Predicting structure–function relationships using natural genetic variation

2018-10-18

The way genes and genomes have evolved can provide information about which genes are critical to an organisms fitness, which genes faciliate adaptation, and how robust genes and gene networks are to perturbation. Thousands of genomes have now been sequenced representing a large swath of plant species and genetic variation within many species which has only begun to be explored in the last few years. However, much of this data remains inaccessible to researchers without significant bioinformatics experience. Here, we have developed a webtool ViVa (Visualization of Variation) to facilitate analysis, by any researcher, of The 1001 *Arabidopsis* Genomes Project dataset on a gene, gene family, or gene network level. We We have demonstrated the utility and accessibility of this tool by analyzing the natural variation in the well-studied nuclear auxin signaling pathway as a group of researchers from an array of experience levels. Our analysis has provided further confirmation of existing knowledge of this well studied pathway and has also helped generate new hypotheses. These results suggest there is much to be learned about less studied gene families and networks from similar analysis, especially when paired with biochemical and genetic characterization. This application provides access for every researcher allowing the community to extract maximum value from this large and expensive-to-generate dataset. ViVa is also extensible to any other database of genetic variation such as The 3,000 Rice Genomes Project.

## List of working titles

* Predicting structure–function relationships using natural genetic variation
* Accelerating structure–function mapping using natural genetic variation
* A comprehensive analysis of natural sequence variation within the Arabidopsis thaliana nuclear auxin signaling pathway
* **Add your suggestions**

# Introduction

The first genome sequence of *Arabidopsis thaliana* facilitated rapid advancement of plant biology through molecular genetics. Since this initial genome, massive scaling of sequencing technology has allowed a pioneering group to survey the genomic variation in natural *Arabidopsis thaliana* populations. 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. Outside of genome-wide studies however, this valuable resequencing dataset has seen little use.

Natural genomic variation also provides a set of feasible functional variation at the gene, gene family and gene network level. Characterizing this variation can provide insights into the function and evolution of genes, families and networks (Delker et al. [2010](#ref-delker_natural_2010); R. C. Wright et al. [2017](#ref-wright_insights_2017)) . Here, we present ViVa a web application and R-package for Visualizing Variation within genes, gene families and gene networks. We apply this to analysis of the 1001 genomes dataset [@. This application allows plant molecular biologists without bioinformatics experience to use genomic variation to formulate hypotheses as to the sequence/function/phenotype relationships within their gene, family, or network of their interest.

This application allows easy access to  
1. lists of missense polymorphisms to facilitate biochemical assays of variant effects (Starita et al. [2017](#ref-starita_variant_2017)),  
2. family-wise alignments of variants to facilitate de novo functional domain identification (Melamed et al. [2015](#ref-melamed_combining_2015)),  
3. lists of accessions containing missense (or any type of) polymorphisms to facilitate segregation analysis and measurment of the distribution of variant effects on phenotype (Park et al. [2017](#ref-park_distributions_2017)),  
4. as well as the study of gene, gene family and network evolution (Delker et al. [2010](#ref-delker_natural_2010); Kliebenstein [2008](#ref-kliebenstein_role_2008)).

Here, we demonstrate the utility of this tool through comprehensive analysis of genomic variation in the nuclear auxin signaling pathway. We identify potential functional variation within the coding sequence of the genes within this network and formulate hypotheses regarding the functional implications of this variation and the evolution of these gene families. Our analysis provides further confirmation of much of the existing knowledge of these genes demonstrating the validity of this approach. We also identify natural alleles of several genes and accessions bearing these alleles to facilitate future characterization structure/function, genotype/phenotype, and epistatic relationships.

# Methods

## Data Sources

### Variant Data

Variant data was queried from the 1001 genomes project website (<http://1001genomes.org>) via URL requests to an (undocumented) API service that provides subset files of the whole-genome VCF. The whole-genome VCF file can be found on the project’s website at <http://1001genomes.org/data/GMI-MPI/releases/v3.1/>

### Accession Information

Details on each of the 1135 accessions including CS stock numbers, and geographic location where the samples were collected were taken from the 1001 genomes website at <http://1001genomes.org/accessions.html>, via the download link at the bottom of the page. This data file has been embedded in the R package.

### Gene and transcript information

Information on the genes and transcripts including chromosomal coordinates, start and end location, and transcript length were taken from either the TAIR10 or Araport11 databases. The TAIR10 database, found at <http://arabidopsis.org>, was accessed via the biomart protocol, using the R package biomaRt. The araport11 database, which can also be found on the arabidopsis.org website (<https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FGenes%2FAraport11_genome_release>), has been embedded in the R package.

## Nucleotide Diversity Calculation

Nei defined the nucleotide diversity statistic in his original paper as: “…the average number of nucleotide differences per site between two randomly chosen DNA sequences.” (Nei, 1979) And provided the equation:

Where Where is the frequency of the ith sequence in the population and 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:

where is the nucleotide (A, T, C or G) at position k on the ith sequence of the population. is the length of the sequence (the same for all sequences in the population, assumes only SNPs exist, no indels). n is the total number of sequences in the population.

from this form we can re-arrange summations to the form below:

here 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 for each site, then average those over the sequence length to calculate , the nucleotide diversity of the sequence.

The function Nucleotide\_diversity in the r1001genomes package calculates for each position in the gene or region that contains a variant. note, is equal to 0 at all locations without variants. This is also what is displayed in the Diversity Plot tab of the web app.

## Detailed calculaiton simplification.

The formula for 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’s original formula (1), modifying it slightly, replacing with , being the number of sequences in the population of the form at location k.

$$\begin{equation}
\displaystyle \pi\_k = \sum\_{ij} \frac{n\_i}{n}\frac{n\_j}{n} \pi\_{ij} =
\frac{1}{n^2}\sum\_{ij} n\_i n\_j \pi\_{ij}\\
\pi\_{ij(k)} = \left\{\begin{array}{l}
1 \quad \textrm{if} \quad i \neq j \\
0 \quad \textrm{if} \quad i = j
\end{array}\right.
\qquad(4)
\end{equation}$$

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 as the number of sequences different from i at k.

we can see the summed term will be zero if , and if so:

next we substitute our definition of

then distribute and split summation:

finally, summing is equal to n:

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

## Software

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

**biomaRt:** The biomaRt package was used for accessing the TAIR10 database on arabidopsis.org

**vcfR:** vcfR was used to read in the VCF files in a flat “tidy” format for easy manipulaiton

**BSgenome:** The BSgenome package BSgenome.Athaliana.TAIR.TAIR9 was used as the source for the complete DNA string of the reference genome (Col-0).

