997). ARF1, a transcription factor that binds to auxin response elements. *Science* **276**, 1865-1868. doi:10.1126/science.276.5320.1865

Vernoux, T., Brunoud, G., Farcot, E., Morin, V., Van den Daele, H., Legrand, J., Oliva, M., Das, P., Larrieu, A., Wells, D. et al. (2011). The auxin signalling network translates dynamic input into robust patterning at the shoot apex. *Mol. Syst. Biol.* **7**, 508. doi:10.1038/msb.2011.39

Wang, R. and Estelle, M. (2014). Diversity and specificity: auxin perception and signaling through the TIR1/AFB pathway. *Curr. Opin. Plant Biol.* **21**, 51-58. doi:10.1016/j.pbi.2014.06.006

Wang, C., Liu, Y., Li, S.-S. and Han, G.-Z. (2015). Insights into the origin and evolution of the plant hormone signaling machinery. *Plant Physiol.* **167**, 872-886. doi:10.1104/pp.114.247403

Weigel, D. and Mott, R. (2009). The 1001 genomes project for Arabidopsis thaliana. *Genome Biol.* **10**, 107. doi:10.1186/gb-2009-10-5-107

Winkler, M., Niemeyer, M., Hellmuth, A., Janitza, P., Christ, G., Samodelov, S. L., Wilde, V., Majovsky, P., Trujillo, M., Zurbriggen, M. D. et al. (2017). Variation in auxin sensing guides AUX/IAA transcriptional repressor ubiquitylation and destruction. *Nat. Commun.* **8**, 15706. doi:10.1038/ncomms15706

Woodward, A. W. and Bartel, B. (2005). Auxin: regulation, action, and interaction. *Ann. Bot.* **95**, 707-735. doi:10.1093/aob/mci083

Wright, R. C., Zahler, M. L., Gerben, S. R. and Nemhauser, J. L. (2017). Insights into the evolution and function of auxin signaling F-Box proteins in arabidopsis thaliana through synthetic analysis of natural variants. *Genetics* **207**, 583-591. doi:10.1534/genetics.117.300092

Yu, H., Zhang, Y., Moss, B. L., Bargmann, B. O. R., Wang, R., Prigge, M., Nemhauser, J. L. and Estelle, M. (2015). Untethering the TIR1 auxin receptor from the SCF complex increases its stability and inhibits auxin response. *Nat Plants* **1**, 14030. doi:10.1038/nplants.2014.30

CHAPTER 2 PREAMBLE

In this work, Clay Wright and Morgan Hamm developed ViVa (Visualizing Variation), a tool that allows for biologists without expertise in computational biology and data processing to make use of the 1001 Accessions *Arabidopsis thaliana* genomic data. My primary contribution to this work was as a graduate tool tester where I used the ViVa tool to pull 1001 Accessions data on the natural variation found within the coding sequences of the *Arabidopsis TIR1/AFBs*. You can access this work in section "Natural variation in the *TIR1/AFB* genes," pages 12 and 13 of this publication. In this section, I reviewed the literature to compile what was known about the fundamental functions of the TIR1/AFBs in auxin signaling, and more specifically known differences in function between different *Arabidopsis* TIR1/AFBs. I created hypotheses about what types of genetic variation I should see in different genes based on whether they had been characterized as having important roles in auxin signaling, what domains were characterized as most important to function, and more. I then used the tool to align and visualize *Arabidopsis* population-level variation in these genes, explaining what differences in variant accumulation and type among different genes and across different domains told us about the functions of these proteins. Finally, I explained whether my initial hypotheses about gene function were supported by variant data and created new hypotheses about the differences in functions between these genes and their evolutionary trajectory within *Arabidopsis.*

To summarize my findings, I observed low variation of auxin and degron-binding regions in all TIR1/AFBs tested, supporting the hypotheses that these regions were critical to gene function. I also observed low variability and low π_N/π_S ratios in genes previously reported to be the most important in auxin signaling, supporting their importance. I surprisingly found AFB1 to be among these well-conserved genes, which at the time was known to be unable to preform a transcriptional canonical receptor function. Today we know AFB1 to have an important role in a non-transcriptional rapid auxin response, supporting this new hypothesis that this gene also had an important role in *Arabidopsis* auxin responses.

Received: 6 December 2018 | Revised: 20 April 2019 | Accepted: 29 April 2019

DOI: 10.1002/pld3.147

ORIGINAL RESEARCH

Accelerating structure-function mapping using the ViVa webtool to mine natural variation

Morgan O. Hamm[1] | Britney L. Moss[2] | Alexander R. Leydon[1] | Hardik P. Gala[1] | Amy Lanctot[1] | Román Ramos[1] | Hannah Klaeser[2] | Andrew C. Lemmex[1] | Mollye L. Zahler[1] | Jennifer L. Nemhauser[1] | R. Clay Wright[3]

[1]Department of Biology, University of Washington, Seattle, Washington

[2]Department of Biology, Whitman College, Walla Walla, Washington

[3]Biological Systems Engineering, Virginia Tech, Blacksburg, Virginia

Correspondence
R. Clay Wright, Biological Systems Engineering, Virginia Tech, Blacksburg, VA.
Email: wrightrc@vt.edu

Funding information
National Institute of Health, Grant/Award Number: R01-GM107084; National Science Foundation, Grant/Award Number: IOS-1546873 and DBI-1402222; Howard Hughes Medical Institute; M.J. Murdock Charitable Trust; Simons Foundation Fellow of the Life Sciences Research Foundation; NSF Graduate Research Fellowship, Grant/Award Number: DGE-1256082

Abstract

Thousands of sequenced genomes are now publicly available capturing a significant amount of natural variation within plant species; yet, much of these data remain inaccessible to researchers without significant bioinformatics experience. Here, we present a webtool called ViVa (Visualizing Variation) which aims to empower any researcher to take advantage of the amazing genetic resource collected in the *Arabidopsis thaliana* 1001 Genomes Project (http://1001genomes.org). ViVa facilitates data mining on the gene, gene family, or gene network level. To test the utility and accessibility of ViVa, we assembled a team with a range of expertise within biology and bioinformatics to analyze the natural variation within the well-studied nuclear auxin signaling pathway. Our analysis has provided further confirmation of existing knowledge and has also helped generate new hypotheses regarding this well-studied pathway. These results highlight how natural variation could be used to generate and test hypotheses about less-studied gene families and networks, especially when paired with biochemical and genetic characterization. ViVa is also readily extensible to databases of interspecific genetic variation in plants as well as other organisms, such as the 3,000 Rice Genomes Project (http://snp-seek.irri.org/) and human genetic variation (https://www.ncbi.nlm.nih.gov/clinvar/).

KEYWORDS

accessibility, *Arabidopsis thaliana*, genome diversity, genotype-phenotype, natural variation, structure-function

1 | INTRODUCTION

The sequencing of the first *Arabidopsis thaliana* genome ushered in a new era of tool development and systematic functional annotation of plant genes (The Arabidopsis Genome Initiative 2000). Since that landmark effort, massive scaling of sequencing technology has allowed for the survey of genomic variation in natural *A. thaliana* populations (Borevitz et al., 2007; Nordborg et al., 2005; Weigel & Mott, 2009).

This manuscript was previously deposited as a BioRxiv preprint at https://doi.org/10.1101/488395

This valuable population genetics resource has led to several associations of genetic loci with phenotypic traits and provided insights into how selective pressure has influenced the evolution of plant genomes (Atwell et al., 2010; Clark et al., 2007; Long et al., 2013).

Beyond its utility in gene discovery and understanding genome evolution, natural genetic variation provides a catalog of permissible polymorphisms that can facilitate the connection of genotype to phenotype at the gene, gene family, and network scales (Joly-Lopez, Flowers, & Purugganan, 2016; Nieduszynski & Liti, 2011). This is an especially critical resource for studying large gene families where loss of function in individual genes may have little or no phenotypic effect (Dharmasiri et al., 2005; Guo, 2013; Moore & Purugganan, 2005) and directed allele replacement remains time and resource-intensive (Chen, Wang, Zhang, Zhang, & Gao, 2019). Natural variation datasets provide novel alleles and germplasm which can be examined with biochemical and genetic approaches to map sequence to function and genotype to phenotype. In human clinical medicine, massively parallel assays of variant effects stand to revolutionize genetic diagnostics and personalized medicine (Gasperini, Starita, & Shendure, 2016; Matreyek, Stephany, & Fowler, 2017; Starita et al., 2017). Similarly, we envision the use of plant natural variation datasets as a tool to revolutionize the breeding and genetic engineering of crop plants by rapidly advancing our understanding of genotype/function/phenotype relationships. A proof-of-principle survey of a relatively small subset of natural variants paired with a synthetic assay of gene function successfully mapped critical functional domains of auxin receptors and identified new alleles which affect plant phenotype (Wright, Zahler, Gerben, & Nemhauser, 2017).

Why is the survey of natural variants not as routine as a BLAST search or ordering T-DNA insertion mutants? One reason may be the current requirement for a fairly high level of bioinformatics expertise to extract the desired information from whole-genome resequencing datasets. While existing resources such as the 1001 Proteomes website (Joshi et al., 2012) and ePlant (Waese et al., 2017) facilitate access to these data at the gene scale, they do not provide comparative summaries or visualizations of variation at the gene family scale. To address this concern, we created ViVa: a webtool and Rpackage for Visualizing Variation, which allows plant molecular biologists of any level access to gene-level data from the 1001 Genomes database. Using ViVa, researchers may: (a) Identify polymorphisms to facilitate biochemical assays of variant effects (Starita et al., 2017; Wright et al., 2017); (b) produce family-wise alignments of variants to facilitate de novo functional domain identification (Melamed, Young, Miller, & Fields, 2015); (c) generate lists of accessions containing polymorphisms to facilitate phenotypic analysis of gene variant effects (Park et al., 2017); and (d) quantify metrics of genetic diversity to facilitate the study of gene, gene family, and network evolution (Delker et al., 2010; Kliebenstein, 2008). Here we present a summary of the functionality of ViVa and an analysis of the natural variation in the nuclear auxin signaling network using ViVa. To succinctly demonstrate the use of ViVa, we focus on the analysis of the *Aux/IAA* family; similar analyses were performed for the other nuclear auxin signaling gene families and are provided as Section 5 for the interested reader.

2 | METHODS

2.1 | Data sources

2.1.1 | Variant data

Variant data were queried from the 1001 genomes project (http://1001genomes.org) via URL requests to their API service (http://tools.1001genomes.org/api/index.html). These queries returned subsets of the whole-genome variant call format (VCF) file as SnpEFF VCF files. The whole-genome VCF file can be found on the project's website at http://1001genomes.org/data/GMI-MPI/releases/v3.1/.

2.1.2 | Germplasm accession information

A dataset of each of the 1,135 accessions including CS stock numbers and geographic locations where the samples were collected was retrieved from the 1001 Genomes website at http://1001genomes.org/accessions.html, via the download link at the bottom of the page. These data file have been embedded in the R package as accessions.

2.1.3 | Gene and transcript accession information

Information on the genes and transcripts including chromosomal coordinates, start and end location, and transcript length were downloaded from Araport11 as a general feature format (GFF) file (Cheng et al., 2017). The Araport11 full genome general feature format file, which can also be found on the TAIR website (https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FGenes%2FAraport11_genome_release), has been embedded in the R package as GRanges object, gr. The TAIR10 database, found at http://arabidopsis.org, was accessed via the biomart function, using the biomaRt R package. Gene identifiers used in this study are in Table 1.

2.2 | Ranking of variant functional effects

Alignments were colored according to the strongest effect variant allele occurring at any frequency at that position as reported in the SnpEFF "effect" field, per the scale in Figure 1.

2.3 | Nucleotide diversity calculation

Nei and Li defined the nucleotide diversity statistic in their original paper as: "the average number of nucleotide differences per site between two randomly chosen DNA sequences" (Nei & Li, 1979), and provided the equation:

$$\pi = \sum_{ij} x_i x_j \pi_{ij}. \quad (1)$$

TABLE 1 Full list of genes used in this study, by identifier, symbol, and classification

Arabidopsis AGI locus identifier	Gene symbol	Clade/class
AT3G62980	TIR1	NA
AT4G03190	AFB1	NA
AT3G26810	AFB2	NA
AT1G12820	AFB3	NA
AT4G24390	AFB4	NA
AT5G49980	AFB5	NA
AT2G39940	COI1	NA
AT4G14560	IAA1	A
AT3G23030	IAA2	A
AT1G04240	IAA3	A
AT5G43700	IAA4	A
AT1G15580	IAA5	A
AT1G52830	IAA6	A
AT3G23050	IAA7	A
AT2G22670	IAA8	A
AT5G65670	IAA9	A
AT1G04100	IAA10	B
AT4G28640	IAA11	B
AT1G04550	IAA12	B
AT2G33310	IAA13	B
AT4G14550	IAA14	A
AT1G80390	IAA15	A
AT3G04730	IAA16	A
AT1G04250	IAA17	A
AT1G51950	IAA18	B
AT3G15540	IAA19	A
AT2G46990	IAA20	C
AT3G16500	IAA26	B
AT4G29080	IAA27	A
AT5G25890	IAA28	B
AT4G32280	IAA29	C
AT3G62100	IAA30	C
AT3G17600	IAA31	C
AT2G01200	IAA32	C
AT5G57420	IAA33	C
AT1G15050	IAA34	C
AT1G15750	TPL	NA
AT1G80490	TPR1	NA
AT3G16830	TPR2	NA
AT5G27030	TPR3	NA
AT3G15880	TPR4	NA
AT1G59750	ARF1	B
AT5G62000	ARF2	B
AT2G33860	ARF3	B
AT5G60450	ARF4	B
AT1G19850	ARF5	A
AT1G30330	ARF6	A
AT5G20730	ARF7	A
AT5G37020	ARF8	A
AT4G23980	ARF9	B
AT2G28350	ARF10	C
AT2G46530	ARF11	B
AT1G34310	ARF12	B
AT1G34170	ARF13	B
AT1G35540	ARF14	B
AT1G35520	ARF15	B
AT4G30080	ARF16	C
AT1G77850	ARF17	C
AT3G61830	ARF18	B
AT1G19220	ARF19	A
AT1G35240	ARF20	B
AT1G34410	ARF21	B
AT1G34390	ARF22	B
AT1G43950	ARF23	B

(Continues)

TABLE 1 (Continued)

where x_i is the frequency of the ith sequence in the population and π_{ij} is the number of sites that are different between the ith and jth sequence divided by sequence length.

A more general form that treats each sequence in the population as unique can be written as follows:

$$\pi = \frac{1}{L * n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{L} \pi_{i,j,k} \quad \pi_{ijk} = \begin{cases} 1 \text{ if } N_{ik} \neq N_{jk} \\ 0 \text{ if } N_{ik} = N_{jk} \end{cases} \tag{2}$$

where N_{ik} is the nucleotide (A, T, C, or G) at position k on the ith sequence of the population. L is the length of the sequence. Indels are excluded from the diversity calculation leading to a single L for the population. n is the total number of sequences in the population.

From this form, we can re-arrange summations to the form the below equation:

$$\pi = \frac{1}{L} \sum_{k=1}^{L} \pi_k \quad \pi_k = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_{ijk} \tag{3}$$

where π_k can be thought of as the site-wise nucleotide diversity at position k, and is equal to the nucleotide diversity of a sequence of length 1 at location k. We can calculate π_k for each site, then average those over the sequence length to calculate π, the nucleotide diversity of the sequence.

The function Nucleotide_diversity in the r1001genomes package calculates π_k for each position in the gene or region that contains a variant. Note, π_k is equal to 0 at all locations without variants. This is also what is displayed in the Diversity Plot tab of the webtool.

2.4 | Detailed π_k calculation simplification

The formula for π_k above requires comparing every sequence to every other sequence at location *k*; however, we know there are only a few variant forms at each individual location.

So, we can revert back to using Nei and Li's original formula 1, modifying it slightly, replacing x_i with $\frac{n_i}{n}$, n_i being the number of sequences in the population with nucleotide N_i at location *k*:

$$\pi_k = \sum_{ij} \frac{n_i}{n}\frac{n_j}{n}\pi_{ij} = \frac{1}{n^2}\sum_{ij} n_i n_j \pi_{ij} \quad \pi_{ij(k)} = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{4}$$

Note that in Equation 1, subscripts *i* and *j* are summed over all sequences in the population; however, in Equation 4*i* and *j* are only summed over unique variants at a particular location *k*.

We will define $n_{!j} = n - n_i$ as the number of sequences different from *i* at position *k*. We can also see that the summed term will be zero if $i = j$, and $n_i n_j$ if $i \neq j$. Therefore:

$$\pi_k = \frac{1}{n^2}\sum_i n_i n_{!i} \tag{5}$$

Next we substitute our definition of $n_{!i}$:

$$\pi_k = \frac{1}{n^2}\sum_i n_i(n - n_i) \tag{6}$$

Distributing and splitting summation yields:

$$\pi_k = \frac{1}{n^2}\left(n\sum_i n_i - \sum_i n_i^2\right) \tag{7}$$

Finally, summing $\Sigma_i n_i$ is equal to *n*:

$$\pi_k = \frac{1}{n^2}\left(n^2 - \sum_i n_i^2\right) \tag{8}$$

This simplified form for π_k is used by the app, because the counts of unique variants at a single nucleotide location can easily be summarized in R.

2.5 | Software

The r1001genomes package has many software dependencies on other R packages, a few of the key bioinformatics packages used are listed below.

biomaRt: used for accessing the TAIR10 database on arabidopsis.org.

vcfR: used to read in the VCF files in a flat "tidy" format for easy manipulation.

BSgenome: used as the source for the complete DNA string of the reference genome (Col-0).

DECIPHER: used to align nucleotide and amino acid sequences of homologous genes.

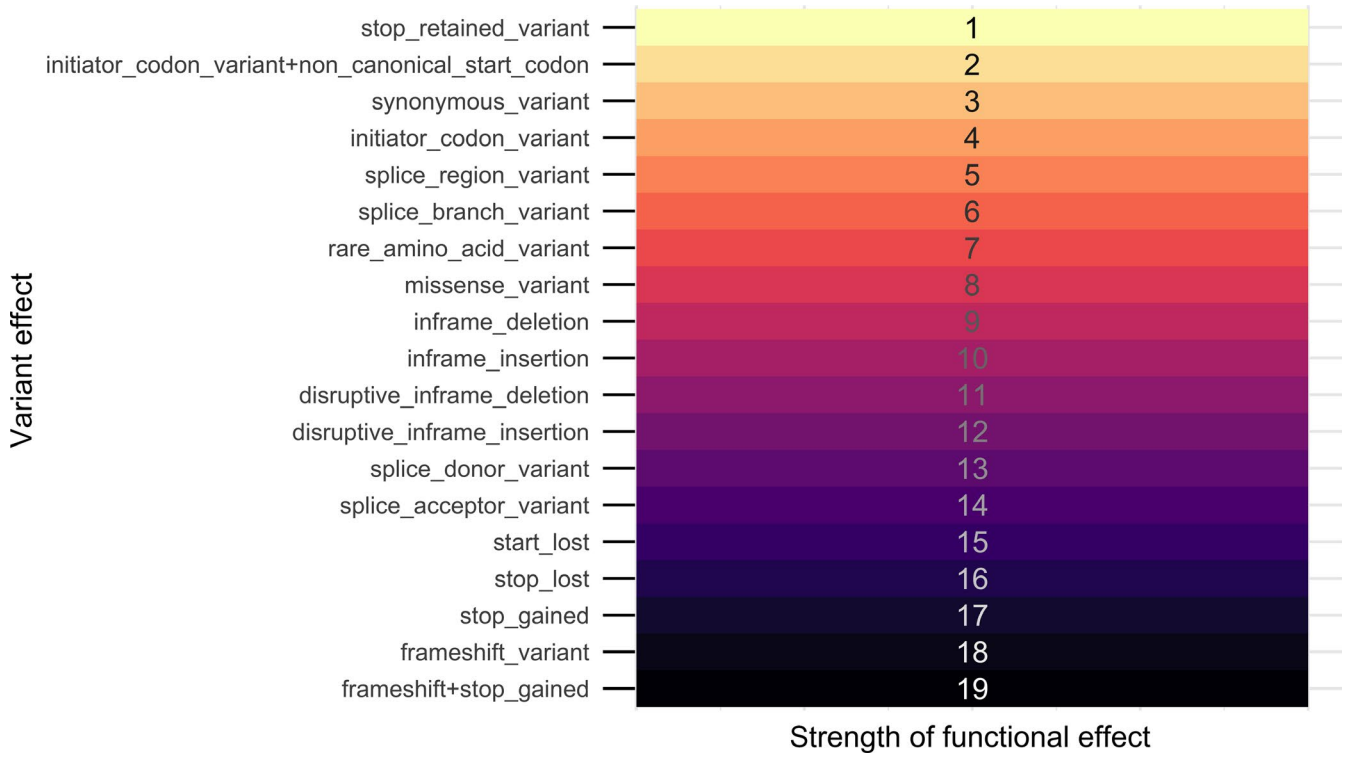


FIGURE 1 Rank order of the strength of functional effects variant effect classes were ordered by subjective prediction of average strength of effect on gene function. Strength was then assigned to each effect on an integer scale

GenomicFeatures: used for handling sequence annotations.

Biostrings: provides the underlying framework for the sequence manipulations used for generating and aligning sequences with BSgenome, Decipher, and GenomicFeatures.

Other packages that were critical to building ViVa and/or writing this document include the following: (Allaire, Ushey, & Tang, 2018; Allaire et al., 2018; Aphalo, 2018a,2018b; Arnold, 2018; Bache & Wickham, 2014; Garnier, 2018a,2018b; Hamm & Wright, 2018; Heibl, 2014; Henry & Wickham, 2018; Ihaka et al., 2019; Lang, 2018; Müller, 2018; Müller & Wickham, 2019; Müller, Wickham, James, & Falcon, 2018; Neuwirth, 2014; Pagès & Aboyoun, 2018; Pagès, Aboyoun, Gentleman, & DebRoy, 2019; Pagès, Aboyoun, & Lawrence, 2018; Pagès, Lawrence, & Aboyoun, 2018; Paradis et al., 2018; R Core Team 2018; Team 2018; Wagih, 2017; Wickham, 2016, 2017a,2017b, 2018a,2018b,2018c; Wickham, François, Henry, & Müller, 2018; Wickham & Henry, 2018; Wickham, Hester, & Francois, 2018; Wickham et al., 2018; Wright, 2019; Xie, 2018a,2018b; Xie, Cheng, & Tan, 2018; Yu, 2018; Yu & Lam, 2019).

2.6 | Testing

A group consisting of undergraduate and graduate students, postdoctoral, and postbaccalaureate researchers (the authors of this publication) were assembled to test the functionality of ViVa. Testers were asked to use the ViVa web interface to analyze the genetic variation at the clade or whole-gene family level of the auxin nuclear signaling pathway. Testers were provided with a brief vignette on how to use ViVa to formulate new hypotheses or support existing hypotheses about gene function and evolution, which has been expanded to the "An Overview of ViVa" section. The testers had background knowledge of these genes as members of laboratories studying auxin nuclear signaling. Testers met weekly with the developers to discuss their user experience and their findings. Tester experiences, issues, and suggestions were incorporated into the ViVa software. The results collected by these analyses are summarized in the "Visualizing Variation within the auxin signaling pathway" section and Section 5.

3 | RESULTS AND DISCUSSION

3.1 | An overview of ViVa

ViVa, in this first iteration, is meant to visualize natural variation in the coding sequences of genes or gene families. Noncoding sequence variation is intentionally excluded from the analysis tools. This reflects the challenges both in alignment of noncoding sequences and the increased difficulty in assessing the effects of variation in these regions (Alexandre et al., 2018).

The stable version of ViVa is hosted at https://www.plantsynbiolab.bse.vt.edu/ViVa/. The development version of ViVa can be accessed as a Docker container https://hub.docker.com/r/wrightrc/r1001genomes/or as an R package at https://github.com/wrightrc/r1001genomes.

3.2 | Gene select and annotation files

At the top of the ViVa webtool are two collapsible panels used for entering the genes to query and custom annotations for those genes (Figures 2a and 3a). The Gene Select panel permits gene input by either typing in or uploading a ".csv" file of AGI/TAIR locus identifiers. The Annotation Files panel is optionally used to upload an annotation file containing coordinates of domains, mutations, or any other sequence knowledge that can be plotted on some of the tabs of the ViVa analysis tabs.

Below the data input section, the rest of the webtool is divided into several tabs containing interactive output.

3.3 | SNP stats: Summary of gene information, structure, and diversity

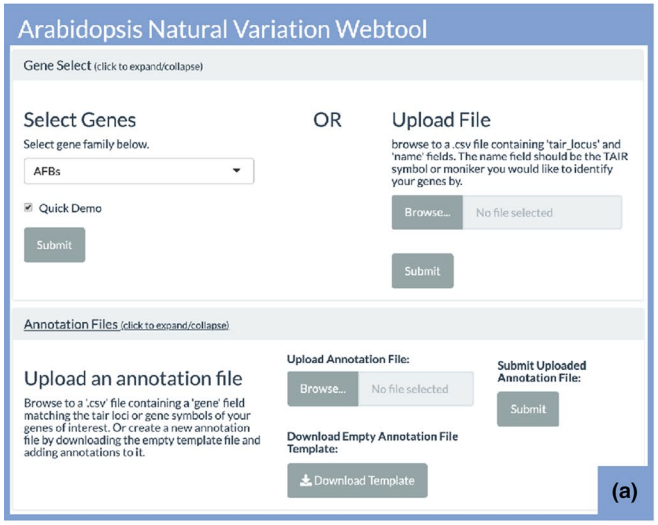
The SNP Stats tab provides general information on the gene transcripts being queried, as well as calculated counts/statistics on the content of variants found in the sample population (Figure 2b). The first table of this section is the basic information about the transcripts, including TAIR locus and symbol, the chromosomal start and end position, and the transcript length. This information was collected from the Araport11 Official Release (06/2016) annotation dataset (Cheng et al., 2017).

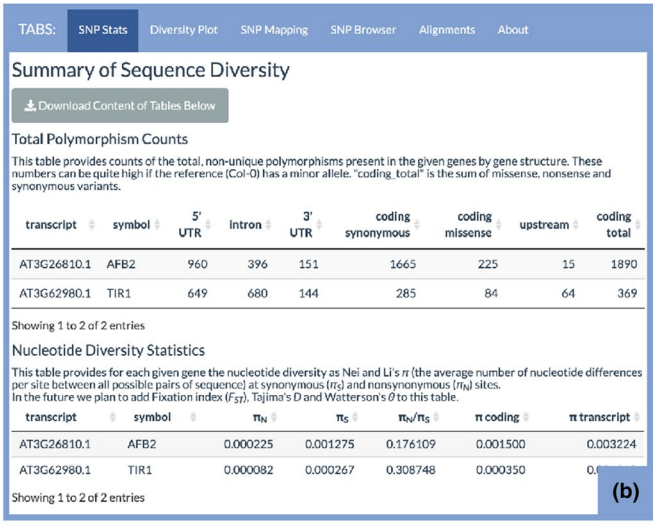
The next two tables provide counts of SNPs across the gene body for each transcript. The Total Polymorphism Counts tab provides the total number of variant observations of nonreference alleles categorized by type and location (the Col-0 accession TAIR9 genome is the reference genome for this dataset). The Unique Allele Count tab only counts the number of unique variant alleles within the population of accessions (e.g., if multiple accessions have the same variant, these are counted as a single allele at that position).

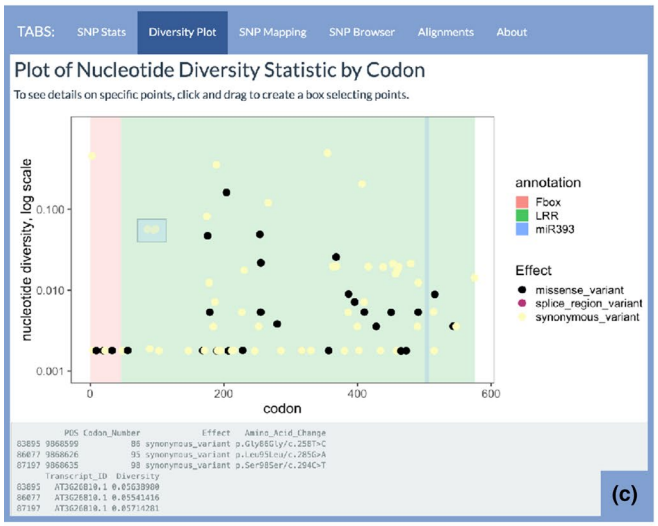
The Nucleotide Diversity Statistics tab provides nucleotide diversity statistic (π) values for the transcript and the coding sequence of each gene (Nei & Li, 1979). Given a set of nucleotide sequences from a population, π is the average number of nucleotide differences per site. Nucleotide diversity is also calculated for the set of only synonymous (π_S) and only nonsynonymous polymorphisms (π_N). As nonsynonymous polymorphisms are more likely to give rise to functional change than synonymous polymorphisms, the ratio of the presence of nonsynonymous to synonymous polymorphisms provides a measure of the potential for functional diversity (Firnberg & Ostermeier, 2013; Whitehead et al., 2012). We present π_N/π_S, the ratio of nonsynonymous to synonymous diversity, here as a correlate for functional diversity throughout ViVa (Hughes, Green, Garbayo, & Roberts, 2000; Nelson, Moncla, & Hughes, 2015). While imperfect, this metric may be suggestive of functional constraint when $\pi_N/\pi_S \gg 1$ (Hughes, 1999).

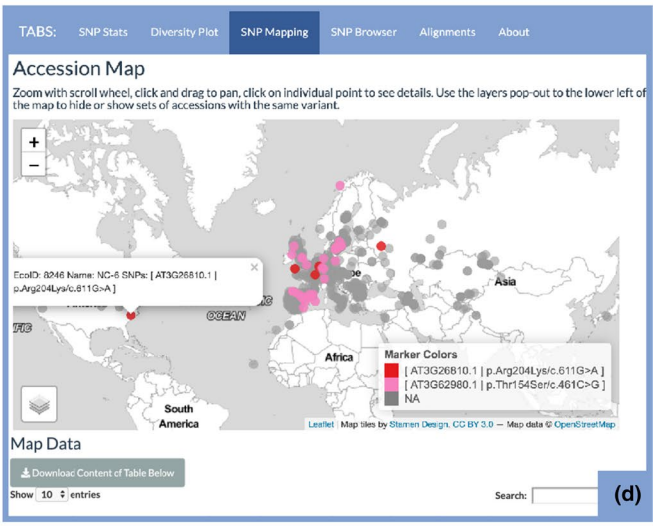
3.4 | Diversity Plot: Visualize allelic diversity across the coding sequence

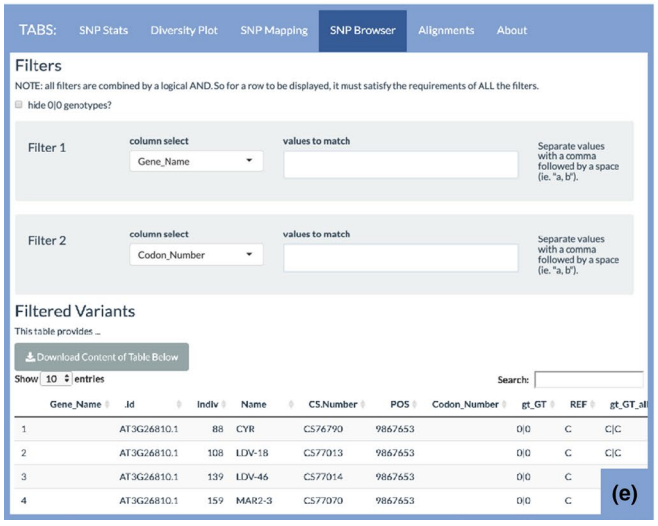
The Diversity Plot tab shows the nucleotide diversity of each variant in the coding region of a selected gene (Figures 2c and 4). Although











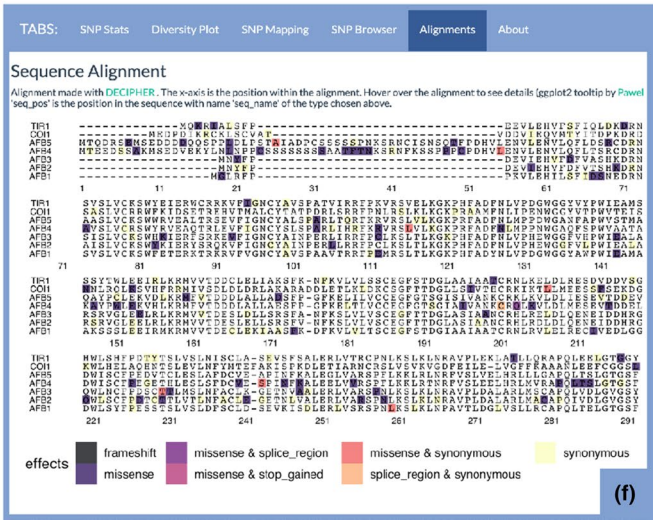


FIGURE 2 Key elements of the webtool (a) The first section contains two collapsible panels, gene select and annotation files, which are used to input information about the genes to be investigated. (b) The SNP Stats tab provides gene-structure level counts and statistics on SNPs. (c) The Diversity Plot tab plots the nucleotide diversity of SNP sites along the length of the coding region of a selected gene. (d) The SNP Mapping tab plots accessions on a world map colored according to the selected set of SNPs. (e) The SNP Browser tab allows variants and accessions to be filtered by any combination of text and numeric fields. (f) The Alignments tab aligns DNA and amino acid sequences of homologous genes and colors sequence elements based on SNPs and annotations

the *X*-axis is marked by codon number from the N-terminus for interpretability, the diversity values are based on single-nucleotide sites. The colors of markers on the plot identify the effect of the polymorphism. If annotation files are provided, the background of the plot is color-coded by the annotated regions. If points on the plot are selected by clicking and dragging a box over them, the data for the selected points appear in the grey box below the plot. Below these are the complete data table containing all points on the plot which can be downloaded as a ".csv" file. This tab allows users to identify regions of high diversity as well as isolate polymorphisms that may affect gene function and exist in multiple accessions, facilitating phenotypic analysis.

3.5 | SNP Mapping: View distributions of SNPs across the globe

The SNP Mapping tab plots the accessions' collection locations on a world map and colors the points based on selected variant alleles (Figure 2d). After the user selects genes and filters based on the SNP type and level of nucleotide diversity, a group of checkboxes becomes available to select variant alleles to display on the map (Figures 3b and 5). The variant alleles are labeled with the Transcript_ID and Amino_Acid_Change fields, in the form [Transcript_ID|Amino_Acid_Change]. After selecting the variant alleles and updating the map, the accessions are plotted on the map colored by each unique combination of the selected alleles. Below the map is a table containing the accession details for all mapped accessions. This tab may help users formulate hypotheses about the relatedness of accessions sharing a common allele and environments in which that allele may be favorable.

3.6 | SNP Browser: Filter and search for variants

The SNP Browser tab provides a way to search and filter the variant data by different fields (Figure 2e). After selecting the transcripts to include, a number of filters can be applied to the dataset to match text values (e.g., gene name or variant effect) or set minimum and maximum limits on the values of numeric fields (e.g., nucleotide diversity; 3c). When these filters are applied, the table below is updated to only contain rows meeting the criteria for all filters. This tab can be useful for identifying all accessions with a particular allele, or any non-reference alleles in a particular region of a gene that may not have been easily accessible in another tab.

3.7 | Alignments: Visualize SNPs on alignments of homologous genes

The Alignments tab provides DNA and amino acid sequence alignments of selected genes, colored according to the variant allele with the strongest functional effect at each position (Figures 2f and 6, for a full description of the color scale see Figure 1). The content of this tab is most useful if the selected genes are all family members or have significant sequence homology. If annotation files are uploaded, open boxes will be overlaid on the alignment, colored by annotation. Hovering the cursor over variants will provide additional details about the alleles present at that locus. This tab facilitates family-wise analysis of functional conservation, allowing users to identify potential functional regions and alleles which may be useful in deciphering this function.

3.8 | Gene Tree: Visualize functional diversity and sequence divergence of a gene family

The Gene Tree tab provides a neighbor joining tree (or uploaded tree created by the user) for the selected genes with the tips of the tree mapped with predicted functional diversity as represented by π_N/π_S in the 1001 Genomes dataset (Figure 7). This tab allows users to generate hypotheses regarding functional diversity and redundancy within the context of the predicted evolution of the gene family.

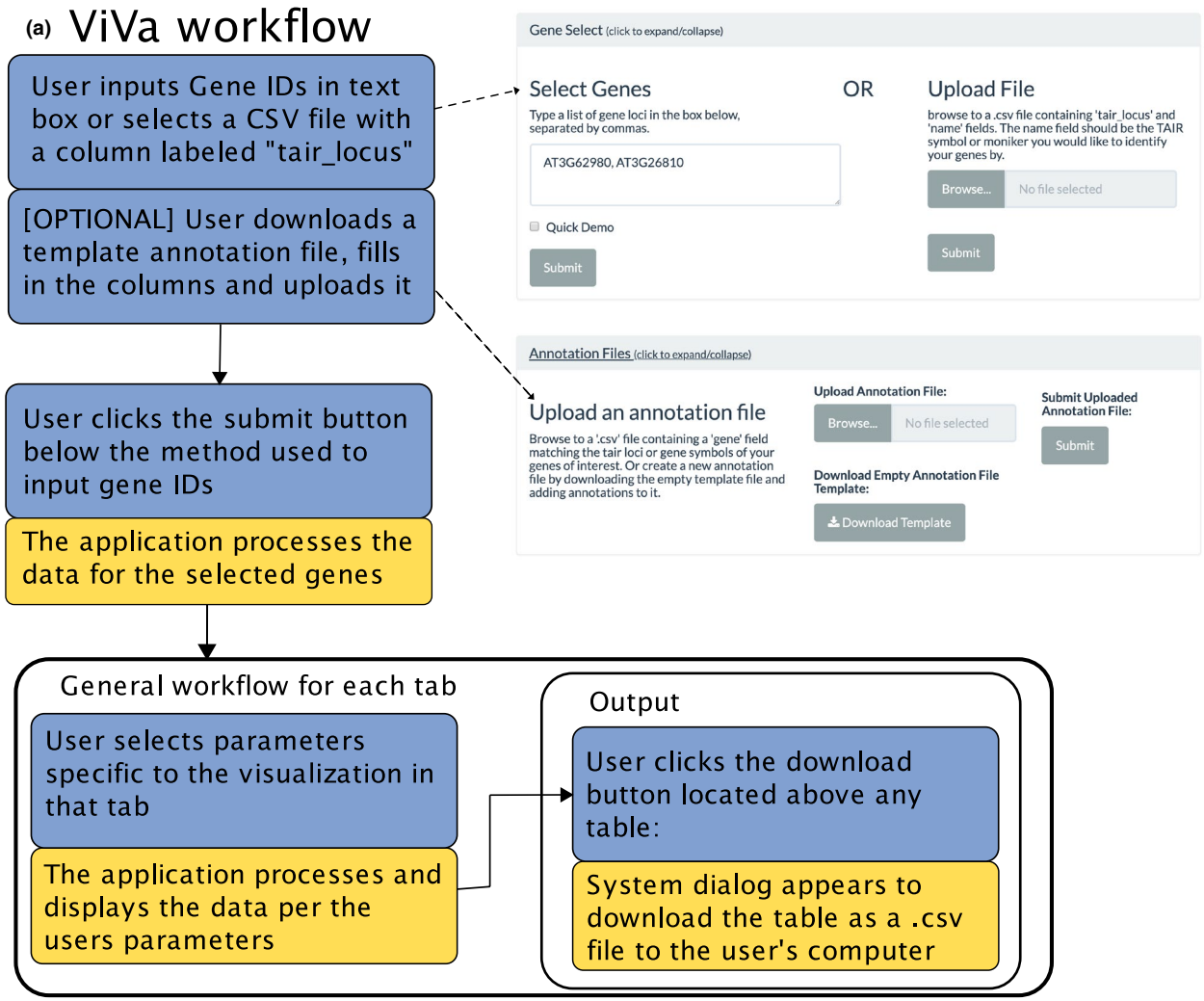
3.9 | ViVa R package: Programmatic access to ViVa's functionality

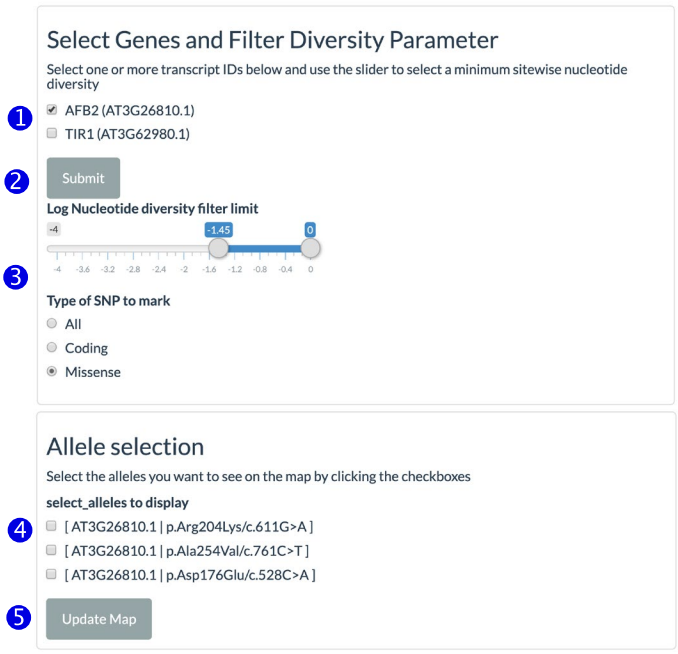
All of the functionalities of the ViVa webtool are implemented through functions within the ViVa R package. In addition to being able to generate the same sets of figures and tables as in the webtool, users of the R package also gain direct access to the underlying data structures, providing greater control over parameters when processing and visualizing the data. The ViVa R package is intended for users familiar with R programming who want to extend the capabilities of the webtool. The ViVa R package can be found at https://github.com/wrightrc/r1001genomes and can be installed in your R environment via the devtools package: remotes::install_github("wrightrc/r1001genomes").

3.10 | Visualizing Variation within the auxin signaling pathway

To test the usability and accessibility of ViVa, we assembled a group of alpha testers comprising postdoctoral, graduate, and undergraduate researchers at a research university (University of Washington) and at a primarily undergraduate institution (Whitman College). Our testers focused their investigation of natural variation on the nuclear auxin signaling pathway. We selected this signaling pathway for multiple reasons including a wealth of functional data and solved structures of several domains or entire proteins. Using this existing knowledge, we were able to qualitatively assess the predictive ability of the ViVa modules. Below, we describe the results for detailed analysis of the *Aux/IAA* family in more detail as a brief vignette of ViVa use. Analysis of each gene family in the nuclear auxin signaling network can be found in the Section 5.

The Aux/IAA family plays a critical role in transmitting auxin signals. Aux/IAA degradation is triggered by auxin accumulation (Zenser, Ellsmore, Leasure, & Callis, 2001) and is mediated by ubiquitination via a SKP-Cullin-F-box ubiquitin ligase complex containing an Auxin-signaling F-box (AFB) auxin receptor protein (Gray,


(a) ViVa workflow
User inputs Gene IDs in text box or selects a CSV file with a column labeled "tair_locus"
[OPTIONAL] User downloads a template annotation file, fills in the columns and uploads it
User clicks the submit button below the method used to input gene IDs
The application processes the data for the selected genes
Gene Select (click to expand/collapse)
Select Genes
Type a list of gene loci in the box below, separated by commas.
AT3G62980, AT3G26810
Quick Demo
Submit
OR
Upload File
browse to a .csv file containing 'tair_locus' and 'name' fields. The name field should be the TAIR symbol or moniker you would like to identify your genes by.
Browse...
No file selected
Submit
Annotation Files (click to expand/collapse)
Upload an annotation file
Browse to a '.csv' file containing a 'gene' field matching the tair loci or gene symbols of your genes of interest. Or create a new annotation file by downloading the empty template file and adding annotations to it.
Upload Annotation File:
Browse...
No file selected
Submit Uploaded Annotation File:
Submit
Download Empty Annotation File Template:
Download Template
General workflow for each tab
User selects parameters specific to the visualization in that tab
The application processes and displays the data per the users parameters
Output
User clicks the download button located above any table:
System dialog appears to download the table as a .csv file to the user's computer


(b) Mapping Tab parameters
Select Genes and Filter Diversity Parameter
Select one or more transcript IDs below and use the slider to select a minimum sitewise nucleotide diversity
1
AFB2 (AT3G26810.1)
TIR1 (AT3G62980.1)
2
Submit
Log Nucleotide diversity filter limit
3
Type of SNP to mark
All
Coding
Missense
Allele selection
Select the alleles you want to see on the map by clicking the checkboxes
select_alleles to display
4
[AT3G26810.1 | p.Arg204Lys/c.611G>A]
[AT3G26810.1 | p.Ala254Val/c.761C>T]
[AT3G26810.1 | p.Asp176Glu/c.528C>A]
5
Update Map

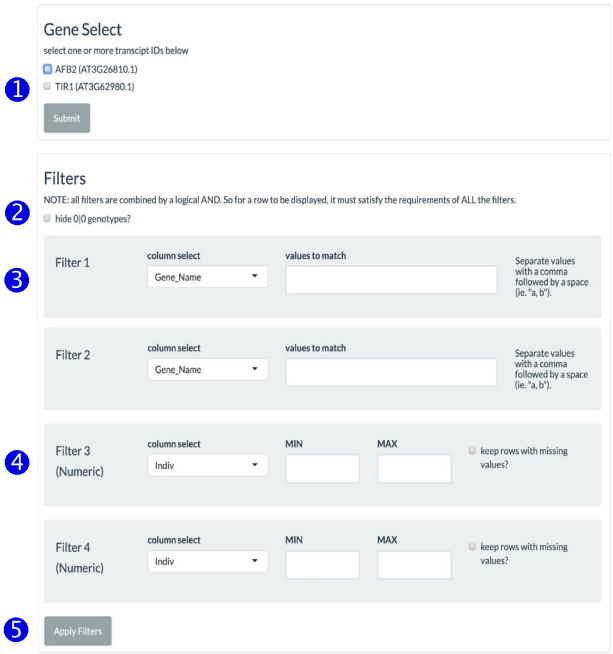
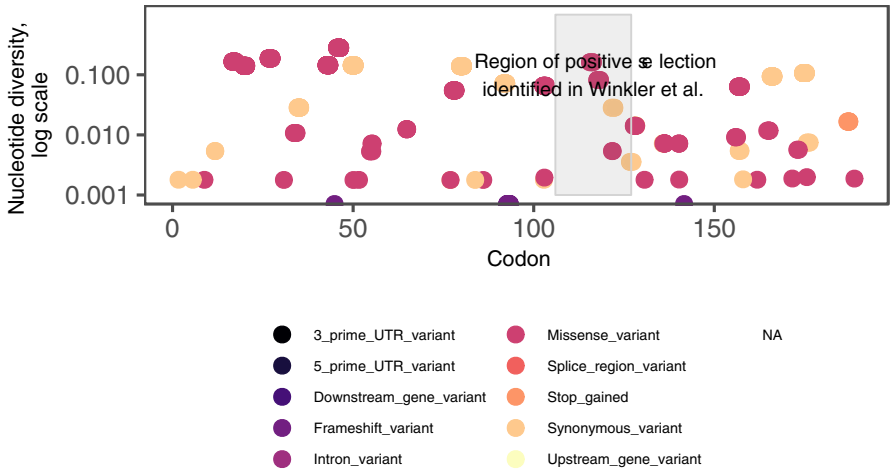

(c) Browser tab parameters
Gene Select
select one or more transcipt IDs below
1
AFB2 (AT3G26810.1)
TIR1 (AT3G62980.1)
Submit
Filters
NOTE: all filters are combined by a logical AND. So for a row to be displayed, it must satisfy the requirements of ALL the filters.
2
hide 0|0 genotypes?
3
Filter 1
column select
Gene_Name
values to match
Separate values with a comma followed by a space (ie. "a, b").
Filter 2
column select
Gene_Name
values to match
Separate values with a comma followed by a space (ie. "a, b").
4
Filter 3 (Numeric)
column select
Indiv
MIN
MAX
keep rows with missing values?
Filter 4 (Numeric)
column select
Indiv
MIN
MAX
keep rows with missing values?
5
Apply Filters

FIGURE 3 ViVa workflow (a) Workflow diagram of ViVa. Blue indicates user actions, yellow indicates processing steps performed by the application. (b) Detailed look at mapping tab parameters: User selects which genes to look at (1) then clicks the Submit button (2). The "Allele selection" panel is then filled in with all non-reference variants meeting the criteria. The user can adjust the range of nucleotide diversity and the type of SNP with a slider and radio buttons (3) to make the list of variants a manageable size. The list of variants is updated as changes are made to these controls. The user then selects variants to display on the map (4). Clicking the Update Map button (5) populates the map below with points located at the collection coordinates of each accession and colored by the selected alleles. (c) Detailed look at the browser tab options: The user first selects which genes to analyze and clicks submit (1). The "hide 0|0 genotype?" checkbox (2) removes rows from the table containing the reference allele. Four configurable filters of two types are provided. The first filter type is text matching (3); the user selects a column to filter from a drop-down menu then enters one or more text strings to match in that column. The second type of filter is numeric range matching (4); the user again selects a column, then specifies a range to match by typing or selecting minimum and maximum values. After configuring the filters, the user clicks the Apply Filters button (5) to update the table. The table is updated such that only rows that meet all filter conditions are displayed

FIGURE 4 IAA6 Diversity Plot. Nucleotide diversity of variant positions throughout the IAA6 coding sequence are plotted and colored according to the effect of the variant alleles at each position. The region of positive selection identified by Winkler et al. is highlighted



Kepinski, Rouse, Leyser, & Estelle, 2001). Variation in the *AFB* family is presented in Section 5 Figures 5, 8, 9). The Aux/IAAs repress transcription of Auxin Response Factor (ARF) bound genes (Tiwari, Hagen, & Guilfoyle, 2004), via recruitment of TOPLESS (TPL) and TOPLESS-related (TPR) co-repressors (Szemenyei, Hannon, & Long, 2008). Variation in the *TPL/TPR* family is presented in Section 5 Figures 10 and 11). Thus Aux/IAA degradation relieves this repression allowing ARFs to activate transcription of auxin response genes (Tiwari, Hagen, & Guilfoyle, 2003). Variation in the *ARF* family is presented in Section 5 Figures 12 and 13).

To predict the functional impact of variation in gene coding sequences, ViVa uses the frequency of nonsynonymous and synonymous polymorphisms. In most cases, nonsynonymous polymorphisms in critical functional domains are likely to have deleterious effects on gene function (Hughes et al., 2000). Therefore, domains which are critical to plant fitness will accumulate fewer nonsynonymous polymorphism than regions which are noncritical domains. Thus, we reasoned that scanning gene coding sequences for regions of relatively low nonsynonymous diversity should highlight functional domains. This general principle can be seen clearly in the analysis of the *Aux/IAA* family of transcriptional co-repressors/co-receptors. Most Aux/IAAs have three major domains. Domain I contains an EAR motif that facilitates interaction with TPL/TPR transcriptional co-repressors (Szemenyei et al., 2008; Tiwari et al., 2004). Domain II, the degron, facilitates interactions with the TIR1/AFB receptors in the presence of auxin (Tan et al., 2007). Domain III (which was originally considered domains III and IV) is a PB1 domain and facilitates interactions with the ARF transcription factors (Guilfoyle & Hagen, 2012; Korasick et al., 2014; Nanao et al., 2014; Ulmasov, Murfett, Hagen, & Guilfoyle, 1997).

We focused our analysis on the most conserved A class Aux/IAAs, which possess all three of these domains. We began by mapping the natural genetic variation onto an alignment of the coding sequences via the ViVa Alignment tab. A similar visualization of the full gene family is presented in the Section 5 (Figure 14). In the alignment of A class Aux/IAAs, the EAR motif and degron can be readily identified by the drop in nonsynonymous variation, as visualized by the lack of strong variant functional effects (Figure 6). The PB1 domain is not as readily identified, perhaps because the multiple key residues are spread out in linear sequence space. It is worth noting that these key residues that facilitate electrostatic PB1-PB1 interactions show little variation.

Natural variation also provides a means to study how gene families are evolving. To visualize this, we used ViVa to map the diversity at nonsynonymous variant sites relative to synonymous variant sites (π_N/π_S) onto the Aux/IAA phylogenetic tree (Figure 7). This visualization enables straightforward comparison of rates of recent functional divergence as predicted by diversity within natural variation in the context of the sequence divergence throughout the history of a gene family. Clades and individual genes exhibiting low-nonsynonymous diversity are likely functionally conserved. Conversely, genes with high-nonsynonymous diversity

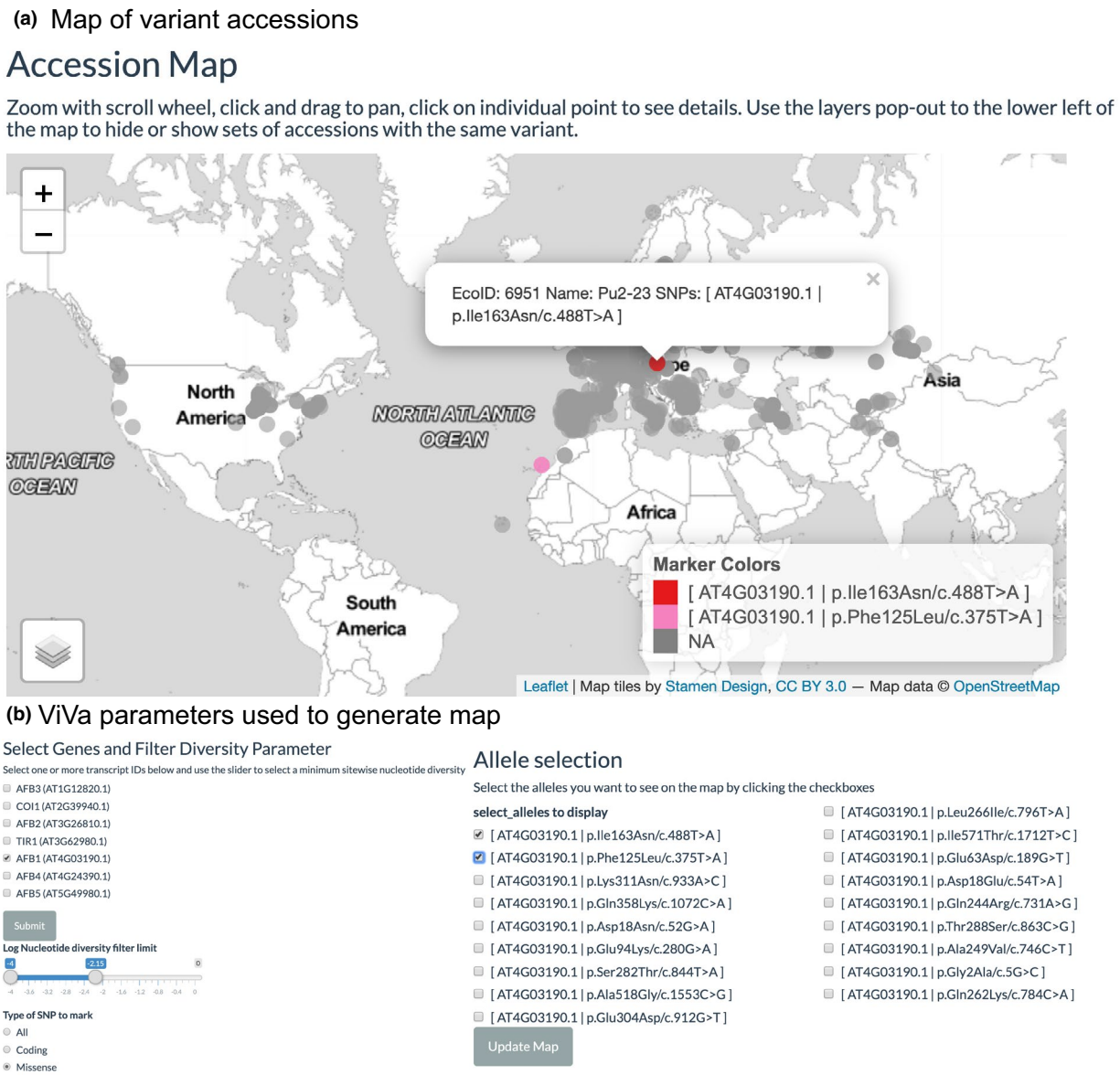


FIGURE 5 Map of AFB1 Oligomerization domain variant accessions. (a) Map showing the two accessions with variants in the AFB1 oligomerization domain (b) The user selectable parameters of the mapping tab used to generate the map are provided as an example of using the SNP Mapping tab in the ViVa webtool. See Figure 3 for details on filling in the parameters

are more likely to be under relaxed functional selection, indicating the possibility of functional drift, emergence of novel function, or pseudogenization.

Previous research has found evidence of both broad genetic redundancy among the *Aux/IAAs* and also specificity within closely related pairs or groups of Aux/IAA proteins (Overvoorde et al., 2005; Winkler et al., 2017). For example, the *iaa8-1 iaa9-1* double mutant and the *iaa5-1 iaa6-1 iaa19-1* triple mutant have wild-type phenotypes (Overvoorde et al., 2005), yet the *IAA6/IAA19* sister pair has significant differences in expression patterns, protein abundances, and functions suggesting they have undergone functional specialization since their divergence (Winkler et al., 2017). A closer examination of the *IAA19* and *IAA6* pair within Brassicaceae found evidence for positive selection and subfunctionalization of *IAA6* relative to *IAA19* (Winkler et al., 2017). Consistent with these results, ViVa revealed higher conservation (i.e., lower ratio of nonsynonymous to synonymous diversity) for *IAA19* (π_N/π_S = 0.55) compared to *IAA6* (π_N/π_S = 2.3; Figure 7), and also detected high nonsynonymous diversity within the same regions of *IAA6* as seen by Winkler et al. (Figure 4). This pattern—one sister showing high nonsynonymous diversity while the other sister was more conserved—was observed frequently across the *Aux/IAA* as well as the *AFB* and *ARF* families (Figures 8 and 12), suggesting this could be a recurring feature in the evolution of these families supporting the large diversity in auxin functions.

The *Aux/IAA* phylogeny clusters into two distinct clades represented by the A and B classes (Remington, Vision, Guilfoyle, & Reed, 2004). The C class Aux/IAAs are missing one or more of the canonical Aux/IAA domains. We found notable exceptions to the

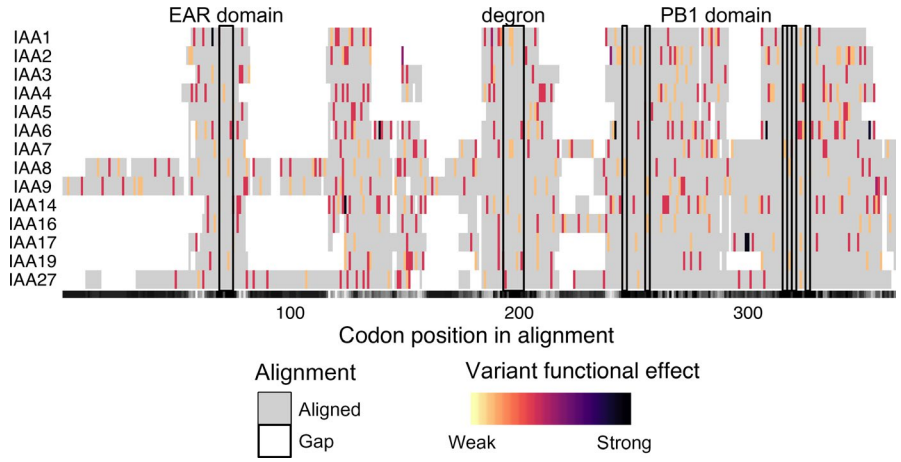


FIGURE 6 Critical functional domains of the conserved Aux/IAA genes show low nonsynonymous variation compared to regions of unknown functional importance. Protein sequences were aligned with DECIPHER (Wright, 2015) and variants were mapped to this alignment and colored according to the predicted functional effect of the allele of strongest effect at that position, with light colors having weaker effects on function and darker colors stronger effects. Red indicates missense variants. Color scale is further explained in Section 2. Alignment consensus is shown in grayscale underneath the plot as measured by evolutionary trace (Płuciennik et al., 2018; Wilkins, Erdin, Lua, & Lichtarge, 2012), with high consensus positions in black and low consensus in white. Key functional domains are outlined in black and labeled above the plot. The EAR domain spans codon alignment positions 70–74, corresponding to IAA1 amino acids LRLGL, 14–18. The degron domain spans alignment positions 194–201, corresponding to IAA1 amino acids QIVGWPPV, 55–62. The charged residues of the PB1 domain correspond to alignment positions 246, 256, 316, 318, 320, and 326 corresponding to IAA1 amino acids K77, R88, D133, D135, D137, and D143

pattern of diversification and conservation between sister pairs within the Class B *Aux/IAA* genes. The *IAA10/IAA11*, *IAA18/IAA26*, and *IAA20/IAA30* pairs showed similar levels of nonsynonymous diversity. For example, *IAA10* and *IAA11* both showed functional conservation (π_N/π_S of 0.80 and 0.67, respectively). The *A. thaliana* ePlant browser indicates that *IAA10* and *IAA11* have almost identical expression patterns (Waese et al., 2017). Together this evidence suggests a strong dosage requirement for these genes or that they have taken on novel functions since their emergence, with both genes contributing similar to plant fitness. In support of novel function, expression and mutant analysis during embryogenesis suggest that IAA10 is required for suspensor-hypophysis transition, while IAA11 is involved in later cell fate transitions (Rademacher et al., 2012).

4 | CONCLUSION

ViVa has allowed our team of testers from various skill levels and backgrounds to meaningfully access and mine the 1001 genomes dataset. The visualizations of natural variation further supported much of the existing structure-function knowledge of the well-studied nuclear auxin signaling pathway and facilitated the generation of new hypotheses. Gene and gene family analyses can be combined, as was done here for nuclear auxin signaling, to understand variation within gene networks. Visualizations of variation within whole-gene networks are planned for future iterations of ViVa. Application of ViVa to less-studied genes and gene families promises to yield more novel hypotheses, which can be evaluated via genetic and functional assays to glean novel structure/function knowledge from this rich dataset.

ViVa results are intended to inform and inspire hypothesis generation, not be taken as absolute evidence of trends in gene or gene family evolution. Among the cautions worth noting in interpreting results are limitations of short-read sequencing that lead to regions of missing data where low-read quality may have prevented variant calls. We have assumed these missing variants are reference alleles, leading to undercounting in ViVa's diversity estimations. Visualizations of this uncertainty will be added to a future version of ViVa. Recent advances in sequencing technologies have been combined to generate extremely high-quality genomes (Michael et al., 2018), and will reduce this source of uncertainty in future resequencing datasets. Another limitation is that the geographic coverage of accessions in the 1001 Genomes dataset is far from uniform, and thus diversity scores may not accurately reflect the allelic distributions of the global *A. thaliana* population.

We hope that ViVa will advance understanding of genotype-phenotype relationships by allowing all researchers access to large resequencing datasets. In the future, we intend to expand ViVa beyond the plant genetics workhorse, *A. thaliana*, to more agriculturally relevant species with existing resequencing projects, such as rice (Wang et al., 2018) and soybean (Zhou et al., 2015). Indeed, the ViVa framework is readily adaptable to any source of targeted resequencing data. If François Jacob's metaphor holds true, and evolution is indeed a tinkerer and not an engineer (Jacob, 1977), it is only by examining the largest possible number of nature's solutions that we may eventually decipher the principles constraining innovations in form and function.

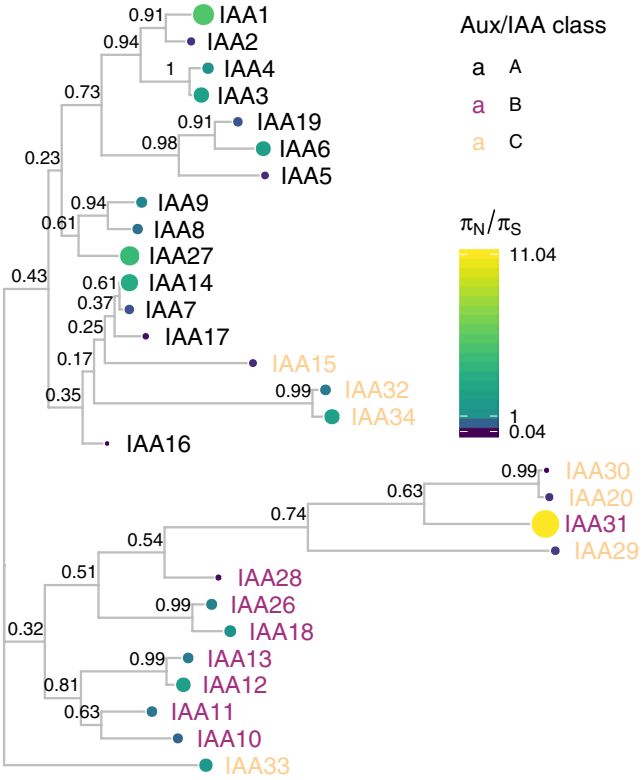


FIGURE 7 IAA protein sequence phylogenetic tree mapped with π_N/π_S reveals patterns of sister pair diversity/conservation. Protein sequences were aligned with DECIPHER (Wright, 2015) and low-information content regions were masked with Aliscore (Kück et al., 2010) prior to inferring a phylogeny with MrBayes (Ronquist & Huelsenbeck, 2003). Tips of the tree are mapped with circles of color and diameter proportional to π_N/π_S. π_N/π_S statistic provides a prediction of functional diversity. Nodes are labeled with the posterior probability of monophyly, a measure of confidence in the branch assignment, with one representing high confidence and zero, low confidence. There are two distinct clades of Aux/IAAs represented by the majority of the A and B classes. C class Aux/IAAs are missing one or more of the canonical Aux/IAA domains. These classes are represented by the text color of the gene name

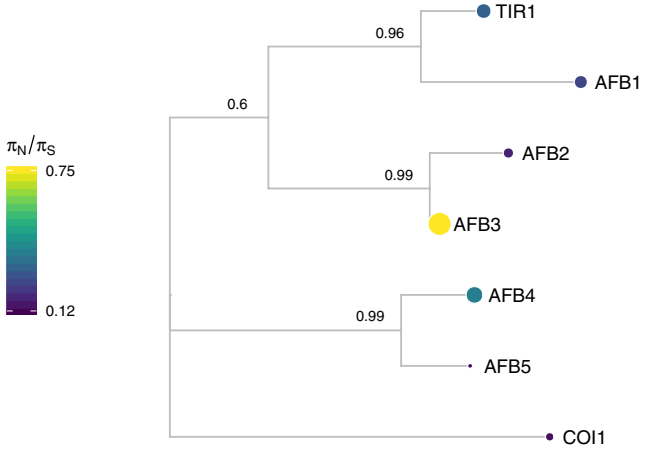


FIGURE 8 Auxin-signaling F-box protein sequence tree mapped with π_N/π_S. Protein sequences were aligned with DECIPHER (Wright, 2015) and low-information content regions were masked with Aliscore (Kück et al., 2010) prior to inferring a phylogeny with MrBayes (Ronquist & Huelsenbeck, 2003). Tips of the tree are mapped with circles of diameter proportional to π_N/π_S. Nodes are labeled with the posterior probability of monophyly

5 | SUPPLEMENTAL ANALYSES

5.1 | Additional natural variation in the *Aux/IAA* genes

For simplicity we have included only the alignment of the class A Aux/IAAs in the main manuscript. We include here the complete alignment of the family (Figure 14).

5.2 | Natural variation in the *TIR1/AFB* genes

Auxin acts by binding to receptors (Auxin-signaling F-Boxes, or AFBs) that in turn target co-repressors (Aux/IAAs) for degradation. The six auxin receptor genes in the model plant *A. thaliana*, *TIR1* and *AFB1-5*, evolved through gene duplication and diversification early in the history of vascular plants (Parry et al., 2009). The rate of co-repressor degradation is determined by the identity of both the receptor and co-repressor (Havens et al., 2012), and this rate sets the pace of lateral root development (Guseman et al., 2015).

All members of this family have been shown to bind auxin and Aux/IAA proteins. However, AFB1 has a drastically reduced ability to assemble into an SCF complex, due to the substitution E8K in its F-box domain, preventing it from inducing degradation of Aux/IAAs (Yu et al., 2015). This lack of SCF formation may allow for the high and ubiquitous AFB1 accumulation observed in *Arabidopsis* tissues (Parry et al., 2009). Higher order receptor mutants in the family containing *afb1* mutants suggest that *AFB1* has a moderate positive effect on auxin signaling (Dharmasiri et al., 2005). Additionally, AFB4 and AFB5 have been shown to preferentially and functionally bind the synthetic auxin picloram, while other family members preferentially bind indole-3-acetic acid (Prigge et al., 2016). Interestingly, the strength and rate with which TIR1/AFBs are able to bind and mark Aux/IAAs for degradation are variable (Calderón Calderón Villalobos et al., 2012; Havens et al., 2012). AFB2 induces the degradation of certain Aux/IAA proteins at a faster rate than TIR1, suggesting some functional specificity has arisen since the initial duplication between the *TIR1/AFB1* and *AFB2/AFB3* clades.

Examining the natural sequence variation across the *AFB* family revealed that *TIR1* and *AFB1* both had very low nonsynonymous diversity (Figure 8), hinting at their likely functional importance and bringing in to question the inconclusive role of *AFB1* in auxin signaling. *AFB3* and *AFB4* had higher nonsynonymous diversity, while their sister genes, *AFB2* and *AFB5* were more conserved. This matches our current understanding of *AFB3* as playing a minor role in the auxin signaling pathway (Dharmasiri et al., 2005). Two frameshift variants and one stop-gained (nonsense) variant were observed in *AFB4* supporting its pseudogenization, suggesting that *AFB4* may

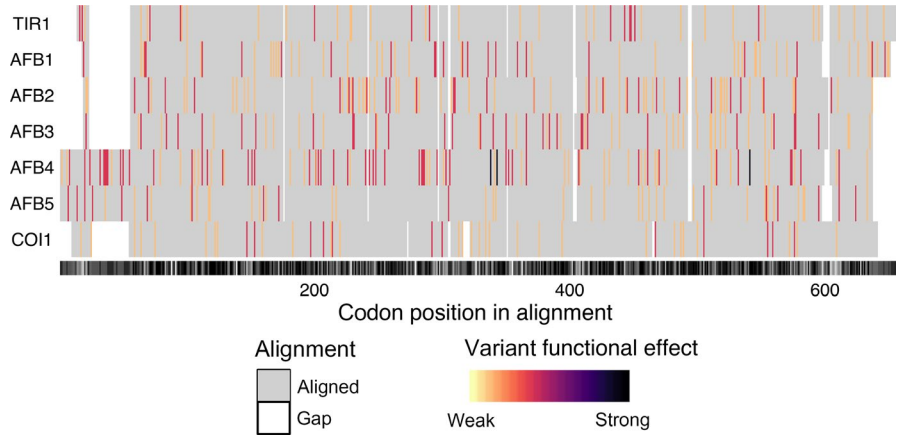


FIGURE 9 Alignment of the auxin-signaling F-box family protein sequences were aligned with DECIPHER (Wright, 2015) and variants were mapped to this alignment and colored according to the predicted functional effect of the allele of strongest effect at that position, with light colors having weaker effects on function and darker colors stronger effects. Red indicates missense variants. Color scale is explained in Section 2. In grayscale underneath the plot, alignment consensus is shown as measured by evolutionary trace (Płuciennik et al., 2018; Wilkins et al., 2012), with high-consensus positions in black and low consensus in white

be undergoing pseudogenization, especially when paired with its low-expression levels (Prigge et al., 2016). AFB4 and AFB5 have an N-terminal extension prior to their F-box domains. This extension had very-high nonsynonymous diversity (Figure 9), suggesting that this extension does not play an important functional role in these proteins.

Although most known functional regions are highly conserved in *AFB1*, there are some nonsynonymous polymorphism in the oligomerization domain that are only present in single accessions (F125E in Can-0 and I163N in Pu2-23, as shown in Figure 5). Mutations in this domain of *TIR1* frequently have a semidominant effect on root phenotypes (Dezfulian et al., 2016; Wright et al., 2017). Characterization of this allele and accession may help determine the role of *AFB1* in this pathway.

5.3 | Natural variation in the *TPL/TPR* genes

The auxin signaling pathway utilizes the TOPLESS (TPL) and TOPLESS-related (TPR) family of Gro/TLE/TUP1 type co-repressor proteins to maintain auxin responsive genes in a transcriptionally repressed state in the absence of auxin (Szemenyei et al., 2008). In *A. thaliana* the five member *TPL/TPR* family includes *TPL* and *TPR1-4*. The resulting proteins are comprised of three structural domains: an N-terminal TPL domain and two WD-40 domains (Long, Ohno, Smith, & Meyerowitz, 2006). TPL/TPR proteins are recruited to the AUX/IAA proteins through interaction with the conserved ethylene-responsive element binding factor-associated amphiphilic repression (EAR) domain (Szemenyei et al., 2008). Canonical EAR domains have the amino acid sequence LxLxL, as found in most AUX/IAAs (Overvoorde et al., 2005). TPL/TPR co-repressors bind EAR domains via their C-terminal to LisH (CTLH) domains found near their N-termini (Long et al., 2006). Recent structural analyses of the TPL N-terminal domain have highlighted the precise interaction interface between TPL and AUX/IAA EAR domains, as well as the TPL-TPL

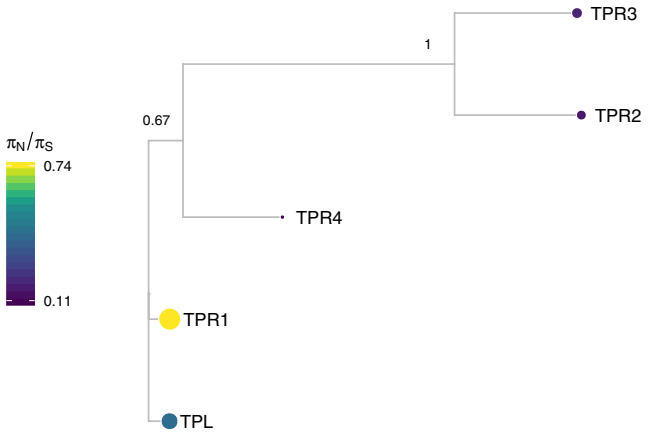


FIGURE 10 TPL protein sequence tree mapped with π_N/π_S. Protein sequences were aligned with DECIPHER (Wright, 2015) and low-information content regions were masked with Aliscore (Kück et al., 2010) prior to inferring a phylogeny with MrBayes (Ronquist & Huelsenbeck, 2003). Tips of the tree are mapped with circles of diameter proportional to π_N/π_S and also are colored according to π_N/π_S. Nodes are labeled with the poster probability of monophyly

dimerization and tetramerization motifs (Ke et al., 2015; Martin-Arevalillo et al., 2017). The residues required for higher order multimers of TPL tetramers have also been identified (Ma et al., 2017). Additional interactions with transcriptional regulation and chromatin-modifying machinery are likely mediated by two tandem beta propeller domains of TPL/TPRs.

The TOPLESS co-repressor family generally exhibits a high level of sequence conservation at the amino acid sequence level across resequenced *A. thaliana* accessions, with all π_N/π_S values below 1 (Figure 10). The closely related *TPL* and *TPR1* have the highest π_N/π_S values, suggesting that these two related genes tolerate a higher degree of sequence and potentially functional diversity compared to *TPR2/3/4*. The N-terminal TPL domain of

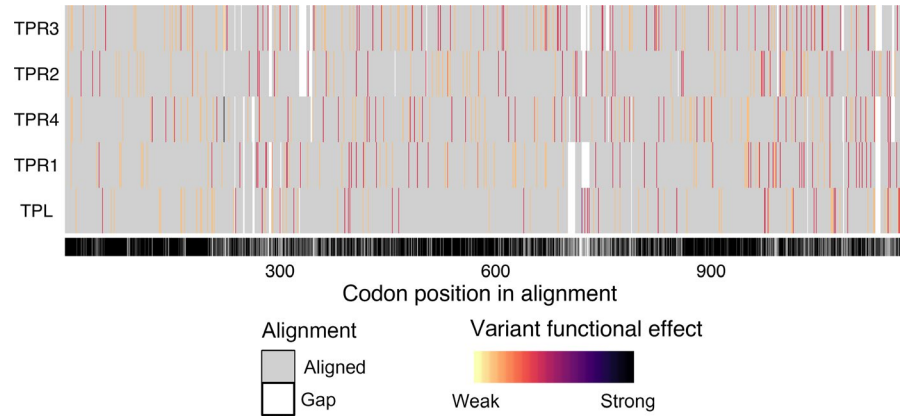


FIGURE 11 Alignment of the TPL/TPR family. Protein sequences were aligned with DECIPHER (Wright, 2015) and variants were mapped to this alignment and colored according to the predicted functional effect of the allele of strongest effect at that position, with light colors having weaker effects on function and darker colors stronger effects. Red indicates missense variants. Color scale is explained in Section 2. In grayscale underneath the plot, alignment consensus is shown as measured by evolutionary trace (Płuciennik et al., 2018; Wilkins et al., 2012), with high-consensus positions in black and low consensus in white

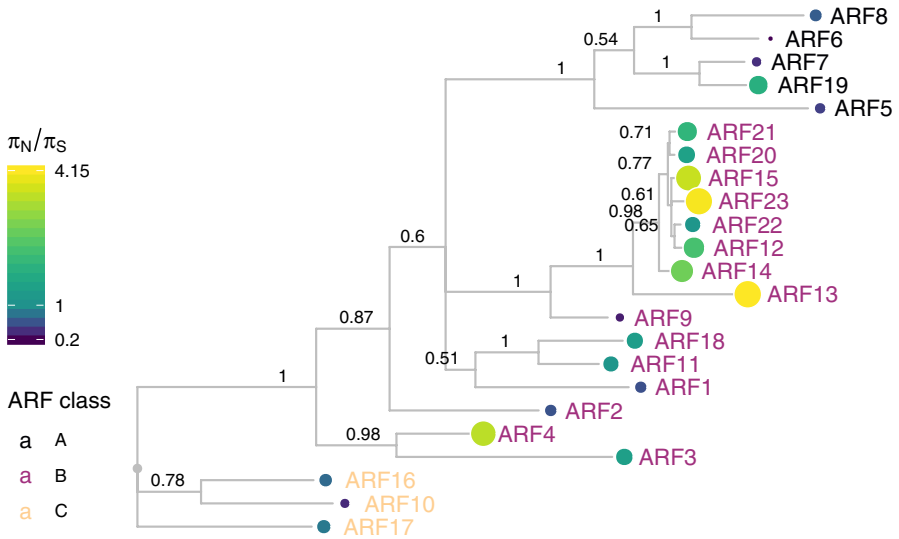


FIGURE 12 Auxin response factor protein sequence tree mapped with π_N/π_S. Protein sequences were aligned with DECIPHER (Wright, 2015) and low-information content regions were masked with Aliscore (Kück et al., 2010) prior to inferring a phylogeny with MrBayes (Ronquist & Huelsenbeck, 2003). Tips of the tree are mapped with circles of diameter proportional to π_N/π_S and also are colored according to π_N/π_S. Nodes are labeled with the poster probability of monophyly

the TPL/TPR family is particularly conserved. All nonsynonymous polymorphisms observed in this region are either in the coils between helices or are highly conservative mutations within helices (i.e., valine to isoleucine), which would be predicted to exhibit little effect on folding and function.

The high degree of conservation in the entire N-terminal domain underscores its importance in TPL/TPR function (Figure 11). For example, the initial *tpl-1* mutation (N176H) in the ninth helix is a dominant gain-of-function allele (Long et al., 2006), which is capable of binding wild-type TPL protein and inducing protein aggregation (Ma et al., 2017). It is therefore understandable that this helix had very low diversity as nonsynonymous variants in this domain could act in a dominant negative fashion.

5.4 | Natural variation in the *ARF* genes

Auxin response is mediated by the auxin responsive transcription factors (ARFs). There are 23 *ARF*s in *A. thaliana* that are divided into three phylogenetic classes. Class A ARFs (ARF5, ARF6, ARF7, ARF8, and ARF19) activate transcription. These ARFs have a glutamine-rich region in the middle of the protein that may mediate activation (Guilfoyle & Hagen, 2007). It has recently been shown that the middle region of ARF5 interacts with the SWI/SNF chromatin remodeling ATPases BRAMA and SPLAYED, possibly to reduce nucleosome occupancy and allow for the recruitment of transcription machinery (Wu et al., 2015). Additionally, ARF7 interacts with Mediator subunits, directly tethering transcriptional activation machinery to its

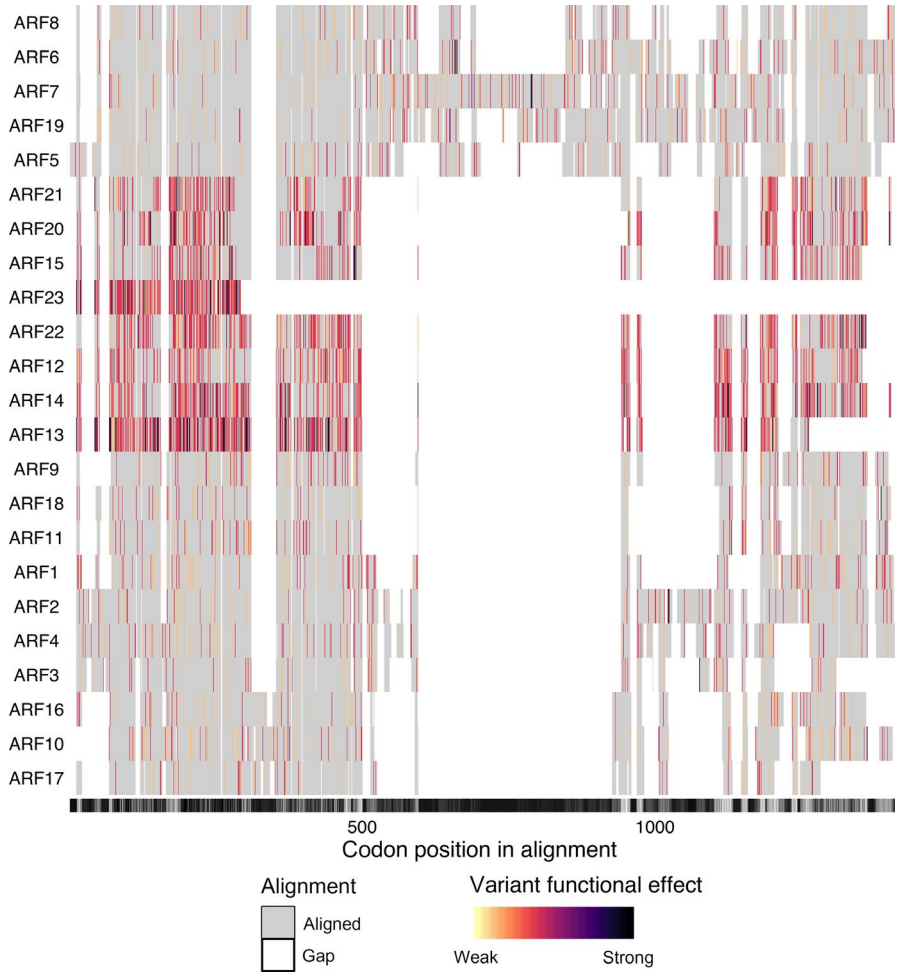


FIGURE 13 Alignment of the full auxin response factor family. Protein sequences were aligned with DECIPHER (Wright, 2015) and variants were mapped to this alignment and colored according to the predicted functional effect of the allele of strongest effect at that position, with light colors having weaker effects on function and darker colors stronger effects. Red indicates missense variants. Color scale is explained in Section 2. In grayscale underneath the plot, alignment consensus is shown as measured by Evolutionary trace (Płuciennik et al., 2018; Wilkins et al., 2012), with high-consensus positions in black and low consensus in white

binding sites in the chromosome (Ito et al., 2016). Class B and C *ARF*s are historically categorized as repressor *ARF*s, although the mechanism through which they confer repression has not been identified. Their middle regions tend to be proline- and serine-rich.

Canonical ARFs are comprised of three major domains. Recent crystallization of these domains have informed structure-function analysis of the ARFs (Boer et al., 2014; Korasick et al., 2014; Nanao et al., 2014). These domains are conserved throughout land plants (Mutte et al., 2018). ARFs share an N-terminal B3 DNA binding domain. Flanking this DNA-binding domain is a dimerization domain, which folds up into a single "taco-shaped" domain to allow for dimerization between ARFs. There is an auxiliary domain that immediately follows and interacts with the dimerization domain. The middle region is the most variable between ARFs, as mentioned above, but is characterized by repetitive units of glutamine (class A), serine, or proline residues (classes B and C). The C-terminal domain of canonical ARFs is a PB1 protein-protein interaction domain mediating interactions among ARFs, between ARFs and other transcription factors, and between ARFs and the Aux/IAA repressors. This interaction domain was recently characterized as a Phox and Bem1 (PB1) domain, which is comprised of a positive and negative face with conserved basic and acidic residues, respectively (Korasick et al., 2014; Nanao et al., 2014). The dipolar nature of the PB1 domain may mediate multimerization by the pairwise interaction of these faces on different proteins as the ARF7 PB1 domain was crystallized as a multimer (Korasick et al., 2014). However, it is unclear whether ARF multimerization occurs or plays a significant role in vivo. Interfering with ARF dimerization in either the DNA-binding proximal dimerization domain or the PB1 domain decreases the ability of class A ARFs to activate transcription in a heterologous yeast system (Pierre-Jerome, Moss, Lanctot, Hageman, & Nemhauser, 2016).

While domain architecture is broadly conserved among the ARFs, there are exceptional cases. Three ARFs do not contain a PB1 domain at all, ARF3, ARF13, and ARF17, and several more have lost the conserved acidic or basic residues in the PB1 domain, suggesting they may be reduced to a single interaction domain. Several ARFs additionally have an expanded conserved region within the DNA-binding domain, of unknown function. The majority of domain variation among ARFs occurs in the large B-class subfamily. The liverwort *Marchantia polymorpha* has a single representative ARF of each class (Flores-Sandoval, Magnus Eklund, & Bowman, 2015). The expansion of these classes in flowering plants is the result of both whole genome and tandem duplication events (Remington et al., 2004). The growth of the ARF family may have allowed for the expansion of the quantity and complexity of loci regulated by the ARFs and subsequent expansion in their regulation of developmental processes (Mutte et al., 2018).

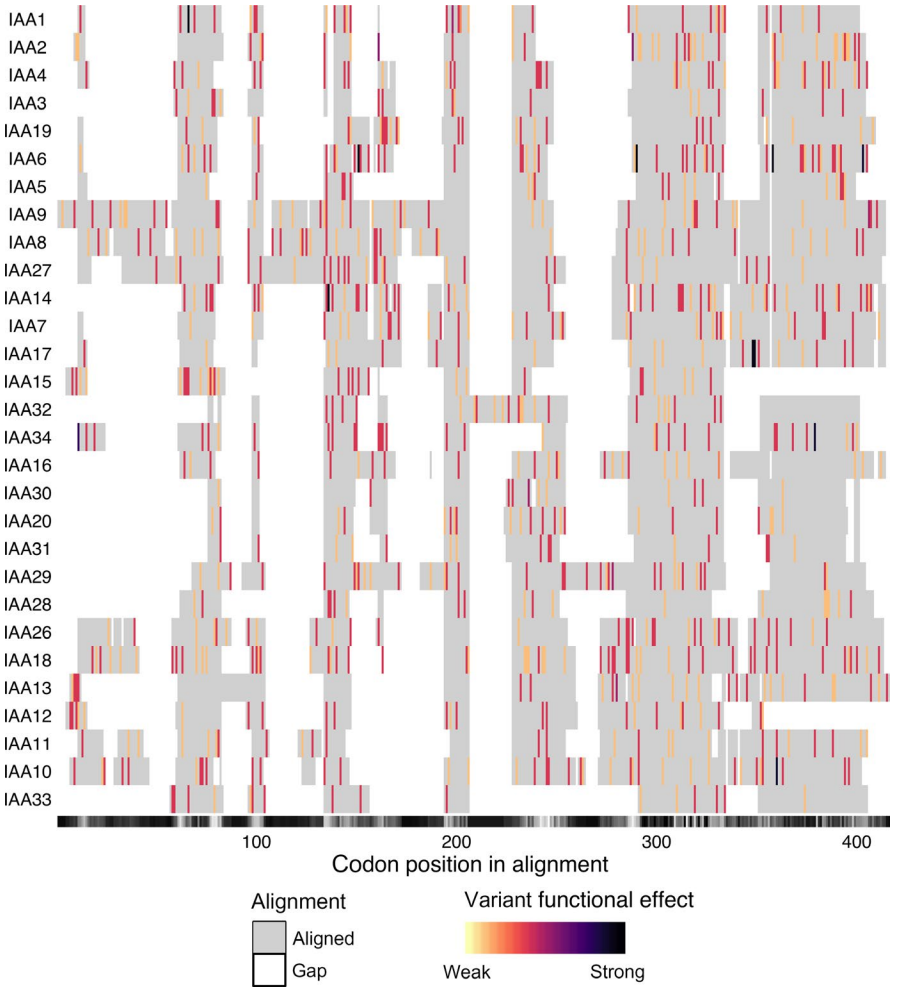


FIGURE 14 Alignment of the complete Aux/IAA family. Protein sequences were aligned with DECIPHER (Wright, 2015) and variants were mapped to this alignment and colored according to the predicted functional effect of the allele of strongest effect at that position, with light colors having weaker effects on function and darker colors stronger effects. Red indicates missense variants. Color scale is further explained in Section 2. Alignment consensus is shown in grayscale underneath the plot as measured by evolutionary trace (Płuciennik et al., 2018; Wilkins et al., 2012), with high-consensus positions in black and low consensus in white

Class A ARFs are the most well-studied ARF subfamily—the five family members all act as transcriptional activators and have well-characterized, distinct developmental targets. Overall the diversity of class A *ARF*s was generally low, especially compared to the classes B and C *ARF*s (Figure 12), suggesting that class A *ARF*s are central to auxin signal transduction and plant development. Analysis of class A *ARF* nonsynonymous diversity suggests that the majority of these *ARF*s are highly functionally conserved, with π_N/π_S values much lower than 1 with the exception of *ARF19*, with π_N/π_S value of 1.8. Comparing diversity within sister pairs, there is a similar trade-off as seen in most *IAA* sister pairs, with one sister being highly conserved and the other more divergent. *ARF19* and *ARF8* are the more divergent class A *ARF*s, with π_N/π_S values at least three time those of their sisters, *ARF7* and *ARF6*, respectively. This may suggest that ARF6 and ARF7 serve more essential purposes in plant development.

For all class A *ARF*s, the middle region of the protein was the predominant high-diversity region (Figure 13). In the analyzed natural variation, ARF7 had several expansions of polyglutamine sequences in the middle region. Polyglutamine regions are known to readily expand and contract throughout evolutionary time due to replication error, and variation in polyglutamine length can have phenotypic consequences and be acted on by natural selection (Press, Carlson, & Queitsch, 2014). The ARF DNA-binding domain had very few, low-diversity missense mutations, as did the critical residues of the PB1 domain. Considering the necessity of their conserved functions, the low level of variation in these key DNA and protein-protein interaction domains is expected.

Many of the class B *ARF*s have very high π_N/π_S ratios relative to the other *ARF*s. ARF23 has a truncated DNA-binding domain and had a high π_N/π_S value of 4.1. ARF13 has many high-diversity nonsense variants and lacks a C-terminal PB1 domain. This high level of diversity, prevalance of high-frequency nonsense variants and frequent loss of critical domains, may suggest that several genes in this class are undergoing pseudogenization.

There are also a few highly conserved class B *ARF*s. The high conservation of *ARF1* and *ARF2* is expected as they play critical, redundant roles in senescence and abscission (Ellis et al., 2005). Little is known about *ARF9*, however, and its low nonsynonymous diversity maybe worthy of investigation.

Class C ARFs show low-nucleotide diversity scores, with all π_N/π_S values substantially lower than 1. *ARF16* was the most conserved, whereas the other clade members (*ARF10*, *ARF17*) had scores at least four times higher. Structurally, all three members of Class C ARFs contain a canonical B3 DNA-binding domain, but only ARF10 and ARF16 contain a PB1 domain. The DNA-binding

domains exhibited overall low diversity. Of the PB1 domain containing class C ARFs, ARF16's PB1 domain exhibited several missense variants which are sporadically distributed, in contrast to the conserved PB1 domain of ARF10. This conservation in the PB1 domain of ARF10 and the DBD of ARF16 may suggest subfunctionalization in this class.

ACKNOWLEDGMENTS

The authors would like to thank Oghenemega Okolo for assistance testing the ViVa software, and Song Li and Bo Zhang for helpful comments on the manuscript. This work was supported by the National Institute of Health (R01-GM107084), the National Science Foundation (IOS-1546873), and the Howard Hughes Medical Institute. R.C.W. received fellowship support from the National Science Foundation (DBI-1402222). B.L.M. and H.K. received support from the M.J. Murdock Charitable Trust. A.R.L. is a Simons Foundation Fellow of the Life Sciences Research Foundation. A.L. was supported by an NSF Graduate Research Fellowship DGE-1256082.