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

**GenomicFeatures:** The GenomicFeatures package was used for handling sequence annotations.

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

# *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 *Arabidopsis thaliana* , *TIR1* and *AFB1-5*, evolved through gene duplication and diversification early in the history of vascular plants (G. Parry et al. [2009](#ref-parry_complex_2009)). The rate of co-repressor degradation is determined by the identity of both the receptor and co-repressor (Havens et al. [2012](#ref-havens_synthetic_2012)), and this rate sets the pace of lateral root development (Guseman et al. [2015](#ref-guseman_auxininduced_2015)).

All members of this family have been shown to bind auxin and Aux/IAA proteins. However, AFB1 has 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(H. Yu et al. [2015](#ref-yu_untethering_2015)). This lack of complexation may allow observed high ubiquitous AFB1 accumulation (G. Parry et al. [2009](#ref-parry_complex_2009)). Higher order mutants in the family containing *afb1* mutants suggest that *AFB1* has a moderate positive effect on auxin signaling. 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 (M. J. Prigge et al. [2016](#ref-prigge_arabidopsis_2016)). Interestingly, the strength and rate with which TIR1/AFBs are able to bind and mark Aux/IAAs for degradation are variable (Calderón Villalobos et al. [2012](#ref-calderonvillalobos_combinatorial_2012); Havens et al. [2012](#ref-havens_synthetic_2012)). AFB2 induced 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.

Although it is unclear what unique roles each receptor plays in growth and development, a number of studies have pointed out differences in the ways the six different receptors in A. thaliana differ in biochemical function and expression domain(N. Dharmasiri et al. [2005](#ref-dharmasiri_plant_2005); G. Parry et al. [2009](#ref-parry_complex_2009); M. J. Prigge et al. [2016](#ref-prigge_arabidopsis_2016)). *TIR1/AFB* genes are expressed ubiquitously in A. thaliana tissues, TIR1, AFB2, and AFB3 have been shown to accumulate in the shoot and root meristems and leaf tissues, with slightly different expression patterns for TIR1 (G. Parry et al. [2009](#ref-parry_complex_2009)). Additionally, the expression of AFB5 is strongly circadian-regulated (Covington and Harmer [2007](#ref-covington_circadian_2007)) and AFB3 is more highly expressed in the roots in the presence of nitrate, facilitating increased lateral root formation (Vidal, Álvarez, and Gutiérrez [2014](#ref-vidal_nitrate_2014)), suggesting more broad environmental regulation of this gene family may exist.

All members of this family have been shown to bind auxin and Aux/IAA proteins. However, AFB1 has 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(H. Yu et al. [2015](#ref-yu_untethering_2015)). This lack of complexation may be the cause of high, ubiquitous AFB1 accumulation (G. Parry et al. [2009](#ref-parry_complex_2009)). Higher order mutants in the family containing *afb1* mutants suggest that *AFB1* has a moderate positive effect on auxin signaling. Additionally, AFB4 and AFB5 have been shown to preferentially bind the synthetic auxin picloram, while other family members preferentially bind indole-3-acetic acid (M. J. Prigge et al. [2016](#ref-prigge_arabidopsis_2016)). Interestingly, the strength and rate with which each TIR1/AFB is able to bind and mark an Aux/IAAs for degradation is variable (Calderón Villalobos et al. [2012](#ref-calderonvillalobos_combinatorial_2012); Havens et al. [2012](#ref-havens_synthetic_2012)).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **transcript** | **symbol** | **πN** | **πS** | **πN/πS** | **π coding** | **π transcript** |
| TIR1 | AT3G62980.1 | 0.000082 | 0.000267 | 0.308748 | 0.000350 | 0.001312 |
| AFB1 | AT4G03190.1 | 0.000156 | 0.000627 | 0.248598 | 0.000783 | 0.001320 |
| AFB2 | AT3G26810.1 | 0.000225 | 0.001275 | 0.176109 | 0.001500 | 0.003224 |
| AFB3 | AT1G12820.1 | 0.000353 | 0.000470 | 0.750053 | 0.000823 | 0.001868 |
| AFB4 | AT4G24390.1 | 0.000853 | 0.002208 | 0.386292 | 0.003060 | 0.004496 |
| AFB5 | AT5G49980.1 | 0.000199 | 0.001666 | 0.119402 | 0.001865 | 0.003393 |
| COI1 | AT2G39940.1 | 0.000255 | 0.001768 | 0.144011 | 0.002019 | 0.002351 |

Examining the natural sequence variation across the *AFB* family ?? revealed that *TIR1* and *AFB1* both had very low nonsynonymous diversity, 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 and *AFB4* perhaps undergoing pseudogenization.

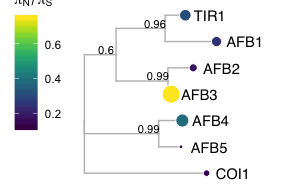
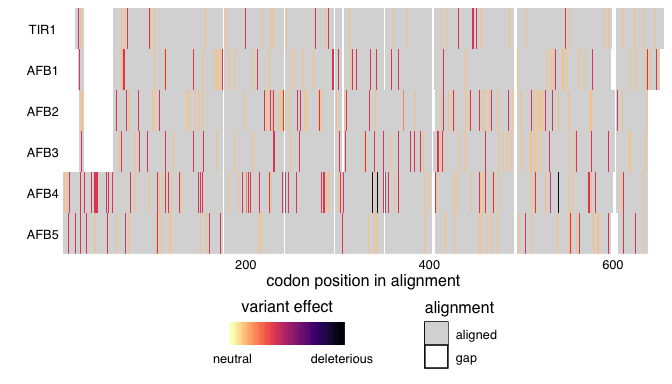


Figure 1 **AFB protein sequence tree mapped with .** Protein sequences were aligned (E. S. Wright [2015](#ref-wright_decipher_2015)) and low information content regions were masked (Kück et al. [2010](#ref-kuck_parametric_2010)) prior to infering a phylogeny (Ronquist and Huelsenbeck [2003](#ref-ronquist_mrbayes_2003)). Tips of the tree are mapped with circles of diameter proportional to and also are colored according to . Nodes are labeled with the poster probability of monophyly.



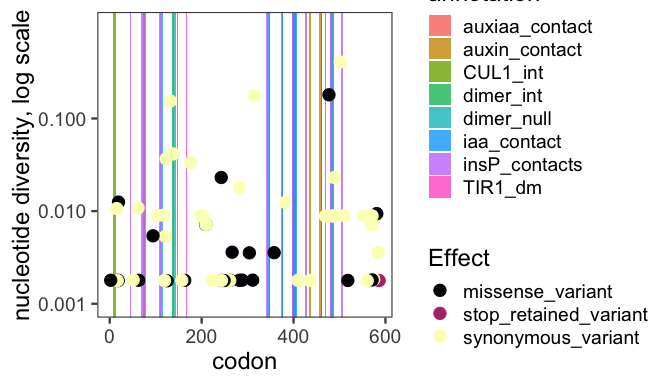
(ref:AFBstructure) A TIR1 structure (2P1Q) in gold in complex with ASK1 in grey and auxin, inositol phosphate and an Aux/IAA degron in teal. TIR1 residues or analagous residues containing nonsynonymous polymorphisms in TIR1 and AFB1 colored in purple and in AFB2 and AFB3 colored in blue.



Figure 2 (ref:AFBstructure)

Although most known functional regions are highly conserved in *AFB1*, there is a nonsynonymous polymorphism in the oligomerization domain, only found in the Can-0 accession. Mutations in this domain of *TIR1* frequently have a semidominant effect on root phenotypes (Dezfulian et al. [2016](#ref-dezfulian_oligomerization_2016); R. C. Wright et al. [2017](#ref-wright_insights_2017)). Characterization of this allele and accession may help determine the role of *AFB1* in this pathway.

**not sure this is very informative, and a map doesn’t seem very exciting, perhaps a domain scale alignment plot would be better here, or above, and would allow us to comment more on where the functional diversity lies within the family.**



The AFB4 and AFB5 receptors have an N-terminal extension prior to the F-box domains. This extension had very high nonsynonymous diversity 3, suggesting that this extension does not play an important functional role in these proteins. Additionally, two frameshift variants and one stop-gained variant were observed in *AFB4* supporting its pseudogenization.

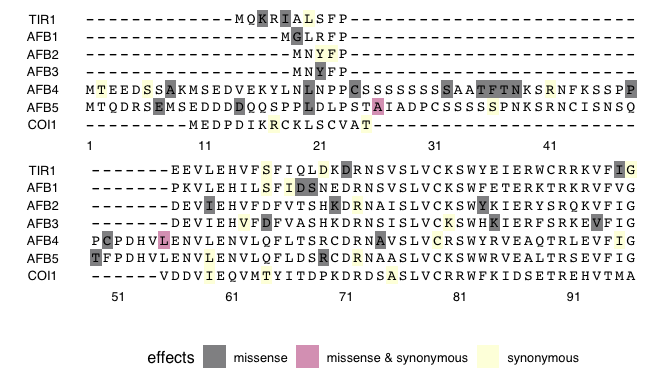


Figure 3 **Alignment of A. thaliana TIR1/AFB F-box domains showing variants.** Cyan marks the F-box binding domain. Yellow, green, and magenta, mark missense and synonymous, synonymous only, and missense only variant locations respectively. Yellow squared E’s show Cul1-interacting residues, differing in AFB1.