AUTHOR CONTRIBUTIONS

RCW and JLN designed the research; MOH and RCW developed the software; MOH, BLM, ARL, HPG, AL, RR, HK, ACL, MLZ, and RCW tested the software, performed the research, and analyzed the data; MOH, BLM, ARL, HPG, AL, RR, HK, ACL, MLZ, JLN, and RCW interpreted the data and wrote the paper.

CONFLICT OF INTEREST

The authors declare no conflict of interest associated with the work described in this manuscript.

REFERENCES

Alexandre, C. M., Urton, J. R., Jean-Baptiste, K., Huddleston, J., Dorrity, M. W., Cuperus, J. T., … Queitsch, C. (2018). Complex relationships between chromatin accessibility, sequence divergence, and gene expression in *Arabidopsis thaliana*. *Molecular Biology and Evolution*, *35*(4), 837–854. https://doi.org/10.1093/molbev/msx326

Allaire, J. J., Ushey, K., & Tang, Y. (2018). *Reticulate: Interface to 'Python'*. Retrieved from https://CRAN.R-project.org/package=reticulate

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., … Iannone, R. (2018). *Rmarkdown: Dynamic documents for R*. Retrieved from https://CRAN.R-project.org/package=rmarkdown

Aphalo, P. J. (2018a). *Gginnards: Explore the innards of 'Ggplot2' objects*. Retrieved from https://CRAN.R-project.org/package=gginnards

Aphalo, P. J. (2018b). *Ggpmisc: Miscellaneous extensions to 'Ggplot2'*. Retrieved from https://CRAN.R-project.org/package=ggpmisc

Arnold, J. B. (2018). *Ggthemes: Extra themes, scales and geoms for 'Ggplot2'*. Retrieved from https://CRAN.R-project.org/package=ggthemes

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., … Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, *465*(7298), 627–631. https://doi.org/10.1038/nature08800

Bache, S. M., & Wickham, H. (2014). *Magrittr: A forward-pipe operator for R*. Retrieved from https://CRAN.R-project.org/package=magrittr

Boer, D. R., Freire-Rios, A., van den Berg, W. A., Saaki, T., Manfield, I. W., Kepinski, S., … Coll, M. (2014). Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell*, *156*(3), 577–589. https://doi.org/10.1016/j.cell.2013.12.027

Borevitz, J. O., Hazen, S. P., Michael, T. P., Morris, G. P., Baxter, I. R., Hu, T. T., … Ecker, J. R. (2007). Genome-wide patterns of single-feature polymorphism in *Arabidopsis Thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(29), 12057–12062. https://doi.org/10.1073/pnas.0705323104

Calderón Villalobos, L. I., Lee, S., De Oliveira, C., Ivetac, A., Armitage, L., Sheard, L. B., … Estelle, M. (2012). A combinatorial TIR1/AFB-Aux/IAA co-receptor system for differential sensing of auxin. *Nature Chemical Biology*, *8*(5), 477–485. https://doi.org/10.1038/nchembio.926

Chen, K., Wang, Y., Zhang, R., Zhang, H., & Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annual Review of Plant Biology*, *70*(1), 667–697. https://doi.org/10.1146/annurev-arplant-050718-100049

Cheng, C.-Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: A complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal*, *89*(4), 789–804. https://doi.org/10.1111/tpj.13415

Clark, R. M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., … Weigel, D. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science*, *317*(5836), 338–342. https://doi.org/10.1126/science.1138632

Delker, C., Pöschl, Y., Raschke, A., Ullrich, K., Ettingshausen, S., Hauptmann, V., … Quint, M. (2010). Natural variation of transcriptional auxin response networks in *Arabidopsis thaliana*. *The Plant Cell*, *22*(7), 2184–2200. https://doi.org/10.1105/tpc.110.073957

Dezfulian, M. H., Jalili, E., Roberto, D. K., Moss, B. L., Khoo, K., Nemhauser, J. L., & Crosby, W. L. (2016). Oligomerization of SCF TIR1 is essential for Aux/IAA degradation and auxin signaling in Arabidopsis. *PLoS Genetics*, *12*(9), e1006301. https://doi.org/10.1371/journal.pgen.1006301

Dharmasiri, N., Dharmasiri, S., Weijers, D., Lechner, E., Yamada, M., Hobbie, L., … Estelle, M. (2005). Plant development is regulated by a family of auxin receptor F box proteins. *Developmental Cell*, *9*(1), 109–119. https://doi.org/10.1016/j.devcel.2005.05.014

Ellis, C. M., Nagpal, P., Young, J. C., Hagen, G., Guilfoyle, T. J., & Reed, J. W. (2005). AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence and floral organ abscission in *Arabidopsis thaliana*. *Development*, *132*(20), 4563–4574. https://doi.org/10.1242/dev.02012

Firnberg, E., & Ostermeier, M. (2013). The genetic code constrains yet facilitates Darwinian evolution. *Nucleic Acids Research*, *41*(15), 7420–7428. https://doi.org/10.1093/nar/gkt536

Flores-Sandoval, E., Magnus Eklund, D., & Bowman, J. L. (2015). A simple auxin transcriptional response system regulates multiple morphogenetic processes in the liverwort *Marchantia polymorpha*. *PLoS Genetics*, *11*(5), e1005207. https://doi.org/10.1371/journal.pgen.1005207

Garnier, S. (2018a). *Viridis: Default color maps from 'Matplotlib'*. Retrieved from https://CRAN.R-project.org/package=viridis

Garnier, S. (2018b). *ViridisLite: Default color maps from 'Matplotlib' (lite version)*. Retrieved from https://CRAN.R-project.org/package=viridisLite

Gasperini, M., Starita, L., & Shendure, J. (2016). The power of multiplexed functional analysis of genetic variants. *Nature Protocols*, *11*(10), 1782–1787. https://doi.org/10.1038/nprot.2016.135

Gray, W. M., Kepinski, S., Rouse, D., Leyser, O., & Estelle, M. (2001). Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. *Nature*, *414*(6861), 271–276. https://doi.org/10.1038/35104500

Guilfoyle, T. J., & Hagen, G. (2007). Auxin response factors. *Current Opinion in Plant Biology*, *10*(5), 453–460. https://doi.org/10.1016/j.pbi.2007.08.014

Guilfoyle, T. J., & Hagen, G. (2012). Getting a grasp on domain III/IV responsible for auxin response FactorIAA protein interactions. *Plant Science*, *190*, 82–88. https://doi.org/10.1016/j.plantsci.2012.04.003

Guo, Y.-L. (2013). Gene family evolution in green plants with emphasis on the origination and evolution of *Arabidopsis thaliana* genes. *The Plant Journal*, *73*(6), 941–951. https://doi.org/10.1111/tpj.12089

Guseman, J. M., Hellmuth, A., Lanctot, A., Feldman, T. P., Moss, B. L., Klavins, E., … Nemhauser, J. L. (2015). Auxin-induced degradation dynamics set the pace for lateral root development. *Development*, *142*(5), 905–909. https://doi.org/10.1242/dev.117234

Hamm, M. O., & Wright, R. C. (2018). *R1001genomes: Access and analyze the 1001 genomes Arabidopsis*. Resequencing Dataset.

Havens, K. A., Guseman, J. M., Jang, S. S., Pierre-Jerome, E., Bolten, N., Klavins, E., & Nemhauser, J. L. (2012). A synthetic approach reveals extensive tunability of auxin signaling. *Plant Physiology*, *160*(1), 135–142. https://doi.org/10.1104/pp.112.202184

Heibl, C. (2014). *Ips: Interfaces to phylogenetic software in R*. Retrieved from https://CRAN.R-project.org/package=ips.

Henry, L., & Wickham, H. (2018). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr.

Hughes, A. L. (1999). *Adaptive evolution of genes and genomes*. Oxford: Oxford University Press.

Hughes, A. L., Green, J. A., Garbayo, J. M., & Roberts, R. M. (2000). Adaptive diversification within a large family of recently duplicated, placentally expressed genes. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(7), 3319–3323. https://doi.org/10.1073/pnas.97.7.3319

Ihaka, R., Murrell, P., Hornik, K., Fisher, J. C., Stauffer, R., Wilke, C. O., … Zeileis, A. (2019). *Colorspace: A toolbox for manipulating and assessing colors and palettes*. Retrieved from https://CRAN.R-project.org/package=colorspace

Ito, J., Fukaki, H., Onoda, M., Li, L., Li, C., Tasaka, M., & Furutani, M. (2016). Auxin-dependent compositional change in Mediator in ARF7- and ARF19-mediated transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 6562–6567. https://doi.org/10.1073/pnas.1600739113

Jacob, F. (1977). Evolution and tinkering. *Science*, *196*(4295), 1161–1166. https://doi.org/10.1126/science.860134

Joly-Lopez, Z., Flowers, J. M., & Purugganan, M. D. (2016). Developing maps of fitness consequences for plant genomes. *Current Opinion in Plant Biology*, *30*, 101–107. https://doi.org/10.1016/j.pbi.2016.02.008

Joshi, H. J., Christiansen, K. M., Fitz, J., Cao, J., Lipzen, A., Martin, J., … Heazlewood, J. L. (2012). 1001 Proteomes: A functional proteomics portal for the analysis of *Arabidopsis thaliana* accessions. *Bioinformatics*, *28*(10), 1303–1306. https://doi.org/10.1093/bioinformatics/bts133

Ke, J., Ma, H., Gu, X., Thelen, A., Brunzelle, J. S., Li, J., … Melcher, K. (2015). Structural basis for recognition of diverse transcriptional repressors by the TOPLESS family of corepressors. *Science Advances*, *1*(6), e1500107. https://doi.org/10.1126/sciadv.1500107

Kliebenstein, D. J. (2008). A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PLoS ONE*, *3*(3), e1838. https://doi.org/10.1371/journal.pone.0001838

Korasick, D. A., Westfall, C. S., Lee, S. G., Nanao, M. H., Dumas, R., Hagen, G., … Strader, L. C. (2014). Molecular basis for AUXIN RESPONSE FACTOR protein interaction and the control of auxin response repression. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(14), 5427–5432. https://doi.org/10.1073/pnas.1400074111

Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B. M., Wägele, J. W., & Misof, B. (2010). Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Frontiers in Zoology*, *7*, 10. https://doi.org/10.1186/1742-9994-7-10

Lang, D. T., & CRAN Team (2018). *XML: Tools for parsing and generating Xml within R and S-Plus*. Retrieved from https://CRAN.R-project.org/package=XML

Long, J. A., Ohno, C., Smith, Z. R., & Meyerowitz, E. M. (2006). TOPLESS regulates apical embryonic fate in Arabidopsis. *Science*, *312*(5779), 1520–1523. https://doi.org/10.1126/science.1123841

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., … Nordborg, M. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics*, *45*(8), 884–890. https://doi.org/10.1038/ng.2678

Ma, H., Duan, J., Ke, J., He, Y., Gu, X., Xu, T. H., … Melcher, K. (2017). A D53 repression motif induces oligomerization of TOPLESS corepressors and promotes assembly of a corepressor-nucleosome complex. *Science Advances*, *3*(6), e1601217. https://doi.org/10.1126/sciadv.1601217

Martin-Arevalillo, R., Nanao, M. H., Larrieu, A., Vinos-Poyo, T., Mast, D., Galvan-Ampudia, C., … Parcy, F. (2017). Structure of the Arabidopsis TOPLESS corepressor provides insight into the evolution of transcriptional repression. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 8107–8112. https://doi.org/10.1073/pnas.1703054114

Matreyek, K. A., Stephany, J. J., & Fowler, D. M. (2017). A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Research*, *45*(11), e102. https://doi.org/10.1093/nar/gkx183

Melamed, D., Young, D. L., Miller, C. R., & Fields, S. (2015). Combining natural sequence variation with high throughput mutational data to reveal protein interaction sites. *PLoS Genetics*, *11*(2), e1004918. https://doi.org/10.1371/journal.pgen.1004918

Michael, T. P., Jupe, F., Bemm, S., Motley, S. T., Sandoval, J. P., Lanz, C., … Ecker, J. R. (2018). High contiguity arabidopsis thaliana genome assembly with a single nanopore flow cell. *Nature Communications*, *9*(1), 541. https://doi.org/10.1038/s41467-018-03016-2

Moore, R. C., & Purugganan, M. D. (2005). The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology*, *8*(2), 122–128. https://doi.org/10.1016/j.pbi.2004.12.001

Müller, K. (2018). *Bindrcpp: An 'Rcpp' interface to active bindings*. Retrieved from https://CRAN.R-project.org/package=bindrcpp

Müller, K., & Wickham, H. (2019). *Tibble: Simple data frames*. Retrieved from https://CRAN.R-project.org/package=tibble

Müller, K., Wickham, H., James, D. A., & Falcon, S. (2018). *RSQLite: 'SQLite' interface for R*. Retrieved from https://CRAN.R-project.org/package=RSQLite

Mutte, S. K., Kato, H., Rothfels, C., Melkonian, M., Wong, G. K.-S., & Weijers, D. (2018). Origin and evolution of the nuclear auxin response system. *eLife*, *7*, e33399. https://doi.org/10.7554/elife.33399

Nanao, M. H., Vinos-Poyo, T., Brunoud, G., Thévenon, E., Mazzoleni, M., Mast, D., … Dumas, R. (2014). Structural basis for oligomerization of auxin transcriptional regulators. *Nature Communications*, *5*, 3617. https://doi.org/10.1038/ncomms4617

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nelson, C. W., Moncla, L. H., & Hughes, A. L. (2015). SNPGenie: Estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics*, *31*(22), 3709–3711. https://doi.org/10.1093/bioinformatics/btv449

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer palettes*. Retrieved from https://CRAN.R-project.org/package=RColorBrewer

Nieduszynski, C. A., & Liti, G. (2011). From sequence to function: Insights from natural variation in budding yeasts. *Biochimica et Biophysica Acta*, *1810*(10), 959–966. https://doi.org/10.1016/j.bbagen.2011.02.004

Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., … Bergelson, J. (2005). The pattern of polymorphism in *Arabidopsis*

thaliana. *PLoS Biology*, *3*(7), e196. https://doi.org/10.1371/journal.pbio.0030196

Overvoorde, P. J., Okushima, Y., Alonso, J. M., Chan, A., Chang, C., Ecker, J. R., … Theologis, A. (2005). Functional genomic analysis of the AUXIN/INDOLE-3-ACETIC ACID gene family members in *Arabidopsis thaliana*. *The Plant Cell*, *17*(12), 3282–3300. https://doi.org/10.1105/tpc.105.036723

Pagès, H., & Aboyoun, P. (2018). *XVector: Representation and manipulation of external sequences*. Retrieved from https://doi.org/10.18129/B9.bioc.XVector

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2019). *Biostrings: Efficient manipulation of biological strings*. Retrieved from https://doi.org/10.18129/B9.bioc.Biostrings

Pagès, H., Aboyoun, P., & Lawrence, M. (2018). *IRanges: Infrastructure for manipulating intervals on sequences*. Retrieved from https://doi.org/10.18129/B9.bioc.IRanges

Pagès, H., Lawrence, M., & Aboyoun, P. (2018). *S4Vectors: S4 implementation of vector-like and list-like objects*. Retrieved from https://doi.org/10.18129/B9.bioc.S4Vectors

Paradis, E., Blomberg, S., Bolker, B., Brown, J., Claude, J., Cuong, H. S., … D. de Vienne. (2018). *Ape: Analyses of phylogenetics and evolution*. Retrieved from https://CRAN.R-project.org/package=ape

Park, B., Rutter, M. T., Fenster, C. B., Symonds, V. V., Ungerer, M. C., & Townsend, J. P. (2017). Distributions of mutational effects and the estimation of directional selection in divergent lineages of *Arabidopsis thaliana*. *Genetics*, *206*(4), 2105–2117. https://doi.org/10.1534/genetics.116.199190

Parry, G., Calderon-Villalobos, L. I., Prigge, M., Peret, B., Dharmasiri, S., Itoh, H., … Estelle, M. (2009). Complex regulation of the TIR1/AFB family of auxin receptors. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(52), 22540–22545. https://doi.org/10.1073/pnas.0911967106

Pierre-Jerome, E., Moss, B. L., Lanctot, A., Hageman, A., & Nemhauser, J. L. (2016). Functional analysis of molecular interactions in synthetic auxin response circuits. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(40), 11354–11359. https://doi.org/10.1073/pnas.1604379113

Płuciennik, A., Stolarczyk, M., Bzówka, M., Raczyńska, A., Magdziarz, T., & Góra, A. (2018). BALCONY: An R package for MSA and functional compartments of protein variability analysis. *BMC Bioinformatics*, *19*(1), 300. https://doi.org/10.1186/s12859-018-2294-z

Press, M. O., Carlson, K. D., & Queitsch, C. (2014). The overdue promise of short tandem repeat variation for heritability. *Trends in Genetics: TIG*, *30*(11), 504–512. https://doi.org/10.1016/j.tig.2014.07.008

Prigge, M. J., Greenham, K., Zhang, Y., Santner, A., Castillejo, C., Mutka, A. M., … Estelle, M. (2016). The arabidopsis auxin receptor F-box proteins AFB4 and AFB5 are required for response to the synthetic auxin picloram. *G3: Genes, Genomes, Genetics*, *6*(5), 1383–1390. https://doi.org/10.1534/g3.115.025585

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rademacher, E. H., Lokerse, A. S., Schlereth, A., Llavata-Peris, C. I., Bayer, M., Kientz, M., … Weijers, D. (2012). Different auxin response machineries control distinct cell fates in the early plant embryo. *Developmental Cell*, *22*(1), 211–222. https://doi.org/10.1016/j.devcel.2011.10.026

Remington, D. L., Vision, T. J., Guilfoyle, T. J., & Reed, J. W. (2004). Contrasting modes of diversification in the Aux/IAA and ARF gene families. *Plant Physiology*, *135*(3), 1738–1752. https://doi.org/10.1104/pp.104.039669

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Starita, L. M., Ahituv, N., Dunham, M. J., Kitzman, J. O., Roth, F. P., Seelig, G., … Fowler, D. M. (2017). Variant interpretation: Functional assays to the rescue. *The American Journal of Human Genetics*, *101*(3), 315–325. https://doi.org/10.1016/j.ajhg.2017.07.014

Szemenyei, H., Hannon, M., & Long, J. A. (2008). TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis. *Science*, *319*(5868), 1384–1386. https://doi.org/10.1126/science.1151461

Tan, X., Calderon-Villalobos, L. I., Sharon, M., Zheng, C., Robinson, C. V., Estelle, M., & Zheng, N. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature*, *446*(7136), 640–645. https://doi.org/10.1038/nature05731

Team, The Bioconductor Dev. (2018). *BiocGenerics: S4 generic functions for bioconductor*. Retrieved from https://doi.org/10.18129/B9.bioc.BiocGenerics

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis Thaliana*. *Nature*, *408*(6814), 796–815. https://doi.org/10.1038/35048692

Tiwari, S. B., Hagen, G., & Guilfoyle, T. (2003). The roles of auxin response factor domains in auxin-responsive transcription. *The Plant Cell*, *15*(2), 533–543. https://doi.org/10.1105/tpc.008417

Tiwari, S. B., Hagen, G., & Guilfoyle, T. J. (2004). Aux/IAA proteins contain a potent transcriptional repression domain. *The Plant Cell*, *16*(2), 533–543. https://doi.org/10.1105/tpc.017384

Ulmasov, T., Murfett, J., Hagen, G., & Guilfoyle, T. J. (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *The Plant Cell Online*, *9*(11), 1963–1971. https://doi.org/10.1105/tpc.9.11.1963

Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., … Provart, N. J. (2017). ePlant: Visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *The Plant Cell*, *29*(8), 1806–1821. https://doi.org/10.1105/tpc.17.00073

Wagih, O. (2017). *Ggseqlogo: A 'Ggplot2' extension for drawing publication-ready sequence logos*. Retrieved from https://CRAN.R-project.org/package=ggseqlogo

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., … Mansueto, L. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*, *557*(7703), 43. https://doi.org/10.1038/s41586-018-0063-9

Weigel, D., & Mott, R. (2009). The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biology*, *10*(5), 107. https://doi.org/10.1186/gb-2009-10-5-107

Whitehead, T. A., Chevalier, A., Song, Y., Dreyfus, C., Fleishman, S. J., De Mattos, C., … Baker, D. (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nature Biotechnology*, *30*(6), 543–548. https://doi.org/10.1038/nbt.2214

Wickham, H. (2016). *Plyr: Tools for splitting, applying and combining data*. Retrieved from https://CRAN.R-project.org/package=plyr

Wickham, H. (2017a). *Reshape2: Flexibly reshape data: A reboot of the reshape package*. Retrieved from https://CRAN.R-project.org/package=reshape2

Wickham, H. (2017b). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wickham, H. (2018a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2018b). *Scales: Scale functions for visualization*. Retrieved from https://CRAN.R-project.org/package=scales

Wickham, H. (2018c). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., & Woo, K. (2018). *Ggplot2: Create elegant data visualisations using the grammar of graphics*. Retrieved from https://CRAN.R-project.org/package=ggplot2

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2018). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & Francois, R. (2018). *Readr: Read rectangular text data*. Retrieved from https://CRAN.R-project.org/package=readr

Wilkins, A., Erdin, S., Lua, R., & Lichtarge, O. (2012). Evolutionary trace for prediction and redesign of protein functional sites. *Methods in Molecular Biology (Clifton, N.J.)*, *819*, 29–42. https://doi.org/10.1007/978-1-61779-465-0_3

Winkler, M., Niemeyer, M., Hellmuth, A., Janitza, P., Christ, G., Samodelov, S. L., … Calderón Villalobos, L. I. A. (2017). Variation in auxin sensing guides AUX/IAA transcriptional repressor ubiquitylation and destruction. *Nature Communications*, *8*, 15706. https://doi.org/10.1038/ncomms15706

Wright, E. S. (2015). DECIPHER: Harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics*, *16*, 322. https://doi.org/10.1186/s12859-015-0749-z

Wright, E. (2019). *DECIPHER: Tools for curating, analyzing, and manipulating biological sequences*. Retrieved from https://doi.org/10.18129/B9.bioc.DECIPHER

Wright, R. C., Zahler, M. L., Gerben, S. R., & Nemhauser, J. L. (2017). Insights into the evolution and function of auxin signaling F-Box proteins in *Arabidopsis thaliana* through synthetic analysis of natural variants. *Genetics*, *207*(2), 583–591. https://doi.org/10.1534/genetics.117.300092

Wu, M.-F., Yamaguchi, N., Xiao, J., Bargmann, B., Estelle, M., Sang, Y., & Wagner, D. (2015). Auxin-regulated chromatin switch directs acquisition of flower primordium founder fate. *eLife*, *4*, e09269. https://doi.org/10.7554/elife.09269

Xie, Y. (2018a). *Bookdown: Authoring books and technical documents with R markdown*. Retrieved from https://CRAN.R-project.org/package=bookdown

Xie, Y. (2018b). *Knitr: A general-purpose package for dynamic report generation in R*. Retrieved from https://CRAN.R-project.org/package=knitr

Xie, Y., Cheng, J., & Tan, X. (2018). *DT: A wrapper of the Javascript library 'Datatables'*. Retrieved from https://CRAN.R-project.org/package=DT

Yu, G. (2018). *Treeio: Base classes and functions for phylogenetic tree input and output*. Retrieved from https://guangchuangyu.github.io/software/treeio

Yu, G., & Lam, T. T.-Y. (2019). *Ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data*. Retrieved from https://guangchuangyu.github.io/software/ggtree

Yu, H., Zhang, Y., Moss, B. L., Bargmann, B. O., Wang, R., Prigge, M., … Estelle, M. (2015). Untethering the TIR1 auxin receptor from the SCF complex increases its stability and inhibits auxin response. *Nature Plants*, *1*(3), 14030. https://doi.org/10.1038/nplants.2014.30

Zenser, N., Ellsmore, A., Leasure, C., & Callis, J. (2001). Auxin modulates the degradation rate of Aux/IAA proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11795–11800. https://doi.org/10.1073/pnas.211312798

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., … Tian, Z. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, *33*(4), 408–414. https://doi.org/10.1038/nbt.3096

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Hamm MO, Moss BL, Leydon AR, et al. Accelerating structure-function mapping using the ViVa webtool to mine natural variation. *Plant Direct*. 2019;3:1–20. https://doi.org/10.1002/pld3.147

Research Report

A Synthetic Approach Allows Rapid Characterization of the Maize Nuclear Auxin Response Circuit[1][OPEN]

Román Ramos Báez,[a] Yuli Buckley,[b] Han Yu,[b] Zongliang Chen,[c] Andrea Gallavotti,[c,d] Jennifer L. Nemhauser,[a] and Britney L. Moss[b,2,3]

[a]University of Washington, Department of Biology, Seattle, Washington 98105
[b]Whitman College, Department of Biology, Walla Walla, Washington 99362
[c]Waksman Institute of Microbiology, Rutgers University, Piscataway, New Jersey 08854-8020
[d]Department of Plant Biology, Rutgers University, New Brunswick, New Jersey 08901

ORCID IDs: 0000-0002-0934-5342 (R.R.B.); 0000-0002-2404-6842 (Y.B.); 0000-0002-1901-2971 (A.G.); 0000-0002-8909-735X (J.L.N.); 0000-0002-1037-5717 (B.L.M.).

Auxin plays a key role across all land plants in growth and developmental processes. Although auxin signaling function has diverged and expanded, differences in the molecular functions of signaling components have largely been characterized in Arabidopsis (*Arabidopsis thaliana*). Here, we used the nuclear Auxin Response Circuit recapitulated in yeast (*Saccharomyces cerevisiae*) system to functionally annotate maize (*Zea mays*) auxin signaling components, focusing on genes expressed during the development of ear and tassel inflorescences. All 16 maize auxin/indole-3-acetic acid repressor proteins were degraded in response to auxin with rates that depended on both receptor and repressor identities. When fused to the maize TOPLESS homolog RAMOSA1 ENHANCER LOCUS2, maize auxin/indole-3-acetic acids were able to repress AUXIN RESPONSE FACTOR transcriptional activity. A complete auxin response circuit comprising all maize components, including the ZmAFB2/3 b1 maize AUXIN SIGNALING F-BOX (AFB) receptor, was fully functional. The ZmAFB2/3 b1 auxin receptor was more sensitive to hormone than AtAFB2 and allowed for rapid circuit activation upon auxin addition. These results validate the conserved role of predicted auxin response genes in maize as well as provide evidence that a synthetic approach can facilitate broader comparative studies across the wide range of species with sequenced genomes.

Auxin is an ancient molecule, and its role as a phytohormone dates back to the earliest diverging land plants (Mutte et al., 2018). In recent years, mounting evidence has emerged for both transcriptional and nontranscriptional cellular auxin response pathways that paint an increasingly complex picture of how auxin signals are detected and transmitted in plant cells. (Leyser, 2018). The gene families involved in the best characterized nuclear auxin response pathway show evidence of expansion that parallels increasing complexity in plant form. For example, the moss *Physcomitrella patens* has three members of the auxin/indole-3-acetic acid (*Aux/IAA*) gene family (Prigge et al., 2016), while the eudicot Arabidopsis (*Arabidopsis thaliana*) has 29 and the monocot maize (*Zea mays*) has 34 (Ludwig et al., 2013; Luo et al., 2018; Matthes et al., 2019). The retention of large gene families involved in this nuclear auxin response following genome duplications, in combination with the central role of auxin in plant development, has led to the hypothesis that functional diversification in this auxin response circuit underpins much of the structural and functional innovations during land plant evolution (Mutte et al., 2018). Connecting natural variation to functional divergence remains a major challenge. This problem becomes even more complicated if selection is operating on the amplitude or dynamics of a network rather than the function of any one component, a premise that itself has been quite difficult to test.

The nuclear auxin response is among the best understood signaling pathways in plants and thus is an excellent model to tackle questions about functional divergence in gene families within a single species and in network functions across multiple species. This nuclear auxin response circuit consists of three modules: (1) activation of auxin-responsive genes by AUXIN

[1]This work was supported by the National Science Foundation (grant no. IOS-1546873 to J.L.N., A.G., and B.L.M. and grant no. IOS-1456950 to A.G.), the M.J. Murdock Charitable Trust (to B.M.), Whitman College (to B.M.), and the Howard Hughes Medical Institute (Gilliam Fellowship no. GT11355 to R.R.B.).

[2]Author for contact: mossbl@whitman.edu.
[3]Senior author.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Britney L. Moss (mossbl@whitman.edu).

B.L.M. and J.L.N. conceived the original research plans; B.L.M., A.G., J.L.N., and R.R.B. supervised the experiments; R.R.B. and B.L.M. designed and performed experiments and analyzed the data; Y.B. and H.Y. performed the repression and full-circuit assays; Z.C. performed the in situ hybridizations; B.L.M., R.R.B., and J.L.N. wrote the article with input from all of the authors.

[OPEN]Articles can be viewed without a subscription.
www.plantphysiol.org/cgi/doi/10.1104/pp.19.01475

RESPONSE FACTOR (ARF) transcription factors; (2) repression of ARFs in the absence of auxin by recruitment of TOPLESS (TPL)/TOPLESS-RELATED (TPR) corepressors via Aux/IAA proteins; and (3) degradation of Aux/IAA proteins via their auxin-induced association with TRANSPORT INHIBITOR RESPONSE1 (TIR1)/AUXIN SIGNALING F-BOX (AFB) receptors (Leyser, 2018). There is some evidence to support a model where distinct combinations of these proteins acting in different cell types lead to the variety of cellular auxin responses observed across plant tissues (Vernoux et al., 2011; Rademacher et al., 2012; Bargmann et al., 2013). However, coexpression of components, feedback regulation, and interactions with other signaling pathways (Piya et al., 2014) have made it virtually impossible to conclusively support or refute this model.

The wealth of genetic, genomic, and biochemical tools in Arabidopsis have made it possible to rapidly build a strong foundational understanding of auxin response. However, to both explore the extent of shared auxin signaling properties across plants and fully interrogate the connection between natural variation in protein sequences and functional innovations in plant development, this nuclear auxin response must be examined in more species. Understanding differences between eudicots and monocots is of particular interest, as the molecular mechanism that explains the differential impacts of widely used auxinic herbicides remains a mystery, where eudicot weeds are killed while grasses are often unaffected (McSteen, 2010). In recent years, functional studies in maize have made inroads in delineating auxin response components. For example, the maize Aux/IAA protein ROOTLESS WITH UNDETECTABLE MERISTEMS1 (RUM1) is known to be functionally important in the development of the root in the embryo as well as in root branching in seedlings (von Behrens et al., 2011). BIF1 and BIF4 are Aux/IAAs known to have important functions in the development of maize inflorescences (Galli et al., 2015). Mutations in REL2, a functional homolog of TPL in Arabidopsis, can rescue TPL mutants in Arabidopsis and have pleiotropic auxin developmental effects in maize, rice (*Oryza sativa*), and Arabidopsis (Gallavotti et al., 2010; Yoshida et al., 2012; Liu et al., 2019). Maize ARFs show similar preferences for auxin response elements to Arabidopsis ARFs (Galli et al., 2018), and the specificity of activator ARFs within maize is not explained by differences in their promoter preferences (Lanctot et al., 2020).

Auxin response circuits built in yeast (*Saccharomyces cerevisiae*) cells (ARC^{Sc}) make it possible to analyze individual and combinatorial functions of Aux/IAA, TIR1/AFB, ARF, and TPL/TPR proteins (Havens et al., 2012; Pierre-Jerome et al., 2014). This heterologous system has several advantages, including precise control over the amount and duration of auxin input, highly quantitative fluorescence outputs, and the ability to study auxin signaling modules with defined connectivity and in the absence of other plant signaling pathways. Studies using Arabidopsis components in ARC^{Sc} have confirmed and extended previous studies showing that Aux/IAAs exhibit a range of tunable auxin-induced degradation rates (Ramos et al., 2001; Dreher et al., 2006; Calderón Villalobos et al., 2012; Havens et al., 2012; Moss et al., 2015) and that this variation in Aux/IAA degradation is central to controlling auxin transcriptional response dynamics and the rate of developmental events in planta (Havens et al., 2012; Pierre-Jerome et al., 2014; Guseman et al., 2015; Moss et al., 2015). The yeast system has also enabled functional characterization of genetic natural variation in auxin receptors (Wright et al., 2017) as well as functional annotation of two putative maize auxin repressor proteins (Galli et al., 2015) and several maize class A activator ARFs (Lanctot et al., 2020).

Here, we have used the ARC^{Sc} system to functionally annotate additional maize auxin signaling components, focusing on genes that are expressed during inflorescence development. The maize Aux/IAAs (ZmIAAs) tested degraded in response to auxin with rates that depended on both receptor and repressor identity. ZmIAAs were able to repress ARF-mediated transcription when fused to a truncated form of the maize TPL homolog REL2. The maize auxin receptor ZmAFB2/3 b1 was more sensitive to auxin than was observed with the fastest acting Arabidopsis receptor, AtAFB2, allowing for more rapid degradation of Aux/IAAs. Finally, a complete auxin response circuit composed of all maize components was fully functional, allowing very sensitive activation of a transcriptional reporter following the addition of auxin. These results validate the conserved role of predicted nuclear auxin response genes in maize as well as provide evidence that a synthetic approach can facilitate broader comparative studies that incorporate the wide range of species with sequenced genomes.

RESULTS

ZmIAAs Are Functional in Auxin Degradation and Repression Modules

The maize B73 genome contains 34 Aux/IAA repressor genes, 16 of which were selected for further analysis here because they are expressed in developing maize inflorescences (Davidson et al., 2011; Bolduc et al., 2012; Eveland et al., 2014) and represent members of six of the eight Aux/IAA clades found in monocots (there is a ninth clade of IAAs in Arabidopsis; Matthes et al., 2019). Out of 13 *ZmIAAs* that were tested by in situ hybridizations, six gave detectable and specific expression patterns in immature tassel inflorescence meristems and spikelet meristems (Fig. 1, A and B). In general, most of these *ZmIAA* genes revealed highly restricted expression patterns marking emerging primordia in inflorescence meristems (*ZmIAA2*, *ZmIAA5*, *ZmIAA9*, *ZmIAA14*, and *ZmIAA28*) as well as vasculature (*ZmIAA5* and *ZmIAA28*), and all showed strong expression in spikelet meristems (Fig. 1B). These

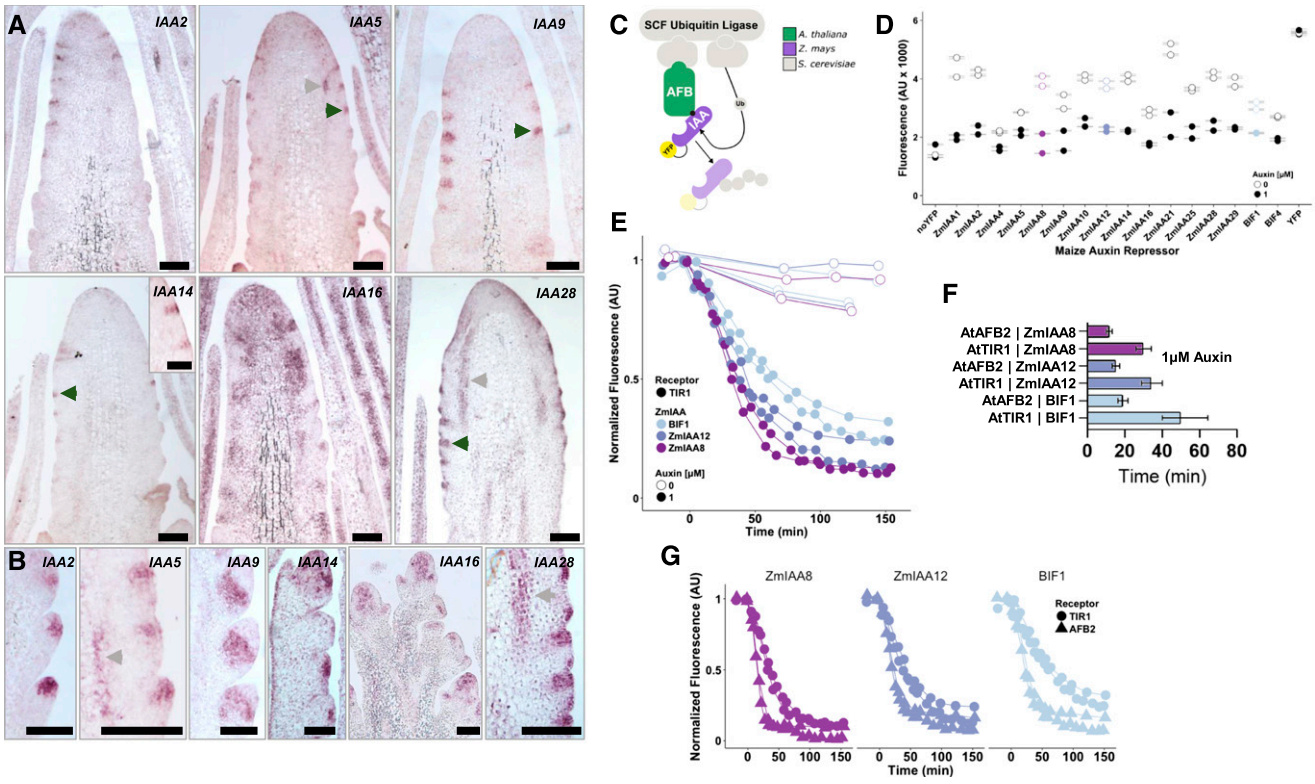


Figure 1. Auxin repressors expressed in maize inflorescence exhibited variable auxin-induced degradation dynamics. A and B, Expression of maize *Aux/IAA* genes by mRNA in situ hybridizations in immature tassels. Expression patterns are shown in inflorescence meristems (A) and spikelet meristems (B). Gray arrowheads indicate vasculature, and green arrowheads indicate axillary meristems. Bars = 100 μm and 50 μm in the inset image. C, The nuclear auxin degradation module in yeast consists of ZmIAAs (purple) tagged with YFP and coexpressed with an Arabidopsis auxin receptor (green). The F-box domain of the receptor facilitates complex formation with the yeast SCF ubiquitin (Ub) ligase machinery (gray). When yeast are exposed to auxin, shown as a black circle, the hormone acts as molecular glue that brings the coreceptor complex together and leads to ubiquitination and proteasomal degradation of the YFP-tagged ZmIAA. This results in a decrease in fluorescence over time. D, The 16 ZmIAAs coexpressed with Arabidopsis TIR1 were degraded in response to auxin. Fluorescence measurements were obtained 2 h post-auxin exposure on a flow cytometer. Data from two replicates are shown; error bars represent SE. AU, Absorbance units. E and G, ZmIAAs degrade at different rates that are dependent upon both repressor (E) and receptor (G) identities. Yeast strains expressing YFP-tagged ZmIAAs and either Arabidopsis TIR1 or AFB2 auxin receptor were exposed to 1 μM auxin or mock treatment (95% [v/v] ethanol) at 0 min, and fluorescence measurements were acquired on a flow cytometer. Data from two replicates are shown. F, YFP:ZmIAA degradation half-lives were calculated from cytometry data in E and G and are presented with 95% confidence intervals.

expression patterns strongly resembled those previously reported for *BIF1* and *BIF4* (Galli et al., 2015) and suggest a high degree of functional redundancy in this family of transcriptional regulators. However, it is not known whether auxin-dependent degradation dynamics or repression strengths may reveal differences in molecular function among individual *Aux/IAAs*. To assess the auxin sensitivity of each ZmIAA, we fused these inflorescence-expressed ZmIAAs to YFP and coexpressed them with an Arabidopsis auxin receptor in yeast (Fig. 1C). The F-box domain of these receptors interacts with the yeast *S*-Phase Kinase-Associated Protein1 (SKP1), Cullin, F-box protein (SCF) complex, allowing for the ubiquitination of Aux/IAAs and subsequent degradation by the yeast 26S proteasome. All ZmIAA proteins were expressed in yeast cells and degraded in response to auxin treatment (Fig. 1D). The ARC^{Sc} degradation module also allowed us to measure functional differences in auxin-induced degradation rates by measuring changes in fluorescence following auxin exposure (Havens et al., 2012). We selected three ZmIAAs that were expressed at similar levels and showed different auxin sensitivities (ZmIAA8, ZmIAA12, and BIF1) and measured auxin-induced degradation dynamics (Fig. 1, E–G). As seen previously for AtIAAs (Dreher et al., 2006; Havens et al., 2012), ZmIAA degradation rates varied between different ZmIAAs coexpressed with AtTIR1 (BIF1, ZmIAA8, and ZmIAA12 half-lives being 49.8, 29.7, and 33.8 min, respectively; Fig. 1, E and F) and were generally more rapid in cells expressing AtAFB2 (11.6, 15.1, and 18.8 min, respectively; Fig. 1, E and G). Previous studies demonstrated that residue substitutions within the core of the highly conserved degron can

slow the rate of Aux/IAA degradation (Ramos et al., 2001; Guseman et al., 2015). However, none of the ZmIAAs expressed in developing inflorescence had variations in the core degron other than V > I in ZmIAA12 and ZmIAA21 (Supplemental Fig. S1A). When paired with AtAFB2, auxin-induced degradation of ZmIAA12 was not the slowest (Fig. 1E).

In addition to degradation in response to auxin, the other major function of Aux/IAAs is repression of ARF-mediated transcriptional regulation, a process that is facilitated by TPL/TPR corepressors (Causier et al., 2012). The maize genome has four TPL-like corepressors, the REL2/REL2-like family, of which REL2 has been shown to have pleiotropic phenotypes associated with meristem maintenance and initiation in maize (Liu et al., 2019). In the Arabidopsis ARC^{Sc} (AtARC^{Sc}), an N-terminal fusion of the first 100 amino acids of TPL (TPLN100) was functional if directly fused to IAAs, facilitating transcriptional repression of ARFs (Pierre-Jerome et al., 2014). We confirmed that REL2 can also confer repression of ZmIAAs by fusing ZmIAA8 to either TPLN100 or REL2N91, a fragment of REL2 that is structurally analogous to TPL100 (Fig. 2A; Supplemental Fig. S1B). Based on new structural information, we designed REL2N91 to include only the first five helices that encompass the LisH and CTLH domains (Martin-Arevalillo et al., 2017). The ARC^{Sc} strains used for repression assays contained Arabidopsis AFB2 and ARF19, which was the strongest and fastest activating ARF in AtARC^{Sc} (Pierre-Jerome et al., 2014).

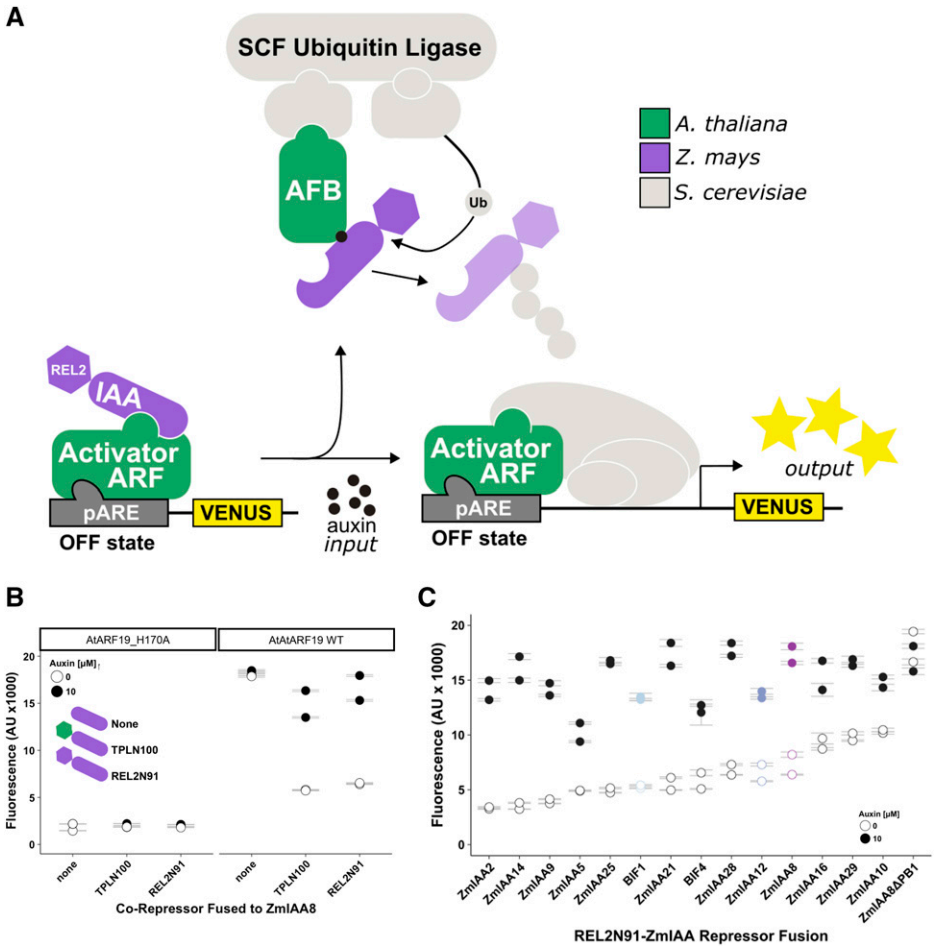


Figure 2. The TPL homolog REL2 enabled ZmIAAs to repress ARFs. A, The auxin repression module in yeast consisted of ZmIAA repressors fused to N-terminal fragments of either the Arabidopsis TPL or maize REL2 corepressors; these were coexpressed with an auxin receptor (Arabidopsis AFB2) and activator transcription factor (Arabidopsis ARF19). Auxin-induced derepression of ARF19 results in activation of the auxin response element-containing promoter (pARE) and expression of VENUS fluorescent protein. B, The N-terminal 91 amino acids of maize REL2 assist ZmIAA8 in conferring transcriptional repression on AtARF19, and this repression is relieved upon the addition of auxin. The REL2N91 fragment was directly compared with the analogous Arabidopsis TPLN100. The AtARF19_H170A mutant is unable to bind DNA, so the auxin response stays off. Strains labeled none contain a ZmIAA8 that has not been fused to a corepressor. C, ZmIAAs fused to REL2N91 exhibited different patterns of auxin-responsive gene activation, independent of their basal repression strength and degradation rate. Two ZmIAAs were unable to repress the auxin response (Supplemental Fig. S1D). All fluorescence measurements in B and C were made 5 h post-treatment. Data from two replicates are shown; error bars represent SE. AU, Absorbance units; WT, wild type.