# *Aux/IAA* genes

The Aux/IAA proteins (IAAs) mediate both auxin perception and transcriptional co-repression of auxin responsive genes. Auxin perception is carried out along with TIR1/AFB proteins, which together with IAAs act as co-receptors to bind and perceive nuclear auxin. Interaction of IAAs with auxin and TIR1/AFBs is conferred by the highly conserved 13 amino acid degron motif (Tan et al. [2007](#ref-tan_mechanism_2007)). Formation of this auxin co-receptor complex triggers ubiquitination and subsequent degradation of IAAs. This degradation in turn relieves their repression upon ARF transcription factors. Stabilizing mutations within the degron can lead to dominant phenotypes (Berleth, Krogan, and Scarpella [2004](#ref-berleth_auxin_2004); Yang et al. [2004](#ref-yang_iaa1_2004)). The *Arabidopsis thaliana* Aux/IAA family has 29 members, most of which possess a canonical auxin degron. Notable exceptions are the auxin-insensitive IAA20 and IAA30 that lack a degron sequence entirely, while the closely related IAA31 retains a semi-functional degron sequence with a very slow degradation rate (Dreher et al. [2006](#ref-dreher_arabidopsis_2006); Havens et al. [2012](#ref-havens_synthetic_2012)). Additionally, IAA degradation rate is influenced by rate motifs found outside the degron, such as the KR motif (Dreher et al. [2006](#ref-dreher_arabidopsis_2006); Moss et al. [2015](#ref-moss_rate_2015)). IAA-mediated repression of ARFs depends upon both an EAR domain, N-terminal of the degron, that recruits co-repressors of the TOPLESS (TPL) family, and a Phox and Bem1 (PB1) domain that enables interaction via hetero-oligomerization with ARF PB1 domains (Nanao et al. [2014](#ref-nanao_structural_2014); Kagale and Rozwadowski [2011](#ref-kagale_ear_2011)).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **transcript** | **symbol** | **πN** | **πS** | **πN/πS** | **π coding** | **π transcript** |
| IAA1 | AT4G14560.1 | 0.000978 | 0.000177 | 5.536235 | 0.001155 | 0.004657 |
| IAA2 | AT3G23030.1 | 0.001417 | 0.004439 | 0.319222 | 0.005855 | 0.009922 |
| IAA3 | AT1G04240.1 | 0.001100 | 0.000448 | 2.454859 | 0.001548 | 0.002976 |
| IAA4 | AT5G43700.1 | 0.003256 | 0.003191 | 1.020354 | 0.006447 | 0.016298 |
| IAA5 | AT1G15580.1 | 0.000507 | 0.002014 | 0.251828 | 0.002522 | 0.003851 |
| IAA6 | AT1G52830.1 | 0.002628 | 0.001174 | 2.237910 | 0.003773 | 0.006030 |
| IAA7 | AT3G23050.1 | 0.001397 | 0.002600 | 0.537310 | 0.003997 | 0.005808 |
| IAA8 | AT2G22670.1 | 0.001117 | 0.001464 | 0.763161 | 0.002580 | 0.006646 |
| IAA9 | AT5G65670.1 | 0.000806 | 0.000884 | 0.911648 | 0.001690 | 0.005921 |
| IAA10 | AT1G04100.1 | 0.000492 | 0.000738 | 0.666431 | 0.001228 | 0.001552 |
| IAA11 | AT4G28640.1 | 0.000542 | 0.000676 | 0.801761 | 0.001217 | 0.002179 |
| IAA12 | AT1G04550.1 | 0.001220 | 0.000577 | 2.115788 | 0.001797 | 0.003425 |
| IAA13 | AT2G33310.1 | 0.000187 | 0.000667 | 0.280662 | 0.000855 | 0.003270 |
| IAA14 | AT4G14550.1 | 0.000664 | 0.000203 | 3.262467 | 0.000867 | 0.006260 |
| IAA15 | AT1G80390.1 | 0.000412 | 0.001512 | 0.272837 | 0.001921 | 0.005510 |
| IAA16 | AT3G04730.1 | 0.000062 | 0.001763 | 0.035302 | 0.001825 | 0.008730 |
| IAA17 | AT1G04250.1 | 0.000050 | 0.000542 | 0.092649 | 0.000592 | 0.002904 |
| IAA18 | AT1G51950.1 | 0.001467 | 0.001285 | 1.141626 | 0.002746 | 0.008301 |
| IAA19 | AT3G15540.1 | 0.002366 | 0.004267 | 0.554322 | 0.006633 | 0.014151 |
| IAA20 | AT2G46990.1 | 0.000398 | 0.001312 | 0.303653 | 0.001710 | 0.006862 |
| IAA26 | AT3G16500.1 | 0.000443 | 0.000489 | 0.907194 | 0.000932 | 0.003849 |
| IAA27 | AT4G29080.1 | 0.000464 | 0.000104 | 4.466882 | 0.000541 | 0.001822 |
| IAA28 | AT5G25890.1 | 0.000065 | 0.000700 | 0.093506 | 0.000765 | 0.005168 |
| IAA29 | AT4G32280.1 | 0.000353 | 0.000890 | 0.396379 | 0.001243 | 0.002508 |
| IAA30 | AT3G62100.1 | 0.000104 | 0.001789 | 0.058011 | 0.001893 | 0.005801 |
| IAA31 | AT3G17600.1 | 0.000707 | 0.000064 | 11.040566 | 0.000771 | 0.002623 |
| IAA32 | AT2G01200.1 | 0.000336 | 0.000407 | 0.825696 | 0.000743 | 0.004400 |
| IAA33 | AT5G57420.1 | 0.000571 | 0.000350 | 1.633507 | 0.000921 | 0.002605 |
| IAA34 | AT1G15050.1 | 0.002507 | 0.001065 | 2.353071 | 0.003566 | 0.004110 |

### Sister Pair Diversification

Previous research has found evidence of both broad genetic redundancy and specificity within closely related pairs or groups of IAA proteins (Overvoorde et al. [2005](#ref-overvoorde_functional_2005)). 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](#ref-overvoorde_functional_2005)), yet the *IAA6/19* 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](#ref-winkler_variation_2017)).

In our analysis of the *IAA* gene family, we frequently observed that one member of most *IAA* sister pairs had high nonsynonymous diversity while the other sister was more conserved 4. In an interspecific comparison between *A. thaliana* and *A. lyrata*, *IAA19* is more conserved, while *IAA6* has regions of increased sequence divergence upstream of the degron and in the N-terminal half of the PB1 domain (Winkler et al. [2017](#ref-winkler_variation_2017)). Consistent with these results, our analysis of intraspecific natural variation revealed higher conservation for *IAA19* ( = 0.55) compared to *IAA6* ( = 2.3) 4, and also detected high diversity within the same regions of *IAA6* as seen in Winkler et al. (Supp Fig IAA6 diversity plot).

**I think IAA20, 30, and 15 should also be in class C. So there are really two large clades A and B and then there are truncations/loss of function/dominant variants scattered throughout, Class C.**

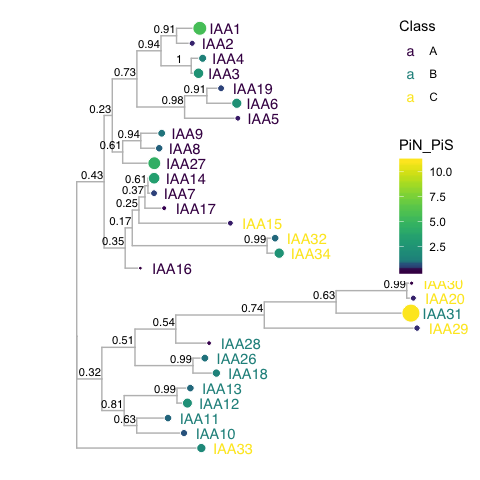


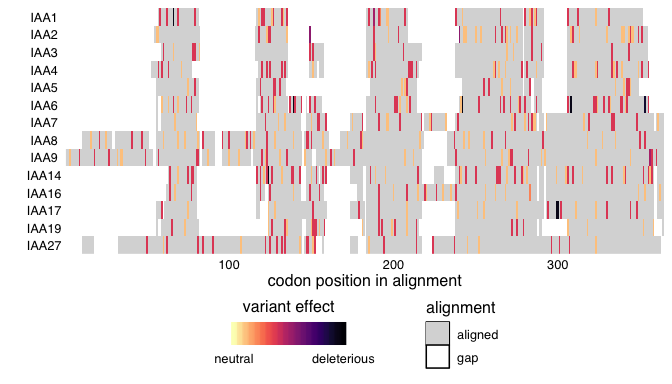
Figure 4 **IAA protein sequence tree mapped with .** Protein sequences were aligned (E. S. Wright [2015](#ref-wright_decipher_2015)) and low information content regions were masked (Kück et al. [2010](#ref-kuck_parametric_2010)) prior to infering a phylogeny (Ronquist and Huelsenbeck [2003](#ref-ronquist_mrbayes_2003)). Tips of the tree are mapped with circles of diameter proportional to and also are colored according to . Nodes are labeled with the poster probability of monophyly.