Each corepressor conferred a similar degree of repression, and in the presence of auxin that repression was relieved and transcription was activated to similar degrees (Fig. 2B). Both the degree of repression and auxin-induced activation dynamics varied greatly across ZmIAAs, and repression level did not necessarily predict activation level (Fig. 2C; Supplemental Fig. S1C). Two ZmIAAs were unable to repress the nuclear auxin response (Supplemental Fig. S1D), possibly due to poor expression or inability to interact with ARF19.

Maize Auxin Receptor ZmAFB2/3 b1 Is More Auxin Sensitive Than AtAFB2

In the maize genome, there are eight members of the TIR1/AFB auxin receptor family (Matthes et al., 2019), four of which appear to be related to the Arabidopsis AFB2/3 clade. Two of these AFB2/3-like maize auxin receptors (ZmAFB2/3 a and b1) were highly expressed in inflorescence meristems and so were tested for activity within an ARC^{Sc} degradation module (Bolduc et al., 2012; Eveland et al., 2014). The ZmAFB2/3 b1 receptor was the only one able to induce the degradation of ZmIAAs in the presence of auxin (Fig. 3A). This receptor is also the most highly expressed in developing inflorescences (Eveland et al., 2014). Alignments and pairwise comparisons between ZmAFB2/3 b1 and Arabidopsis auxin receptors showed that they share 60% to 67% sequence similarity across both the F-box and LRR domains (Supplemental Fig. S2, A–D). Auxin-induced degradation carried out by ZmAFB2/3 b1 was even faster than AtAFB2, with ZmIAA10 half-lives of 4.9 and 9.9 min, respectively (Fig. 3, A and B). We next assessed the expression of the auxin receptors in the yeast strains and found that all were expressed and accumulated to similar levels (Fig. 3C; Supplemental Fig. S2, E and F), suggesting that differences in expression level are not likely to explain differences in degradation module activity. Furthermore, the amount of AtAFB2 receptor accumulation in yeast (within a twofold range) was previously shown to have no effect

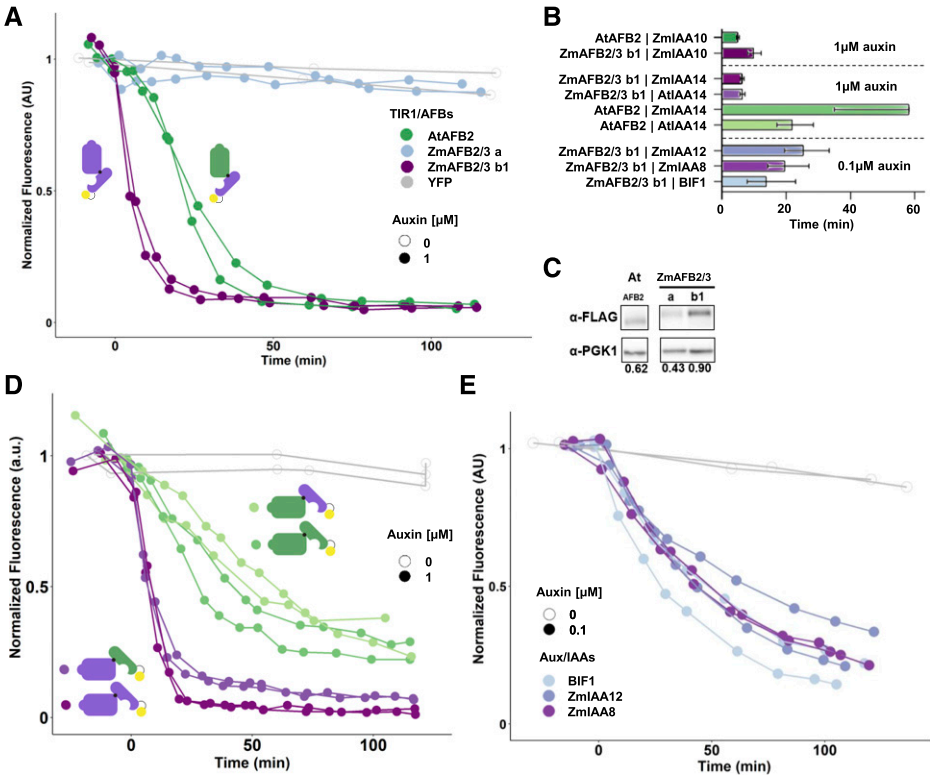


Figure 3. The maize auxin receptor ZmAFB2/3 b1 exhibited a higher basal activity compared with the Arabidopsis ortholog. A, ZmAFB2/3 b1 was sensitive to auxin and exhibited faster degradation than AtAFB2. ZmAFB2/3 a was not sensitive to auxin. Yeast coexpressing an Arabidopsis or maize auxin receptor and YFP-ZmIAA10 were treated with 1 μM auxin or mock (0), and fluorescence was measured on a flow cytometer. B, ZmIAA half-lives were calculated with 95% confidence intervals for data shown in A, D, and E. C, The auxin receptor expression level in yeast does not necessarily correlate with functionality. Yeast lysates were probed with anti-FLAG (for receptors) or anti-PGK1 (loading control) antibodies. Fold expression values shown below the bands were calculated by using ImageJ to quantify the intensity of each band and dividing the intensity of the receptor band by the intensity of the PGK1 band. D, ZmAFB2/3 b1 always exhibited faster IAA degradation in response to auxin, whether paired with an Arabidopsis or a maize Aux/IAA. All combinations of ZmAFB2/3 b1 or AtAFB2 paired with AtIAA14 or ZmIAA14 were tested within an auxin degradation module. E, When paired with ZmAFB2/3 b1, three maize Aux/IAAs (BIF1, ZmIAA8, and ZmIAA12) exhibited different orders of degradation speed than when paired with AtAFB2 (Fig. 1). AU, Absorbance units.

on auxin-induced repressor degradation rate (Havens et al., 2012). Hypermorphic variations like those described in AtAFB2 natural variants might explain the increased activity of ZmAFB2/3 b1 (Wright et al., 2017). Although the known AtAFB2 hypermorphic variant D176E was not present in ZmAFB2/3 b1, the T491R substitution is at a known hypermorphic site (T491V variant found in AtAFB2) and might explain differences in activity (Supplemental Fig. S2A).

The auxin sensitivity of ZmAFB2/3 b1 and AtAFB2 was next tested with Aux/IAAs from each species to determine which component was primarily responsible for driving differences in speed of the degradation module. The maize receptor was faster than AtAFB2 when paired with either a maize or an Arabidopsis Aux/IAA (Fig. 3, B and D). We also noticed that preauxin Aux/IAA accumulation levels were lower in strains expressing ZmAFB2/3 b1 compared with AtAFB2 (Supplemental Fig. S3A). We hypothesized that the maize receptor might be interacting with the repressor independently of auxin. Treatment with auxinole, a compound shown to block the interaction of rice TIR1 (OsTIR1) and auxin repressors (Yesbolatova et al., 2019), resulted in greater accumulation of Aux/IAAs, particularly in yeast strains expressing ZmAFB2/3 b1 (Supplemental Fig. S3A). This suggested that at least some of the increased auxin sensitivity of ZmAFB2/3 b1 may be due to auxin-independent degradation. Auxin sensitivity analyses showed that ZmAFB2/3 b1 is responsive to auxin over 2 orders of magnitude (0.01–1 μM; Supplemental Fig. S3B). We next assessed whether ZmAFB2/3 b1 degradation modules were functional with each of the 16 ZmIAAs. All ZmIAAs were degraded by ZmAFB2/3 b1 (Fig. 3E; Supplemental Fig. S3C). Furthermore, some ZmIAAs exhibited different patterns of auxin-induced degradation dynamics than were observed when coexpressed with AtAFB2: for example, BIF1 paired with ZmAFB2/3 b1 degraded faster (13.9 min) than ZmIAA8 (19.7 min) or ZmIAA12 (25.4 min) when paired with AtAFB2 (Figs. 1D and 3E; note also the 10-fold difference in auxin sensitivities). This is reminiscent of the different patterns of AtIAA degradation rates observed for Arabidopsis AFB2 and TIR1 auxin receptors (Havens et al., 2012). The auxin repressors that possess V > I variation in the core degron (ZmIAA12 and ZmIAA21; Supplemental Fig. S1A) exhibited consistently slow degradation rates compared to most other ZmIAAs when paired with ZmAFB2/3 b1 (Fig. 3E; Supplemental Fig. S3C), in contrast with the AtAFB2 data (Fig. 1E). Thus, maize auxin degradation modules constituted in yeast are more sensitive to hormone, and degradation dynamics are dependent upon both receptor and repressor identity.

The Maize Auxin Response Circuit Is Tuned to Respond to Low Auxin Concentrations

Having determined that maize receptors, repressors, and corepressors are functional in varying combinations, the final step was to express them with a maize auxin transcription factor to reconstitute an entire maize auxin response circuit (ZmARC^{Sc}; Fig. 4A). ZmARF27 is orthologous to AtARF19, the ARF used in the AtARC^{Sc}, and is expressed strongly in immature tassel (Eveland et al., 2014; Matthes et al., 2019). ZmARC^{Sc} yeast strains containing various ZmIAAs were able to repress ZmARF27 transcriptional activity to varying degrees (Fig. 4B), conferring as strong or stronger repression on ZmARF27 than on AtARF19 (Supplemental Fig. S4A). The ZmARC^{Sc} strains were all derepressed by the addition of a high concentration of auxin (Fig. 4B; Supplemental Fig. S4A). Given that ZmAFB2/3 b1 is more sensitive to auxin than AtAFB2 in auxin degradation modules (Fig. 3), we hypothesized that this would also result in ZmARC^{Sc} transcriptional responses activating at lower doses of auxin. For each ZmARC^{Sc} strain tested, the transcriptional response was activated at lower auxin levels compared with strains in which the ZmAFB2/3 b1 was replaced with AtAFB2 (Fig. 4C; Supplemental Fig. S4B). Thus, the maize auxin response circuit is exquisitely sensitive to auxin hormone levels, and response dynamics can be tuned by altering ZmIAA identity.

DISCUSSION

The synthetic auxin response circuit used in this study is impressively modular, with the ability to mix and match components from distantly related plant species. Maize ZmIAAs expressed in developing inflorescence tissues degraded in response to auxin, with variable degradation rates that depended on both ZmIAA and receptor identity (Fig. 1). Furthermore, these ZmIAAs were able to repress transcription when paired with the maize corepressor REL2 (Fig. 2). Maize auxin degradation modules were more sensitive to auxin than orthologous Arabidopsis degradation modules in yeast (Fig. 3). Finally, a maize auxin response circuit featuring ZmARF27 was fully functional and more sensitive to auxin than circuits containing an Arabidopsis receptor (Fig. 4). This work represents a broadly useful strategy for employing synthetic biology approaches to functionally annotate and characterize genetic diversity within conserved signaling networks.

The decreasing costs of genome sequencing have led to a dramatic increase in genetic natural variation information available. However, functional annotation of this new genetic information is lagging behind. The ARC^{Sc} platform allows for rapid functional validation and characterization of large gene families, as seen previously for auxin signaling modules within Arabidopsis (Havens et al., 2012; Pierre-Jerome et al., 2014, 2016; Moss et al., 2015; Wright et al., 2017) and now for maize (Figs. 1 and 2). Of particular note, the yeast system enabled multidimensional functional annotation (degradation and repression behaviors) of 16 different Aux/IAAs expressed in developing maize inflorescences. We observed variable degradation and repression behaviors across the ZmIAAs studied, providing

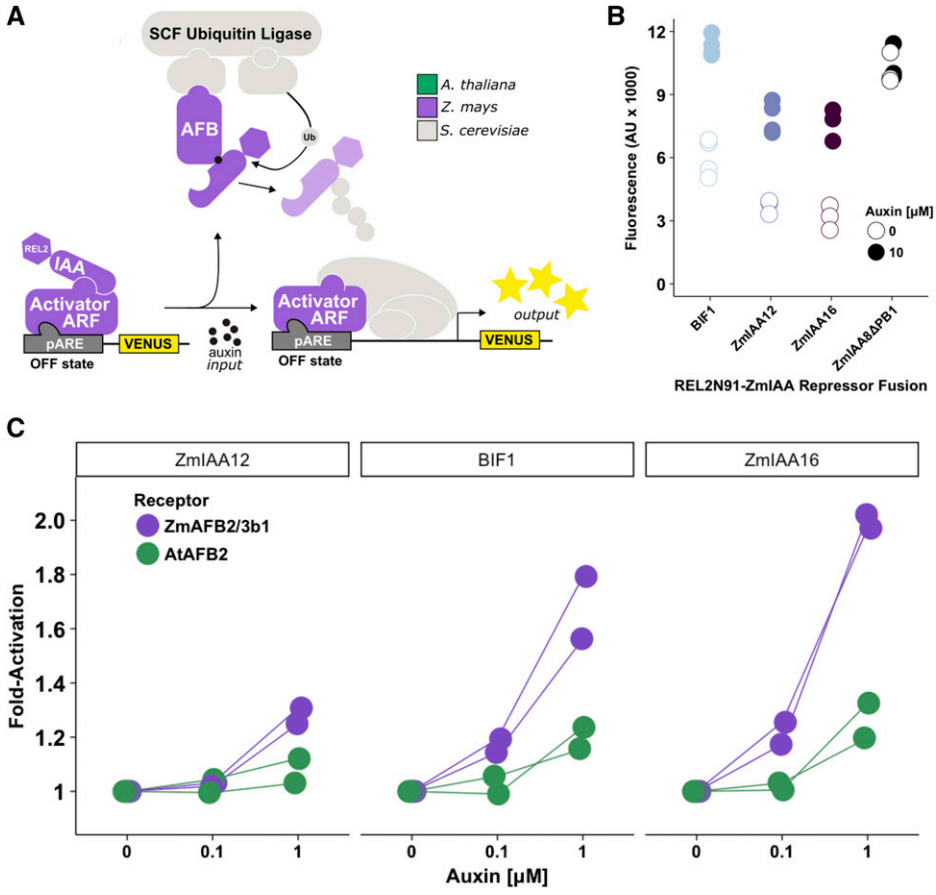


Figure 4. A maize auxin response circuit featuring ZmAFB2/3 b1 is functional and more sensitive to auxin than a circuit with AtAFB2. A, An auxin response circuit in yeast utilizing maize components ($ZmARC^{Sc}$) was assembled as shown with ZmARF27. B, The $ZmARC^{Sc}$ is functional and responds to auxin. Data represent three to four replicates of fluorescence measurements taken 5 h following auxin hormone treatment. The ZmIAA8 strain with a deleted PB1 domain represents an always-on state. AU, Absorbance units. C, The ZmAFB2/3 b1 auxin receptor confers higher auxin sensitivity than AtAFB2 in $ZmARC^{Sc}$ yeast strains. Strains shown in B and their AtAFB2 cognates were treated with 0, 0.1, or 1 μM auxin for 5 h before fluorescence measurements. Data are plotted as fold untreated.

evidence of biochemical differences that introduce another layer of complexity on top of known phylogenetic diversity (Ludwig et al., 2013) and tissue-specific expression patterns (Matthes et al., 2019). In agreement with previous studies (von Behrens et al., 2011; Ludwig et al., 2013; Zhang et al., 2016), ZmIAA2, ZmIAA10/RUM1, ZmIAA20/BIF4, and ZmIAA29/RUL1 were functional within auxin degradation modules (Figs. 1 and 3) and capable of repressing transcription (Fig. 2). Furthermore, by showing that ZmIAAs can repress the auxin circuit with the assistance of REL2, we provide an additional piece of evidence to support existing genetic and biochemical data demonstrating that REL2 is a functional homolog of TPL (Liu et al., 2019). Thus, our functional annotation and characterization of maize auxin signaling modules is in agreement with the few existing genetic and biochemical studies on maize Aux/IAAs and REL2 and extends functional characterization to many previously uninvestigated maize auxin signaling components for which there are no known mutants. The recent recapitulation of the Arabidopsis abscisic acid hormone signaling pathway in yeast underscores the feasibility of performing similar functional annotation efforts for putative orthologous proteins across many plant signaling pathways (Ruschhaupt et al., 2019). Admittedly these minimal signaling modules do not encompass the full complexity present in vivo, although the ARC^{Sc} has been used to investigate competition between corepressors and the functional role of IAA and ARF PB1 domain oligomerization (Pierre-Jerome et al., 2014, 2016) as well as ARF promoter architecture preferences (Lanctot et al., 2020).

The ARC^{Sc} system is a valuable platform for continued dissecting of the differences in auxin response between monocots and dicots. For example, ARC^{Sc} should be useful in elucidating the mechanisms of auxinic herbicide tolerance and resistance from the perspectives of chemical biology, evolution, and genetic engineering; however, further work will be needed to overcome the observed limitation that 2,4-dichlorophenoxyacetic acid and other auxinic herbicides are actively transported out of yeast cells (Cabrito et al., 2011). In monocots, responses to most auxinic herbicides are much weaker than in eudicots, which die when these herbicides are applied to them (McSteen, 2010). Auxin metabolism and transport diverge between monocots and eudicots in ways that could contribute to a divergence in their auxin response (Zhao et al., 2013; Lu et al., 2015; He et al., 2018, 2019). Monocot differences in auxin signaling and reception components might also contribute to these different responses. Because maize and Arabidopsis both have similar numbers of members in nuclear auxin signaling gene families (TIR1/AFBs, Aux/IAAs, ARFs, etc.), we believe that any auxin response differences present at

this level would more likely be driven by differences in protein function than by differences in pathway protein stoichiometries. Our results bring up exciting new questions about the role of auxin in maize. In the yeast system, ZmAFB2/3 b1 is more sensitive to lower auxin dosages than AtAFB2 and can lead to Aux/IAA turnover even in the absence of auxin (Fig. 3). Analysis of relative auxin levels during tassel and ear development, perhaps using the DII-VENUS reporter (Mir et al., 2017), could be used to validate these findings in planta and provide possible clues to the impact of increased sensitivity on feedback within the auxin system and/or on the dynamics of downstream responses.

Understanding how hormone signaling protein functions have changed (or remained similar) throughout evolution can help connect protein function to growth and developmental processes and inform the design of synthetic proteins. ARC^{Sc} has revealed the functional ramifications of the evolution of auxin signaling components within Arabidopsis, allowing for the functional comparison of gene variants across accessions (Wright et al., 2017). By allowing for quantitative comparisons between Arabidopsis and maize auxin signaling pathways, ARC^{Sc} is a promising tool to study the evolution of molecular function more broadly across distantly related plant species, in this case diverging 250 million years ago. It also provides the ability to directly and quickly study multiple functional protein interactions to help elucidate the evolution of widespread developmental phenomena such as inflorescence development. From a synthetic biology perspective, platforms like ARC^{Sc} can enable the rapid design, characterization, and comparison of highly divergent proteins for use in engineering synthetic signaling machinery within plants or other biological systems.

The types of experiments and analyses described here that utilize a synthetic hormone signaling system in yeast are readily carried out by novice scientists in research laboratories and in course settings. Many of the experiments described in this article were piloted or executed by 45 undergraduate students at a primarily undergraduate institution, Whitman College. Most of these students were participants in several semesters of course-based undergraduate research experiences. This type of undergraduate-focused synthetic biology research experience offers an exciting way to acquaint young scientists with plant synthetic biology and could be readily adapted to focus on the standardization and characterization of parts for use in synthetic biology, an area of continuing concern (Decoene et al., 2018).

MATERIALS AND METHODS

In Situ Hybridizations in Maize Inflorescences

For mRNA in situ hybridizations, 0.2- to 0.4-cm tassel primordia from the maize (*Zea mays*) B73 inbred line were dissected and fixed in cold paraformaldehyde acetic acid solution as described previously (Galli et al., 2015). Hybridizations were carried out at 59°C. Antisense in situ probes for all 13 *Aux/IAA* genes (*IAA2*, *IAA4*, *IAA5*, *IAA8*, *IAA9*, *IAA10*, *IAA12*, *IAA14*, *IAA16*, *IAA21*, *IAA25*, *IAA28*, and *IAA29*) were synthesized by in vitro transcription (T7 RNA polymerase; Promega) of the entire or partial coding sequences cloned in pENTR vector or pGEM T-easy (Promega) and digested with respective endonuclease enzymes. The vectors and primers used for probe design are listed in Supplemental Table S1.

Plasmid Construction

Plasmid and primer design was performed in Benchling. Maize *Aux/IAA* and *ARF* sequences were obtained from the Grassius Database or synthesized by Integrated DNA Technologies. The maize REL2 fragment was cloned from plasmids generated by Liu et al. (2019). The maize *TIR1/AFB* gene sequences were obtained from Paula McSteen. These sequences were ordered from Integrated DNA Technologies with codon optimization for *Saccharomyces cerevisiae* and then cloned into pCR-BLUNT plasmid using the Zero Blunt TOPO PCR Cloning Kit (James et al., 2000). The *ZmIAA* genes were inserted into pGP4GY plasmids (Havens et al., 2012) via Gateway cloning (Life Technologies) or by Gibson cloning (Gibson et al., 2009). *ZmARF27* was cloned into the pGP8G vector by Gibson cloning. Auxin receptors (*TIR1* and *AFB2*s) were cloned into pGP8GF plasmid containing a 3X-FLAG-6X-HIS tandem affinity purification tag. Each PCR amplification was performed using Q5 High-Fidelity DNA Polymerase (New England Biolabs), and the products were purified using the EZNA Cycle Pure Kit or the NEB Monarch PCR & DNA Cleanup Kit and confirmed by sequencing (MCLab or Genewiz). All plasmids used in this research are listed in Supplemental Table S2, and a subset of these have been made available at Addgene.

Yeast Methods

The yeast *S. cerevisiae* were was grown in either yeast peptone dextrose (YPD) or Synthetic Complete (SC) medium made according to standard recipes and supplemented with 80 mg mL^{-1} adenine. Yeast transformation of linearized plasmids was performed using a standard lithium acetate protocol (Gietz and Woods, 2002). Matings were performed by coinoculation of MATa and MATα strains at low density in YPD medium with shaking overnight at 30°C. All yeast transformations and matings were streaked onto SC plates lacking appropriate auxotrophic compounds for selection, followed by isostreaking onto yeast peptone adenine dextrose (YPAD) plates prior to glycerol stocking. The full yeast strain list is included in Supplemental Table S3.

Auxin Degradation Module Assays

Auxin degradation module assays were carried out as previously described (Pierre-Jerome et al., 2017). In brief, yeast colonies from YPAD plates were used to inoculate 3 mL of SC medium in culture tubes. Cell density (in events μL^{-1}) was estimated by flow cytometry, and cultures were diluted into 4 mL of fresh SC medium in culture tubes such that cells were in log-phase growth 16 h later and for the duration of the assay (typically in the range of 0.5 to 2.5 events μL^{-1} was sufficient). All growth was at 30°C in an incubated shaker at 250 rpm. Fluorescence measurements were taken on the flow cytometer prior to the addition of auxin to establish a baseline. Auxin (indicated concentrations, IAA) or mock (95% [v/v] ethanol) treatments were added, and cultures were returned to the incubator. Fluorescence measurements were acquired at 10-min intervals for the first 1 h after auxin addition and every 15 min thereafter until the fluorescence level in most strains had plateaued (approximately 2 h). Control strains were measured every 1 h for the duration of the experiment. For auxinole assays, yeast were grown overnight in 2 mL of SC medium and diluted 1:200 into fresh SC medium containing auxinole (50 μM in ethanol) or mock (ethanol); fluorescence measurement was taken 6 h following treatment.

Repression/Activation Module Assays

Auxin repression/activation module assays were carried out as previously described (Pierre-Jerome et al., 2017). Briefly, yeast cells inoculated from colonies on a YPAD plate were grown overnight in culture tubes containing 2 mL of SC medium at 30°C in an incubated shaker at 250 rpm. In the morning, the cells were diluted 1:200 into 2 mL of fresh SC medium in culture tubes and returned to the incubator. After 2 h, auxin (indicated concentrations) or mock treatment (95% [v/v] ethanol) were added and strains continued to incubate at 30°C for 5 h. Fluorescence measurements were taken on a flow cytometer.

Flow Cytometry Settings and Flow Set Data Analysis

Fluorescence measurements were taken with a BD Accuri C6 flow cytometer and CSampler plate adapter using an excitation wavelength of 488 and an emission detection filter at 533 nm. A total of 10,000 to 20,000 events above a 400,000 FSC-H threshold (to exclude debris) were collected for each sample, and data were exported as FCS 3.0 files for processing using the flowCore R software package and custom R scripts (Havens et al., 2012; Pierre-Jerome et al., 2017). These scripts enable gating for live, diploid, singlet yeast cells; for time-course data, they also facilitate the subtraction of background fluorescence and fold initial normalization. Data from at least two independent replicates were combined and plotted in R. Half-life values with 95% confidence intervals were calculated by importing fold-initialized auxin degradation time series data to GraphPad Prism 8 and using nonlinear regression analysis (plateau followed by one-phase decay with X0 set to the time of auxin addition). All data can be found in Flow Repository.

Immunoblotting

Yeast were grown overnight in 5 mL of liquid YPD medium at 30°C with shaking at 225 rpm. The next morning, cultures were diluted to an OD of 0.2 into fresh YPD medium and grown to an OD of ~1. Cells were then pelleted and resuspended in 1% (w/v) SDS, 8 M urea, 10 mM MOPS, pH 6.8, 10 mM EDTA, and 0.01% (w/v) bromophenol blue buffer with 1 mM phenylmethylsulfonyl fluoride. Cell pellets were lysed by glass bead disruption in lysis buffer containing protease inhibitors. Aliquots of 7 μL of yeast proteins in buffer were loaded onto a 10% SDS-PAGE gel (Bio-Rad). Anti-FLAG antibody (Cell Signaling Technology) was used to probe for TIR1/AFBs-3X-FLAG-6X-HIS, and anti-PGK1 antibody (Invitrogen) was used as a loading control.

Protein Sequence Alignments

All protein sequence alignments were performed using ClustalW in MegaX software. Alignments were visualized using CLC Workbench.

Accession Numbers

All names and accession numbers of major genes and proteins used in our experiments are listed in Supplemental Table S2.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. The N-terminal fragment of REL2 fused to ZmIAAs repressed transcription.

Supplemental Figure S2. The maize auxin AFB2/3 b1 receptor protein is orthologous to those in other species and expressed at similar levels in yeast.

Supplemental Figure S3. ZmAFB2/3 b1 degraded ZmIAAs more readily in the absence of auxin and was more sensitive to auxin.

Supplemental Figure S4. Auxin response circuit behavior depended on the identity of the ARF, TIR1/AFB, and Aux/IAA.

Supplemental Table S1. Vectors and primers used for *Aux/IAA* cloning and expression analysis.

Supplemental Table S2. Gene identifiers and plasmids used in yeast transformations.

Supplemental Table S3. Yeast strains used in this study.

ACKNOWLEDGMENTS

We thank members of the Nemhauser and Imaizumi labs for helpful discussion and guidance on experimental design and execution, Alex Leydon for assistance with REL2N91 designs, Paula McSteen and Norman Best for providing maize receptor sequences and for advice on the article, and especially the students in B.L.M.'s Synthetic Cell Biology classes, Tristan Cates, Kristina Jackson, Kathleen Daly-Jensen, Hannah Klaeser, and Alex Koriath, for initial experimental work.

Received November 27, 2019; accepted February 22, 2020; published March 2, 2020.

LITERATURE CITED

Bargmann BOR, Vanneste S, Krouk G, Nawy T, Efroni I, Shani E, Choe G, Friml J, Bergmann DC, Estelle M, et al (2013) A map of cell type-specific auxin responses. Mol Syst Biol **9:** 688

Bolduc N, Yilmaz A, Mejia-Guerra MK, Morohashi K, O'Connor D, Grotewold E, Hake S (2012) Unraveling the KNOTTED1 regulatory network in maize meristems. Genes Dev **26:** 1685–1690

Cabrito TR, Teixeira MC, Singh A, Prasad R, Sá-Correia I (2011) The yeast ABC transporter Pdr18 (ORF YNR070w) controls plasma membrane sterol composition, playing a role in multidrug resistance. Biochem J **440:** 195–202

Calderón Villalobos LIA, Lee S, De Oliveira C, Ivetac A, Brandt W, Armitage L, Sheard LB, Tan X, Parry G, Mao H, et al (2012) A combinatorial TIR1/AFB-Aux/IAA co-receptor system for differential sensing of auxin. Nat Chem Biol **8:** 477–485

Causier B, Ashworth M, Guo W, Davies B (2012) The TOPLESS interactome: A framework for gene repression in Arabidopsis. Plant Physiol **158:** 423–438

Davidson RM, Hansey CN, Gowda M, Childs KL, Lin H, Vaillancourt B, Sekhon RS, de Leon N, Kaeppler SM, Jiang N, et al (2011) Utility of RNA sequencing for analysis of maize reproductive transcriptomes. Plant Genome **4:** 191–203

Decoene T, De Paepe B, Maertens J, Coussement P, Peters G, De Maeseneire SL, De Mey M (2018) Standardization in synthetic biology: An engineering discipline coming of age. Crit Rev Biotechnol **38:** 647–656

Dreher KA, Brown J, Saw RE, Callis J (2006) The Arabidopsis Aux/IAA protein family has diversified in degradation and auxin responsiveness. Plant Cell **18:** 699–714

Eveland AL, Goldshmidt A, Pautler M, Morohashi K, Liseron-Monfils C, Lewis MW, Kumari S, Hiraga S, Yang F, Unger-Wallace E, et al (2014) Regulatory modules controlling maize inflorescence architecture. Genome Res **24:** 431–443

Gallavotti A, Long JA, Stanfield S, Yang X, Jackson D, Vollbrecht E, Schmidt RJ (2010) The control of axillary meristem fate in the maize ramosa pathway. Development **137:** 2849–2856

Galli M, Khakhar A, Lu Z, Chen Z, Sen S, Joshi T, Nemhauser JL, Schmitz RJ, Gallavotti A (2018) The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. Nat Commun **9:** 4526

Galli M, Liu Q, Moss BL, Malcomber S, Li W, Gaines C, Federici S, Roshkovan J, Meeley R, Nemhauser JL, et al (2015) Auxin signaling modules regulate maize inflorescence architecture. Proc Natl Acad Sci USA **112:** 13372–13377

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA III, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods **6:** 343–345

Gietz RD, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol **350:** 87–96

Guseman JM, Hellmuth A, Lanctot A, Feldman TP, Moss BL, Klavins E, Calderón Villalobos LI, Nemhauser JL (2015) Auxin-induced degradation dynamics set the pace for lateral root development. Development **142:** 905–909

Havens KA, Guseman JM, Jang SS, Pierre-Jerome E, Bolten N, Klavins E, Nemhauser JL (2012) A synthetic approach reveals extensive tunability of auxin signaling. Plant Physiol **160:** 135–142

He Q, Yang L, Hu W, Zhang J, Xing Y (2018) Overexpression of an auxin receptor OsAFB6 significantly enhanced grain yield by increasing cytokinin and decreasing auxin concentrations in rice panicle. Sci Rep **8:** 14051

He Y, Yan L, Ge C, Yao XF, Han X, Wang R, Xiong L, Jiang L, Liu CM, Zhao Y (2019) *PINOID* is required for formation of the stigma and style in rice. Plant Physiol **180:** 926–936

James KD, Hughes MA, Williams PA (2000) Cloning and expression of *ntnD*, encoding a novel NAD(P)$^+$-independent 4-nitrobenzyl alcohol

dehydrogenase from *Pseudomonas* sp. Strain TW3. J Bacteriol **182:** 3136–3141

Lanctot A, Taylor-Teeples M, Oki EA, Nemhauser J (2020) Specificity in auxin responses is not explained by the promoter preferences of activator ARFs. Plant Physiol **182:** 1533–1536

Leyser O (2018) Auxin signaling. Plant Physiol **176:** 465–479

Liu X, Galli M, Camehl I, Gallavotti A (2019) RAMOSA1 ENHANCER LOCUS2-mediated transcriptional repression regulates vegetative and reproductive architecture. Plant Physiol **179:** 348–363

Lu G, Coneva V, Casaretto JA, Ying S, Mahmood K, Liu F, Nambara E, Bi YM, Rothstein SJ (2015) OsPIN5b modulates rice (Oryza sativa) plant architecture and yield by changing auxin homeostasis, transport and distribution. Plant J **83:** 913–925

Ludwig Y, Zhang Y, Hochholdinger F (2013) The maize (*Zea mays* L.) AUXIN/INDOLE-3-ACETIC ACID gene family: Phylogeny, synteny, and unique root-type and tissue-specific expression patterns during development. PLoS ONE **8:** e78859

Luo J, Zhou JJ, Zhang JZ (2018) Aux/IAA gene family in plants: Molecular structure, regulation, and function. Int J Mol Sci **19:** E259

Martin-Arevalillo R, Nanao MH, Larrieu A, Vinos-Poyo T, Mast D, Galvan-Ampudia C, Brunoud G, Vernoux T, Dumas R, Parcy F (2017) Structure of the *Arabidopsis* TOPLESS corepressor provides insight into the evolution of transcriptional repression. Proc Natl Acad Sci USA **114:** 8107–8112

Matthes MS, Best NB, Robil JM, Malcomber S, Gallavotti A, McSteen P (2019) Auxin EvoDevo: Conservation and diversification of genes regulating auxin biosynthesis, transport, and signaling. Mol Plant **12:** 298–320

McSteen P (2010) Auxin and monocot development. Cold Spring Harb Perspect Biol **2:** a001479

Mir R, Aranda LZ, Biaocchi T, Luo A, Sylvester AW, Rasmussen CG (2017) A DII domain-based auxin reporter uncovers low auxin signaling during telophase and early G1. Plant Physiol **173:** 863–871

Moss BL, Mao H, Guseman JM, Hinds TR, Hellmuth A, Kovenock M, Noorassa A, Lanctot A, Villalobos LIAC, Zheng N, et al (2015) Rate motifs tune auxin/indole-3-acetic acid degradation dynamics. Plant Physiol **169:** 803–813

Mutte SK, Kato H, Rothfels C, Melkonian M, Wong GKS, Weijers D (2018) Origin and evolution of the nuclear auxin response system. eLife **7:** e33399

Pierre-Jerome E, Jang SS, Havens KA, Nemhauser JL, Klavins E (2014) Recapitulation of the forward nuclear auxin response pathway in yeast. Proc Natl Acad Sci USA **111:** 9407–9412

Pierre-Jerome E, Moss BL, Lanctot A, Hageman A, Nemhauser JL (2016) Functional analysis of molecular interactions in synthetic auxin response circuits. Proc Natl Acad Sci USA **113:** 11354–11359

Pierre-Jerome E, Wright RC, Nemhauser JL (2017) Characterizing auxin response circuits in Saccharomyces cerevisiae by flow cytometry. Methods Mol Biol **1497:** 271–281

Piya S, Shrestha SK, Binder B, Stewart CN Jr., Hewezi T (2014) Protein-protein interaction and gene co-expression maps of ARFs and Aux/IAAs in Arabidopsis. Front Plant Sci **5:** 744

Prigge MJ, Greenham K, Zhang Y, Santner A, Castillejo C, Mutka AM, O'Malley RC, Ecker JR, Kunkel BN, Estelle M (2016) The Arabidopsis auxin receptor F-box proteins AFB4 and AFB5 are required for response to the synthetic auxin picloram. G3 (Bethesda) **6:** 1383–1390

Rademacher EH, Lokerse AS, Schlereth A, Llavata-Peris CI, Bayer M, Kientz M, Freire Rios A, Borst JW, Lukowitz W, Jürgens G, et al (2012) Different auxin response machineries control distinct cell fates in the early plant embryo. Dev Cell **22:** 211–222

Ramos JA, Zenser N, Leyser O, Callis J (2001) Rapid degradation of auxin/indoleacetic acid proteins requires conserved amino acids of domain II and is proteasome dependent. Plant Cell **13:** 2349–2360

Ruschhaupt M, Mergner J, Mucha S, Papacek M, Doch I, Tischer SV, Hemmler D, Chiasson D, Edel KH, Kudla J, et al (2019) Rebuilding core abscisic acid signaling pathways of Arabidopsis in yeast. EMBO J **38:** e101859

Vernoux T, Brunoud G, Farcot E, Morin V, Van den Daele H, Legrand J, Oliva M, Das P, Larrieu A, Wells D, et al (2011) The auxin signalling network translates dynamic input into robust patterning at the shoot apex. Mol Syst Biol **7:** 508

von Behrens I, Komatsu M, Zhang Y, Berendzen KW, Niu X, Sakai H, Taramino G, Hochholdinger F (2011) Rootless with undetectable meristem 1 encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and post-embryonic lateral root initiation in maize. Plant J **66:** 341–353

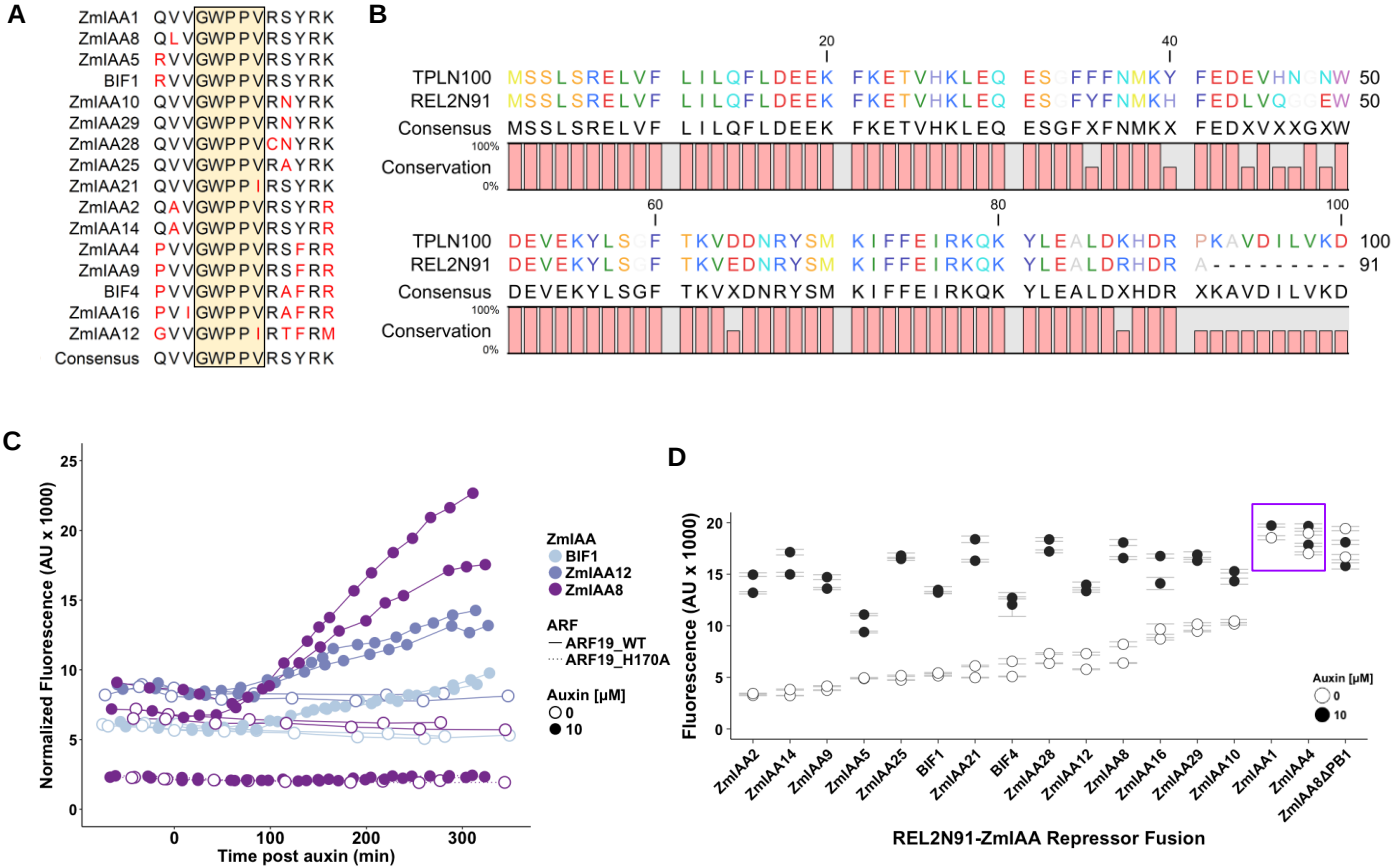
Wright RC, Zahler ML, Gerben SR, Nemhauser JL (2017) Insights into the evolution and function of auxin signaling F-box proteins in *Arabidopsis thaliana* through synthetic analysis of natural variants. Genetics **207:** 583–591

Yesbolatova A, Natsume T, Hayashi KI, Kanemaki MT (2019) Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole. Methods **164–165:** 73–80

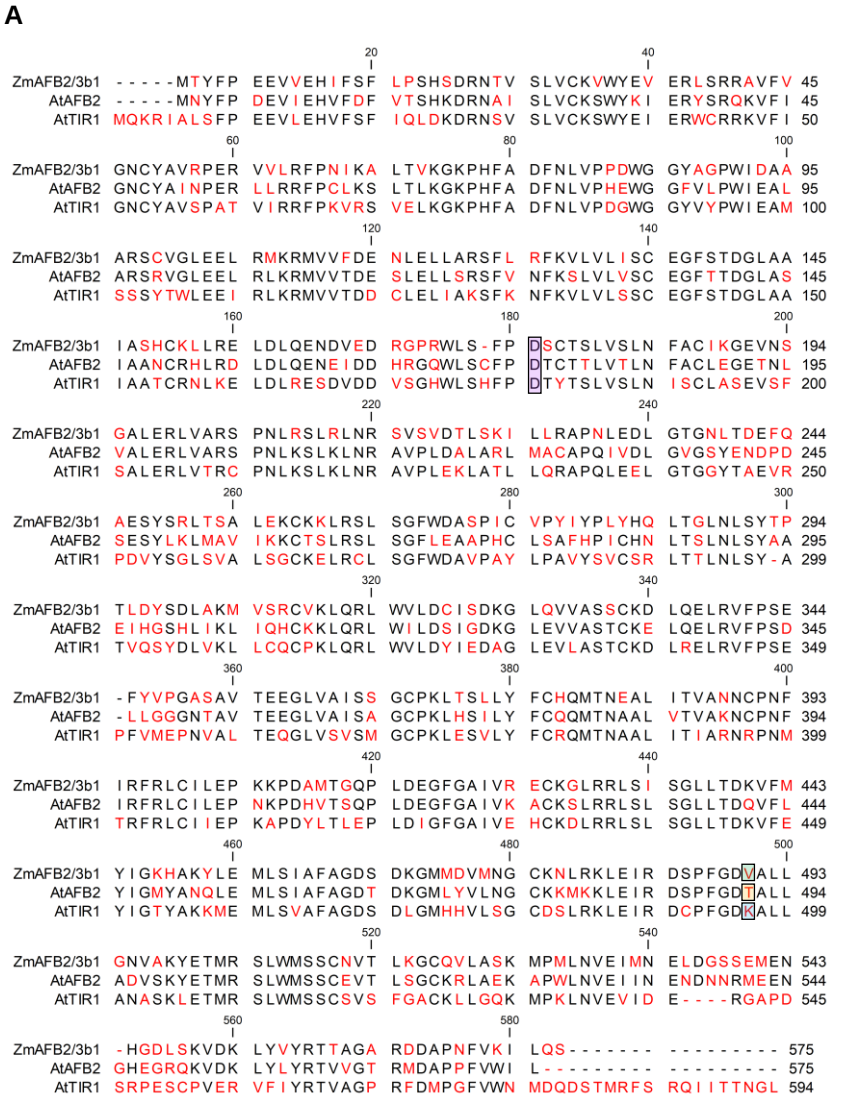
Yoshida A, Ohmori Y, Kitano H, Taguchi-Shiobara F, Hirano HY (2012) Aberrant spikelet and panicle1, encoding a TOPLESS-related transcriptional co-repressor, is involved in the regulation of meristem fate in rice. Plant J **70:** 327–339

Zhang Y, Marcon C, Tai H, von Behrens I, Ludwig Y, Hey S, Berendzen KW, Hochholdinger F (2016) Conserved and unique features of the homeologous maize Aux/IAA proteins ROOTLESS WITH UNDETECTABLE MERISTEM 1 and RUM1-like 1. J Exp Bot **67:** 1137–1147

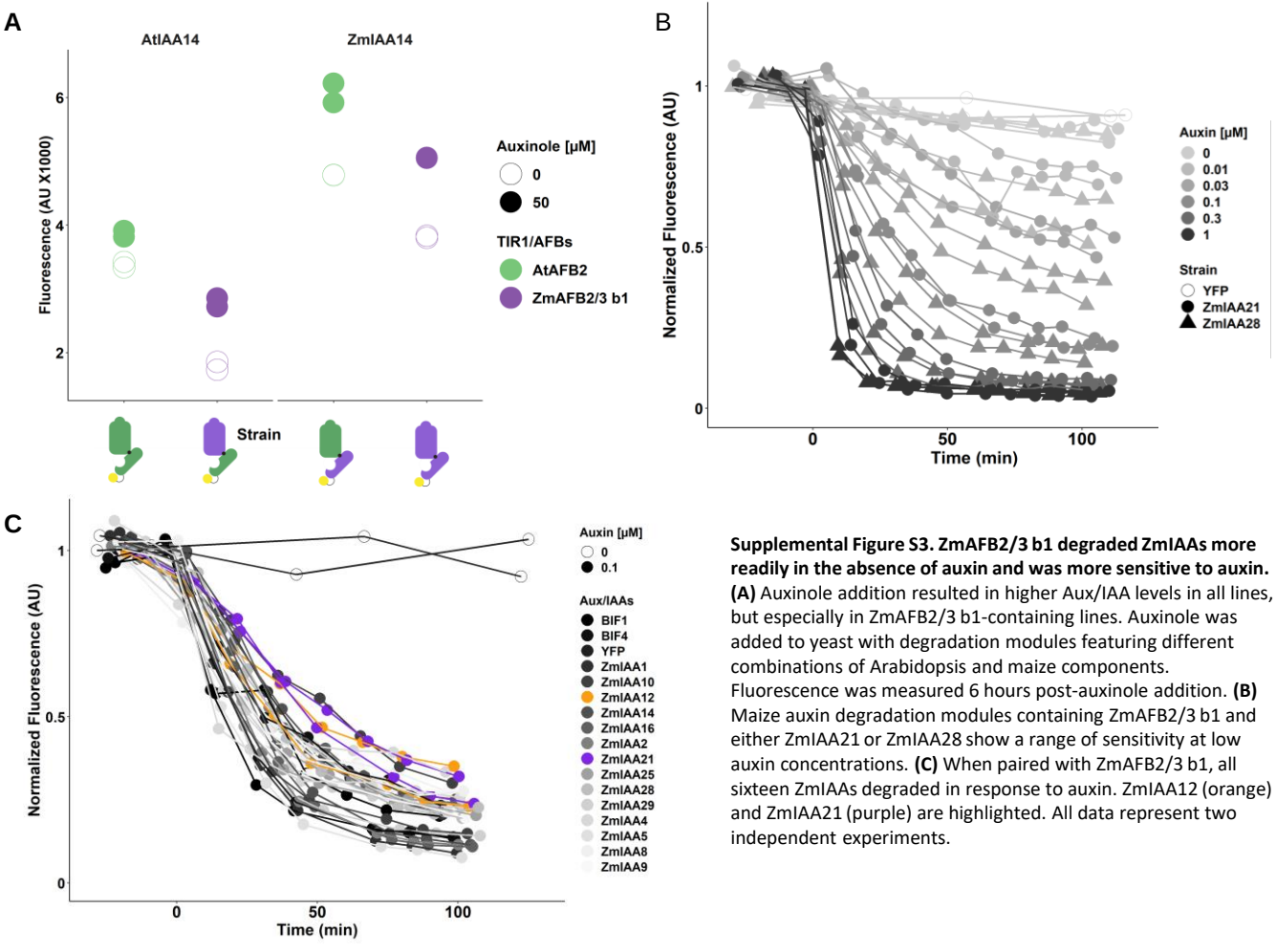
Zhao Z, Zhang Y, Liu X, Zhang X, Liu S, Yu X, Ren Y, Zheng X, Zhou K, Jiang L, et al (2013) A role for a dioxygenase in auxin metabolism and reproductive development in rice. Dev Cell **27:** 113–122



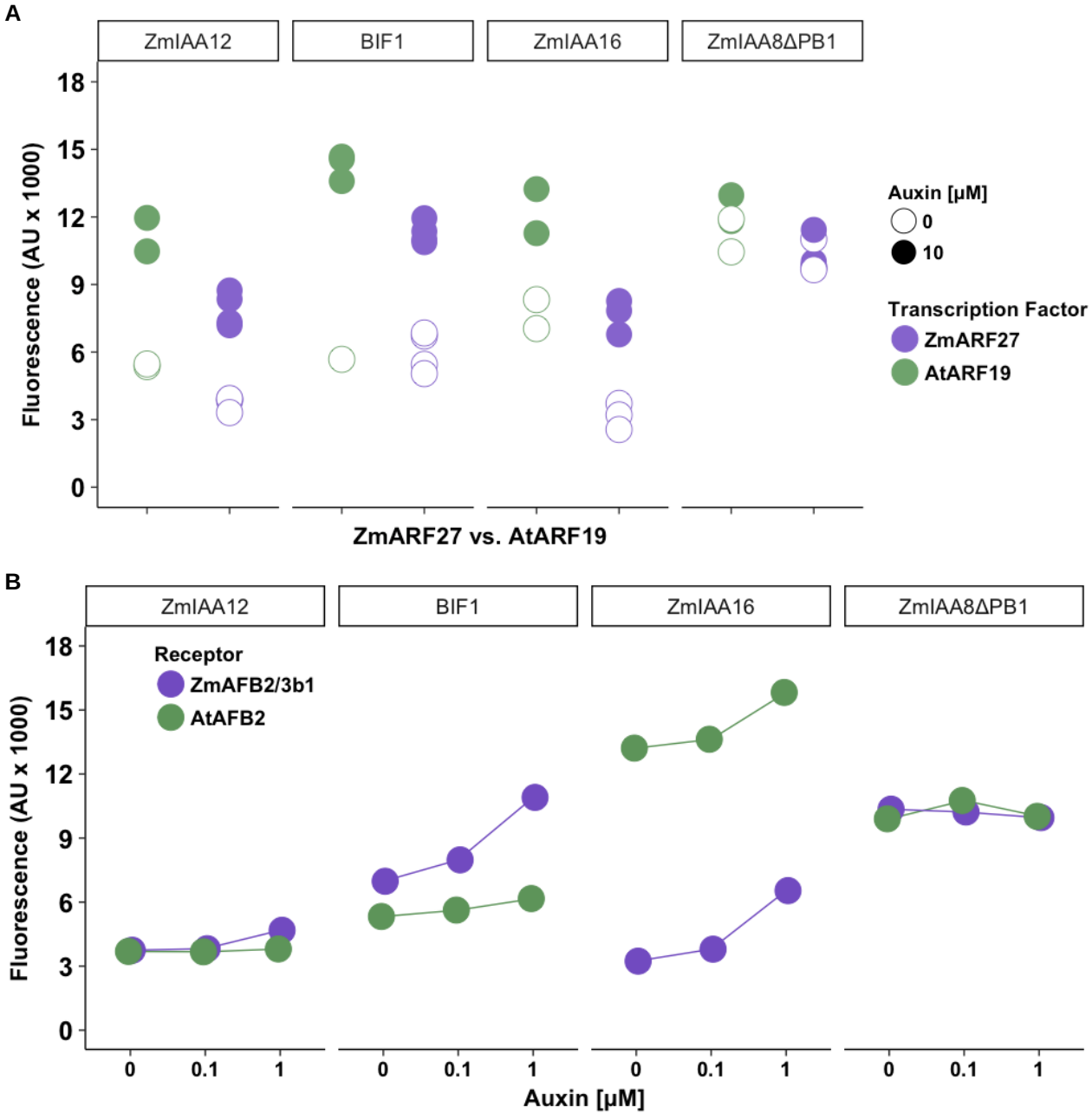
Supplemental Figure S1. The N-terminal fragment of REL2 fused to ZmIAAs repressed transcription. (A) Alignment of the degron consensus sequences, as defined in Ramos et al. 2001, for the 16 ZmIAAs tested. The core degron sequence (yellow) and variants (red) are highlighted. **(B)** TPLN100 and REL2N91 were aligned using MegaX software. They show high sequence similarity. Only positions 40 and 43-48 are different. **(C)** Timecourse of transcriptional activation in strains with REL2N91-ZmIAA fusions. **(D)** ZmIAAs fused to REL2N91 were inserted into the repression module. These showed a range of activation and repression levels, as measured by flow cytometry for accumulation of and auxin-responsive UbiVENUS. The purple box highlights the two ZmIAAs that did not repress transcription. Auxin or mock treatment (95% Ethanol) was added at t = 0 min and fluorescence measurements were made on a flow cytometer.



Supplemental Figure S2. The maize auxin AFB2/3 b1 receptor proteins is orthologous to those in other species and expressed at similar levels in yeast. (A) Alignment of the amino acid sequences of Arabidopsis TIR1 and AFB2 and maize AFB2/3 b1 with variant residues (red) highlighted. Known hypermorphic variations in *Arabidopsis* in *AFB2* (T491R and D176E). Residues highlighted if small and nonpolar (yellow), hydrophobic (green), polar (pink), negatively charged (purple) or positively charged (blue). **(B)** The sequence similarities are conserved across both the F-box domains **(C)**, and the LRR domains **(D)**. **(E,F)** Two representative Western blots showing **(E)** PGK1 protein loading controls and **(F)** anti-FLAG blots showing different FLAG-tagged AFB2/3 protein quantities in yeast lines. Positive control was yeast expressing OsAFB2-3xFLAG6xHIS, which had previously been shown to be functional and expressed in our degradation module. Negative control was yeast without an integrated transgene.



Supplemental Figure S3. ZmAFB2/3 b1 degraded ZmIAAs more readily in the absence of auxin and was more sensitive to auxin. **(A)** Auxinole addition resulted in higher Aux/IAA levels in all lines, but especially in ZmAFB2/3 b1-containing lines. Auxinole was added to yeast with degradation modules featuring different combinations of Arabidopsis and maize components. Fluorescence was measured 6 hours post-auxinole addition. **(B)** Maize auxin degradation modules containing ZmAFB2/3 b1 and either ZmIAA21 or ZmIAA28 show a range of sensitivity at low auxin concentrations. **(C)** When paired with ZmAFB2/3 b1, all sixteen ZmIAAs degraded in response to auxin. ZmIAA12 (orange) and ZmIAA21 (purple) are highlighted. All data represent two independent experiments.



Supplemental Figure S4. Auxin response circuit behavior depended on the identity of the ARF, TIR1/AFB, and Aux/IAA. (A) Circuits expressing different ARFs showed different levels of repression, as well as different levels of activation post auxin-addition. Experiments performed as in Fig. 4B. **(B)** Circuits expressing AtAFB2 or ZmAFB2/3 b1 had different sensitivities to auxin, and different basal levels of repression. Data shown as the same as in Fig. 4C, but without fold-activation normalization and including ZmIAA8 strain with a deleted PB1 domain that represents an always "on" state.

Supplemental Table S1: Vectors and primers used for *AUX/IAA* cloning and expression analysis

Gene name	IAAs	Vector	Fwd primer 5′-3′	Rev primer 5′-3′	Enzyme	note
GRMZM2G159285	ZmIAA2	pENTR	none	none	SalI	construct ordered from GRASSIUS (pUT5743)
GRMZM2G004696	ZmIAA5	pENTR	none	none	SacI	construct ordered from GRASSIUS (pUT6443)
GRMZM2G057067	ZmIAA9	pENTR	none	none	BglII	construct ordered from GRASSIUS (pUT6812)
GRMZM2G077356	ZmIAA14	pENTR223-Sfi	GAATTCGGCCGTCAAGGCCAATGGCTGGCGCCGA(	AGTCGACGGCCCATGAGGCCTCAGCTCTTGTTCTTC	AscI	none
GRMZM2G121309	ZmIAA16	pENTR	none	none	BglII	construct ordered from GRASSIUS (pUT1132)
GRMZM2G035465	ZmIAA28	pGEM T-Easy	ATGTCGCCTCCACTCGAGCT	TGACCCTGCTTGGCATCACCA	XhoI	none
GRMZM2G104176	ZmIAA4	pENTR	none	none	XhoI	construct ordered from GRASSIUS (pUT6756)
GRMZM2G167794	ZmIAA8	pENTR	none	none	SacI	construct ordered from GRASSIUS (pUT5549)
GRMZM2G037368	ZmIAA10	pGEM T-Easy	CACGACTACATCGGCCTCTC	GGGTATTCTTCCGGTAGTTGC	NcoI	none
GRMZM2G142768	ZmIAA12	pENTR	none	none	BamHI	construct ordered from GRASSIUS (pUT5582)
GRMZM2G147243	ZmIAA21	pENTR223-Sfi	GAATTCGGCCGTCAAGGCCAATGTCGCCCCCACTC	AGTCGACGGCCCATGAGGCCCTAGTTGCGATTCTTC	SacII	none
GRMZM2G115357	ZmIAA25	pENTR223-Sfi	GAATTCGGCCGTCAAGGCCAATGGAGCTCGACGC(	AGTCGACGGCCCATGAGGCCTCACCTGGCCTCAGA(	BsaI	none
GRMZM2G163848	ZmIAA29	pENTR	none	none	XhoI	construct ordered from GRASSIUS (pUT5762)

Supplemental Table S2: Gene identifiers and plasmids used for yeast transformations

Gene Identifier	Gene Name	Plasmid Number	Plasmid Name	Reference	Deposited to AddGene
TIR1/AFBs					
GRMZM2G155849	ZmAFB2/3 a	pNL2303	pGP8GF-ZmAFB2/3 a		yes
GRMZM5G848945	ZmAFB2/3 b1	pNL2280	pGP8GF-ZmAFB2/3 b1		yes
GRMZM2G137451	ZmAFB2/3 b2	pNL2281	pGP8GF-ZmAFB2/3 b2		yes
AT3G62980	AtTIR1	pNL1692	pGP5G-AtTIR1	Havens et al., 2012	
""	""	pNL2279	pGP8GF-AtTIR1		yes
AT3G26810	AtAFB2	pNL1330	pGP5G-AtAFB2	Havens et al., 2012	
""	""	pNL2301	pGP8GF-AtAFB2		yes
ARFs					
GRMZM2G160005	ZmARF27	pNL2480	pGP8A-ZmARF27		
AT1G19220	AtARF19	pNL666	pGP8A-AtARF19	Pierre-Jerome et al., 2014	
IAAs					
GRMZM2G159285	ZmIAA2	pML64	pENTR-ZmIAA2		yes
""	""	pML71	pGP4GY-ZmIAA2		
""	""	pML147	pGP4G-REL2N91:ZmIAA2		
GRMZM2G104176	ZmIAA4	pML93	pENTR-ZmIAA4		yes
""	""	pML96	pGP4GY-ZmIAA4		
""	""	pML155	pGP4G-REL2N91:ZmIAA4		
GRMZM2G004696	ZmIAA5	pML65	pENTR-ZmIAA5		yes
""	""	pML72	pGP4GY-ZmIAA5		
""	""	pML156	pGP4G-REL2N91:ZmIAA5		
GRMZM2G167794	ZmIAA8	pML83	pENTR-ZmIAA8		yes
""	""	pML86	pGP4GY-ZmIAA8		
""	""	pNL2474	pGP4G2-ZmIAA8		
""	""	pML109	pGP4G-TPLN100:ZmIAA8		
""	""	pML140	pGP4G-REL2N91:ZmIAA8		
""	ZmIAA8ΔPB1	pML144	pENTR-ZmIAA8ΔPB1		yes
""	""	pML145	pGP4G-REL2N91:ZmIAA8ΔPB1		
GRMZM2G057067	ZmIAA9	pML84	pENTR-ZmIAA9		yes
""	""	pML87	pGP4GY-ZmIAA9		
""	""	pML157	pGP4G-REL2N91:ZmIAA9		
GRMZM2G037368	ZmIAA10	pML66	pENTR-ZmIAA10		yes
""	""	pML73	pGP4GY-ZmIAA10		
""	""	pML158	pGP4G-REL2N91:ZmIAA10		
GRMZM2G142768	ZmIAA12	pML67	pENTR-ZmIAA12		yes
""	""	pML74	pGP4GY-ZmIAA12		
""	""	pML148	pGP4G-REL2N91:ZmIAA12		
GRMZM2G077356	ZmIAA14	pML89	pENTR-ZmIAA14		yes
""	""	pML91	pGP4GY-ZmIAA14		
""	""	pML159	pGP4G-REL2N91:ZmIAA14		
GRMZM2G121309	ZmIAA16	pML85	pENTR-ZmIAA16		yes
""	""	pML88	pGP4GY-ZmIAA16		
""	""	pML149	pGP4G-REL2N91:ZmIAA16		
GRMZM2G147243	ZmIAA21	pML94	pENTR-ZmIAA21		yes
""	""	pML97	pGP4GY-ZmIAA21		
""	""	pML160	pGP4G-REL2N91:ZmIAA21		
GRMZM2G115357	ZmIAA25	pML95	pENTR-ZmIAA25		yes
""	""	pML98	pGP4GY-ZmIAA25		
""	""	pML161	pGP4G-REL2N91:ZmIAA25		
GRMZM2G035465	ZmIAA28	pML70	pENTR-ZmIAA28		yes
""	""	pML77	pGP4GY-ZmIAA28		
""	""	pML162	pGP4G-REL2N91:ZmIAA28		
GRMZM2G163848	ZmIAA29	pML133	pENTR-ZmIAA29 (yeast codon optimized)		yes
""	""	pML103	pGP4GY-ZmIAA29		
""	""	pML163	pGP4G-REL2N91:ZmIAA29		
GRMZM2G130953	BIF1	pML131	pENTR-BIF1		yes
""	""	pML105	pGP4GY-BIF1		
""	""	pML150	pGP4G-REL2N91:BIF1		
GRMZM5G864847	BIF4	pML130	pENTR-BIF4		yes
""	""	pML106	pGP4GY-BIF4		
""	""	pML164	pGP4G-REL2N91:BIF4		
AT4G14550	AtIAA14	pNL1316	pGP4GY-AtIAA14	Havens et al., 2012	
Cloning Vectors					
N/A	N/A	pNL531	pGP5G-ccdB	Havens et al., 2012	
N/A	N/A	pNL1334	pGP4GY-ccdB	Havens et al., 2012	
AT1G15750	TPL	pNL687	pGP4G-TPLN100-ccdB	Pierre-Jerome et al., 2014	
GRMZM2G042992	REL2	pML124	pGP4G-REL2N91-ccdB		yes
N/A	N/A	pNL2194	pGP8GF-ccdb		yes

Supplemental Table S3: Yeast strains used in this study

Strain Name	Figure	Strain Number
W303-1A: MATa leu2-3,112 trp1-1 can1-100 ura3-1 his3-1,115 ybp1-1		yML1
W814-29B: MATα ade2-1 trp1-1 can1-100 ura3-1 leu2-3,112 his3-1,115		yML2
5G-AtTIR1 | 4GY-BIF4_STOP	1	yML209
5G-AtTIR1 | 4GY-ZmIAA29_STOP	1	yML198
5G-AtTIR1 | 4GY-ZmIAA4_STOP	1	yML180
5G-AtTIR1 | 4GY-ZmIAA25_STOP	1	yML188
5G-AtTIR1 | 4GY-ZmIAA16_STOP	1	yML186
5G-AtTIR1 | 4GY-ZmIAA14_STOP	1	yML184
5G-AtTIR1 | 4GY-ZmIAA21_STOP	1	yML182
5G-AtTIR1 | 4GY-ZmIAA1_STOP	1	yML144
5G-AtTIR1 | 4GY-ZmIAA10_STOP	1	yML121
5G-AtTIR1 | 4GY_ZmIAA28_STOP	1	yML125
5G-AtTIR1 | 4GY_ZmIAA9_STOP	1	yML133
5G-AtTIR1 | 4GY-ZmIAA5-STOP	1	yML120
5G-AtTIR1 | 4GY-ZmIAA2-STOP	1	yML119
5G-AtTIR1 | 4GY-BIF1_STOP	1	yML207
5G-AtTIR1 | 4GY_ZmIAA8_STOP	1	yML132
5G-AtTIR1 | 4GY-ZmIAA12-STOP	1	yML122
5G-AtTIR1 | 4G-IAA17 (diploid, autofluorescence control)	1,3,S3,S4	yML12
5G-AtTIR1 | 4GY-Stop (diploid, unfused YFP control)	1,S4	yML11
5G-AtAFB2 | 4GY-BIF1_STOP	1	yML208
5G-AtAFB2 | 4GY_ZmIAA8_STOP	1	yML134
5G-AtAFB2 | 4GY-ZmIAA12-STOP	1	yML128
5G-AFB2, 4GTPLN100-ZmIAA8 | 6P3_2x_UbiVenus-stop, 8A_ARF19_H170A	2	yML211
5G-AFB2, 4GTPLN100-ZmIAA8 | 6P3_2x_UbiVenus-stop, 8A-ARF19	2	yML213
5G-AFB2 | 6P3_2x_UbiVenus-stop, 8A_ARF19_H170A	2	yML217
5G-AFB2 | 6P3_2x_UbiVenus-stop, 8A-ARF19	2	yML219
5G-AFB2, 4G-ZmREL2N91-ZmIAA8-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A_ARF19_H170A	2	yML246
5G-AFB2, 4G-ZmREL2N91-ZmIAA8-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML247
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA8deltaPB1-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML267
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA2-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML269
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA12-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML271
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA16-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML273
5G-AFB2, pGP4G-ZmREL2N91-BIF1-STOP-new_linker | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML275
5G-AFB2, pGP4G-ZmREL2N91-BIF4-STOP-new_linker | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML306
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA29sc-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML305
5G-AFB2, 4G-ZmREL2N91-ZmIAA28-STOP | 6P3_2x_UbiVenus-stop, 8A-ARF19	2,S1	yML304
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA25-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML303
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA21-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML302
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA14-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML301
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA4-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML300
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA1-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML299
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA10-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML298
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA9-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML297
5G-AFB2, pGP4G-ZmREL2N91-ZmIAA5-STOP | pGP6P3_2x_UbiVenus-stop, pGP8A-ARF19	2,S1	yML296
8GF-AtAFB2 | 4GY-ZmIAA10	3	yNL3465
8GF-ZmAFB2/3 a | 4GY-ZmIAA10	3	yNL3473
8GF-ZmAFB2/3 a	3,S2	yNL3397
8GF-ZmAFB2/3 b1	3,S2	yNL3931
8GF-AtAFB2	3,S2	yNL3462
8GF-ZmAFB2/3 b2	3,S2	yNL3398
8GF-OsAFB2	S2	yNL3461
8GF-AtAFB2 | 4GY-AtIAA14	3,S3	yNL3466
8GF-AtAFB2 | 4GY-ZmIAA14	3,S3	yNL3483
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA29	S3	yNL3841
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA25	S3	yNL3840
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA21	S3	yNL3839
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA10	3,S3	yNL3838
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA5	S3	yNL3837
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA4	S3	yNL3836
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA1	3,S3	yNL3835
8GF-ZmAFB2/3 b1 | 4GY-BIF4	S3	yNL3834
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA16	S3	yNL3817
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA2	S3	yNL3816
8GF-ZmAFB2/3 b1 | 4GY-AtIAA14	3,S3	yNL3932
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA14	3,S3	yNL3708
8GF-ZmAFB2/3 b1 | 4GY-BIF1	3,S3	yNL3704
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA8	3,S3	yNL3705
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA12	3,S3	yNL3707
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA28	S3	yNL3708
8GF-ZmAFB2/3 b1 | 4GY-ZmIAA21	S3	yNL3839
6P32x-UbiVenus-STOP, 8A-ZmARF27, 4G-ZmIAA8ΔPB1:REL2N91, 8G-ZmAFB2/3 b1-3F6H	4	yML392
6P32x-UbiVenus-STOP, 8A-ZmARF27, 4G-ZmIAA12:REL2N91, 8GF-ZmAFB2/3b1-3F6H	4,S4	yML458
6P32x-UbiVenus-STOP, 8A-ZmARF27, 4G-BIF1:REL2N91, 8GF-ZmAFB2/3b1-3F6H	4,S4	yML456
6P32x-UbiVenus-STOP, 8A-ZmARF27, 4G-ZmIAA16:REL2N91, 8GF-ZmAFB2/3b1-3F6H	4,S4	yML448
6P32x-UbiVenus-STOP, 8A-ZmARF27, 4G-ZmIAA12:REL2N91, 8G-AtAFB2-3F6H	4,S4	yML459
6P32x-UbiVenus-STOP, 8A-ZmARF27, 4G-BIF1:REL2N91, 8G-AtAFB2-3F6H	4,S4	yML457
6P32x-UbiVenus-STOP, 8A-ZmARF27, 4G-ZmIAA16:REL2N91, 8G-AtAFB2-3F6H	4,S4	yML451
6P32x-UbiVenus-STOP, 8A-ZmARF27, 4G-ZmIAA8ΔPB1:REL2N91, 8GF-AtAFB2-3F6H	S4	yML450
6P32x-UbiVenus-STOP, 8A-ARF19, 4G-ZmIAA8ΔPB1:REL2N91, 8G-ZmAFB2/3 b1-3F6H	S4	yML461
6P32x-UbiVenus-STOP, 8A-ARF19, 4G-ZmIAA12:REL2N91, 8GF-ZmAFB2/3b1-3F6H	S4	yML462
6P32x-UbiVenus-STOP, 8A-ARF19, 4G-BIF1:REL2N91, 8GF-ZmAFB2/3b1-3F6H	S4	yML464
6P32x-UbiVenus-STOP, 8A-ARF19, 4G-ZmIAA16:REL2N91, 8GF-ZmAFB2/3b1-3F6H	S4	yML463

CHAPTER 4

PNAS

Main Manuscript for

A single helix repression domain is functional across eukaryotes

Alexander R. Leydon[1*], Román Ramos Baez[1*], Jennifer L. Nemhauser[1]

[1] University of Washington, Department of Biology, Seattle, WA 98105-1800, USA

* These authors contributed equally to this work.

Jennifer L Nemhauser

Email: jn7@uw.edu

Author Contributions: Experimental design and analysis by ARL, RRB, and JLN. Research performed by ARL and RRB. Manuscript written by ARL, RRB, and JLN.

Competing Interest Statement: Authors declare no competing interests.

Classification: Biological Sciences/Plant Biology

Keywords: LisH, co-repressor, TPL, transcriptional repression

This DOCX file includes:

Main Text

Support

This work was supported by the National Institute of Health (R01- GM107084) and a Faculty Scholar Award from the Howard Hughes Medical Institute. ARL was supported as a Simons Foundation Fellow of the Life Sciences Research Foundation. RRB was supported as a Gilliam Fellow of the Howard

Hughes Medical Institute.

Abstract

The corepressor TOPLESS (TPL) and its paralogs coordinately regulate a large number of genes critical to plant development and immunity. As in many members of the larger pan-eukaryotic Tup1/TLE/Groucho corepressor family, TPL contains a Lis1 Homology domain (LisH), whose function is not well understood. We have previously found that the LisH in TPL—and specifically the N-terminal 18 amino acid alpha-helical region (TPL-H1) —can act as an autonomous repression domain. We hypothesized that homologous domains across diverse LisH-containing proteins could share the same function. To test that hypothesis, we built a library of H1s that broadly sampled the sequence and evolutionary space of LisH domains, and tested their activity in a synthetic transcriptional repression assay in *Saccharomyces cerevisiae*. Using this approach, we found that repression activity was highly conserved and likely the ancestral function of this motif. We also identified key residues that contribute to repressive function. We leveraged this new knowledge for two applications. First, we tested the role of mutations found in somatic cancers on repression function in two human LisH-containing proteins. Second, we validated function of many of our repression domains in plants, confirming that these sequences should be of use to synthetic biology applications across eukaryotes.

Significance Statement

The LisH domain is found in nearly 25,000 proteins with diverse functions, including several corepressors. Previously, we found that the first helix (H1) of the LisH domain in a plant corepressor functioned as a short (18 amino acids), modular repression domain. Here, we extended this analysis by surveying the function of LisH-H1s representing over 1,000 proteins from fungi, animals and plants. The majority of the tested sequences repressed reporter activity in yeast, and many retained this function when

tested in plants, pointing to repression as an evolutionarily conserved role for LisH domains. We further adapted our assay to study the impact on repressive activity of cancer-associated mutations in two LisH-domain containing human proteins: TBL1 and DCAF1.

Main Text

Introduction

Transcriptional repression is enacted through a diverse array of mechanisms, which are often directed by a group of proteins known as corepressors (1–4). Corepressors do not directly bind DNA, but instead recruit inhibitory machinery to specific loci via interactions with transcription factors. Among the best-studied corepressors are: animal SMRT (silencing mediator of retinoic acid and thyroid hormone receptor) and NCoR (nuclear receptor corepressor) complexes (5, 6); yeast Tup1 (7–9) and its homologs *Drosophila* Groucho (Gro) and mammalian transducing-like enhancer (TLE) (10); plant TOPLESS/TOPLESS-RELATED1-4 (TPL/TPR1-4), LEUNIG/ LEUNIG_HOMOLOG (LUG/LUH), and HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 15 (HOS15) (11–16). Despite knowing the identity of many corepressor proteins, much is left to uncover about how these complexes integrate input signals to initiate, sustain, and relieve transcriptional repression.

Plant corepressor families share a general structural similarity, where the N-terminus contains protein-protein interaction domains followed by an unstructured linker to a C-terminal WD40 beta-propeller domain (13, 14). While the full N-terminal structure is not highly conserved, all plant corepressors share a Lis1 Homology domain (LisH), which is generally known as a protein multimerization interface (17–25). In TPL, the LisH enables homodimerization of TPL/TPRs (26, 27). The LisH is followed by a C-terminal to LisH (CTLH) domain that binds partner proteins via an ETHYLENE-RESPONSIVE ELEMENT BINDING FACTOR-ASSOCIATED AMPHIPHILIC REPRESSION (EAR) motif (12, 28). The N-terminal domain of TPL also contains a CT11-RanBPM

(CRA) domain, which acts as a second homo-multimerization interface that folds back over to stabilize the LisH domain (26, 27). In previous work, we performed a structure-function analysis of the first 188 amino acids of the TPL N-terminal domain (TPLN188). We found that the N-terminal domain of TPL contains two distinct repression domains, one of which is the LisH domain (29). Specifically, the first of two alpha helical regions within the LisH, termed hereafter Helix 1 (TPL-H1), was sufficient to repress transcription in yeast when recruited to a promoter (29). We therefore defined the TPL-H1 peptide sequence as an autonomous repressor domain, meaning that it is a delimited sequence sufficient to repress transcription when brought to DNA via fusion to another protein.

The 33-residue LisH motif is found in many proteins across eukaryotes; currently, there are more than 25,000 unique LisH sequence entries in the SMART protein database (30). These proteins have a broad range of annotated functions, including: cytoskeleton-interacting proteins, ubiquitin ligase complexes, and transcriptional regulation. The founding member LIS1 regulates microtubule function and is required for proper neurodevelopment (31). While LIS1 has been broadly studied in its cytoplasmic context, recent work has also demonstrated a nuclear role in gene expression (32). Several E3 ubiquitin-ligase-associated proteins carry LisH domains, such as DDB1–Cul4-associated factor 1 (DCAF1, (33)), which is involved in myriad pathways contributing to development and disease (34). The glucose-induced-degradation (GID) E3 ligase complex is assembled by intermolecular LisH interactions (24). Other LisH-containing proteins have well characterized roles in human health and disease such as the oncogene Transducin-beta like 1 (35). TBL1 is a component of the SMRT/NCoR complex (6), and acts as an exchange factor, facilitating the conversion of SMRT/NCoR repressed loci into sites of active transcription (36). TBL1's LisH domain is required for its transcriptional activity (22). Understanding conserved functions of LisH domains has the potential to simultaneously shed light on multiple core cell and developmental processes and provide insights into diseases that result from dysregulation of LisH-containing proteins.

Previously, we have recapitulated the auxin response pathway in yeast by transferring essential components from plants (*Arabidopsis thaliana* Auxin Response Circuit in *Saccharomyces cerevisiae*, *At*ARC^{Sc}; (37)). In *At*ARC^{Sc}, an auxin-responsive transcription factor (ARF) binds to a promoter driving expression of a fluorescent reporter. ARF activity is repressed by interaction with a fusion protein comprised of an Aux/IAA protein and a portion of the N-terminal domain of TPL. Reporter activation can be quantified after addition of auxin by flow cytometry (37). In this way it is possible to test the direct effect of various mutations in TPL, or other transcriptional repressors, at an orthogonal, synthetic locus in a quantitative manner.

Here, we modified the *At*ARC^{Sc} to better understand the repressive function and evolutionary history of LisH-H1s. We first interrogated the critical residues within AtTPL-H1 that are required for robust transcriptional repression. Next, we built a library of H1 sequences from proteins with diverse annotated functions across the eukaryotic lineage to test the extent of H1 repressive function. We then focused our attention on two applications. We used our yeast assay to survey the effect of documented somatic cancer variants of two human LisH-containing proteins (HsTBL1 and HsDCAF1) on protein stability and repressive function, and we tested our most repressive H1 sequences as candidates for synthetic transcriptional repressors in plants. Together, our findings uncovered the ancestral transcriptional repression ability of the LisH domain, and showcased how this system can be used to understand disease states, as well as be incorporated into synthetic biology applications.

Results

The TPL LisH domain is a short transcriptional repression domain.

By directly fusing corepressor fragments to an auxin co-receptor (IAA3 in these studies) and thereby bringing it near the synthetic auxin-regulated locus, the *At*ARC^{Sc} makes it possible to sensitively measure repressive activity even in fragments that would not be recruited to DNA on their own. Here, we

focused on the small modular LisH domain which we previously demonstrated is sufficient to repress transcription in *At*ARC^{Sc} assays (29, 37), even when truncated to its first 18 amino acids (H1, Figure 1A-B). We first identified solvent-facing amino acids in AtTPL-H1, as these residues were less likely to be involved in stabilizing the hydrophobic interactions between intra-AtTPL helical domains and might be available to interact with partner proteins (26, 27). Six of these candidate residues were mutated to alanine (Fig. 1A, pink residues) in the context of H1-IAA3, and assayed for repression activity. The amino acids on either end of the helix (R6 and E18) were required for repression (Fig. 1B), as we observe high reporter expression when these are replaced with alanines. A mutation of E7, the immediate neighbor of R6, slightly increased reporter expression (Fig. 1B) and lowered the final activation level after auxin addition when compared with wild type AtTPL-H1. This result suggests it likely plays a smaller, supporting role in repressive function (Fig. 1C). Consistent with this interpretation, the R6A, E7A double mutant behaved similarly to R6A alone. Likewise, the D17A, E18A double mutation did not enhance the effect of E18A alone. Q14A and D17A were indistinguishable from wild type H1 (Fig. 1C). In contrast to the other mutations which had negative if any effects on H1 repression activity, F10A strengthened the durability of repression H1, converting it into an auxin-insensitive repression domain (Fig. 1C). One explanation for this stabilizing effect is that in the context of the full-length AtTPL protein F10 is buried in a hydrophobic cluster (along with F163, F33, F34 and L165, (26, 27)). Truncations of AtTPL left F10 solvent-exposed, rendering it less stable overall. This situation was at least partially remedied by substituting a smaller, less hydrophobic residue (F10A).

To rapidly test an expanded library of mutations in LisH H1 sequences, we designed a new auxin response circuit (ARC) that includes an epitope tag to allow quantification of repressor protein levels (Fig. 1D). All parts of the circuit that are held constant across experiments were integrated at the *URA3* locus. The H1-1xHA-IAA3 fusion protein was both detectable in western blots and showed minimal interference with repression activity (Fig. 1E, Supplement 1). Our next step was to perform an amino acid swap at AtTPL-H1 residues R6 and F10 (Fig. 1B-C). In the case of R6, we observed broad tolerance of

amino acid swap (Fig. 1G), with the exceptions of charge inversion (R6D and R6E), and an unsurprising reduction of repression in R6P, which is likely to interfere with its alpha-helical structure. The unexpected result that R6A has only a mild decrease in repression in this experimental design is likely due to higher relative expression of a plasmid-based AtTPL-H1 compared to the other components. In the context of this altered stoichiometry of circuit components, repression strength is slightly less sensitive to loss of function in AtTPL-H1. Several amino acid swaps had a negative effect on protein accumulation (R6E, R6N, and R6V), likely explaining the observed loss of repression. In all experiments we compared these mutations to a well characterized alpha-helical linker sequence (α-helix-HA-IAA3) as a control, as well as IAA3 alone (None) (Fig. 1F,G).

In the case of F10, we also observed a broad tolerance of amino acid substitutions (Fig. 1G). A small number of substitutions had negative effects on protein accumulation (F10N, F10V, and F10L) which likely explained their reduced repressive activity. Several substitutions (E, Q, and R) showed similar repression strength to wild type, despite their different physicochemical character. We introduced these variants into the fully integrated *At*ARC^{Sc} to test for sensitivity to auxin treatment. Similarly to F10A, substitution of a negative (E) or polar (Q) residue in TPL-H1 resulted in an increased durability of repression; in contrast, the substitution of a positive charge (R) had a milder effect (Fig. 1G inset). These results support the hypothesis that the exposure of a hydrophobic residue (F) by truncation negatively impacts TPL-H1 activity. In addition, applications using the TPL-H1 as an autonomous repression domain should incorporate one of the stronger repression variants assayed here.

Defining the LisH H1 sequence

Although the repressive function of most LisH-H1 domains have not been directly tested, many LisH-containing proteins are known transcriptional regulators (16, 38–41). Many other LisH-containing proteins are primarily cytoplasmic or have well-studied primary functions in ubiquitination or cytoskeletal

dynamics, making it unlikely anyone would have tested their impact on transcription. Recent work on Lis1, which contains the founding LisH domain and has been thought to be exclusively cytoplasmic, suggests that this assumption about functions for these proteins may be misplaced. New evidence suggests that Lis1 moonlights as a transcriptional repressor (32), raising the possibility that the same might be true for other LisH-containing proteins.

To better understand the diversity of LisH-H1 sequences and the proteins that carry them, we performed Maximum Likelihood (ML) reconstruction (42) of sequences sampled from over 1000 diverse LisH-containing proteins across eukaryotes (Fig. 2A, Supplement. 2). While the 18 amino acids of the LisH-H1 sequences is too short to produce optimal bootstrap values (43), our reconstruction allows us to make associations based on sequence similarity. The sequences cluster into five main clades, defined by nodes I-V. As some LisH-H1 sequences were identical across orthologous genes (Supplement 3), we selected the most well-studied gene containing each sequence as a representative, and added annotations for the listed full-length protein's functions and cellular localization (Fig. 2B, Supplement 3).We observed a related cluster of genes with published roles in transcriptional activation (Fig. 2B, blue circles), as well as a related cluster of H1 sequences from published repressors (Fig. 2B, red circles). Sequence alignments highlight residues of interest across genes and clades (Fig. 2C). For example, there are several highly conserved residues (L8, N9, L11, I12, L16, and Y15) that comprise the inward-facing dimerization interface in known structures.

LisH repressive function appears to be widely conserved and ancestral

The modifications to *At*ARC^{SC} (Fig. 1E) allowed us to directly compare repressive activity across distantly related LisH-H1 sequences. To do so, we created a plasmid library using representative LisH-H1 sequences from across our reconstruction (Fig. 2A). We also tested protein levels for all fusion proteins by western blot. One not very surprising trend is that proteins that are detected at lower levels are

generally poor at repression; however, the converse is not true: high accumulation of a fusion protein was not well-correlated with strong repression (Fig. 2D, Supplement 5).

While Clade I had high sequence diversity (Supplement 6), the overall similarity to TPL-H1 led us to hypothesize that these domains would also have a repressive function. This is indeed what we observed (Fig. 2D). Clade I LisH-H1 sequences from proteins involved in protein ubiquitination SmRANB9, DCAF1 (from *Homo sapiens* and *Arabidopsis thaliana*), and LuDDB1 (44) had repressive function, as did the splicing factor HsSMU1 (45). In addition, we found that H1s that belong to genes characterized as nuclear-localized transcriptional repressors, such as ScSIF2, AtHOS15, and AtLUG robustly repressed reporter activity to similar levels as AtTPL (Fig. 2D). However, many of the strongest repressors were found in genes across the tree without previously characterized roles in transcriptional repression (e.g., HsSMU1, SpADN2, and NcSSH4).

We observed that Clade II was characterized by a slightly lower sequence variability than Clade I and a high incidence of R14 and E18 residues (Fig. 2C, Supplement 5). LisH-H1s from the repressor proteins ScSIF2 (46) and AtHOS15 (40) repress strongly, while H1s closer in sequence to HsTBL1X are somewhat weaker. HsTBL1 recruits the repressive SMRT/NCoR complex, but upon stimulus facilitates the recruitment of transcriptional activators to the locus (36). Perhaps HsTBL1-H1, and closely related domains, have retained weaker repressor function to permit better exchange activity.

Clade III LisH-H1 sequences encompassed all predicted members of the multiprotein E3 ligase GID complex (24), also known as the CTLH (carboxy-terminal to LisH) complex in mammals (47). Clade III sequences were well conserved, with a high conservation of M13, and N14 (Fig. 2C, Supplement 6), which are on the solvent facing side, suggesting that these residues might lower affinity with repressive partner proteins. Most clade III LisH-H1s were expressed at levels higher than AtTPL, and sequences with both E6 and D7 showed the lowest capacity to repress.

Clade IV H1s included the nuclear localized transcriptional activators SpADN2, SpADN3 and ScMSS11, as well as the plant corepressors AtLUG and AtLUH (Fig. 2B). This clade was marked by a high incidence of Y11, Y13, and K18 residues and a strikingly low level of protein accumulation (Fig. 2C, Supplement 6). The L13Y variation may contribute to this loss, as L13 is a highly conserved residue elsewhere. Despite the low level of SpADN2-H1 and ScMSS11-H1, they were highly repressive. The LisH-H1 sequences representing SpADN2 and ScTAF5 were repressive despite the full-length proteins being annotated as transcriptional activators. This finding suggests that other domains within these proteins and/or interactions with specific partners are able to modify, or even nullify, LisH-H1 repressive function.

Finally, Clade V included highly diverse sequences belonging to genes coding for both nuclear and non-nuclear localized proteins with no annotated transcriptional functions (Fig. 2B). The HsLIS1-H1 sequence is both well expressed and repressive, pointing to a possible role for this sequence in mediating repression, consistent with other recent findings (32). Interestingly, unlike ScGID8, H1s from other GID/CTLH complex members such as SmMAEA and HsRANB9 retain repressive ability (48).

To further investigate these trends, we reconstructed predicted ancestral LisH-H1 sequences for nodes of interest across the phylogeny (Fig. 2 – Supplement 7). All reconstructed sequences were good repressors except for the Clade IV sequence which was poorly expressed and like most other members of its clade with weak repressive function, had a Y11 variation.

It is striking how few residues within the LisH-H1 sequence are required for repressive function. Repressive function is found in LisH-H1s across all clades measured, as well as basal LisH-H1s such as CaFLO8-H1 and SpYC5C-H1. The most repressive sequences have conserved hydrophobic amino acids L8,11,16 and I12 on the multimerization interface. (Supplement 6). We added the eight most well-conserved residues of these top repressors onto an alpha helix backbone and found it to be capable of robust repression (Supplement 7). This suggests that residues outside of this conserved core LisH-H1 motif serve mainly to tune the repressive function. This indicates that the most important determinant of

repression is the multimerization interface and suggests that most LisH-H1 sequences should retain this activity if localized to chromatin. It is worth noting that in these assays we are only testing the LisH-H1 sequence, which cannot homo-dimerize on its own (29). The persistence of repressive function in non-nuclear proteins may suggest that there are moonlighting functions for more of these proteins.

LisH domains are important for human disease.

The human oncogene HsTBL1 is a transcriptional regulator and exchange factor involved in repression and activation (4, 22, 35, 36). Its dysregulation is implicated in the progression of multiple cancers (38, 50, 51). All human TBL1 isoforms (TBL1X, TBL1XR1, TBL1Y) contain a LisH domain, and the N-terminal region of TBL1X (residues 1-76, Fig. 3A) could replace TPL in our synthetic circuit (Figure 3B, dashed lines). As with TPL, TBL1-H1 repressed to a similar degree as the whole TBL1 N-terminus, and created a circuit that was more sensitive to auxin-induced degradation (Fig. 3B, solid lines).

The Catalogue of Somatic Mutations in Cancer (COSMIC) (52) has recorded five non-synonymous mutations in HsTBL1-H1 (pooled mutations from data for TBL1X, TBL1XR1, TBL1Y whose H1s are identical in sequence, Fig. 3C). We tested the functional impact of these mutations by engineering them into the HsTBL1-H1-substituted version of *At*ARC^{Sc}. All tested variants increased HsTBL1 repressive function. HsTBL1-H1^{Y64C} was the strongest repressor and accumulated to the highest level, suggesting that this variant may potentiate oncogenesis by stabilizing HsTBL1. Both HsTBL1-H1^{R65Q} and HsTBL1-H1^{R14W} had lower protein abundance than wild-type HsTBL1-H1 yet higher repression rates. Together, these results imply that the level of repression by HsTBL1-H1 variants could be a useful, high throughput test for identifying the correlation between a specific variant and its disease-causing potential. Further investigation with many more variants, and associated clinical outcomes, is needed to follow-up on that prediction.

A number of LisH containing proteins in our phylogeny are components of E3 ubiquitin ligase complexes, one of which is a substrate receptor for Cullin RING ligase 4 (CRL4) and is named DDB1 (DNA damage-binding protein 1) and CUL4-associated factor 1 (DCAF1, (34). HsDCAF1 regulates diverse cell processes, and has been implicated in cancer (34) as well as subversion by HIV viral accessory proteins (33). The DCAF1 LisH has been implicated in both dimerization (21) and transcriptional repression, where it has been demonstrated to inhibit p53's transcriptional activity through binding of hypoacetylated Histone 3 tails (53, 54). Following the same methodology as described for TBL1-H1, we engineered four known, non-synonymous cancer-associated mutations into the HsDCAF1-H1-substituted version of *At*ARC^{Sc} (Fig. 3E). HsDCAF1-H1 had a strong repressive function (Fig. 3F). Activity and fusion protein accumulation were affected minimally in HsDCAF1-H1^{H856Y} and HsDCAF1-H1^{I853M}. However, HsDCAF1-H1^{L851F} dramatically reduced levels of the fusion protein and repressive function, while HsDCAF1-H1^{R854Q} was one of the strongest repressors that we tested and accumulated to a high level. More information is needed to determine whether any variant-induced change in HsDCAF1-H1 and HsTBL1-H1 measured in the *At*ARC^{Sc} is relevant in the context of the full-length proteins and in its native cellular context,. The performance metrics of the variants tested for both HsTBL1-H1 and HsDCAF1-H1 could be applied to future design of synthetic repression domains, and screening for small molecule agonists/antagonists.

LisH-H1s are effective synthetic repressor domains in plants

As LisH-H1s from distantly related species repressed transcription in yeast, we wondered whether they would work in other organisms as well. As a first test, we established a transient plant repression assay using essentially the same components as the *At*ARC^{Sc}. The small number of modifications that we made to transfer the assay to plants included replacing yeast regulatory sequences for each gene with ones known to work in plants, switching to the well-characterized and highly sensitive DR5 auxin response reporter (56), and engineering mutations in IAA3 in the critical contact residues in the EAR motif (LxLxL

to AxAxA) to prevent recruitment of endogenous TPL/TPR proteins (Fig. 4A, (29)). We found that, as in yeast, AtTPL-H1 was an effective repressor in the plant assays, and that similar trends were observed in the AtTPL-H1 variants tested in both systems (Fig. 4, Fig. 4 – Figure Supplement 8). Of particular interest for design of novel synthetic repressors, AtTPL-H1^{F10A} showed stronger repression than wild-type AtTPL-H1. We also tested sixteen of the most effective H1 repressors from our yeast assay, and successfully detected repression activity from nearly all of them (Fig 4). The few exceptions (e.g., Pf350509, DmCG5614) highlight the on-going challenge of standardizing parts across organisms. The overall success in transferring so many short repression domains from yeast to plants underscores the deep conservation of repression mechanisms across eukaryotes. It is also an excellent indicator that the LisH-H1 library characterized here—with its high tolerance for sequence diversity and range of repression strengths for tunability—can be a rich source of short, sequence-orthogonal repression domains for a diversity of engineering biology specifications.

Conclusions

Synthetic signaling approaches aim to shed light on mechanisms of core cellular functions in natural systems, and to apply this knowledge to novel, beneficial interventions. The use of the *At*ARC^{Sc} in this study was successful for both objectives. First, we have isolated a single, short alpha helical domain within AtTPL that can function as an autonomous repression domain. Second, we have found that homologous domains found in thousands of proteins and diverse eukaryotes share this repression activity, and that repression was likely the function of the ancestral LisH-H1. Third, we were able to successfully apply knowledge of LisH-H1 function to the exploration of function of human disease-associated variants, as well as validate a number of these sequences as effective synthetic repression domains in plants. Future efforts should be directed at identifying the precise mechanism of transcriptional repression by AtTPL-H1, as well as other LisH-H1 sequences. These added insights would likely allow for a more directed approach to engineering LisH-H1 activity, as well as leading to a deeper understanding of what

role LisH-H1 is playing in proteins that do not normally function in transcriptional regulation. A deepening of the connection between structure with function should also be useful for adapting the *At*ARC^{Sc} assays modified with human proteins for clinical applications, and optimizing the use of LisH-H1-based repression domains in synthetic circuits.