Within the Group B *IAA* genes, there are four sister pairs—*IAA10*/*IAA11*, *IAA12*/*IAA13*, *IAA18*/*IAA26*, and *IAA20*/*IAA30* 4 (Remington et al. [2004](#ref-remington_contrasting_2004)). Interestingly, among these sister pairs only *IAA12* and *IAA13* have a noticeable difference in nonsynonymous diversity. The *Arabidopsis thaliana* EFP browser shows that *IAA10* and *IAA11* have almost identical expression patterns, with the exception of *IAA10* being more highly expressed in the developing seed, and *IAA11* being more highly expressed in root tissue during lateral root development **need to cite original data source here**. In our analysis, *IAA10* and *IAA11* both showed functional conservation ( of 0.80 and 0.67 respectively) in support of *IAA10* and *IAA11* playing redundant roles. Accessions and associated variants identified here provide a potentially expedited route to elucidating the extent of redundancy between family members such as *IAA10* and *IAA11*.

### New Loss-of-Function/Gain-of-Function alleles

Accessions with functional variants can be potentially used to investigate the extent of redundancy/specialization within sister pairs. For example, three accessions have a missense variant affecting the EAR domain of IAA10 (HR-5, UKSE06-118 and UKSW06-240, all from southern UK). This variant (L49S) is in the second leucine of the LxLxL motif and likely reduces the ability of IAA10 to recruit the TOPLESS co-repressor (M.-S. Lee, An, and Cho [2016](#ref-lee_biological_2016)). The nonsense variant (IAA10-Y219\*) resulting in truncated PB1 domain may be more indicative of *IAA10* null phenotype, as it likely prevents IAA10 from interacting with an ARF. This variant is only found in one accession from northeastern Spain (IP-Svi-0). These accessions and potential loss-of-function alleles may facilitate exploration of the function of *IAA10*.

Within the available natural sequence variation, we found several missense polymorphisms affecting the EAR domain or degron of several canonical clade A IAAs. However, no missense polymorphisms were present in the KR motif or residues critical for PB1-PB1 domain interactions (5). The majority of missense mutations observed were in the variable regions between the conserved domains.

(ref:A-IAA-alignment) 

Several examples of possible gain-of-function missense variants were found in the degron of *IAA2*, *IAA3*, *IAA6*, *IAA7*, *IAA8*, *IAA13*, *IAA15*, *IAA18*, *IAA19*, *IAA27*, *IAA28* and *IAA29*. Most degron variants were very minor alleles in the population, only represented in one or two accessions. Interestingly, degron missense variants occured more frequently in *IAA* genes that have a closely related sister, which may indicate redundant function within the sister-pairs. **I think this is an interesting finding, but I’m not sure about this hypothesis of redundant function. If pairs are acting redundantly and we assume the degron variants are dominant, we would expect the auxin signal they transmit (their function) to be lower, in sum, as a pair. So I would actually hypothesize that there has been some divergence in function as a pair.**

*IAA6* has two potentially dominant missense mutations found within the degron (Supp Fig. alignment), one previously believed to lead to the *shy1-1* phenotype (Kim et al. [1996](#ref-kim_two_1996)). Interestingly, this variant is widely represented in 31 accessions across Europe (Supplemental Map Fig). This finding supports the proposition that iaa6-C78R may not be the mutation causing the *shy1-1D* phenotype but rather represents natural variation in the degron sequence ([note from Mike Prigge in TAIR](https://www.arabidopsis.org/servlets/TairObject?type=notepad&id=11247)). These accessions could be used to confirm whether *iaa6-C78R* is responsible for the *shy1-1D* phenotype. Additionally, an accession was identified with the IAA6-V77M variant, also in the degron, that may be useful to further explore *iaa6* gain-of-function phenotypes.

### Identifying novel functional sequences in IAAs

An exciting possible use for the Natural Variation Webtool is to search for conserved, low-diversity sites that may represent possible sites of functional import, such as post-translational modification or protein-protein interaction. For example, recent detailed biochemical analysis identified putative lysine ubiquitination sites in members of the *IAA6/19* sister pairs (Winkler et al. [2017](#ref-winkler_variation_2017)); Supp Fig IAA6/19 alignment). These sites are conserved across natural sequence variation, while several other lysine residues vary (Supp Figs alignment and map). In IAA6, there are 17 lysines of which ~6 are putative ubiquitination sites~ \*\*I think this should be 5, K3, K32, K33, K91, and K97. Missense variants occured in K34Q, K140N \*. One is proximal to the KR motif, while two others show natural variation. In IAA19, there are 15 total lysines, 8 of which are putative ubiquitination sites (K3, K25, K68, L87, K93, K100, K111, and K141). Two of these ubiquitination sites are proximal to the KR motif and degron. **This sentence is not clear. Only 1 of the ubiquitination sites varies? or one lysine, I think you mean this only 1 lysine that varies, but it is not ubiquitinated.** Natural sequence diversity may be used to narrow down a subset of lysines to assess for functional ubiquitination.

The members of the *IAA8/9/27* clade have a unique N-terminal extension with a highly conserved region common to orthologs in several other species (Dreher et al. [2006](#ref-dreher_arabidopsis_2006)). *IAA8* and *IAA9* play redundant roles in leaf serration (Koenig et al. [2009](#ref-koenig_auxin_2009)), while *IAA8* also plays an important role in lateral root development (Arase et al. [2012](#ref-arase_iaa8_2012)). IAA8 has been reported to uniquely localize in the cytosol (Arase et al. [2012](#ref-arase_iaa8_2012)) where it may interact with LSD1 protein (N. S. Coll, Epple, and Dangl [2011](#ref-coll_programmed_2011)). The N-terminal extension has low natural sequence diversity in *IAA8* and *IAA9* (Supp Fig IAA8/9 alignment) that may represent a site of novel functionality, possibly mediating interaction with LSD1. The few accessions with missense alleles in this region in both *IAA8* and *IAA9* may be valuable in probing the functionality of this region (Supplemental Fig IAA8/9 map).

Members of the *IAA17/7/14/16* clade have a second EAR motif (LxLxL) with no observed missense variants in *IAA7*, *IAA16*, or *IAA17* (Supp Fig IAA7/14/16/17 alignment). This observation supports the experimental evidence for IAA7 that the second EAR motif is functional, *i.e.* it can interact with TPR1 co-repressor and has a minor repressive role in auxin-related developmental phenotypes [lee\_biological\_2016]. For *IAA14*, there were two missense variants that may be of interest to probe functionality of this second EAR motif. One variant, L41F, in the central lysine is represented in only one accession, and the other in the second “x” (K40N) is found in twelve accessions around Copenhagen (Supp Fig. IAA14 map).

# *TPL/TPR* genes

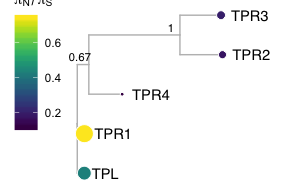
The Auxin signaling pathway utilizes the TOPLESS (TPL) and TOPLESS-related (TPR) family of Gro/TLE/TUP1 type co-repressor proteins to maintain auxin responsive genes in a transcriptionally-repressed state in the absence of auxin (Szemenyei, Hannon, and Long [2008](#ref-szemenyei_topless_2008)). In *Arabidopsis 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 et al. [2006](#ref-long_topless_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, Hannon, and Long [2008](#ref-szemenyei_topless_2008)). Canonical EAR domains have the amino acid sequence LxLxL, as found in most AUX/IAAs (figure xA) (Overvoorde et al. [2005](#ref-overvoorde_functional_2005)). TPL/TPR co-repressors bind EAR domains via their C-terminal to LisH (CTLH) domains found near their N-termini (citations of pre-structure founding papers/reviews). 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 dimerization and tetramerization motifs (Martin-Arevalillo et al. [2017](#ref-martin-arevalillo_structure_2017); Ke et al. [2015](#ref-ke_structural_2015)). The residues required for higher-order multimers of TPL tetramers have also been identified (Ma et al. [2017](#ref-ma_d53_2017)). Additional interactions with transcriptional regulation and chromatin modifying machinery are likely mediated by two tandem beta propeller domains of TPL/TPRs. These domain have not yet been crystallized, however numerous high resolution crystal structures of beta propeller domains allow for structural prediction of the TPL C-terminal beta propeller domains and the locations of likely interaction faces (Figure, X).

The TOPLESS co-repressor family generally exhibits a high level of sequence conservation at the amino acid sequence level across resequenced *Arabidopsis thaliana* accessions, with all values below 1 (Table X). The closely related *TPL* and *TPR1* have the highest values (TPL-0.425, TPR1-0.739), suggesting that these 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 the TPL/TPR family is particularly conserved (Figure X). 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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **transcript** | **symbol** | **πN** | **πS** | **πN/πS** | **π coding** | **π transcript** |
| TPL | AT1G15750.1 | 0.000178 | 0.000418 | 0.425290 | 0.000596 | 0.001011 |
| TPR1 | AT1G80490.1 | 0.000324 | 0.000438 | 0.739744 | 0.000760 | 0.001371 |
| TPR2 | AT3G16830.1 | 0.000078 | 0.000448 | 0.173226 | 0.000526 | 0.001010 |
| TPR3 | AT5G27030.1 | 0.000339 | 0.001737 | 0.195375 | 0.002076 | 0.003516 |
| TPR4 | AT3G15880.1 | 0.000104 | 0.000911 | 0.114674 | 0.001014 | 0.001842 |

The high degree of conservation in the entire N-terminal domain underscores its importance in TPL/TPR function. For example, the initial *tpl-1* mutation (N176H) in the ninth helix is a dominant gain-of-function allele (Long et al. [2006](#ref-long_topless_2006)), which is capable of binding wild-type TPL protein and inducing protein aggregation (Ma et al. [2017](#ref-ma_d53_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. Addtionally, the crystal structure of TPL has revealed that the The LisH domain and CRA domain coordinate TPL dimerization and tetramerization respectively (26601214, 28698367), the CTLH domain is critical for interaction with DNA-binding transcription factors (26601214), and helix nine within the CRA domain may be critical for the formation of higher-order TPL complexes (multimers of TPL tetramers, 28630893).

Compared to the TPL domain, both the proline rich and WD-40 domains exhibited a higher frequency of missense variants. We utilized Phyre2.0 modeling software to predict the structure of the TPL/TPR WD-40 repeats in order to map non-synonymous mutations to protein structure (Figure X). Here we can see both faces of each predicted beta-propeller structure of the WD-40 domains, which are connected by a short peptide linker. This model reveals amino acid conservation in the core and central residues of the WD-40 domains. ~~These observations could be used as the basis for forward mutagenesis strategies to test whether this region of TPL functions similarly to its homolog TLE, which binds its target genes EH1 and the C-terminal WRPW/Y motif (Hairy/Hes/Runx) in the center of the WD-40 domain (pmid 16762837).~~ \*\*Can we say something more general here? There is not enough background to support this. What is TLE from? What is EH1? Is



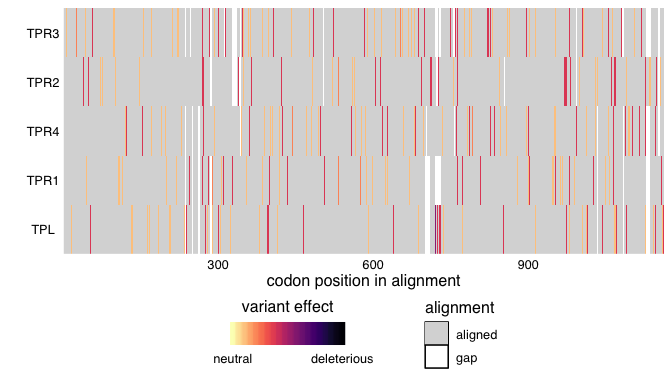


Figure 6 **Alignment of A. thaliana TPL/TPR genes showing variants.** This alignment should probably be paired down to a key region rather than the eniter AA sequence.