Materials and Methods

Alignments and Phylogenetic Reconstruction

LisH domains were identified using UniProt (https://www.uniprot.org/), Pfam (http://pfam.xfam.org/family/PF08513.7) and SMART (http://smart.embl-heidelberg.de/) databases. LisH Helix 1 domains were aligned using Clustal Omega. Tree sequences were selected from the PFAM LisH clade PF08513 by performing an alignment of the representative proteome dataset with a 15% cutoff value (1235 sequences). The evolutionary history was inferred by using the Maximum Likelihood method and Le_Gascuel_2008 model (58). The tree with the highest log likelihood (-2711.40) was used. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 6.6850)). The percentage of replicate trees in which the associated taxa clustered together were calculated via bootstrap using 1000 replicates (59). This analysis involved 143 amino acid sequences. The cladogram was derived from this tree, only showing relationships among the 63 experimentally analyzed sequences. There was a total of 13 positions in the final dataset. Ancestral sequence reconstruction was done with an expanded tree using the same methods. Evolutionary analyses were conducted in MEGA X (42). Logo plots were created with an online tool ((60) https://weblogo.berkeley.edu/logo.cgi).

Cloning

We used the VEGAS adapted method to create different forms of AtARC^{Sc} plasmids (61). We used a plasmid containing LisH-H1 fused to AtIAA13 and expressing URA3 as a backbone in LisH-H1 plasmid library construction. Plasmids were synthesized and confirmed with sequencing by Twist Bioscience (www.twistbioscience.com). All plasmids were transformed into reported haploid strain URA::[pRPS2-AtAFB2-ttCIT1, LEU2, pADH1-ARF19-ttADH1, pP3(2x)-UbiVenus-ttCYC1]. A standard lithium acetate protocol (62) was used for transformations of digested plasmids. All cultures were grown at 30°C with shaking at 220 rpm. For construction of plant vectors we used the MoClo toolkit (63) to design and clone plasmids containing our top 10 most repressive IAA13-H1s into vector pICH86966. Each H1 sequence is identical to yeast H1-HA-IAA3 constructs, except that IAA3 was mutated to prevent recruitment of endogenous TPL/TPR proteins. We transformed these into *A. tumefaciens* strain GV3101 via electroporation. All plasmids and oligos used are listed in Supplement 4.

Library Design

Phylogenetic library contains sequences selected from the Pfam LisH alignment PF08513 using the representative protein database (RP15, 1,235 sequences, http://pfam.xfam.org/family/PF08513). Ancestrally reconstructed sequences contain synthetic sequences predicted at nodes I-V using MEGAX node reconstruction software. HsTBL1-H1 and HsDCAF1-H1 mutational libraries contain somatic mutations found in human cancer cells within these helixes and were identified using COSMIC datasets ((52), https://cancer.sanger.ac.uk/cosmic). AtTPL site-saturation mutational libraries at residues AtTPL-H1 R6 and F10 contain synthetic sequences probing the function of these sites in helix 1. The alpha helix control sequence $(EAAAK)_3$ was created based on well-studied synthetic alpha helix linkers (64).

Flow Cytometry

Fluorescence measurements were taken using a Becton Dickinson (BD) special order cytometer with a 514-nm laser excitation fluorescence that is cut off at 525 nm prior to photomultiplier tube collection (BD, Franklin Lakes, NJ). Events were annotated, subset to singlet yeast using the FlowTime R package (https://github.com/wrightrc/flowTime). A total of 10,000 - 20,000 events above a 400,000 FSC-H threshold (to exclude debris) were collected for each sample and data exported as FCS 3.0 files for processing using the flowCore R software package and custom R scripts (65, 66). Data from at least two independent replicates were combined and plotted in R (https://ggplot2.tidyverse.org/).

Yeast Methods

Standard yeast drop-out and yeast extract–peptone–dextrose plus adenine (YPAD) media were used, with care taken to use the same batch of synthetic dropout (SDO) media for related experiments. Haploid transformants were selected on appropriate prototrophy (SDO -Tryptophan, -Leucine). Yeast were grown at 30°C on selection plates for two days, and in SDO liquid media with 250rpm in a deep well 96-well plate format overnight for cytometry analysis (66). Liquid cultures were diluted 1:200 with fresh SDO the morning of cytometry analysis and measured after 5 hours of growth to a concentration of ~200-500 events/μL.

Western Blot

For yeast expressed proteins, yeast cultures grown to an OD600 of 1. Cells were harvested by centrifugation. Cells were lysed by vortexing for 5 min in the presence of 200 μl of 0.5-mm diameter acid washed glass beads and 200 μl SUMEB buffer (1% SDS, 8 M urea, 10 mM MOPS pH 6.8, 10 mM EDTA, 0.01% bromophenol blue, 1mM PMSF). Lysates were then incubated at 65°C for 10 min and cleared by centrifugation prior to electrophoresis and blotting. Antibodies: anti-HA-HRP (REF-12013819001, Clone 3F10, Roche/Millipore Sigma, St. Louis, MO), anti-PGK1 (ab113687, AbCam).

Protein concentrations were quantified using ImageJ, with PGK1 protein measured in each strain to normalize protein concentrations across strains. To compare protein concentrations to AtTPL H1, these were then normalized to TPL concentration using this equation: ([X-H1]/[PGK1])/[AtTPL-H1]/[PGK1] where X is the H1 variant. AtTPL-H1 normalized protein concentrations were plotted on a Log2 scale. For tobacco expressed proteins, four leaf discs were pooled from one representative experiment and homogenized by bead-beating with two steel ball bearings for 1 minute. 200 µl of protein sample buffer was added to each sample, vortexed for 5 minutes and then boiled for 10 minutes and cleared by centrifugation prior to electrophoresis and blotting. Antibodies: anti-HA-HRP (REF-12013819001, Clone 3F10, Roche/Millipore Sigma, St. Louis, MO).

Plant growth

For synthetic repression assays in tobacco, *Agrobacterium*-mediated transient transformation of *N. benthamiana* was performed as per (67). 5 ml cultures of *Agrobacterium* strains were grown overnight at 30°C shaking at 220 rpm, pelleted, and incubated in MMA media (10 mM MgCl2, 10 mM MES pH 5.6, 100 µM acetosyringone) for 3 hours at room temperature with rotation. Strain density was normalized to an OD600 of 1 for each strain in the final mixture of strains before injection into tobacco leaves. Leaves were removed, and eight different regions were excised using a hole punch, placed into a 96-well microtiter plate with 100 µl of water. Each leaf punch was scanned in a 4 × 4 grid for yellow and red fluorescence using a plate scanner (Tecan Spark, Tecan Trading AG, Switzerland). Fluorescence data was quantified and plotted in R (ggplots2).

Acknowledgements

We thank members of the Nemhauser group including Cassandra Maranas, Eric Yang, and Dr. Sarah Guiziou for constructive discussions and comments on this manuscript; Prof. Grant Brown and Prof.

Maitreya Dunham for advice on yeast genetics and approaches; and Prof. Adam Leaché and Prof. Verónica Di Stilio for advice on constructing and interpreting phylogenies of short sequences.

Figure Legends

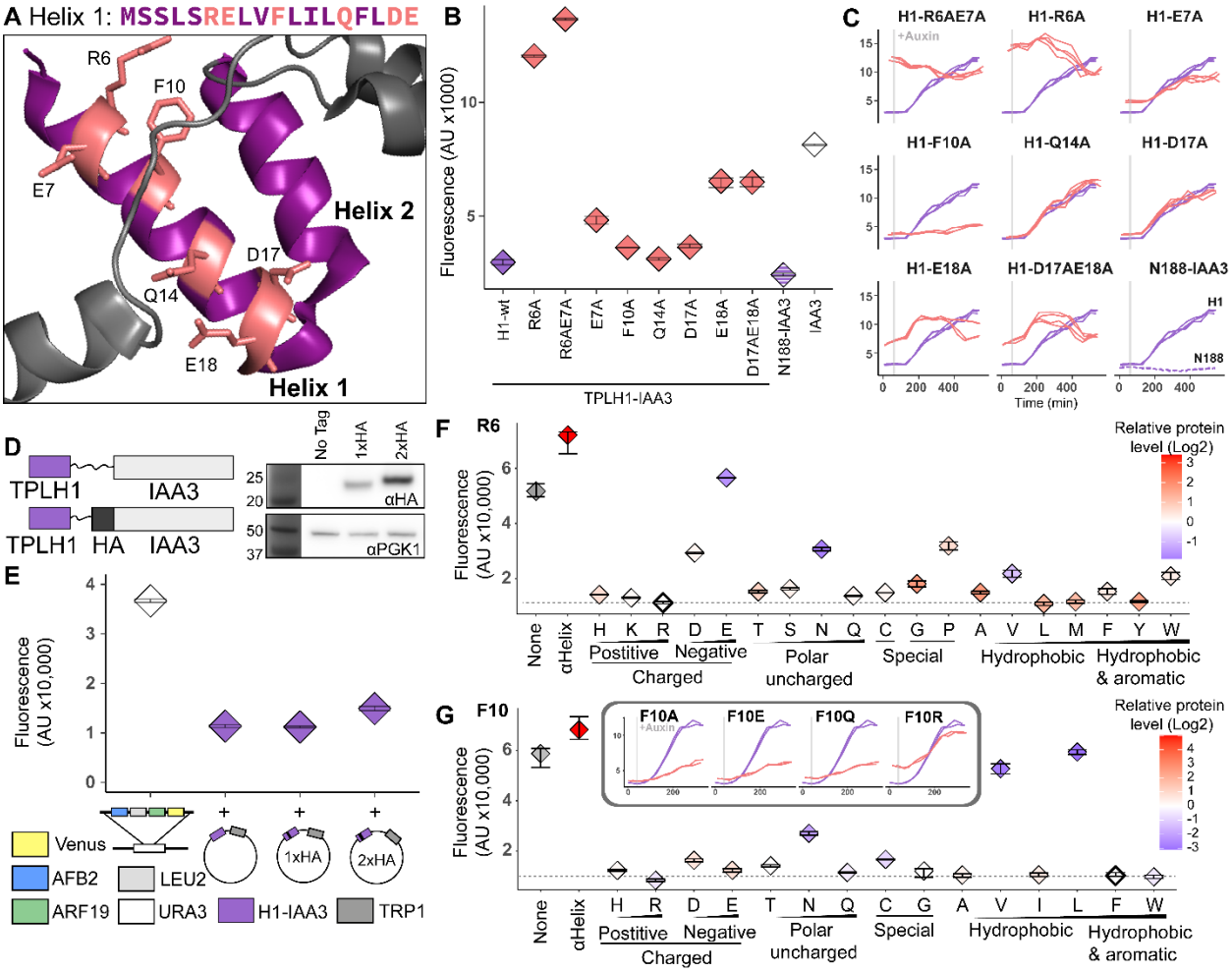


Figure 1. AtTPL LisH H1 is a very short autonomous repression domain

A. Sequence and structure of Helix 1 (AtTPL-H1) (PDB: 5NQS). The LisH domain is colored purple, and amino acids chosen for mutation are highlighted in both the sequence and the structure with pink. **B.** An alanine scan of residues predicted to be solvent-facing. Repression activity of indicated alanine substitutions is in red, and wild-type H1 sequence is in blue. AtTPLN188-IAA3 (blue hatch), the first 188 amino acids of AtTPL fused to AtIAA3, and IAA3 with no corepressor (white) are included for reference. **C.** Time course flow cytometry of selected H1 mutations. Auxin (10μM IAA) was added at the time indicated by the gray bar. **D.** Schematic of HA epitope placement and western blots of tagged constructs. **E.** Repression activity of HA-tagged AtTPL-H1 constructs. All components of the *At*ARC^{Sc} that were held constant across experiments (*auxin promoter::Venus*, *ARF19*, *AFB2*) were integrated at the URA3 locus. The variable H1-HA-IAA3 constructs were expressed from a plasmid carrying the *TRP1* prototrophic gene. **F-G.** Flow cytometry was used to measure the repression activity of AtTPL-H1 constructs with a range of amino acid substitutions at position R6 (**F**) and F10 (**G**), as alanine substitutions at these positions strongly inhibited and enhanced repression activity, respectively**.** Constructs with wild-type residues are indicated with a bold outline, and the dotted lines represent their repression strength. Protein accumulation was assayed by western blot using an a-HA antibody and normalized to yeast PGK1. Levels of AtTPL-H1 mutants are shown relative to wild type AtTPL-H1 with each data point color coded from blue (low) to red (high) expression on a log2 scale. **G.** [Inset] Specified variants were tested in the more sensitive, fully integrated *At*ARC^{Sc}. The blue line in each graph is the wild-type TPL-H1 sequence. **B,C,E,F,G.** Each panel represents two independent time course flow cytometry experiments of the AtTPL-H1 constructs indicated, all fused to IAA3. For all cytometry, every point represents the median fluorescence of at least 10,000 individually measured yeast cells (AU - arbitrary units) and error bars represent 95% confidence intervals.

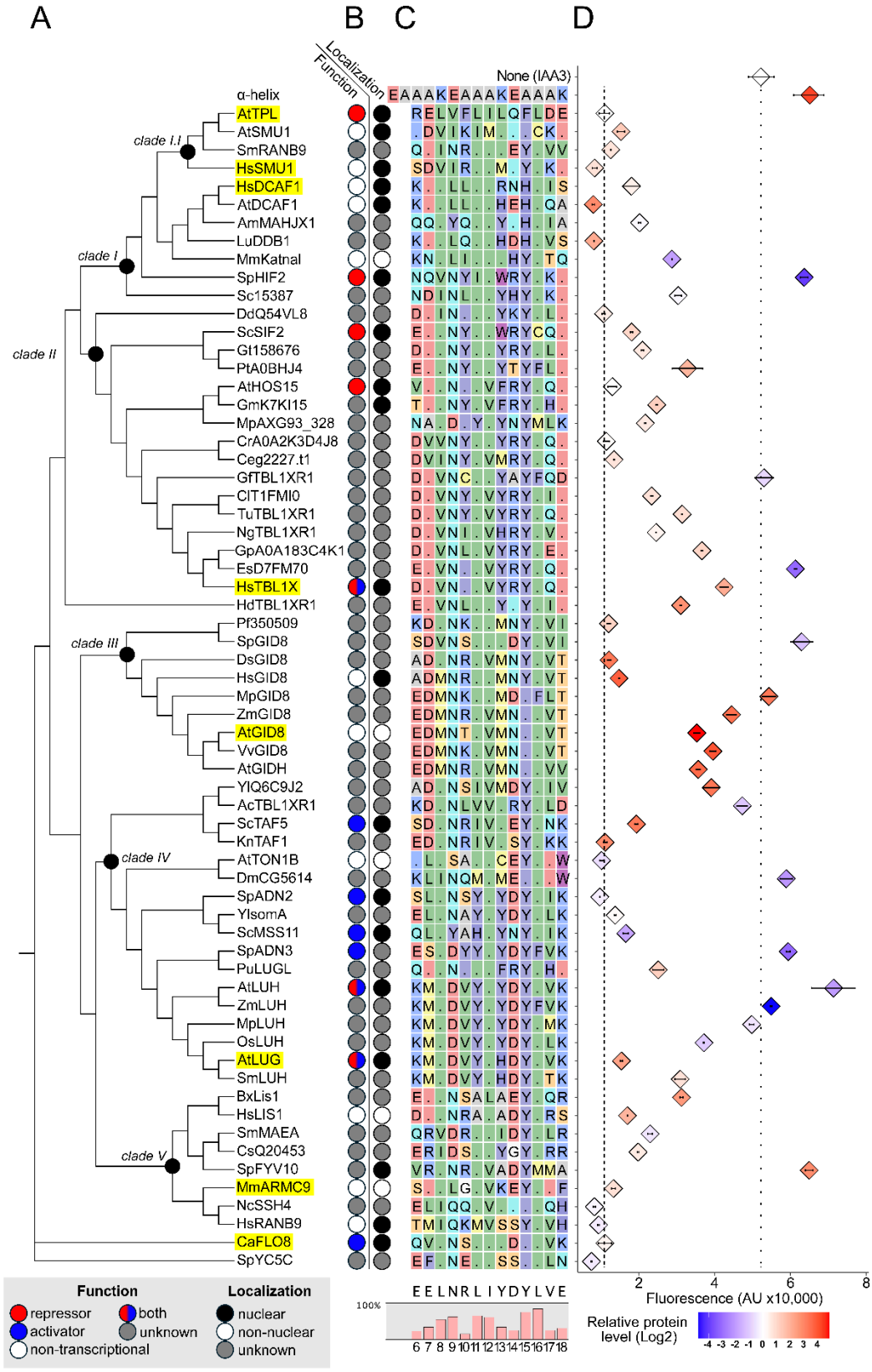

A
B
C
D
Localization
Function
None (IAA3)
α-helix
clade I.I
clade I
clade II
clade III
clade IV
clade V
AtTPL
AtSMU1
SmRANB9
HsSMU1
HsDCAF1
AtDCAF1
AmMAHJX1
LuDDB1
MmKatnal
SpHIF2
Sc15387
DdQ54VL8
ScSIF2
Gt158676
PtA0BHJ4
AtHOS15
GmK7KI15
MpAXG93_328
CrA0A2K3D4J8
Ceg2227.t1
GfTBL1XR1
ClT1FMI0
TuTBL1XR1
NgTBL1XR1
GpA0A183C4K1
EsD7FM70
HsTBL1X
HdTBL1XR1
Pf350509
SpGID8
DsGID8
HsGID8
MpGID8
ZmGID8
AtGID8
VvGID8
AtGIDH
YlQ6C9J2
AcTBL1XR1
ScTAF5
KnTAF1
AtTON1B
DmCG5614
SpADN2
YlsomA
ScMSS11
SpADN3
PuLUGL
AtLUH
ZmLUH
MpLUH
OsLUH
AtLUG
SmLUH
BxLis1
HsLIS1
SmMAEA
CsQ20453
SpFYV10
MmARMC9
NcSSH4
HsRANB9
CaFLO8
SpYC5C
Function
repressor
activator
non-transcriptional
both
unknown
Localization
nuclear
non-nuclear
unknown
100%
E E L N R L I Y D Y L V E
6 7 8 9 10 11 12 13 14 15 16 17 18
2
4
6
8
Fluorescence (AU x10,000)
Relative protein level (Log2)
-4 -3 -2 -1 0 1 2 3 4

Figure 2. The repressive function of the LisH domain is likely ancestral. A. A Maximum Likelihood (46) phylogeny of LisH-H1 sequences from diverse eukaryotes. Ancestral sequences of interest were inferred (42) at nodes of interest (black dot). A representative protein was selected for each sequence. **B.** Published function and subcellular localization for each protein were annotated and sources cited (Supplement 3). The first column marks whether a protein is a transcriptional repressor (red), transcriptional activator (blue), has another function (white), or an uncharacterized function (gray). The second column marks proteins as nuclearly localized (black), non-nuclear (white), or uncharacterized (gray). **C.** LisH-H1 sequences were aligned and residues colored by their physicochemical class (RASMOL color scheme (49)). Residues that are the same as those in the AtTPL-H1 sequence at the top of the alignment are indicated with a period. The consensus sequence for H1, and the relative conservation rate of different residues along the helix, are displayed below the alignment. **D.** Flow cytometry and a modified *At*ARC^{Sc} depicted in Fig. 1E was used to quantify the relative repressive function of different LisH-H1s. We have marked the fluorescence levels detected by the positive AtTPL-H1-HA-IAA3 control (dashed line) and negative IAA3 repression control (dotted line). Protein accumulation was measured by western blot and normalized to yeast PGK1. Levels of protein expression are shown relative to AtTPL-H1 with each data point color coded from blue (low) to red (high) expression on a log2 scale. For all cytometry, every point represents the median fluorescence of at least 10,000 individually measured yeast cells (AU - arbitrary units) and error bars represent 95% confidence intervals.

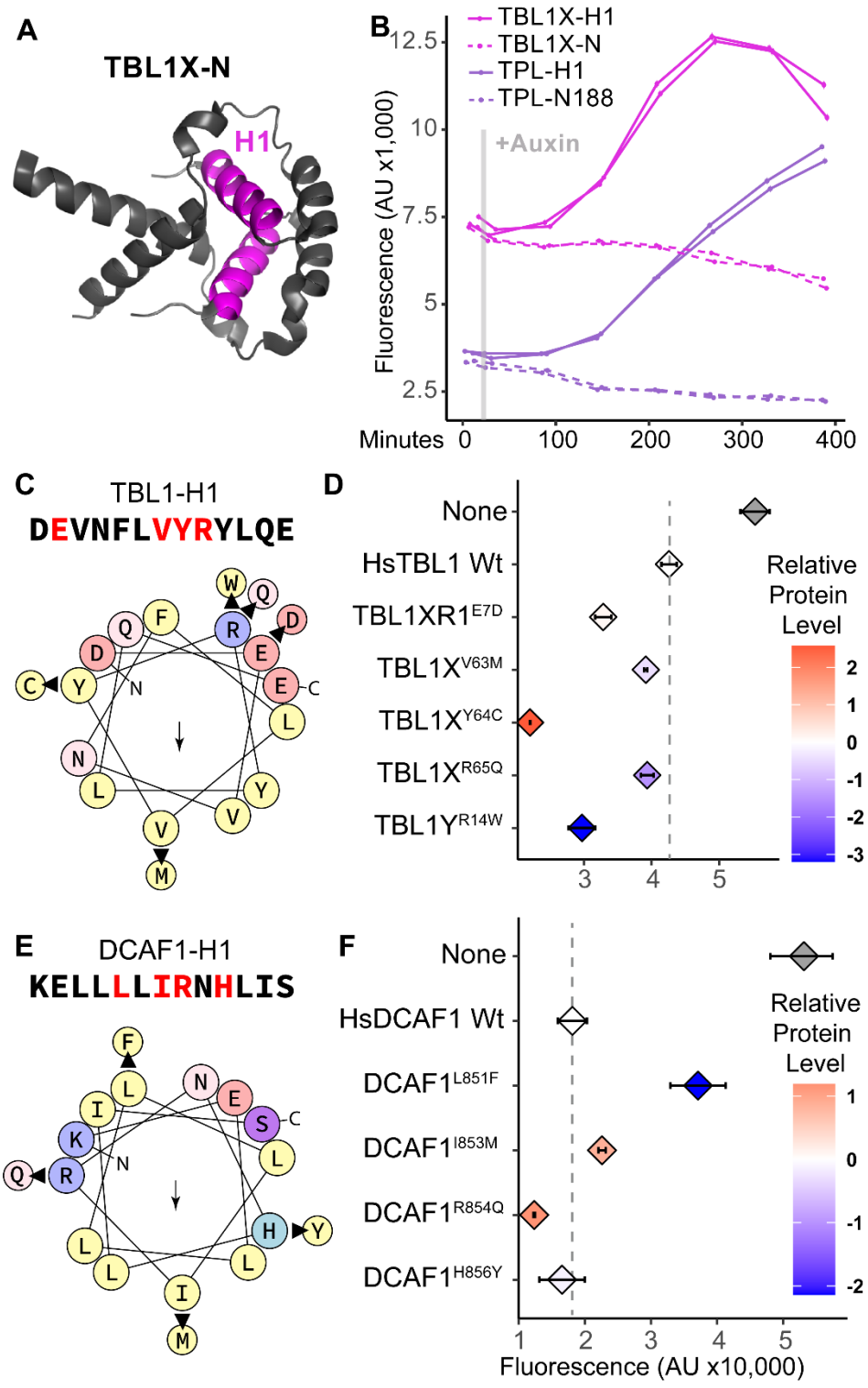

A
TBL1X-N
H1
B
TBL1X-H1
TBL1X-N
TPL-H1
TPL-N188
+Auxin
Fluorescence (AU x1,000)
12.5
10
7.5
5
2.5
Minutes 0
100
200
300
400
C
TBL1-H1
DEVNFLVYRYLQE
D
None
HsTBL1 Wt
TBL1XR1E7D
TBL1XV63M
TBL1XY64C
TBL1XR65Q
TBL1YR14W
Relative Protein Level
2
1
0
-1
-2
-3
3
4
5
E
DCAF1-H1
KELLLLIRNHLIS
F
None
HsDCAF1 Wt
DCAF1L851F
DCAF1I853M
DCAF1R854Q
DCAF1H856Y
Relative Protein Level
1
0
-1
-2
1
2
3
4
5
Fluorescence (AU x10,000)

Figure 3: LisH domains are important for human disease. A. The dimerized protein structure of the N-terminal domain of Hs TBL1X, with H1 highlighted (pink, 2XTD). **B.** Time course flow cytometry of H1 and N-termini of HsTBL1X-IAA3 and AtTPL-IAA3 following auxin addition. Auxin (IAA-10μM) was added at the indicated time (gray bar, + Aux). HsTBL1X (pink) and AtTPL (purple), isolated H1 (solid line), N-terminal region (dotted line). Yeast strains are grown and measured after 400 minutes, with auxin added to some samples (solid lines) at the time indicated by the gray line. **C.** and **E.** Helical wheel depiction of HsTBL1X and HsDCAF1 H1 sequences colored by their physicochemical class (yellow – hydrophobic, red/blue – charged, pink/purple – polar uncharged, arrow indicates hydrophobic face) produced by HeliQuest (55). Arrows show the mutations found in these loci in the catalog of Somatic Mutations in Cancer (COSMIC) library (52), and where they occur. **D,F.** Effects on protein repressive function of these mutations in HsTBL1 and HsDCAF1 sequences are measured using flow cytometry. Each panel represents two independent time course flow cytometry experiments of the H1s indicated. For all cytometry, every point represents the median fluorescence of at least 10,000 individually measured yeast cells (AU - arbitrary units) and error bars represent 95% confidence intervals. Protein accumulation was measured by western blot and normalized to yeast PGK1. Levels of mutant protein constructs are shown relative to wild type H1s (HsTBL1-H1 and HsDCAF1-H1, respectively, dotted line) with each data point color coded from blue (low) to red (high) expression on a log2 scale.

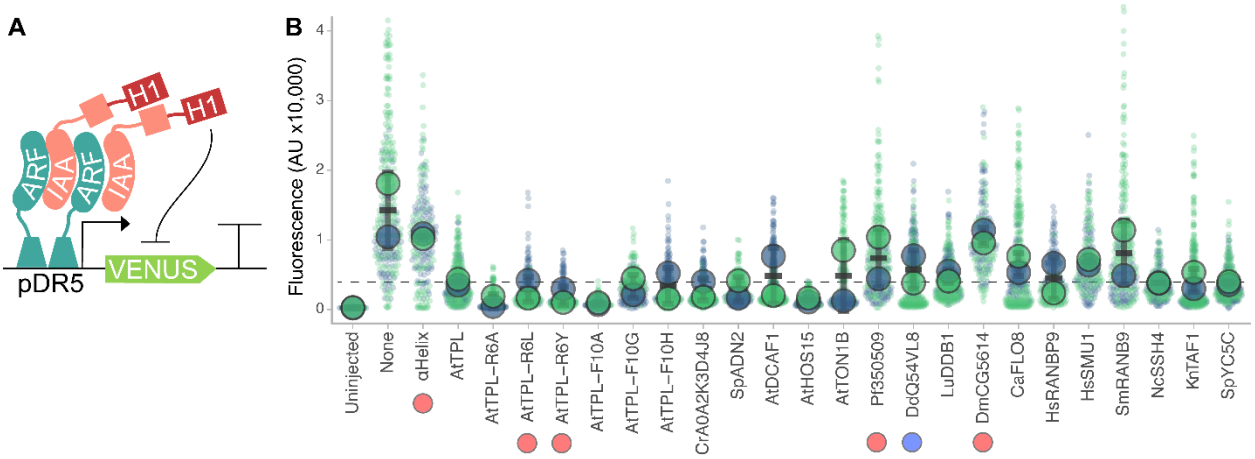


Figure 4. The H1 can act as a synthetic repressor domain *in planta.*

A. Schematic of components injected in transient repression assays in *Nicotiana benthamiana*. pDR5:Venus reporter is based on a highly sensitive synthetic auxin promoter (57). Asterisk indicates EAR motif mutation. **B.** Repression activity of a range of H1 sequences are shown. Reporter activation was measured in four separate leaf injections (biological replicates) in two days of injection (large circles are pooled data from one day). Dashed line indicates AtTPL-H1 repression level. Circles below the construct names indicate protein levels that are 5-fold higher (red) or 5-fold lower (blue) than those of H1-IAA3 constructs.

PNAS

Supplementary Information for

A single helix repression domain is functional across eukaryotes

Alexander R. Leydon[1*], Román Ramos Baez[1*], Jennifer L. Nemhauser[1]

[1] University of Washington, Department of Biology, Seattle, WA 98105-1800, USA

* These authors contributed equally to this work.

Jennifer L Nemhauser

Email: jn7@uw.edu

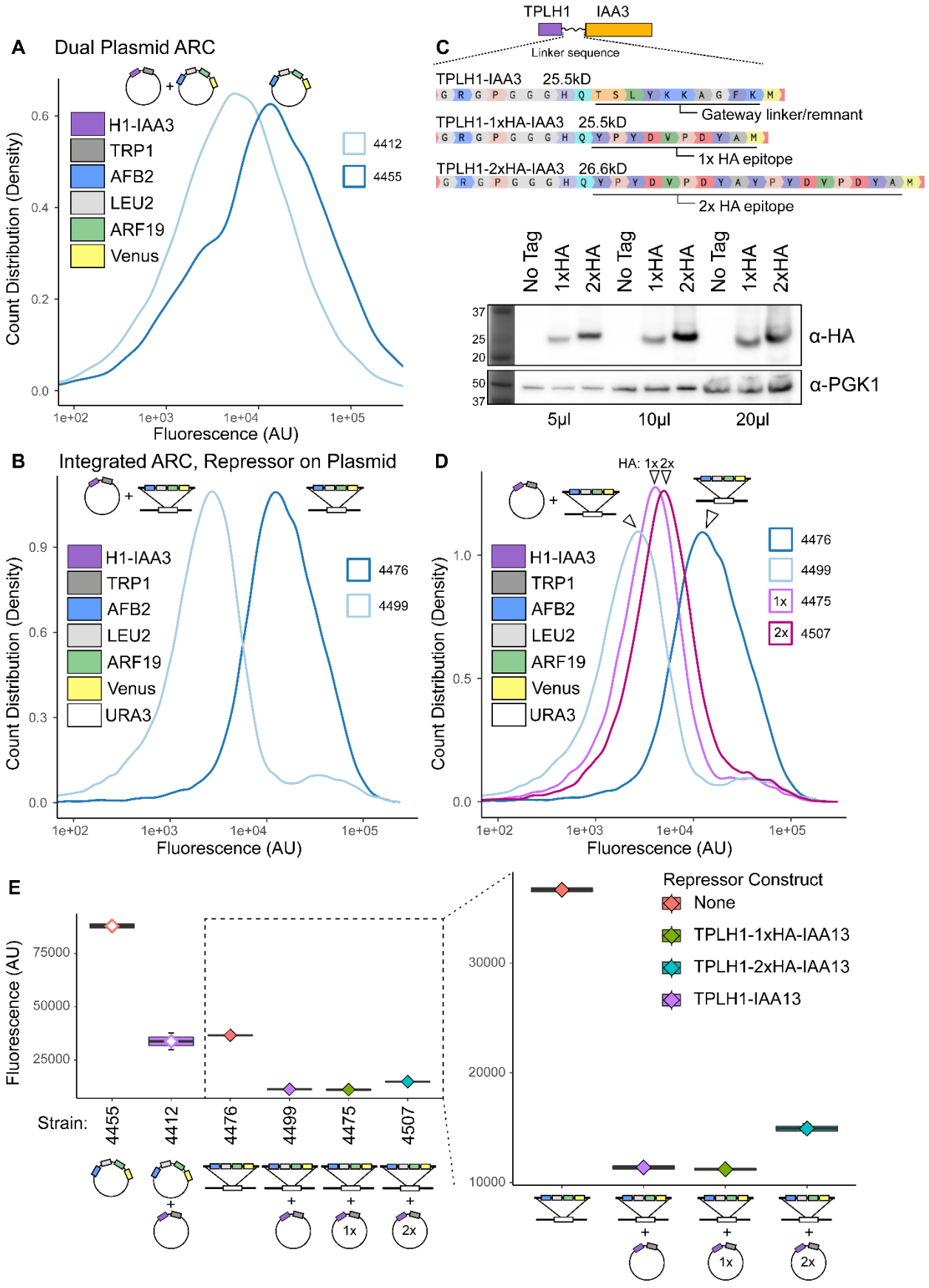
This PDF file includes:

Figures S1 to S2 and S5 to S8

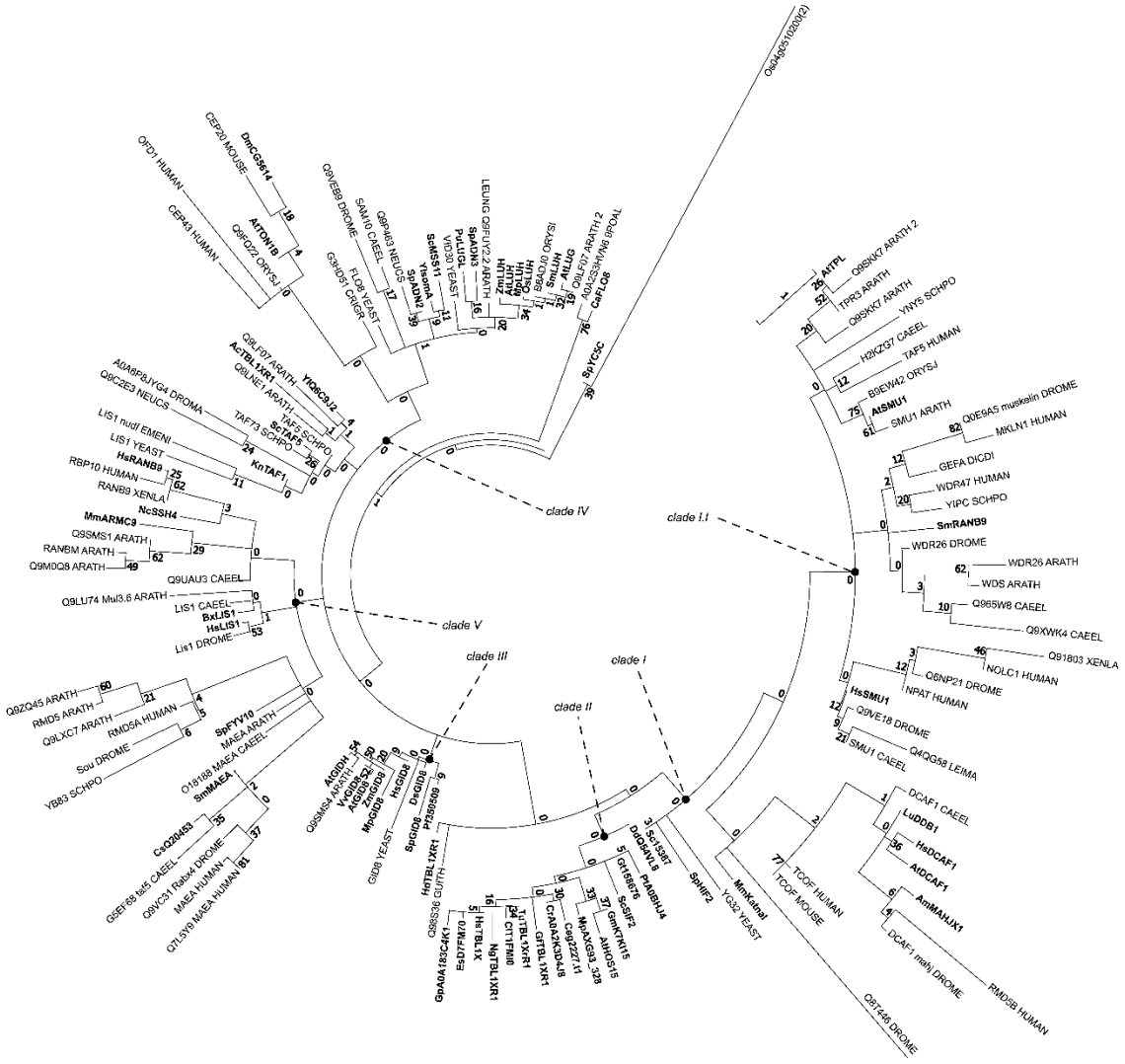
Table S3

Table S4

Legends for S1 to S8


A
Dual Plasmid ARC
H1-IAA3
TRP1
AFB2
LEU2
ARF19
Venus
4412
4455
Count Distribution (Density)
Fluorescence (AU)
B
Integrated ARC, Repressor on Plasmid
URA3
4476
4499
C
TPLH1
IAA3
Linker sequence
TPLH1-IAA3 25.5kD
Gateway linker/remnant
TPLH1-1xHA-IAA3 25.5kD
1x HA epitope
TPLH1-2xHA-IAA3 26.6kD
2x HA epitope
No Tag
1xHA
2xHA
α-HA
α-PGK1
5µl
10µl
20µl
D
HA: 1x 2x
4507
E
Repressor Construct
None
TPLH1-1xHA-IAA13
TPLH1-2xHA-IAA13
TPLH1-IAA13
Strain:
4455
4412
4476
4499
4475
4507

Supplement 1: the single plasmid ARC uses a hybrid integrated/unintegrated yeast auxin response circuit. A. Flow cytometry on strains containing the ARC split into two plasmids with (light blue) or without (dark blue) an H1 repressor. In circuits with an unintegrated reporter, repression was observed, yet there was a wide peak width, limiting the resolution between the repressed and de-repressed response states. **B.** Integration of all components except the repressor led to tighter peak width distributions and increased the resolution of the repressed state when tested by fluorescence flow cytometry. **C.** Schematic of engineered versions of the H1-IAA3 repressor with single (1x) or double (2x) HA epitope tags, and Western blots with antibodies against HA and PGK1. A single HA epitope was sufficient for detection. **D.** Flow cytometry of epitope tagged H1-IAA constructs (1xHA – violet, 2xHA – wine, untagged repressors - light blue, no repressor - dark blue). **E.** Summary of fluorescence flow cytometry. For all flow cytometry each panel represents two independent time course flow cytometry experiments of the TPL helices indicated, all fused to IAA3, every plot represents the average fluorescence of at least 10,000 individually measured yeast cells (AU - arbitrary units). The error bars represent 95% confidence intervals.


clade I
clade I.1
clade II
clade III
clade IV
clade V
Os04g0510200(2)
AtTFL
Q8SNK7 ARATH 2
TPR3 ARATH
Q8SKK7 ARATH
YNY5 SCHPO
H2KZG7 CAEEL
TAF5 HUMAN
B9EW42 ORYSJ
AtSMU1
SMU1 ARATH
Q0E9A5 mueklin DROME
MKLN1 HUMAN
GEFA DICDI
WDR47 HUMAN
YIPC SCHPO
SmRANB9
WDR26 DROME
WDR26 ARATH
WDS ARATH
Q965W8 CAEEL
Q8XWK4 CAEEL
Q91803 XENLA
NOLC1 HUMAN
Q6NP21 DROME
NPAT HUMAN
HsSMU1
Q9VE18 DROME
SMU1 CAEEL
Q4QG58 LEIMA
DCAF1 CAEEL
LuDDB1
HsDCAF1
AtDCAF1
AmMAHJX1
DCAF1 mahj DROME
RMD5B HUMAN
Q87446 DROME
TCOF HUMAN
TCOF MOUSE
MmKatnal
YG32 YEAST
SpNIF2
3 Sc41587
DdGSAV1.8
PfADBHU4
5 Gt158876
ScSIF2
GtNKTKT15
AtHOS15
MpAXG93_328
Cs9g22741
CrAbACKD4.8
GtTBL1XR1
CrTBL1XR1
NgTBL1XR1
HsTBL1X
CrTBL1XR0
GbA0A183C4K1
EsD17W10
Q8S536 GUITH
GtDB TEAST
SpGID8
DsGID8
HsGID8
MpGID8
AtGID8-2
VcGID8-2
AtGID8-1
PmGID8
AtGID8
Q8SMS4 ARATH
SpFYV10
MAEA ARATH
MAEA CAEEL
SmMAEA1
C1B188 MAEA CAEEL
CsGD0453
GSEF69 taf5 CAEEL
Q8VC31 Rabx4 DROME
MAEA HUMAN
GTL5Y9 MAEA HUMAN
Scu DROME
YB83 SCHPO
Q9ZQ45 ARATH
RMD5 ARATH
Q9LXC7 ARATH
RMD5A HUMAN
Q9LU74 Mvl3.6 ARATH
LIS1 CAEEL
BxLIS1
HsLIS1
Lis1 DROME
Q9SMS1 ARATH
RANBM ARATH
Q9M0Q8 ARATH
Q9UAU3 CAEEL
NcSSH4
MmARMC9
RANB9 XENLA
RBP10 HUMAN
HsRANB9
KnTAF1
LIS1 nudf EMENI
LIS1 YEAST
A0A6PBJYG4 DROMA
Q8C2E3 NEUCS
TAF73 SCHPO
ScTAF5
Q9LNE1 ARATH
AcTBL1XR1
Q8LF07 ARATH
YIORC9.1/2
G3MD51 CRIGR
FLO8 YEAST
SAM10 CAEEL
Q8PM93 NEUCS
SpADK2
SdMS51
YnomA
PdCLO43
SpAON3
VIC33 YEAST
LEUNG G3FUY2.2 ARATH
AtLUH
HsTLUH
ZmLUH
GtLUH
SmLUH
AtLUG
Q8LFU7.1 ARATH 2
A0A2S3WYN6 9PEOA
CeFLO8
SpYCS5C
Os04g0510200(2)
CEP20 MOUSE
DmCG5514
AtTON1B
Q9FQZ2 ORYSJ
CEP43 HUMAN
OFD1 HUMAN

Supplement 2: Extended phylogeny for Figure 2. The evolutionary history of LisH-H1 sequences was inferred by using the Maximum Likelihood method and Le Gascuel, 2008 model (58). The tree with the highest log likelihood (-2709.60) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 6.1034)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 143 amino acid sequences. There were a total of 13 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (59). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (42).