Despite the presence of five *TPL/TPR* genes in the genome, which are considered to be redundant (pmid 16763149), very few observed SNPs are predicted to result in loss of function for any *TPL/TPR* gene. Only one frameshift variant was observed, *TPR4-665delG*, in a single accession (Aiell-4 from Italy). This deletion occurs in codon 222, just after the TPL N-terminal domain, and generates an early stop eleven codons downstream from the deletion. It is unclear whether this N-terminal fragment would be expressed at any appreciable level, as it is possible that the generation of an early stop codon would trigger nonsense-mediated decay. However, if expressed, it would be fascinating to examine whether this naturally occurring N-terminal truncation retains activity as a co-repressor. Given the observation that the TPL N-terminal domain is sufficient to bind histones (28630893), and repress transcription via synthetic assays in yeast (24979769), it is possible that this represents a semi-functional TPL variant. There has been no evidence for prioritization of individual *TPL/TPR* family members in the auxin response pathway, but experiments modulating the gene copy number of *TPR1* revealed the importance of maintaining the proper dosage of *TPL/TPR* in order to accurately trigger the salicylic acid innate immune response (20647385). These results, taken together with the absence of *TPL/TPR* loss-of-function alleles in the *Arabidopsis thaliana* accessions suggest that perhaps the entire family is being conserved at the nucleotide level in the population to meet a gene dosage requirement, with little opportunity for neo- or sub-functionalization.

# *ARF* genes

Auxin response is ultimately mediated by the auxin responsive transcription factors (ARFs). There are 23 ARFs in *Arabidopsis 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 (T. J. Guilfoyle and Hagen [2007](#ref-guilfoyle_auxin_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](#ref-wu_auxinregulated_2015)). Additionally, ARF7 interacts with Mediator subunits, directly tethering transcriptional activation machinery to its binding sites in the chromosome (Ito et al. [2016](#ref-ito_auxindependent_2016)). Class B and C ARFs are historically categorized as repressor ARFs, though the mechanism through which they confer repression has not been identified. Their middle regions tend to be proline- and serine-rich (citation, Ulmasov?).

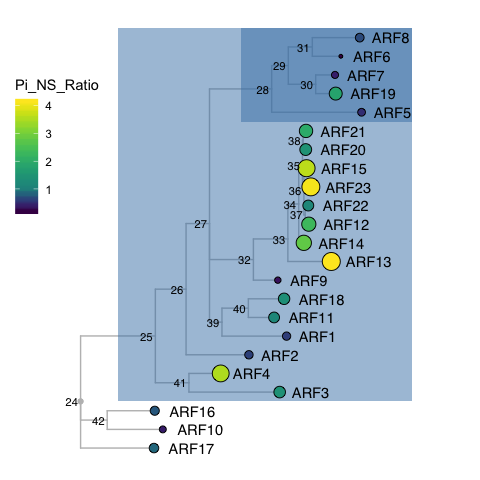
Canonical ARFs are comprised of three major domains. Recent crystallization of these domains have informed structure-function analysis of the ARFs (Boer et al. [2014](#ref-boer_structural_2014); Korasick et al. [2014](#ref-korasick_molecular_2014); Nanao et al. [2014](#ref-nanao_structural_2014)). These domains are conserved throughout land plants (Mutte et al. [2018](#ref-mutte_origin_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).

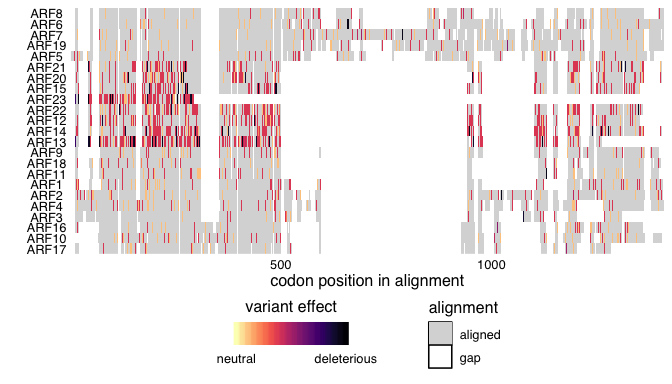
The C-terminal domain of the ARF is a protein-protein interaction domain mediating interactions among ARFs, between ARFs and other transcription factors, and between ARFs and the Aux/IAA repressors. This interaction domain was recently characterized as a Phox and Bem1 (PB1) domain, which is comprised of a positive and negative face with conserved basic and acidic residues, respectively (Korasick et al. [2014](#ref-korasick_molecular_2014); Nanao et al. [2014](#ref-nanao_structural_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](#ref-korasick_molecular_2014)). However, it is unclear whether ARF multimerization occurs or plays a developmental 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 et al. [2016](#ref-pierre-jerome_functional_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, Eklund, and Bowman [2015](#ref-flores-sandoval_simple_2015)). The expansion of these classes in flowering plants is the result of both whole genome and tandem duplication events (Remington et al. [2004](#ref-remington_contrasting_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.

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 *ARFs* was generally low, especially compared to the class B and C *ARFs*, suggesting that class A *ARFs* are central to auxin signal transduction and plant development. Analysis of class A *ARF* nonsynonymous diversity suggests that the majority of these *ARFs* are highly functionally conserved, with values much lower than 1 with the exception of *ARF19*, with 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 *ARFs*, with 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.