#	Sequence	Representative Gene	Function	Localization
1	RELVFLILQFLDE	AtTPL	Repressor	Nuclear
2	RDVIKIMLQFCKE	AtSMU1	Other	Nuclear
3	QEINRLILEYLVV	SmRANB9	Unknown	Unknown
4	SDVIRLIMQYLKE	HsSMU1	Other	Nuclear
5	KELLLLIRNHLIS	HsDCAF1	Other	Nuclear
6	KELLLLIHEHLQA	AtDCAF1	Other	Nuclear
7	QQLYQLIYQHLIA	AmMAHJX1	Unknown	Unknown
8	KELLQLIHDHLVS	LuDDB1	Unknown	Unknown
9	KNLLILILHYLTQ	MmKatnal	Other	Non-nuclear
10	NQVNYIIWRYLKE	SpHIF2	Repressor	Nuclear
11	NDINLLIYHYLKE	Sc15387	Unknown	Unknown
12	DEINFLIYKYLLE	DdQ54VL8	Unknown	Unknown
13	EELNYLIWRYCQE	ScSIF2	Repressor	Nuclear
14	DELNYLIYRYLLE	Gt158676	Unknown	Unknown
15	EELNYLIYTYFLE	PtA0BHJ4	Unknown	Unknown
16	VELNFLVFRYLQE	AtHOS15	Repressor	Nuclear
17	TELNYLVFRYLHE	GmK7KI15	Unknown	Nuclear

#	Sequence	Representative Gene	Function	Localization
18	QELNFLIFRYLHE	MpAXG93_328	Unknown	Unknown
19	DVVNYLIYRYLQE	CrA0A2K3D4J8	Unknown	Unknown
20	DVINYLVMRYLQE	Ceg2227.t1	Unknown	Unknown
21	DEVNCLIYAYFQD	GfTBL1XR1	Unknown	Unknown
22	DEVNYLVYRYLIE	ClT1FMI0	Unknown	Unknown
23	DEVNYLVYRYLQE	TuTBL1XrR1	Unknown	Unknown
24	DEVNILVHRYLVE	NgTBL1XR1	Unknown	Unknown
25	DEVNLLVYRYLEE	GpA0A183C4K1	Unknown	Unknown
26	EEVNFLVYRYLQE	EsD7FM70	Unknown	Unknown
27	DEVNFLVYRYLQE	HsTBL1X	Activator/Repressor	Nuclear
28	EEVNLLIYQYLIE	HdTBL1XR1	Unknown	Unknown
29	KDLNKLIMNYLVI	Pf350509	Unknown	Unknown

#	Sequence	Representative Gene	Function	Localization
30	SDVNSLILDYLVI	SpGID8	Unknown	Unknown
31	ADLNRLVMNYLVT	DsGID8	Unknown	Unknown
32	ADMNRLIMNYLVT	HsGID8	Other	Nuclear
33	EDMNKLIMDFFLT	MpGID8	Unknown	Unknown
34	EDMNRLVMNFLVT	ZmGID8	Unknown	Unknown
35	EDMNTLVMNFLVT	AtGID8	Other	Other
36	EDMNKLVMNFLVT	VvGID8	Unknown	Unknown

#	Sequence	Representative Gene	Function	Localization
37	EDMNRLVMNFLVV	AtGIDH	Unknown	Unknown
38	ADLNSIVMDYLIV	YlQ6C9J2	Unknown	Unknown
39	KDLNLVVLRYLLD	AcTBL1XR1	Unknown	Unknown
40	SDLNRIVLEYLNK	ScTAF5	Activator	Nuclear
41	EDLNRIVLSYLKK	KnTAF1	Unknown	Unknown
42	RLLSALICEYLDW	AtTON1B	Other	Other
43	KLINQMIMEFLDW	DmCG5614	Unknown	Unknown
44	SLLNSYIYDYLIK	SpADN2	Activator	Nuclear
45	ELLNAYIYDYLLK	YlsomA	Unknown	Unknown

#	Sequence	Representative Gene	Function	Localization
46	QLLYAHIYNYLIK	ScMSS11	Activator	Nuclear
47	ESLDYYIYDYFVK	SpADN3	Activator	Unknown
48	NALDFYIYNYMLK	PuLUGL	Unknown	Unknown
49	KMLDVYIYDYLVK	AtLUH	Activator/Repressor	Nuclear
50	KMLDVYIYDYFVK	ZmLUH	Unknown	Unknown
51	KMLDVYIYDYLMK	MpLUH	Unknown	Unknown
52	KMLDVYIYDYLLK	OsLUH	Unknown	Unknown

#	Sequence	Representative Gene	Function	Localization
53	KMLDVYIHDYLVK	AtLUG	Activator/Repressor	Nuclear
54	KMLDVYIHDYLTK	SmLUH	Unknown	Unknown
55	EELNSALAEYLQR	BxLIS1	Unknown	Unknown
56	DELNRAIADYLRS	HsLIS1	Other	Non-nuclear
57	QRVDRLIIDYLLR	SmMAEA	Unknown	Unknown
58	ERIDSLIYGYLRR	CsQ20453	Unknown	Unknown
59	VRLNRLVADYMMA	SpFYV10	Unknown	Nuclear
60	SELLGLVKEYLDF	MmARMC9	Other	Other
61	ELIQQLVLQFLQH	NcSSH4	Unknown	Unknown

#	Sequence	Representative Gene	Function	Localization
62	TMIQKMVSSYLVH	HsRANBP9	Other	Nuclear
63	QVLNSLILDFLVK	CaFLO8	Activator	Nuclear
64	EFLNELISSFLLN	SpYC5C	Unknown	Unknown

Uniprot Searchable	**Species**	**Other Genes**
TPL_ARATH	Arabidopsis thaliana	TmTPL, VvTPR4, AoTPR2, CsTPL, CcTPR1, PaTPL, CmTPR2, AcTPR2, CnTPR3, TuTPR2, over 100 proteins
SMU1_ARATH	Arabidopsis thaliana	GwSMU1, HhSMU1, PpSMU1, HgSMU1, MdSMU1, BsSMU1X2, GaSMU1X1, GvSMU1, over 100 proteins
G4VEU8_SCHMA	*Schistosoma mansoni*	A0A183T2Q9, A0A068YDU5, H2KNK2, A0A183B251 , G4VEU8, A0A183VRA7, G4VEU8_SCHMA
SMU1	Homo sapiens (Human)	Q2TAY7
Q9Y4B6	Homo sapiens (Human)	PtDCAF1, CjDCAF1, NaDACF1, SaDCAF1, PcDACF1, and over 100 proteins
DCAF1_ARATH	Arabidopsis thaliana	A0A178UYS1, Q9M086, DCAF1_ARATH
AgaP_AGAP012134	Anopheles merus	AaMAHJX1, AcMAHJX1, AmMAHJ, AaMAHJX2, XP_320393.4, KFB51317.1
A0A1S3K6T0	*Lingula unguis*	A0A1S3K7Q1, A0A1S3IUX2
KATL2_MOUSE	*Mus musculus*	F7DX12, F1M5A4, Q9D3R6, KATL2_MOUSE
HIF2_SCHPO	Schizosaccharomyces pombe	none
SteCoe_15387	*Stentor coeruleus*	A0A1R2C3W4, A0A1R2CB06, A0A1R2BT11
DDB0206475_DICDI	*Dictyostelium discoideum*	DDB_G0280261, DICPUDRAFT_29050
SIF2_YEAST	*Saccharomyces cerevisiae*	P38262, SIF2_YEAST
GUITHDRAFT_158676	*Guillardia theta*	BoCAD5229314.1
A0BHJ4	*Paramecium tetraurelia*	A0DLD6, A0BHJ4
HOS15_ARATH	Arabidopsis thaliana	Q9FN19, A0A178UQT7, HOS15_ARATH, F2DCR6, A0A287I129, A0A287I0S4, A0A287I0S0
K7KI15_GLYMA	Glycine max	A0A199W458, K7KI15, D8T4U4, D8RCD5

Uniprot Searchable	**Species**	**Other Genes**
AXG93_328s1070	*Marchantia polymorpha subsp. ruderalis*	A0A176VRL0
A0A2K3D4J8_CHLRE	*Chlamydomonas reinhardtii*	TsFbxw7, CHLRE_12g527500v5, CiHXX76_004694, 10 proteins
CEUSTIGMA_g2227.t1	*Chlamydomonas eustigma*	A0A250WW80
A0A1C7LUX2	*Grifola frondosa*	KAH6916628.1
T1FMI0_HELRO	*Helobdella robusta*	none
T1K4V1	*Tetranychus urticae*	A0A2D4BVA7, C1FJI2, T1K4V1
W7TZ27_9STRA	*Nannochloropsis gaditana*	NSK_003728
GLOPA	*Globodera pallida*	MeCAD2177457.1, MgKAF7638710.1, MeCAD2168972.1
ECTSI	*Ectocarpus siliculosus*	A0A024G5P9, F0YLG7
TBL1X_HUMAN	Homo sapiens (Human)	T0QLT6, D0NVN2, A7S5U0, A0A2B4S828, A0A1J1IKJ5, Q95RJ9, A0A1B0FD47, I0YVP9, H9JDI6, B7Q2X9, A0A1S3J4R8, A0A0L8H107, K1RDI6, V4A8I8, Q7Q371, A0A087UPF7, A0A2A4J0P6, A0A1S4EKS5, E9GI91, K7IPS0, A0A158NMP9, T1HP17, E0VC07, A0A2J7QTE6, A0A067R4E4, A0A226E5C7, A0A1W4XAC3, D6WDR4, A0A1D2N2Y2, T1JMU6, A0A2G8JKT3, A0A2G8LR25, C3ZEP1, O60907, S9XQS6, A2RUX0, A5WVU0, A0A1D5PXG9, A5PF33, F6S036, A0A226NGR2, Q9QXE7, G3V6G5, A0A1D5NXL1, Q9BQ87, Q9BZK7, A0A091KAK7, D3ZNF4, Q8BHJ5, M7B1G6, F6WUH3, W4XYV6, F6YCJ1
RAMVA_A0A1D1V3Z7	*Hypsibius dujardini*	RAMVA_A0A1D1V3Z7
BCR36DRAFT_350509	*Piromyces finnis*	KAG4099912.1, OUM64527.1, ORY12276.1,

Uniprot Searchable	**Species**	**Other Genes**
YDED_SCHPO	*Schizosaccharomyces pombe*	none
B4IML9	*Drosophila sechellia*	Q8SZN4
GID8_HUMAN	Homo sapiens (Human)	GID8_CHICK, A0A087TXU7, A0A1S3JEZ4, A0A226NJJ8, C3ZQT3, B7PS79, Q9VWS1, E9QHU7, Q6PC55, M7BF90, T1J9E3, Q9NWU2, Q9D7M1, Q6YDN8, F7CLX0, Q5ZKQ7, V8NIM7, F7DIR3, E7FGY2
MARPO_0028s0043	Marchantia polymorpha	, AXG93_2931s1420
A0A317YIJ7_MAIZE	Zea mays	, DKX38_017052, PR202_ga26473, Sadunf11G0026900, ZmGIH8, ONL95757.1, GFY91837.1, C3L33_15476, GFZ03098.1, C2845_PM01G08500, XP_003558414.1, 89 proteins
GID8_ARATH	Arabidopsis thaliana	BRARA_K00274, CAF2133834.1, CAA7029683.1, Bca52824_086799, HID58_043154, DY000_02041394, BnaA03g60080D, XP_010418067.1, EsGIDH, CsGID8, CrGID8X1, 49 proteins
W1P668	*Amborella trichopoda*	VvGIDH, KAG7019632.1, TEA_010235, E3N88_37174, FEM48_Zijuj01G0148200, Ahy_B03g063295, F2P56_009768, KAA0056956.1, KAE8716948.1, E3N88_41952, CcGID8, CcGIDH, CBI18750.3, 100 proteins

Uniprot Searchable	**Species**	**Other Genes**
O23690_ARATH	*Arabidopsis thaliana*	CAE5957163.1, AXX17_AT1G11330, KAG7591256.1, AlGID81X, KAG7653870.1, NP_001318979.1, EFH66107.1, KAG7645900.1, AAO63342.1, NP_001321853.1, CAD5312385.1, 19 proteins
YALI0_D10791g	*Yarrowia lipolytica*	Q6C9J2
TBL1XR1_ANACO	*Ananas comosus*	CAD1841307.1, AcHOS15
TAF5_YEAST	Saccharomyces cerevisiae	P38129, A0A1E3PKV2, W1Q9X6, A0A1D8PSS1, K0KTJ7, S6ENA4, A0A167EJ71, C4R4L4
A0A1Y1HYL3	*Klebsormidium nitens*	A0A1Y1HYL3
TON1B_ARATH	Arabidopsis thaliana	CIPAW_02G177500, I3760_02G177300, JrTON1A, I3760_02G177300, CiTON1A, JmTON1A, Leryth_016134, CAD5325854.1, CmTON1AX1, AlTON1A, 44 proteins
Q9VF40	*Drosophila melanogaster*	AT18024p
ADN2_SCHPO	*Schizosaccharomyces pombe*	NONE
YALI0_E03102g	*Yarrowia lipolytica*	FOA43_001062, HII12_001016, BKA90DRAFT_133770, B0I71DRAFT_126502, YALI2_E00070g, YALI0E03102p, BRETT_003633, CJU89_1027, 9 proteins

Uniprot Searchable	**Species**	**Other Genes**
MSS11_YEAST	Saccharomyces cerevisiae	CAD6641633.1, AJS95620.1, Samss11p, YJM1443, YJM1389, JM1399, YJM320, AJS90831.1, AJS85596.1, SpMss11, AJS66828.1, AJS89087.1, AJS93898.1, AJS84716.1, 54 proteins
ADN3_SCHPO	*Schizosaccharomyces pombe*	ADN3_SCHPO
AXG93_146s1420	Pyrus ussuriensis	MARPO_0023s0181
LUH_ARATH	Arabidopsis thaliana	A0A287T3L7, A0A287T3B8, A0A287T3E2, A0A287T3D2, A0A287T3C9, A0A287T3M2, A0A287T3B7, A0A199V044, Q10A93, K7K8T5, I1LA80, O48847, A0A178VQH2, A0A2K3P9I9, LUH_ARATH
F2ELW0_HORVV	Zea mays	A0A287L2N7, A0A287L2N6, F2ELW0, A0A287L2M9, A0A287L2Q9, 75 proteins
MARPO_0033s0155	Marchantia polymorpha	COLO4_34959, F8388_006400, DVH24_017861, RZB52406.1, RZB52407.1, PuLUGL, DVH24_029234, POE58772.1, QsLUG, C4D60_Mb07t20460, JHK85_023989, CAG1855999.1, CsLUG1X, 100 proteins
Q0JBT9	*Oryza sativa subsp. japonica*	TuLUG, MQM01692.1, ZmLEGLC5167_041388, CY35_06G035600, BDL97_06G036400, C5167_015243, MQL71672.1, OIW03074.1, XP_019457374.1, LaLUGL, 100 proteins

Uniprot Searchable	**Species**	**Other Genes**
LUG_ARATH	Arabidopsis thaliana	B9RPC5_RICCO, A0A2J6KP98, A0A178UVW2, A0A251SRG4, A0A251UB06, A0A0R0FPP1, A0A0R0JDA7, A0A2K3MNX1, D8SAR0, D8QW58, S8CL12, I1M9M2, S8DZ67, Q9FUY2
SELMO_94317	*Selaginella moellendorffii*	SmLUGX1, X2, X3, X4
A0A1I7STR7	*Bursaphelenchus xylophilus*	, CAD5215296.1
LIS1_HUMAN	Homo sapiens (Human)	Q803D2, LIS1_BOVIN, Q7T394, Q9PTRZ, A0A091JVU8, P43034, P63005, P63004, F6PQP7, V8P437, Q6NZH4
A0A0N5AP93	*Syphacia muris*	VDD86692.1
Q20453_CAEEL	*Caenorhabditis elegans*	H2L2B2, EGT35366.1, NP_001256138.1, NP_001256137.1, NP_505524.2, NP_872162.2, NP_001023935.1
FYV10_SCHPO	Schizosaccharomyces pombe	NONE
ARMC9 MOUSE	Mus musculus	MXQ79359.1, MdARMC9, MbARMC9, KAF4014044.1, XP_037705437.1, XP_037705438.1, XP_045054028.1, XP_045054029.1, XP_022423649.1, NaARMC9X1, PtARMC9X1, DlARMC9X2, 100 proteins
Q7S7G6	Neurospora crassa	GQX73_g5821, CIB48_g11750, PODANS_2_950, KAH6631079.1, PaRBP10, INS49_012262, SMAC_08383, NEUTE1DRAFT_57335, NcRANB, 43 proteins

Uniprot Searchable	**Species**	**Other Genes**
RANB9_HUMAN	Homo sapiens (Human)	F1LVV3, P69566, M7BR83, Q96S59, A0A091KE14, F7D8G8, A0A226NKE1
FLO8_CANAL	*Candida albicans*	XP_002421174.1, W5Q_05038 , AAQ03244.1, MGS_04938, W5O_04937, MG1_04938, MEU_04910, L150_04850, MEY_04890, MGO_04880, MEW_04821, MEM_04918, 29 proteins
YC5C_SCHPO	*Schizosaccharomyces pombe*	O94712

References
DOI: 10.1126/science.1151461
DOI: 10.1104/pp.109.141705
NONE
DOI: 10.1016/j.yexcr.2005.02.017
https://doi.org/10.1016/j.molcel.2012.09.004
DOI: 10.1105/tpc.108.058891
NONE
NONE
DOI: 10.1371/journal.pgen.1007078
DOI: 10.1016/s0960-9822(98)70304-5
NONE
NONE
https://doi.org/10.1016/j.jmb.2005.06.025
NONE
NONE
DOI: 10.1104/pp.18.01156
NONE

References
NONE
NONE
NONE
NONE
NONE
NONE
NONE
NONE
NONE
DOI: 10.1210/jc.2016-2531
NONE
NONE

References
NONE
NONE
DOI: 10.7554/eLife.35528
NONE
NONE
DOI: 10.1186/1471-2229-12-83
NONE

References
NONE
NONE
NONE
DOI: 10.1016/s0092-8674(00)81220-9
NONE
doi: 10.1105/tpc.107.056812
NONE
DOI: 10.1128/EC.00078-09
NONE

References
DOI: 10.1046/j.1365-2958.2003.03247.x
DOI: 10.1371/journal.pgen.1003104
NONE
DOI: 10.1105/tpc.19.00115, DOI: 10.1186/1471-2229-14-54
NONE
NONE
NONE

References
DOI: 10.1073/pnas.230352397
NONE
NONE
https://doi.org/10.1091/mbc.e12-03-0210
NONE
NONE
DOI: 10.1038/nbt1222
DOI: 10.1038/s41588-018-0054-7
NONE

References
DOI: 10.1083/jcb.200801133
DOI: 10.1091/mbc.e05-06-0502
NONE

Supplement 3: H1 sequence information. Each column details information about each LisH-H1 examined in this study, including: its sequence (**A**), the name that appears in figures (**B**), the annotated function of this representative gene (**C**) and any information about its localization (**D**), a Uniprot-searchable name (**E**), the species where this gene was identified (F), other genes with an identical H1 sequence (**G**), and citations for annotated localization and function for each gene ("NONE" for those with uncharacterized function or localization) (**H**).

YEAST PLASMIDS		Corresponding H1 sequences
3455	c1-1_TPL_ARATH	RELVFLILQFLDE
3454	c9-2_DCAF1_ARATH	KELLLLIHEHLQA
3453	c17-3_LUH_ARATH	KMLDVYIYDYLVK
3452	c17-6_LUG_ARATH	KMLDVYIHDYLVK
3451	c14-4_HOS15_ARATH	VELNFLVFRYLQE
3450	c4-3_GID8_ARATH	EDMNTLVMNFLVT
3448	c4-8_O23690_ARATH	EDMNRLVMNFLVV
3447	c1-2_SMU1_ARATH_3	RDVIKIMLQFCKE
3446	c20-1_TON1B_ARATH	RLLSALICEYLDW
3445	c15-9_TBL1X_HUMAN	DEVNFLVYRYLQE
3444	c18-2_LIS1_HUMAN	DELNRAIADYLRS
3443	c4-2_GID8_HUMAN	ADMNRLIMNYLVT
3442	c22-1_RANB9_HUMAN	TMIQKMVSSYLVH
3441	c10-1_KATL2_MOUSE	KNLLILILHYLTQ
3440	c1-3_SMU1_HUMAN	SDVIRLIMQYLKE
3439	c4-1_Dmel_DROME	ADLNRLVMNYLVT
3438	c19-1_MAEA_SYPMU	QRVDRLIIDYLLR
3437	c9-3_AgaP_AGAP012134	QQLYQLIYQHLIA
3436	c15-5_HELRO	DEVNYLVYRYLIE
3435	c18-1_Lis1_BURXY	EELNSALAEYLQR
3434	c9-1_DDB1_LINUN	KELLQLIHDHLVS
3433	c15-6_T1K4V1	DEVNYLVYRYLQE
3432	c11-1_RAMVA	EEVNLLIYQYLIE
3431	c20-2_Q9VF40	KLINQMIMEFLDW
3430	c15-7_GLOPA	DEVNLLVYRYLEE
3429	c3-1_BCR36DRAFT_350509	KDLNKLIMNYLVI
3428	c23-1_YC5C_SCHPO	EFLNELISSFLLN
3427	c14-1_SIF2_YEAST	EELNYLIWRYCQE
3426	c5-1_FLO8_CANAL	QVLNSLILDFLVK
3425	c8-1_TAF5_YEAST	SDLNRIVLEYLNK
3424	c16-3_MSS11_YEAST	QLLYAHIYNYLIK
3423	c17-1_ADN3_SCHPO	ESLDYYIYDYFVK
3422	c12-1_HIF2_SCHPO	NQVNYIIWRYLKE
3421	c16-2_YALI0_E03102g	ELLNAYIYDYLLK
3420	c15-3_Tbl1xr1_RHVIN	DEVNCLIYAYFQD
3419	c16-1_ADN2_SCHPO	SLLNSYIYDYLIK
3418	c6-1_YALI0_D10791g	ADLNSIVMDYLIV
3417	c19-3_FYV10_SCHPO	VRLNRLVADYMMA
3416	c3-2_YDED_SCHPO	SDVNSLILDYLVI
3415	c21-1_SSH4_NUECR	ELIQQLVLQFLQH
3414	c4-6_A8HQD2_CHLRE	EDMNRLVMNFLVT
3413	c15-1_A0A2K3D4J8_CHLRE	DVVNYLIYRYLQE
3412	c4-5_AMTR	EDMNKLVMNFLVT
3411	c4-4_MARPO_0028s0043	EDMNKLIMDFFLT

3410	c17-5_MARPO_0033s0155	KMLDVYIYDYLMK
3409	c7-1_TBL1XR1_ANACO	KDLNLVVLRYLLD
3408	c17-7_SELMO_94317	KMLDVYIHDYLTK
3407	c14-2_GUITH_158676	DELNYLIYRYLLE
3406	c24-1_Q0JBT9	KMLDVYIYDYLLK
3405	c17-4_F2ELW0_HORVV	KMLDVYIYDYFVK
3404	c15-4_tbl1xr1_NAGAD	DEVNILVHRYLVE
3403	c17-2.2_AXG93_328s1070	QELNFLIFRYLHE
3402	c15-2_CHEUS	DVINYLVMRYLQE
3401	c14-5_K7KI15_GLYMA	TELNYLVFRYLHE
3400	c17-2.1_AXG93_146s1420	NALDFYIYNYMLK
3399	c8-2_TAF1_KLENI	EDLNRIVLSYLKK
3398	c15-8_ECTSI	EEVNFLVYRYLQE
3397	c2-1_RANB9_SCHMA	QEINRLILEYLVV
3396	c19-2_CELE_CAEEL	ERIDSLIYGYLRR
3395	c14-3_A0BHJ4	EELNYLIYTYFLE
3394	c13-1_DDB0206475_DICDI	DEINFLIYKYLLE
3393	c12-2_SteCoe_15387	NDINLLIYHYLKE
3392	TBL1XR1_E7D	DDVNFLVYRYLQE
3391	TBL1X_F61V	DEVNVLVYRYLQE
3390	TBL1X_V63M	DEVNFLMYRYLQE
3389	TBL1X_Y64C	DEVNFLVCRYLQE
3388	TBL1X_R65Q	DEVNFLVYQYLQE
3387	TBL1X_R14W	DEVNFLVYWYLQE
3386	HsDCAF1_wt	KELLLLIRNHLIS
3385	HsDCAF1_L851F	KELLFLIRNHLIS
3384	HsDCAF1_I853M	KELLLLMRNHLIS
3383	HsDCAF1_R854Q	KELLLLIQNHLIS
3382	HsDCAF1_H856Y	KELLLLIRNYLIS
3381	TPLH1_R6	RELVFLILQFLDE
3380	TPLH1_R6H	HELVFLILQFLDE
3379	TPLH1_R6K	KELVFLILQFLDE
3378	TPLH1_R6D	DELVFLILQFLDE
3377	TPLH1_R6E	EELVFLILQFLDE
3376	TPLH1_R6S	SELVFLILQFLDE
3375	TPLH1_R6T	TELVFLILQFLDE
3374	TPLH1_R6N	NELVFLILQFLDE
3373	TPLH1_R6Q	QELVFLILQFLDE
3372	TPLH1_R6C	CELVFLILQFLDE
3371	TPLH1_R6G	GELVFLILQFLDE
3370	TPLH1_R6P	PELVFLILQFLDE
3369	TPLH1_R6A	AELVFLILQFLDE
3368	TPLH1_R6I	IELVFLILQFLDE
3367	TPLH1_R6L	LELVFLILQFLDE
3366	TPLH1_R6M	MELVFLILQFLDE
3365	TPLH1_R6F	FELVFLILQFLDE
3364	TPLH1_R6W	WELVFLILQFLDE

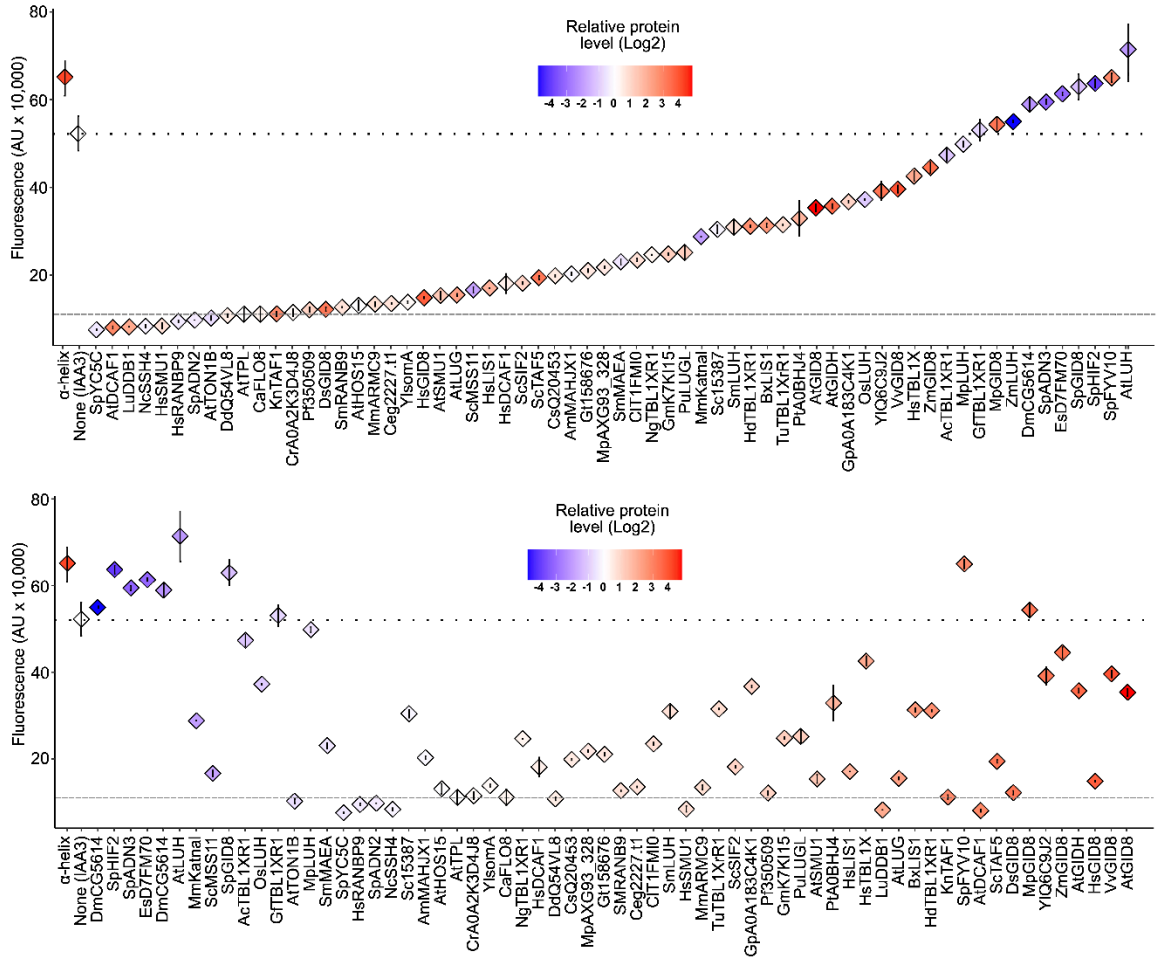
3363	TPLH1_R6Y	YELVFLILQFLDE
3362	TPLH1_R6V	VELVFLILQFLDE
3356	ARMC9_Mm	SELLGLVKEYLDF
3355	alphahelix	EAAAKEAAAKEAAAK
3353	TPLH1_F10H	RELVHLILQFLDE
3352	TPLH1_F10D	RELVDLILQFLDE
3351	TPLH1_F10E	RELVELILQFLDE
3350	TPLH1_F10T	RELVTLILQFLDE
3349	TPLH1_F10N	RELVNLILQFLDE
3348	TPLH1_F10Q	RELVQLILQFLDE
3347	TPLH1_F10C	RELVCLILQFLDE
3346	TPLH1_F10G	RELVGLILQFLDE
3345	TPLH1_F10A	RELVALILQFLDE
3344	TPLH1_F10I	RELVILILQFLDE
3343	TPLH1_F10L	RELVLLILQFLDE
3341	TPLH1_F10W	RELVWLILQFLDE
3340	TPLH1_F10V	RELVVLILQFLDE
3339	TPL_2xH1	RELVFLILQFLDEGGGGSGGGGSRELVFLILQFLDE
3338	TPL_3xH1	RELVFLILQFLDEGGGGSGGGGSRELVFLILQFLDEGGGGSGGGGSRELVFLILQFLDE
3337	TPL_F10A	RELVALILQFLDE
3336	TPL_F10R	RELVRLILQFLDE
3335	TPLH1_REDE>AAAA	AALVFLILQFLAA
3334	TPL_4xH1	RELVFLILQFLDEGGGGSGGGGSRELVFLILQFLDEGGGGSGGGGSRELVFLILQFLDEGGGGSGGGGSRELVFLILQFLDE
4714	Clade I.I	RDVIRLILQYLKE
4713	Clade III	ADLNRLIMNYLVT
4712	Clade IV	NTLNAYIYDYLIK
4711	Clade V	QELNRLIVDYLLR
AGRO PLASMIDS		**Corresponding H1 sequences**
3548	p35SLong_5U:[H1-AtTPL-F10A]-HA-IAA3-3U+Ter-AtuNos	RELVALILQFLDE
3547	p35SLong_5U:[H1-AtTPL-F10G]-HA-IAA3-3U+Ter-AtuNos	RELVGLILQFLDE
3546	p35SLong_5U:[H1-AtTPL-F10H]-HA-IAA3-3U+Ter-AtuNos	RELVHLILQFLDE
3545	p35SLong_5U:[H1-AtTPL-R6A]-HA-IAA3-3U+Ter-AtuNos	AELVFLILQFLDE

3544	p35SLong_5U:[H1-AtTPL-R6Y]-HA-IAA3-3U+Ter-AtuNos	YELVFLILQFLDE
3543	p35SLong_5U:[H1-AtTPL-R6L]-HA-IAA3-3U+Ter-AtuNos	LELVFLILQFLDE
3542	p35SLong_5U:[H1-alphaH]-HA-IAA3-3U+Ter-AtuNos	EAAAKEAAAKEAAAK
3541	p35SLong_5U:[H1-AtHOS15]-HA-IAA3-3U+Ter-AtuNos	VELNFLVFRYLQE
3540	p35SLong_5U:[H1-RANB9-schma]-HA-IAA3-3U+Ter-AtuNos	QEINRLILEYLVV
3539	p35SLong_5U:[H1-DmDMEI]-HA-IAA3-3U+Ter-AtuNos	ADLNRLVMNYLVT
3538	p35SLong_5U:[H1-BCR36DRAFT]-HA-IAA3-3U+Ter-AtuNos	KDLNKLIMNYLVI
3537	p35SLong_5U:[H1-A0A2K3D4J8]-HA-IAA3-3U+Ter-AtuNos	DVVNYLIYRYLQE
3536	p35SLong_5U:[H1-TAF1]-HA-IAA3-3U+Ter-AtuNos	EDLNRIVLSYLKK
3535	p35SLong_5U:[H1-FLO8]-HA-IAA3-3U+Ter-AtuNos	QVLNSLILDFLVK
3534	p35SLong_5U:[H1-AtTPL]-HA-IAA3-3U+Ter-AtuNos	RELVFLILQFLDE
3533	p35SLong_5U:[H1-DDB0206475]-HA-IAA3-3U+Ter-AtuNos	DEINFLIYKYLLE
3532	p35SLong_5U:[H1-AtTON1B]-HA-IAA3-3U+Ter-AtuNos	RLLSALICEYLDW
3531	p35SLong_5U:[H1-ADN2]-HA-IAA3-3U+Ter-AtuNos	SLLNSYIYDYLIK
3530	p35SLong_5U:[H1-HsRANBP9]-HA-IAA3-3U+Ter-AtuNos	TMIQKMVSSYLVH
3529	p35SLong_5U:[H1-HsSMU1]-HA-IAA3-3U+Ter-AtuNos	SDVIRLIMQYLKE
3528	p35SLong_5U:[H1-SSH4]-HA-IAA3-3U+Ter-AtuNos	ELIQQLVLQFLQH

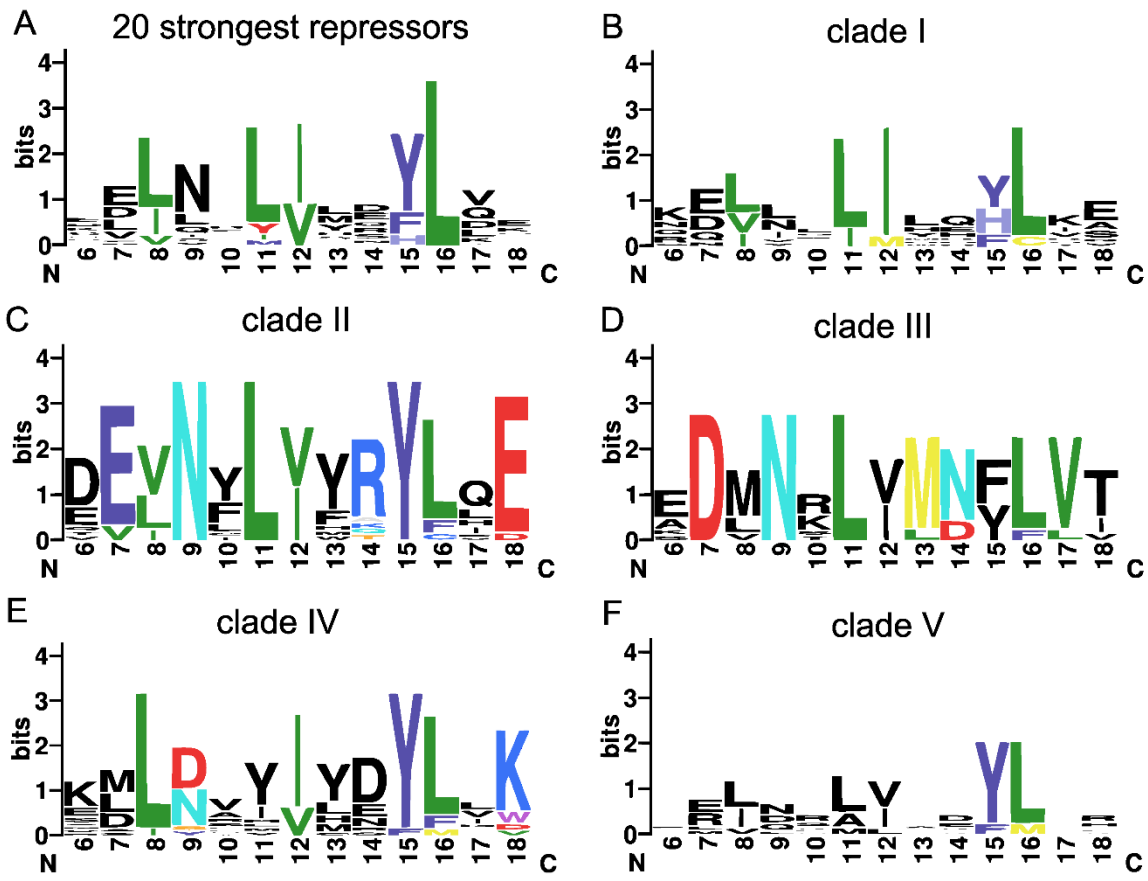
3527	p35SLong_5U:[H1-DDB1]-HA-IAA3-3U+Ter-AtuNos	KELLQLIHDHLVS
3526	p35SLong_5U:[H1-AtDCAF1]-HA-IAA3-3U+Ter-AtuNos	KELLLLIHEHLQA
3525	p35SLong_5U:[H1-YC5C]-HA-IAA3-3U+Ter-AtuNos	EFLNELISSFLLN
Name	**OLIGO SEQUENCE**	**Purpose**
TPLH1_R6A_F	TTCTCTTAGTgctGAGCTCGTTTTCTTG	AtTPL-H1 Alanine mutations
TPLH1_R6A_R	GACATCATTTTGTGATGGATC	AtTPL-H1 Alanine mutations
TPLH1_E7A_F	TCTTAGTAGAgctCTCGTTTTCTTGATC	AtTPL-H1 Alanine mutations
TPLH1_E7A_R	GAAGACATCATTTTGTGATG	AtTPL-H1 Alanine mutations
TPLH1_R6AE7A_F	TTCTCTTAGTgctgctCTCGTTTTCTTGATCTTACAG	AtTPL-H1 Alanine mutations
TPLH1_R6AE7A_R	GACATCATTTTGTGATGGATC	AtTPL-H1 Alanine mutations
TPLH1_F10A_F	AGAGCTCGTTgcttTGATCTTACAGTTTC	AtTPL-H1 Alanine mutations
TPLH1_F10A_R	CTACTAAGAGAAGACATCATTTTG	AtTPL-H1 Alanine mutations
TPLH1_Q14A_F	CTTGATCTTAgctTTTCTCGATGAAGGGAGAGG	AtTPL-H1 Alanine mutations
TPLH1_Q14A_R	AAAACGAGCTCTCTACTAAG	AtTPL-H1 Alanine mutations
TPLH1_D17A_F	ACAGTTTCTCgctGAAGGGAGAG	AtTPL-H1 Alanine mutations
TPLH1_D17A_R	AAGATCAAGAAAACGAGCTC	AtTPL-H1 Alanine mutations
TPLH1_E18A_F	GTTTCTCGATgctGGGAGAGGAC	AtTPL-H1 Alanine mutations
TPLH1_E18A_R	TGTAAGATCAAGAAAACGAG	AtTPL-H1 Alanine mutations
TPLH1_D17AE18A_F	ACAGTTTCTCgctgctGGGAGAGGAC	AtTPL-H1 Alanine mutations
TPLH1_D17AE18A_R	AAGATCAAGAAAACGAGC	AtTPL-H1 Alanine mutations
TPLN_1xHA_F	gccggattatgcgTCTTCTCTTAGTAGAGAGC	AtTPL-H1 HA tag addition

TPLN_1xHA_R	acatcatacggataCATCATTTTGTGATGGATC	AtTPL-H1 HA tag addition
TPLN_2xHA_F	tatccgtatgatgtgccggattatgcgTCTTCTCTTAGTAGAGAGC	AtTPL-H1 HA tag addition
TPLN_2xHA_R	cgcataatccggcacatcatacggataCATCATTTTGTGATGGATC	AtTPL-H1 HA tag addition
3132_F10X_R	CTACTAAGAGAAGACATCATTTTG	AtTPL-H1 F10 mutations
3132_F10E_F	AGAGCTCGTTgaaTTGATCTTACAGTTTC	AtTPL-H1 F10 mutations
3132_F10Q_F	AGAGCTCGTTcaaTTGATCTTACAGTTTC	AtTPL-H1 F10 mutations
3132_F10R_F	AGAGCTCGTTagaTTGATCTTACAGTTTC	AtTPL-H1 F10 mutations

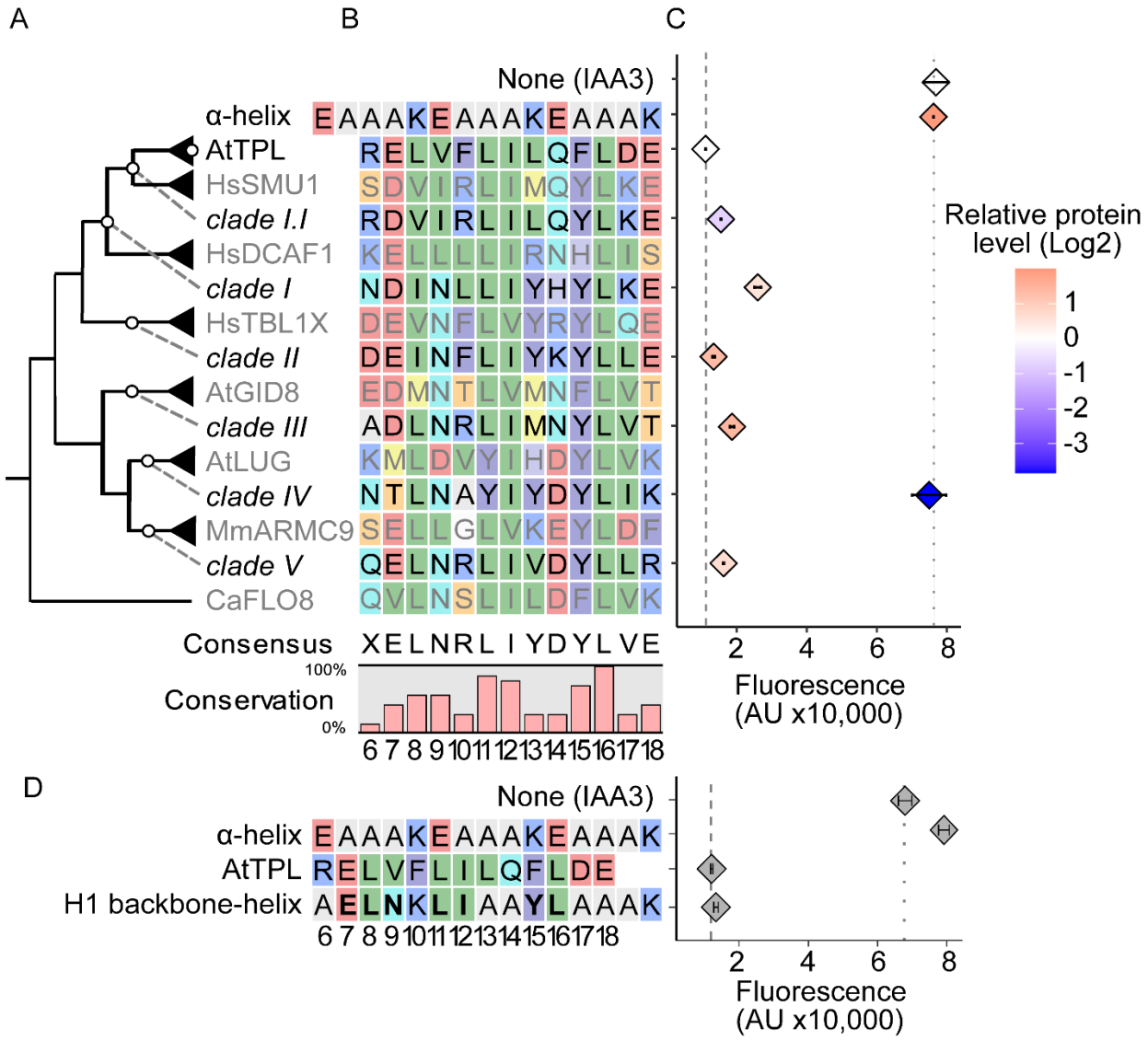
Supplement 4: We list all primers used in these experiments, as well as information about all plasmids used in yeast or plant assays. For plasmids, we list **A.** Nemhauser Lab plasmid numbers, **B.** the names of our sequences, and **C.** the LisH-H1 amino acid sequences. For oligos, we list **A.** their names, **B.** their DNA sequences, and **C.** their experimental purpose.



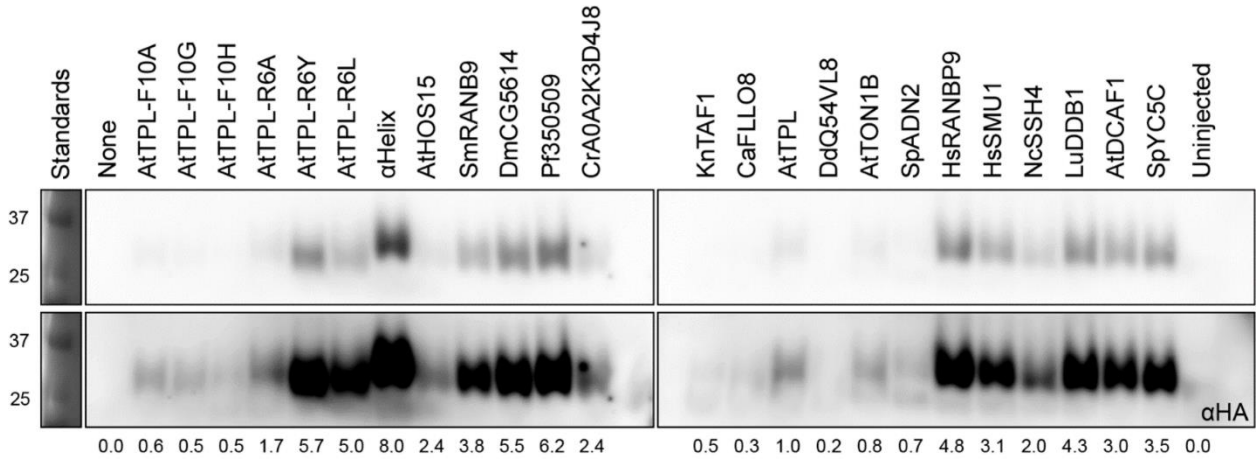
Supplement 5: Repression assay data visualizations. The same cytometry data represented in Figure 2, ordered by A. how much repression we were able to detect, and B. how much protein accumulated. A, B. We have marked the fluorescence levels detected by the positive AtTPL-H1-HA-IAA3 control (dashed line) and negative IAA3 repression control (dotted line). Protein accumulation was measured by western blot and normalized to yeast PGK1. Levels of LisH-H1 proteins are shown relative to AtTPL-H1 with each data point color coded from blue (low) to red (high) expression on a log2 scale. For all cytometry, every point represents the average fluorescence of at least 10,000 individually measured yeast cells (a.u. - arbitrary units). The error bars represent 95% confidence intervals.



Supplement 6: Clade logo plots. Each panel represents the residues found in the H1s of proteins across the **A.** top 20 most repressive sequences, **B.** clade I, **C.** Clade II **D.** clade III, **E.** clade IV, **F.** and clade V. Taller columns represent more conserved residues. Letters appear longer the more commonly they are found at the specified residue. Letters at well-conserved residues are color coded by their physicochemical class. Logo plots were created with an online tool ((60) https://weblogo.berkeley.edu/logo.cgi)).


A
B
C
None (IAA3)
α-helix
AtTPL
HsSMU1
clade I.I
HsDCAF1
clade I
HsTBL1X
clade II
AtGID8
clade III
AtLUG
clade IV
MmARMC9
clade V
CaFLO8
Consensus
Conservation
100%
0%
6 7 8 9 10 11 12 13 14 15 16 17 18
Relative protein level (Log2)
1
0
-1
-2
-3
2
4
6
8
Fluorescence (AU x10,000)
D
None (IAA3)
α-helix
AtTPL
H1 backbone-helix
6 7 8 9 10 11 12 13 14 15 16 17 18
Fluorescence (AU x10,000)

Supplement 7: Ancestral Sequence Reconstruction A. The phylogeny from Supplement 1 was used to infer ancestral sequences at nodes marked in Supplement 1. A simplified cladogram shows these nodes (black dots) next to extant LisH-H1 sequences tested (black) and LisH-H1 sequences used to contextualize nodes (grey). Sequence clade I.I exists within clade representing an ancestral state closer to AtTPL-H1. Reconstructed sequences clade I and clade II are identical to Sc15387 and DdQ54VL8 respectively. **B.** Sequences were aligned and residues colored by their physicochemical class (RASMOL color scheme). The consensus sequence for LisH-H1 is aligned alongside the relative conservation rate of different residues along the helix. **C.** Flow cytometry and a modified *At*ARC^{Sc} depicted in Fig. 1E was used to quantify the relative repressive function of different H1s. For all cytometry, every point represents the average fluorescence of at least 10,000 individually measured yeast cells (a.u. - arbitrary units). The error bars represent 95% confidence intervals. We have marked the fluorescence levels detected by the positive AtTPL-H1-HA-IAA3 control (dashed line) and negative IAA3 repression control (dotted line). Protein accumulation was measured by western blot and normalized to yeast PGK1. Levels of construct expressions are shown relative to AtTPL-H1 with each data point color coded from blue (low) to red (high) expression on a log2 scale. **D.** We test synthetic sequence H1-backbone helix containing the most well conserved residues from the top twenty most repressive sequences (bolded residues), using the same format (A.,B.,C.).



Supplement 8: Protein expression in transient expression assays in tobacco. Transient expression assays from Figure 4 were analyzed by Western blotting to determine the expression levels of H1-HA-IAA3 fusion proteins. Four leaf discs from were pooled from one representative experiment and analyzed by electrophoresis and blotting with antibodies against HA. Two exposures are provided to demonstrate detection of lowly expressed H1-HA-IAA3 variants. Protein normalization was calculated compared to the AtTPL-H1 wild type sample and is listed below each lane.

CHAPTER 5

In 2018, I joined the UW Queer mentorship program through the Q Center in hopes to help a young queer student with some guidance and companionship. I was assigned to work with Perfectious. We talked a few times; I bought them lunch, discussed their goals, and encouraged them to open up about their transness. Although we connected, they became harder and harder to contact as the year went on, showing clear signs that they were struggling. Eventually, they stopped responding, and I was advised by the Q Center not to reach out to them or investigate. I mentored others through the program and moved on, feeling a bit ashamed I couldn't provide them with the mentorship they needed. I googled them in 2020 hoping to find some news about them: whether they had graduated, joined a lab, maybe even published a paper. Instead, I found their name on a King County obituary for the homeless.

I am often frustrated when I have conversations around queerness in academic settings. I find that the experiences shared, and the struggles mentioned by other queer folks are – while perhaps true to me – not the same experiences I think are largely responsible for making academic and professional settings unsafe and inaccessible to many queer and trans people. I want to see my communities making science alongside me. I have dedicated myself to learn how to be a good mentor, care for my local queer community, and to teaching others about the experiences of Black, Brown, and Latine queer and trans people. In this chapter, I will outline some of the efforts that I have led. I dedicate my work in this area to Perfectious.

ICAR 2021 Workshop

Sex Work and Homelessness: Addressing Taboo Issues to Increase Queer and Trans Recruitment and Retention in Biology

Román Ramos Báez, Jonathan Santos-Ramos, Jennifer L. Nemhauser

Homelessness and sex work are very much interconnected with queer and trans identities, where 35% of Black and Latine transfeminine people are involved in sex work[1] and up to 40% of homeless youth identify as queer or trans[2]. Related to these experiences, queer and trans people experience high rates of harassment, physical assault, and sexual violence. A large proportion of Black and Latine queer and trans people in academia are involved in sex work at every level[3]. Covid-19 has had huge impacts in the livelihood of sex workers and has left many homeless, disproportionately burdening these communities.

To address the historical exclusion of queer and trans people from academia, and to increase retention of the few queer and trans folks in academia today we must understand how LGBT and BIPOC communities intersect with sex work and homelessness. This will help us put a stop to our filtering out and unnecessary penalization of queer and trans people in academia. We also need to be prepared to have conversations around these issues if we are to support trainees that have a history of homelessness or sex work. Finally, being in conversation with these communities can help us make Biology more relevant to the interests and needs of queer and trans members of our community.

These issues are rarely addressed, especially in professional communities since:

1. Sex work and homelessness are considered inappropriate to speak of at work and often irrelevant to STEM.
2. Many queer and trans people are already struggling to get into and stay in academia, and so avoid association with these taboo topics for safety.

3. Many queer and trans people who do get into academia have no history of homelessness and sex work because those who do have been largely filtered out.
4. The lack of an intersectional lens means homelessness and involvement in sex work are widely considered separate from identity and therefore not addressed in recruitment and retention efforts.

Addressing these also plays a critical role in empowering women, BIPOC, and our community at large. In this workshop, I provided some practical fundamental skills, kickstarted conversations in this long-ignored topic, point people to resources, and highlight the importance of using an intersectional lens in DEI efforts.

Five key changes to prioritize (as shared in the workshop):

1. Count trans and gender-diverse people
    a. Asking gender: we often don't need to know gender. Allow people not to disclose when necessary. Ask for pronouns instead.
    b. Gender questions should be fill-ins.
    c. Man and woman instead of male and female.
    d. Allow for use of chosen name in all documents and settings.
    e. Provide nonbinary options for titles (i.e. Mx.).
2. Allow people to take care of their mental health
    a. Investigate & advocate (I&A): Does your worker's/student's healthcare include therapy?
    b. I&A: Does your institution list therapists in the area with specific identities and specialties?
    c. Have a mental health day off protocol:
        i. Check-ins so a day doesn't spiral into weeks.
        ii. No need to explain reasoning for break.

3. Allow people to take care of their bodies
    a. Healthcare covers screens for sexually transmitted diseases.
    b. PReP and retroviral drugs.
    c. (I&A): Gender-affirming surgeries, hormones, and therapy
    d. Pay people well so they can house themselves safely, buy their medicine, groom themselves, eat well, etc.
4. Support people in building community
    a. Fund relocation costs
    b. List affinity groups in your website, or link to existing one
    c. Support involvement in affinity groups outside of the institution/academia. Drag, ballroom, nightlife, music, dance
    d. Invite people to play a role in defining their success
    e. Highlight folks for their science.
5. Support people with a history of sex work and homelessness
    a. (I&A): Do not require criminal background checks
    b. Allow coworkers privacy
    c. Be ready to refer to resources
    d. Keep focus OFF people's bodies, and ON their love for and talent for science.

REFERENCES

1. **Jaime M. Grant, Lisa A. Mottet, Justin Tanis, Jack Harrison, Jody L. Herman, and Mara Keisling,** Injustice at Every Turn: A Report of the National Transgender Discrimination Survey, (Washington DC: National Center for Transgender Equality and National LGBTQ Task Force, 2011).
2. **Choi, S.K., Wilson, B.D.M., Shelton, J., & Gates, G.** (2015). Serving Our Youth 2015: The Needs and Experiences of Lesbian, Gay, Bisexual, Transgender, and Questioning Youth Experiencing Homelessness. Los Angeles: The Williams Institute with True Colors Fund.
3. **Erin Fitzgerald, Sarah Elspeth, Darby Hickey** (2015). Meaningful Work: Transgender experiences in the sex trade. *Harper Jean Tobin National Center for Transgender Equality* https://transequality.org/sites/default/files/Meaningful%20Work-Full%20Report_FINAL_3.pdf

Table 1. Brainstorm for workshop discussion

Key point	**Why is this especially relevant to queer and trans people?**	**Solutions at both the institutional level and one-on-one level**
Access to therapy is esp. important to QTPOC (queer and trans people of color).	• High suicide rates • High rates sexual violence • need of therapist to apply for trans healthcare • rejection from family • Disproportionate effects on BIPOC folk	• Pay people well so they can access therapy. • Give lab personnel good access to healthcare. • Allow people to take mental health days.
Homelessness is a queer and trans issue. Access to community/chosen family is important.	• Chosen family and mutual aid • cost of moving • difficulty with stability • high rates poverty • high rates homelessness • Disproportionate effects on BIPOC folks	• Help people with relocation costs. • Support involvement in nightlife, queer and trans support groups, etc. • Accept this does happen to people that want to participate in science.
Sex work is a QTPOC issue	• Queer relationships, sex, and bodies often seen as ‘unprofessional’ . • Queer nightlife’s importance in history of support and community building. • Sex work as means to get housing, autonomy, community. • Criminalization of sex work • Disproportionate effects on BIPOC folk • Sex work also predisposes people to higher levels of harassment and sexual violence.	• Comprehensive healthcare. STIs, birth control, and mental health • Do not require criminal records. Bigger goal is decriminalize sex work. • Be body-positive. We are often fetishized and victims of sexual violence. We are also often shunned. • Do not reject people for “unprofessional attire” in and outside of work. Accept that people in academia right now participate in sex work.

Scientific American – Letter to the NSF and NCSES

By Riley DeHority, Román Ramos Báez, Timber Burnette, Lars Howell (2021) Nonbinary Scientists Want Funding Agencies to Change How they Collect Gender Data. *Scientific American* *https://www.scientificamerican.com/article/nonbinary-scientists-want-funding-agencies-to-change-how-they-collect-gender-data/*

Every year, the National Science Foundation (NSF) and the National Center for Science and Engineering Statistics (NCSES) send a series of surveys to students and researchers around the country. The surveys are used to monitor changing demographics and track levels of financial support for scientific research, and filling them out is required for anyone who receives NSF funding. There are limited gender options in these surveys: male, female and, on some surveys, "do not wish to disclose."

As scientists who exist outside the gender binary, many of us do wish to disclose our gender, but are unable to do so because these terms do not reflect our identities. Nonbinary scientists and other scientists outside the gender binary experience gender beyond the typical man-and-woman dichotomy, and often identify as transgender. Being unable to accurately report our gender precludes accurate data collection for these organizations, and further marginalizes nonbinary scientists. Nonbinary identities are increasingly common; most nonbinary people are under the age of 29, and members of Gen Z are more than twice as likely to identify as nonbinary, genderfluid or nonconforming than older generations. A recent study in Pittsburgh even found that nearly one in 10 young people self-identified as gender diverse. It is time for the NSF and the NCSES to update their policies and language to better quantify and support this growing transgender and gender diverse population.

The lack of information on transgender and gender diverse scientists is both a symptom and cause of exclusion from science at large. By not collecting accurate gender data, the NSF limits our ability to

understand and quantify representation, funding disparities and retention, and ultimately stops us from assessing and addressing any gaps. The STEM Inclusion Study found that LGBTQ+ scientists have fewer resources and career opportunities, experience more harassment and exclusion at work and are more likely to consider leaving their current job positions than their heterosexual, cisgender peers. Further, we know from our involvement in communities like the International Society of Nonbinary Scientists that nonbinary researchers face barriers in gender segregated environments like dorms, conference accommodations and field work. Accurate data from annual NSF surveys will allow us to better understand and advocate for our transgender and gender diverse scientific community.

Determining which multiple-choice gender options to include is a common barrier to updating demographics surveys. We recommend a simple solution: change gender questions to a write-in field and allow all people to accurately describe themselves. Asking “What is your gender?” is feasible at scale, generates accurate results and accommodates language changes over time. We recommend an expanded list of gender options when an open-ended question is unfeasible, such as “agender,” “gender variant,” “two-spirit,” “nonbinary” and “genderqueer” in addition to “man” and “woman.” There should also be an option to specify a chosen name, and that name should be used in all communications. Respecting chosen names, pronouns and gender identity is critical for acceptance and dignity.

As scientists and scientific institutions, we must promote collection of accurate data. Changing methods based on new information is a cornerstone of good science. The NSF is collecting inaccurate data: nonbinary people, who are neither men nor women, must either report their gender falsely or not disclose it. An NSF spokesperson confirmed last year that the agency is considering splitting the question into two parts: “current gender identity” and “assigned sex at birth.” Despite its popularity, using biological sex as a proxy for gendered experience is unscientific, outdated and often motivated by a desire to exclude.

Much of the evidence in favor of a two-step question comes from trans health research, whose findings may not apply outside of medical contexts. One often-cited study about intake forms in a sexual health clinic showed that a two-step question led to 4.8 times more trans people being identified, but it was the addition of a “Nonbinary/Genderqueer" gender option that accounted for over half of the increase.

Context is also critical here: a person who is willing to tell their doctor about their birth assignment may feel differently about the same question on a federal census. Transgender people are often uncomfortable being asked their assigned sex at birth, and nonbinary scientists we know often prefer not to disclose it. Further, while “transgender” is not a stable or truly measurable category, this information can be collected directly and respectfully by asking the optional question “Do you identify as transgender?” Relying on assigned sex at birth to back-calculate who is and isn’t a gender minority just creates another artificial binary and flattens the experiences of scientists in our community.

On May 10, we sent an open letter to officials at the NSF and the NCSES, asking them to improve how they collect gender data on official forms and surveys. Our letter to the NSF and NCSES was signed by 413 transgender and gender diverse scientists and 2,280 cisgender scientists, most of whom are also NSF-funded. Our proposed changes align with the goals of the NSF’s 2018–2022 Strategic Plan, where the agency committed to “attract, retain, and empower a talented and diverse workforce.” Through these changes, NSF and NCSES can respectfully and accurately quantify gender diversity.

This is an opinion and analysis article; the views expressed by the author or authors are not necessarily those of Scientific American.

Román Ramos Báez, Timber Burnette, Lars M. Howell, Emerson Lynch, Ezra Jay Kottler, Parker K Lund, Nick Barts, Riley DeHority

Dear NSF Director Sethuraman Panchanathan, NCSES Division Director Emilda B. Rivers, NSF Program Director John M. Finamore, NSF Office Head Rhonda J Davis, NSF Office Head C. Suzanne Iacono, the Committee on Equal Opportunities in Science and Engineering, and the Office of Science and Technology Policy,

We write today as a collective of nonbinary, genderqueer, and trans scientists to propose changes in the way the National Science Foundation asks for and records gender and gendered information in applications and other forms where gender is requested. Several of the signers have felt discomfort and alienation as a result of the gender binary prevalent throughout NSF forms. While there is an option to not report gender, this diversity should be quantified; by not providing options for those who are not male or female, the NSF is collecting incomplete and incorrect demographic data. Although these changes might feel minor, they are vital to ensuring that gender minorities in our community are not othered, made invisible, or otherwise inaccurately represented in NSF records. Changes like these are an important step in decolonizing science and validating the many genders that have historically existed outside of man and woman in non-western spaces.

Ultimately, we recommend changing gender from a bulleted list to a write-in field — this allows for people of all genders to accurately describe themselves. We recognize this data may be harder to analyze (via open text analysis), but it is the most inclusive approach. If a write-in field is not feasible, we recommend the following changes to questions regarding gender on all forms:

- Replacing gender identities "male" and "female" with "man" and "woman". Male and female are biological terms for sex, not socio-cultural terms for gender.
- Including more gender identities, such as (but not limited to) "agender", "gender variant", "two-spirit", "non-binary", and "genderqueer", besides "man" and "woman".
- Allowing respondents to pick more than one option.
- Keeping the "Do not wish to disclose" option, but only with expanded options. Participants may wish to disclose their gender but do not see it reflected in the presented options.

Additionally, the NSF should provide the option on all forms for individuals to specify a chosen name when names are required. We acknowledge the need for the NSF to have legal names on grant forms and federal applications; however, the process of legally changing one's name is long, involved, and costly. Addressing correspondence and grant information to the member's chosen name will ensure all participants feel welcome, accepted, and recognized while pursuing research through the NSF. Publishers such as American Association for the Advancement of Science (AAAS) and Elsevier are offering trans-inclusive name change policies. The NSF allowing for the inclusion of chosen names on forms would be advancing a scientific culture of inclusion. The NSF adding a location for chosen names and pronouns would demonstrate the organizations respect for trans and nonbinary scientists.

These changes align with the NSF's stated diversity initiatives to actively support increasing workplace diversity and inclusion, which includes recruitment and retention. Further, they will create a more welcoming environment for people of all genders when applying to be a part of the NSF community. These changes will also ensure that NSF records of people's gender identity are accurate, and that people who are neither men nor women are represented.

Sincerely,
413 trans and gender diverse scientists
[Names Omitted from Public Letter]
We, the undersigned, support our trans and nonbinary colleagues and encourage the NSF to implement their recommendations: (n = 2,280 allies + 50 anonymous & name only submissions)
[Names Omitted from Public Letter]

Sensitive Roots

Sensitive Roots is a social media drag project designed to celebrate and talk about queer and trans people of color. This project is a resource for queer and trans people of color working in professional settings on how to survive and thrive, as well as a guide for our coworkers and institutions on how to increase Black, Latine and Indigenous recruitment and retention, making work safer for all along the way. I use cheeky drag to help you build practical skills and knowledge in empathetic mentorship and training of historically excluded groups in academia, especially LGBTQIA+ and BIPOC people! I'll talk about how to prevent harm and conflict, how to be resilient, how to show up in solidarity, and more.

Episode 1: The Who, What, and Why of Sensitive Roots

Link to episode 1: https://youtu.be/KoKsTIXKLqg

Today I'm going to specifically talk about who sensitive roots is for, and what the purpose of this series is.

Who is Sensitive Roots for?

In queer spacess, we often talk about how Black and Brown trans women basically started pride, defined drag makeup, ballroom culture, and really queer culture in general. We have so much to learn from our Black and Brown queer and trans family.

Sensitive Roots is for anyone who wants to learn about those experiences and my experiences, but is designed for the queer and/or trans Black and Brown people trying to not just survive but thrive in professional settings and still be able to show up as their whole selves. It's also for our allies who want to support us in that effort.

Who am I?

I am a nonbinary, feminine, Latine person. I'm a plant molecular biologist and involved in outreach work and in the queer drag and nightlife community outside of academia. You can learn more about me in this video over here. I'm always trying to find creative ways to make these institutions, which have a violent history where people like me are intentionally excluded, work for me.

Why does it matter?

Lately in academia, I've seen more and more people engaging in conversations about diversity, inclusion, equity, and justice. As a queer person of color, I'm really happy to see this. However, I noticed that the conversation never gets to me. Yes, we'll talk about what it's like being LGBTQIA+, but we will center White experiences, American experiences, heteronormative experiences. Yes, we'll talk about the Latine

experience, or Black or Indigenous experiences, but not about what it's like to be queer as well. Yes, we'll talk about the importance of intersectional thinking, but not what that means or what that looks like to be at those intersections. Meanwhile, queer Black, Indigenous, and Latine people are everywhere in the media and playing a huge role in popular culture, and at the same time, we are disproportionately suffering from sexual violence, homelessness, lack of access to mental and physical health, and hate crimes. How can you help and celebrate us if you don't know why we keep falling victim to systemic violence? How can you imagine institutions that don't hurt us if you don't realize how they are hurting us? How can you get us to participate if you don't know why we don't participate. I want to touch on all of that. And for the queer and trans Black and Brown people watching, how do we survive, how do we take care of each other? We often need to participate in these institutions to thrive. How do we come out alive?

Why the name Sensitive Roots?

I chose the name Sensitive Roots because, often growing up I was called sensitive because I was so effeminate. It was always said with a negative connotation. But when I think of a cell being sensitive as a molecular biologist, I think of it as acknowledging that, oh! This molecule exists, and it's here. Now, when I think of roots in a plant, I think of them as being the part of the plant that goes unseen, unexplored and underappreciated. Roots are a huge part of plants and what keep the plant grounded. In Sensitive Roots, I hope to acknowledge what's there, but goes unseen, unheard, and ignored.

Sensitive Roots will look like me talking about my personal experiences, me interviewing others, and me generally providing you with information that might help you to reimagine institutions or imagine new ones where we can flourish, or honestly, at least survive.

Welcome to Sensitive Roots!

Episode 2: How to ask about gender?

Link to episode 2: https://youtu.be/_pwnb9NXnrM

Welcome back to Sensitive Roots with me, Romi, where I talk about whatever isn't being talked about professional settings, especially as it pertains to Black, Latine, and Indigenous Queer and Trans people.

Today I'm wearing the Trans Pride flag colors because I am talking about how to ask for people's genders in professional settings, whether or not we should even do that in the first place, and why our genders matter. Check out my write-up with simple and easy recommendations on how to ask for gender in forms. Steal it and spread it around.

Why does gender matter?

Gender is important because how we identify affects how we see ourselves. It's part of being a person – unless you're agender, which means that you don't have a gender – and that's cool too. There's a big problem in professional settings where people are often only asked whether they are men or women, without being given other options. This means that the statistics available on the experiences of nonbinary and trans people in most settings often goes unaccounted for. This is especially a problem because queer and trans people disproportionately experience harm and discrimination in ways that often the institutions that are asking us about our gender identities are directly responsible for. We should be acknowledging gender more so we can better reimagine these institutions to not be especially harmful to queer and trans folks of color.

What other genders are there?

It is important to acknowledge that there is a colonial history to the constructs of man and woman. Although many societies around the world have an equivalent man and woman gender, many also have well-established third genders. Indigenous communities in the US use "two-spirit" to refer to a

lot of these different genders. These all often fall under the trans umbrella. Although "trans" is a single letter in the initialism LGBTQIA+, it encompasses most genders. Anyone that identifies with a gender that is not their gender assigned at birth is considered trans. Usually in the media we see this as being people that medically transition and identify with a binary gender, "man" or "woman," but trans can also refer to people that are non-binary, non-conforming, gender fluid, two-spirit, and many more. There are lots of ways to be trans. There are lots of genders other than man or woman. For these reasons, the ignoring of these genders is both racist and insensitive to trans folks.

How do you ask folks about gender?

So how do you ask folks at work whether they are, you know, *jokingly motions a limp wrist*? First thing you should ask yourself is whether or not you really need to ask. Gender is complicated. Personally, I don't expect most cis people to understand the nuances of my gender. Perhaps it is important for me to be open about my gender to you. In this case, you should make clear that coming out about being trans is something that you are OK with and will not affect you trans co-worker's standing at work or work relationships. However, if someone simply doesn't feel like sharing, there is good reason for that, and that should be respected.

So you might be wonder, how do I refer to people? How do I find out their pronouns? Ask! Not only is this OK to ask, but you should definitely share yours as well, especially if you are a cis person, since this will signal to trans folks that you understand the importance of asking for pronouns. Moreover, asking someone's gender is not enough information to know what pronouns they use. Many queer men and women today use both "he" or "she" AND "they/them" pronouns. For some binary trans people, however, using gender neutral pronouns might signal that you are uncomfortable with their binary trans identity. Similarly, many non-binary people are assumed to use only "they/them pronouns, but some explicitly use other less common pronouns, or might even be comfortable with any pronouns.

Bottom line:

1. Only ask people's genders if necessary.

2. Always share your pronouns and ask for pronouns from others.

3. Follow my guidelines on how to ask for gender in forms.

4. Continue to make sure people of all genders are being seen, being respected, and accounted for.

And with that said, thank you for watching Sensitive Roots. Bye!

Episode 3: Mediating the connection between outreach and trauma

Link to episode 3: https://youtu.be/uYZLKLitQ90

Hi and welcome back to Sensitive Roots with me: Romi, where I talk about whatever is not being talked about in professional settings, especially as it pertains to Black and Brown queer and trans people. Today I am going to focus on the connection between outreach and trauma:

Why do people that experience continuous abuse end up being the ones that do most of the outreach?

If you are doing this work, how do you protect yourself?

If you want to be in solidarity with people doing this work, what can you do to help?

This is a starter for a conversation so just let me know your thoughts down below.

Defining Trauma

The way I would define it, trauma is a distressing experience. It is something that happened maybe once or is continuously happening that was emotionally damaging. I think of 'continuous abuse' and the trauma that comes with it as being informed by our identities: the way we present ourselves, the resources we had growing up or currently, where we come from, etc. All those can really expose us to constant microaggressions and abuse. People in professional settings from historically excluded communities are very likely to have trauma related to this continuous abuse, and that is something that is important for us to be aware of. An important sidenote: sometimes people assume we have specific types of trauma because of our identities. This is misguided. Just be aware that we've probably experienced some form of continuous trauma.

What you've heard before: Let's not glorify trauma

First let me give you the story you have already heard; that we shouldn't glorify trauma. People from historically excluded backgrounds are often the ones that are paired up with outreach efforts directly related to the trauma that they've experienced. In a way, folks that invite us to do this work are saying

"Hey! You know how to survive this. You've been abused emotionally and physically, yes? Well, we're doing a workshop about that! You should join and open yourself up!"

If it's not obvious why that's an issue, to expose ourselves to this kind of work puts us in a place where we are likely to have to re-expose ourselves to this trauma and think deeply and productively about it. We also still have our own trauma to deal with as well as our regular work. This puts extra work in our hands – very emotional and difficult work.

What we don't really talk about

As a person who loves to do this kind of work, this is a great way for us to feel like we are making a difference in issues that not only matter to us but are critical for our survival in professional settings/the world as it is. This can also be great for our mental health in that it might bring us closer to like-minded people and our community in a workplace where finding that community might be hard. However, sometimes the people leading these efforts are not well equipped to do this work. Often, these leaders are coerced into doing the work because they need to do it to survive or protect their community. This makes for some of the most passionate leaders but being very emotionally invested often means they/we overwork themselves/ourselves. We must be the best version of ourselves to take care of others and do good work, and in this way, we can end up hurting others that we work with when we do not. Being exposed to continuous abuse also means we have likely become habituated to this trauma, and either unknowingly become a part of the cycle of abuse or develop defense mechanisms for protecting ourselves from it that might be good at keeping us safe but harmful to others.

It is hard to come to the realization that people that you look up to and work with might become your abusers, or that you yourself who does such selfless work might become an abuser to those you are working with. It seems counterintuitive: you are in an outreach space! You all know about these issues and the importance of taking care of yourselves. But we forget that some of the ways we have learned to take care of ourselves in abusive situations do not serve us as far as protecting others and being leaders. It is really important to mention that a lot of these efforts are not well supported by the larger community,

organizations, networks, or mutual aids that are pushing for the work to be done in the first place, and this is a big reason why this abuse happens.

So how do you protect yourself?

If you experience some form of harassment or abuse from someone you are doing outreach with, I cannot tell you exactly what to do. This is specific to your experience and your relationship with your abuser. Here are some options:

1. Recognize that this is something that happens. You need to take care of yourself. You are worth it. You do not deserve to be harassed or abused. We often feel that addressing the abuse is distracting from the very important work being done by the team, and that we must stay quiet so the work can get done. The work matters, but you matter too!
2. Distance yourself. Even if it means you cannot continue to do the work or need to find a new way to do the work.
3. Engage in restorative justice. Try to have a mediated conversation with the person that is harassing/abusing you. This can be more powerful than just walking away because if you do walk away, it is likely that someone else will walk replace you and get hurt in the same ways that you were hurt in that position. Addressing this work in a setting where only the people that need to be involved are involved is also a good way to prevent things from escalating further. This method allows you to protect yourself, protect others involved, and potentially help the abuser involved to work on their teamwork and leadership skills and mental health so they will not continue the cycle of hurting you, themselves, and others.
4. Finally, you can involve a bigger organization, or do a callout. Honestly, I think this rarely works and is likely to be a big waste of time and money, a big distraction from the work, and increase ill-feelings for all parties involved. Know that this is an option if other options are not safe or effective.

As a preventative measure, remember to structure your group's involvement in a way where all people have access to breaks, therapy, and that no one is being overworked because of avoidable pressures from within the group.

If what you are getting out of this so far is that marginalized people are ill-equipped to lead and abusive, you are not understanding. Most of this is not the fault of people doing the work and can be avoided when people doing the work have the resources they need.

How to take care of vulnerable people doing the work?

As a leader of a team or as someone that is otherwise creating this work for others, there are a few things you can do to support those that are doing the work:

1. Try to do the work yourself. For this, you must have already done the reading, and you must act based on what people you are addressing are asking for. Make sure you are not just acting based on what is fun, natural, or easy for you. Act in true solidarity, therefore addressing true community needs.
2. Instead of enlisting people to do the work for you, enlist people to advise you in the work. You can ask people who have more experience or are directly affected by your project what they feel about specific actions, whether some information you have is correct or not, etc. This way the work they are putting in is not as time consuming and they get to play the role of consulting from a distance instead.
3. Provide people with resources and money. If you ask people to open themselves up, you must stitch them back together. What do they need? Do they have access to healthcare? Therapy? A community of support? Can they take breaks? Everyone (including you if you are helping to do the work) needs to be compensated.
4. If someone else is doing the work, and its unrelated to your efforts, make sure to give them a boost, whether it is on social media, email chains, etc. Make sure people who are doing the work are being recognized because they will often go unheard, especially Black and Brown queer and

trans people.

So yeah! Those are my thoughts on how outreach is connected to trauma in ways that I think if we recognize we can do a much better job at protecting ourselves, protecting others, and doing more powerful and effective work.

Episode 4: The connection between sex work and queer and trans people

Hi and welcome back to Sensitive Roots with me: Romi, where I talk about whatever is not being talked about in professional settings, especially as it pertains to Black and Brown queer and trans people.

Sex work is probably one of the hardest things to talk about in most professional settings because it's sexual in nature and therefore assumed work inappropriate, irrelevant to work, it's seen as something we're not experts on and therefore have no business discussing, and it's seen as political, the kind of political that's generally thought of as not OK to talk about. And yet, I think some conversation around sex work is necessary if we're going to talk about making work safer and more inclusive to queer and trans people. Like I've mentioned before, the amount of queer and trans Black and Latino people doing sex work is around 35%[1], so it's quite common. There's a lot I want to say so let's break down the conversation into a few parts.

I'll start by defining sex work, then connecting sex work to queer and trans identities, talk about key points to make professional settings more inclusive to sex workers, and finally explain how to productively talk about sex work.

Defining sex work

First, let's define sex work. Sex work or the sex trade is an umbrella term for different types of work, including sexual encounters, escorting, erotic dancing, and erotic content creation. The term was coined by sex worker, artist, and activist Carol Leigh in the 70s who designed this term to unite people in different industries in their pursuit to legitimize and decriminalize their work. Legitimizing sex work means getting people to see and respect it as honest work, and decriminalizing it means abolishing laws that make this work a criminal offense. These are arguably the two most important factors in protecting the safety of sex workers.

Connecting sex work to queer and trans people

So why is this relevant to queer and trans people? Well, queer and trans people are much more likely to participate in sex work. Especially Black, Latino, and Indigenous people. With queerness comes a different less heteronormative relationship with our bodies, our partners, and sex that allow us the freedom to do this work without the same social consequence and shame that a cis-heteronormative person might experience. Being at this identity intersection also means people are more likely to be distant from or disowned by their families, which often disapprove of this work. This lack of support because of participating in sex work AND because of being openly queer and or trans can mean you're more likely to experience poverty, homelessness at a young age, and to struggle with poor mental health. Surviving on your own can also be harder. Being queer or trans is also in many states legally a sufficient reason to get fired from your job, kicked out by your landlord, and at times can be a loophole making it legal for you to be murdered. Yeah, murdered[2]. Being undocumented can also be a reason why you are limited in your job choices. In these situations, engaging in sex work becomes a way to exist from a short list of options, and yet it remains a choice and allows some autonomy. Some people do sex work despite having good health and financial stability. Sex work can be pleasurable and fulfilling work for people and is not intrinsically dangerous in ways that other labor isn't. Most of the dangers come from it being criminalized and demonized. For trans folks, sex work can be a way to feel validated in their gender identity. It is also becoming safer and more popular to participate in through digital content creation. Unfortunately for some of us whether we engage in sex work doesn't matter because our bodies are fetishized. This means we are assumed to be sex workers just because of being trans, having a racialized body, being feminine, or dressing in non-cis-heteronormative ways. This on its own can mean experiencing harassment and abuse. For all those reasons, sex work is relevant to queer and trans experiences.

How do we make professional settings more inclusive to sex workers?

Luckily, a lot of policy that makes professional settings more inclusive to sex workers also benefits everyone!

1. First, **keep focus away from people's bodies and on their work**. Often, it's easier said than done, since a lot of biases we have towards people are subconscious. This looks like allowing people to express themselves freely through fashion, it looks like not commenting on people's attire and bodies, and it looks like not concerning yourself with how people present themselves outside of work. That leads to the second point.
2. **Allow people privacy**. People should not feel pressured to share about their families, relationships, life, hobbies outside of work. You should not be asking people about their involvement in sex work. If people don't feel comfortable sharing certain aspects of their personal life, it's important to respect that. They may have determined it would be unsafe or uncomfortable for them to do so.
3. Be ready to **refer people to resources**. Participating in the sex trade, having a history of sex work, or being assumed a sex worker comes with higher incidence of sexual harassment and abuse. Do you know how to point people to title IX resources? Do you know of the people within your institutions that are safe for them to speak anonymously one-on-one? Does your workplace provide insurance that covers therapy, mental health resources, comprehensive sexual health resources and coverage? Are people's salaries enough for them to live comfortably without needing to supplement their work?
4. Last thing, **say no to employee criminal background checks**. Many types of sex work are criminalized, which means people that participate in it are more likely to have a criminal record. This means sex workers might be labelled as dangerous criminals. It's even more likely for racialized people. It's understandable to not want to work with dangerous people, but it's imperative to remember that people are usually criminalized for minor offenses. These criminal records are racist and sexist and therefore, perhaps ironically, go against workplace safety.

How to talk sex work as non-sex workers

When having conversations about sex work, we want to do so in ways that align with how sex workers would like to be spoken of. Luckily, I found a resource that goes into those guidelines in great detail[3], but here's a main takeaway.

Firstly, we don't want to talk about sex workers as inherently victims. Terms like **prostituted person** can imply this. It's also true that people with good intentions tend to focus on **survival sex workers** and distinguish them from so called **consensual sex workers**. While some have chosen sex work from a very limited list of options, we don't want to disregard their autonomy. If someone is having sex or producing explicit content unconsensually, we don't conflate that with sex work, we call that sexual abuse or sexual assault.

Often times, distinguishing between people that enjoy doing sex work and survival sex workers is done to try to pick out which types of sex work are OK and which are not. It's used to justify policing who can and cannot do this work. It's a way to turn the conversation away from why people are experiencing health insecurity, housing insecurity, etc. and towards blaming sex work as casual of this insecurity and therefore demonizing it.

I see my queer and trans friends and community, Black, Latino and Indigenous community participating in the sex trade, and I see them do so in ways that help sustain them and at times empower them. I also see that they fear the impact this has on their career prospects, and how they are discriminated against at work not just for their identity and presentation, but for their participating in this work. For that reason, I think it's essential that we name this relationship between sex workers and the identities of people historically excluded from academia and other professional settings. Please comment below to tell me your thoughts. Let's continue this conversation and advocacy!

1 Erin Fitzgerald, Sarah Elspeth Patterson, Darby Hickey, Cherno Biko. (2015) Meaningful Work: transgender experiences in the sex trade. The National Transgender Discrimination Survey, Red Umbrella

Project, Best Practices Policy Project, National Center for Transgender Equity. https://transequality.org/sites/default/files/Meaningful%20Work-Full%20Report_FINAL_3.pdf

2 Alexandra Holden. (2020) The Gay/Trans Panic Defense: What It is, and How to End It. American Bar Association. https://www.americanbar.org/groups/crsj/publications/member-features/gay-trans-panic-defense/

3 Chris Bruckert, Anna-Aude Caouette, Jenn Clamen, Kara Gillies, Members of the Downtown Eastside Sex Workers United Against Violence (SWUAV), Sheri Kiselbach, Émilie Laliberté, Tara Santini, Keisha Scott, Emily (2013) Stella: Living and working in safety and with dignity. https://www.nswp.org/sites/nswp.org/files/StellaInfoSheetLanguageMatters.pdf

FUTURE DIRECTIONS

Auxin is important in shaping almost every aspect of growth and development in all plants. This has made it the focus of investigations in plant morphology since the 1800s when Charles Darwin first discovered its role in plant growth. Auxin signaling and the molecular interactions that govern it have been investigated thoroughly enough that we have been able to rebuild heterologous auxin signaling pathways in yeast (Havens et al., 2012). For the last ten years, studies of auxin signaling in yeast coupled with plant synthetic biology experiments have helped us to understand more about how this signaling pathway has diversified within *Arabidopsis* (Hamm et al., 2019) and across distantly related species (Ramos Báez et al., 2020); the important functions of signaling component domains (Guseman et al., 2015; Leydon et al., 2022), the connections between pathway dynamics and plant structure (Guseman et al., 2015; Wright et al., 2017), the functions of promoter sequence and architecture (Lanctot et al., 2020), and more. Nevertheless, questions remain about the more subtle features distinguishing different auxin signaling components and how they inform plant growth and development.

Plants with complex tissues and organs tend to have large, expanded families of signaling component genes, all coding for paralogous proteins predicted to have slightly different functions (Israeli et al., 2019; Mutte et al., 2018). Are differences in paralog function driven by their different expression patterns? What about differences in molecular functions, such as their preferred interacting partners, or their ability to perform all the canonical functions associated with their family? Work I did with the ViVa tools on the population genetics of TIR1/AFBs further support that the differentiated role of AFB1 in rapid auxin response is important to Arabidopsis growth and development. With a recent explosion in studies focusing on the non-canonical rapid response (Cao et al., 2019; Fendrych et al., 2018; Prigge et al., 2019), questions remain about which other F-boxes play a role in this response, and whether other signaling pathway components play a role as well. Although we have made major progress in understanding the role of receptor-IAA interactions in determining the speed of growth responses, other interactions like the one between IAA and co-repressors aren't well studied. Studies focusing on the

functions of B and C-class ARFs have revealed that C-class ARFs tend to not have auxin-responsive functions (E et al., 2018), while B-class ARFs seem to work by competing with A-class ARFs, but exceptions to these rules exist (Kuhn et al., 2020). Clearly, expansion in auxin signaling do allow for expansions in signaling complexity that play important roles in the increased complexities of responses and structure seen in flowering plants.

Most of what we know about auxin signaling has been carried out in Arabidopsis. It is very likely that auxin signaling components operate very differently in Arabidopsis compared to distantly related plants. How are orthologous genes different in their molecular function? How much of what we know about auxin signaling in Arabidopsis can be translated to other plants? How much of the difference that we see in the structures of different plants can be directly linked to differences in auxin signaling function? These questions are especially important in culturally important crop plants. My work with the maize auxin signal transduction pathway helped to answer some of these questions (Ramos Báez et al., 2020). We were able to show that IAAs in maize have different expression patterns in the cob, likely driving differences in their function. We were also able to use yeast to create a system that allows for easy comparison of signaling components from distantly related species. We showed corn and *Arabidopsis* components were well-conserved, and even found conserved differences in function across orthologs of the auxin receptor clades. Further studies need to compare functions of signaling components in other relevant crop species, and across non-flowering plants to see how far back clade-specific molecular functions are conserved, and where they originate. Phylogenetic reconstructions of the expansion of auxin signaling components have never been done thoroughly and would help to better understand the differences between component expansion across species. Tying auxin signaling and evolution at the plant kingdom level has been attempted by many auxin biologists (Kato et al., 2018, 2020; Martin-Arevalillo et al., 2019; Mutte et al., 2018) yet will require more direct collaboration with classically trained evolutionary biologists.

Auxin appears to be a general “ON” signal, activating whatever growth-related molecular processes a cell is primed to undergo. To better understand this requires understanding why auxin signaling is an especially good mechanism for priming growth and developmental functions. How do auxin signaling components and their interactors operate at the molecular level? My work looking at the functions of the LisH-H1 domain revealed the surprisingly well-conserved repressive function of this domain across eukaryotes, hinting at the role of plant co-repressors in a process like mammalian transcriptional priming (Leydon et al., 2022). Could this transcriptional priming function be critical to why auxin responses are so fast, reliable, and versatile? Further work needs to be done to understand the mechanisms of this function and compare it to mammalian pausing and other plant transcriptional responses. What are the plant orthologs of the transcriptional pausing complex proteins? Which are being recruited by LisH-H1? Which auxin signaling components are part of this complexing and at what stages of the auxin response?

Understanding how hormone signaling networks arise is incredibly important to understanding cell biology. The heterologous auxin response circuit in yeast is ideal for studying the evolution and inception of hormone signaling networks. Yeast is great for studying pathways heterologously (Havens et al., 2012; Starr et al., 2017), auxin signaling is well conserved across plants (Lavy et al., 2016; Tao and Estelle, 2018), and this pathways is short, allowing us to use only a few simplified components to go from auxin perception to the activation of a transcriptional response. In my time in Joe Thornton’s lab, I will scour though databases of viridiplantae genomic data to construct a picture of auxin signaling component diversification. I will then use this information to pinpoint nodes of signaling component inception, resuscitate them with ancestral sequence reconstruction, and test their functions in yeast using our heterologous system. From there, I will then trace the evolution of these components in time, mapping out variations in sequence and how they affect the interaction strengths between different components. This work will help to show us which interactions first develop to connect hormone signals to transcriptional changes. It will also help to model how other hormone signaling pathways may arrive, especially those of

closely related pathways like jasmonic acid, karrakin, and strigolactone signaling in plants (Blázquez et al., 2020).

I have come a long way during my graduate career. Actively being a queer person of color and participating in those communities in Berkeley radicalized me, and although I spent lots of time in other plant biology labs, I also spent a lot of time with folks with no science background who went on to have careers in healthcare, public policy, and social justice. I felt guilty for loving and participating in basic biology where felt I wouldn't have direct impacts on the communities that needed help most. On top of that, knowing that I would be working in synthetic biology, engineering plants, and studying corn made me feel that I was setting myself up to work for everyone's favorite villainous mega GMO and herbicide corporation.

I gave the work a chance. The more I understood what I was doing and the possibilities a future in synthetic biology had in store for me, the more I understood why my heart led me here. I learned that I could be an impactful mentor, having multiple people pour their hearts out to me, seeing me as a possibility for positive change and as a role model. I learned how to balance lab work with mutual aid and celebration of queer communities of color outside of work. I practiced being an agent of change within academia. I learned that it was OK to indulge – to unplug from all the injustices of the world and sit with a plant, a yeasty smell, a whirring cytometer – and think about all the neat little proteins buzzing inside a cell that make the world go round. I learned that I could hold both the reality that I would always be a part of violent systems and institutions, and the truth that I can use them as a tool to be an agent of change. I learned that sometimes doing basic biology can open the door to ideas, like those opened by the lisH-H1 project, that have huge potential to make positive and direct impact on people. I learned to think of synthetic biology as just another tool used by humans in their long history of editing plants to have closer and better relationships with the people that depend on them. I learned to think of myself as a little mutation (person) on a protein (institution), recruiting other mutations (awesome people) through epistatic drift (share values), slowly changing the function of the whole (burning it down, as they say). At every

step, I've taken measured risks in joining my science self with other parts of myself to create new ways of being. I've fused ideas from different fields to answer questions in ways I see few tackling them. At every step, I see my path as an academic scientist more clearly and I am more confident in asking for a seat at the table, in defending my work and its value, and in knowing that I am moving in the right direction.

REFERENCES

Blázquez, M.A., Nelson, D.C., and Weijers, D. (2020). Evolution of Plant Hormone Response Pathways. Annual Review of Plant Biology *71*, null. https://doi.org/10.1146/annurev-arplant-050718-100309.

Cao, M., Chen, R., Li, P., Yu, Y., Zheng, R., Ge, D., Zheng, W., Wang, X., Gu, Y., Gelová, Z., et al. (2019). TMK1-mediated auxin signalling regulates differential growth of the apical hook. Nature *568*, 240–243. https://doi.org/10.1038/s41586-019-1069-7.

E, F.-S., Dm, E., Sf, H., Jp, A., Tj, F., Er, L., Jf, G., A, V.-L., T, D., Ss, L., et al. (2018). Class C ARFs evolved before the origin of land plants and antagonize differentiation and developmental transitions in Marchantia polymorpha (New Phytol).

Fendrych, M., Akhmanova, M., Merrin, J., Glanc, M., Hagihara, S., Takahashi, K., Uchida, N., Torii, K.U., and Friml, J. (2018). Rapid and reversible root growth inhibition by TIR1 auxin signalling. Nature Plants *4*, 453–459. https://doi.org/10.1038/s41477-018-0190-1.

Guseman, J.M., Hellmuth, A., Lanctot, A., Feldman, T.P., Moss, B.L., Klavins, E., Villalobos, L.I.A.C., and Nemhauser, J.L. (2015). Auxin-induced degradation dynamics set the pace for lateral root development. Development *142*, 905–909. https://doi.org/10.1242/dev.117234.

Hamm, M.O., Moss, B.L., Leydon, A.R., Gala, H.P., Lanctot, A., Ramos, R., Klaeser, H., Lemmex, A.C., Zahler, M.L., Nemhauser, J.L., et al. (2019). Accelerating structure-function mapping using the ViVa webtool to mine natural variation. Plant Direct *3*, e00147. https://doi.org/10.1002/pld3.147.

Havens, K.A., Guseman, J.M., Jang, S.S., Pierre-Jerome, E., Bolten, N., Klavins, E., and Nemhauser, J.L. (2012). A synthetic approach reveals extensive tunability of auxin signaling. Plant Physiol. *160*, 135–142. https://doi.org/10.1104/pp.112.202184.

Israeli, A., Capua, Y., Shwartz, I., Tal, L., Meir, Z., Levy, M., Bar, M., Efroni, I., and Ori, N. (2019). Multiple Auxin-Response Regulators Enable Stability and Variability in Leaf Development. Curr. Biol. *29*, 1746-1759.e5. https://doi.org/10.1016/j.cub.2019.04.047.

Kato, H., Nishihama, R., Weijers, D., and Kohchi, T. (2018). Evolution of nuclear auxin signaling: lessons from genetic studies with basal land plants. J. Exp. Bot. *69*, 291–301. https://doi.org/10.1093/jxb/erx267.

Kato, H., Mutte, S.K., Suzuki, H., Crespo, I., Das, S., Radoeva, T., Fontana, M., Yoshitake, Y., Hainiwa, E., van den Berg, W., et al. (2020). Design principles of a minimal auxin response system. Nature Plants *6*, 473–482. https://doi.org/10.1038/s41477-020-0662-y.

Kuhn, A., Ramans Harborough, S., McLaughlin, H.M., Natarajan, B., Verstraeten, I., Friml, J., Kepinski, S., and Østergaard, L. (2020). Direct ETTIN-auxin interaction controls chromatin states in gynoecium development. ELife *9*, e51787. https://doi.org/10.7554/eLife.51787.

Lanctot, A., Taylor-Teeples, M., Oki, E.A., and Nemhauser, J. (2020). Specificity in auxin responses is not explained by the promoter preferences of activator ARFs. Plant Physiol. https://doi.org/10.1104/pp.19.01474.

Lavy, M., Prigge, M.J., Tao, S., Shain, S., Kuo, A., Kirchsteiger, K., and Estelle, M. (2016). Constitutive auxin response in Physcomitrella reveals complex interactions between Aux/IAA and ARF proteins. Elife *5*. https://doi.org/10.7554/eLife.13325.

Leydon, A.R., Baez, R.R., and Nemhauser, J.L. (2022). A single helix repression domain is functional across eukaryotes. 2022.05.09.491245. https://doi.org/10.1101/2022.05.09.491245.

Martin-Arevalillo, R., Thévenon, E., Jégu, F., Vinos-Poyo, T., Vernoux, T., Parcy, F., and Dumas, R. (2019). Evolution of the Auxin Response Factors from charophyte ancestors. PLoS Genet *15*. https://doi.org/10.1371/journal.pgen.1008400.

Mutte, S.K., Kato, H., Rothfels, C., Melkonian, M., Wong, G.K.-S., and Weijers, D. (2018). Origin and evolution of the nuclear auxin response system. ELife *7*, e33399. https://doi.org/10.7554/eLife.33399.

Prigge, M.J., Kadakia, N., Greenham, K., and Estelle, M. (2019). Members of the Arabidopsis auxin receptor gene family are essential early in embryogenesis and have broadly overlapping functions. BioRxiv 529248. https://doi.org/10.1101/529248.

Ramos Báez, R., Buckley, Y., Yu, H., Chen, Z., Gallavotti, A., Nemhauser, J.L., and Moss, B.L. (2020). A Synthetic Approach Allows Rapid Characterization of the Maize Nuclear Auxin Response Circuit. Plant Physiol. *182*, 1713–1722. https://doi.org/10.1104/pp.19.01475.

Starr, T.N., Picton, L.K., and Thornton, J.W. (2017). Alternative evolutionary histories in the sequence space of an ancient protein. Nature *549*, 409. https://doi.org/10.1038/nature23902.

Tao, S., and Estelle, M. (2018). Mutational studies of the Aux/IAA proteins in Physcomitrella reveal novel insights into their function. New Phytol *218*, 1534–1542. https://doi.org/10.1111/nph.15039.

Wright, R.C., Zahler, M.L., Gerben, S.R., and Nemhauser, J.L. (2017). Insights into the Evolution and Function of Auxin Signaling F-Box Proteins in Arabidopsis thaliana Through Synthetic Analysis of Natural Variants. Genetics *207*, 583–591. https://doi.org/10.1534/genetics.117.300092.