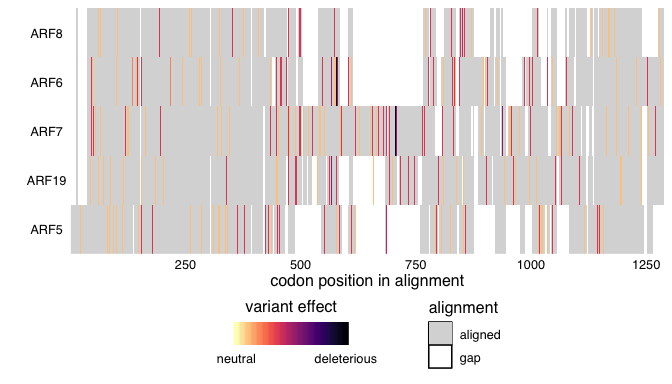
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **transcript** | **symbol** | **πN** | **πS** | **πN/πS** | **π coding** | **π transcript** |
| ARF1 | AT1G59750.1 | 0.001083 | 0.001832 | 0.591220 | 0.002914 | 0.005754 |
| ARF2 | AT5G62000.1 | 0.000335 | 0.000551 | 0.607475 | 0.000884 | 0.001193 |
| ARF3 | AT2G33860.1 | 0.000660 | 0.000478 | 1.380876 | 0.001130 | 0.001128 |
| ARF4 | AT5G60450.1 | 0.000709 | 0.000202 | 3.510435 | 0.000911 | 0.001285 |
| ARF5 | AT1G19850.1 | 0.000185 | 0.000368 | 0.501169 | 0.000493 | 0.000776 |
| ARF6 | AT1G30330.1 | 0.000150 | 0.000755 | 0.198981 | 0.000903 | 0.001108 |
| ARF7 | AT5G20730.1 | 0.000615 | 0.001490 | 0.412529 | 0.002102 | 0.003288 |
| ARF8 | AT5G37020.1 | 0.000317 | 0.000475 | 0.666836 | 0.000791 | 0.001699 |
| ARF9 | AT4G23980.1 | 0.000728 | 0.002227 | 0.327061 | 0.002955 | 0.003703 |
| ARF10 | AT2G28350.1 | 0.000847 | 0.002185 | 0.387555 | 0.003029 | 0.003811 |
| ARF11 | AT2G46530.1 | 0.000673 | 0.000582 | 1.155427 | 0.001228 | 0.001457 |
| ARF12 | AT1G34310.1 | 0.003390 | 0.001473 | 2.301858 | 0.004733 | 0.005396 |
| ARF13 | AT1G34170.1 | 0.009079 | 0.002190 | 4.146384 | 0.010062 | 0.010705 |
| ARF14 | AT1G35540.1 | 0.004417 | 0.001594 | 2.770892 | 0.005487 | 0.006662 |
| ARF15 | AT1G35520.1 | 0.002323 | 0.000641 | 3.621812 | 0.002924 | 0.003574 |
| ARF16 | AT4G30080.1 | 0.000315 | 0.000422 | 0.746242 | 0.000734 | 0.000707 |
| ARF17 | AT1G77850.1 | 0.001458 | 0.001724 | 0.845695 | 0.003182 | 0.003203 |
| ARF18 | AT3G61830.1 | 0.000283 | 0.000212 | 1.338799 | 0.000434 | 0.000926 |
| ARF19 | AT1G19220.1 | 0.000755 | 0.000412 | 1.830943 | 0.001168 | 0.002494 |
| ARF20 | AT1G35240.1 | 0.003138 | 0.002162 | 1.451252 | 0.005220 | 0.005501 |
| ARF21 | AT1G34410.1 | 0.002646 | 0.001338 | 1.978045 | 0.003749 | 0.004031 |
| ARF22 | AT1G34390.1 | 0.002558 | 0.002191 | 1.167182 | 0.004465 | 0.005284 |
| ARF23 | AT1G43950.1 | 0.008468 | 0.002073 | 4.084158 | 0.008494 | 0.010323 |





### Class A *ARFs*

For all class A ARFs, the middle region of the protein was the predominant high diversity region. The DNA-binding domain had very few, low-diversity missense mutations, as did 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.



Class A ARFs all contain the canonical B3 DNA-binding domain, PB1 interaction domain and glutamine-rich middle region (T. J. Guilfoyle and Hagen [2007](#ref-guilfoyle_auxin_2007)). 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 be acted on by natural selection and have phenotypic consequences (Press, Carlson, and Queitsch [2014](#ref-press_overdue_2014)).

A few missense variants in *ARF5* highlighted potential functional variation in the class A ARFs. A high frequency allele *(What are the variants?)*, resulting in substitutions of small amino acids on the positive face of the PB1 domain for bulky phenylalanines, may disrupt interactions with other ARF or Aux/IAA PB1 domains. Interestingly, these variants were always found together in accessions Tu-B2-3, Ru-2, Haes-1, Rd-0, 627ME-13Y1, Le-0, Bu-0 and Bsch-0 **Where are these?**. Another highly represented polymorphism in ARF5 results in change in the 17th amino acid from a leucine to a methionine. This position is in the N-terminal extension of ARF5 which is of unknown function. It is possible that this methionine acts as an alternative start site in the accessions.

### Class B *ARFs*

ARF23 has a truncated DNA-binding domain and had a high value of 4.1 and several high-frequency nonsense variants, as may be expected of a gene that is pseudogenizing.

ARF3, ARF13 and ARF17 all lack a C-terminal PB1 domain, which suggests they may be unable to interact with Aux/IAA repressors, possibly acting as auxin-insensitive competitive inhibitors of ARF-mediated transcription (Fig ARF3\_13\_17 alignment). These ARFs span the range of ratios relative to the other ARFs, ranging from 0.85 to 4.1. ARF17 is more conserved than the activator ARF, ARF19. ARF13 has many nonsense variants of high diversity and may be pseudogenizing.

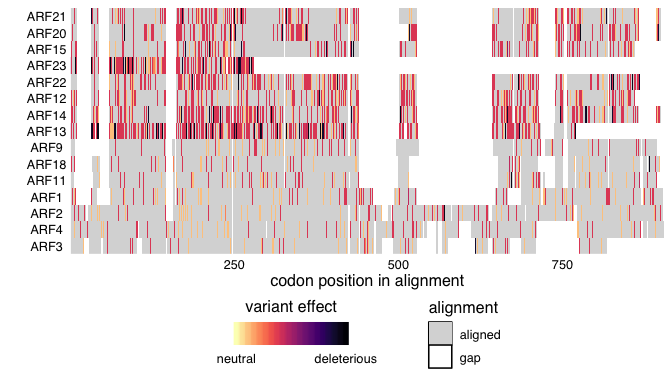


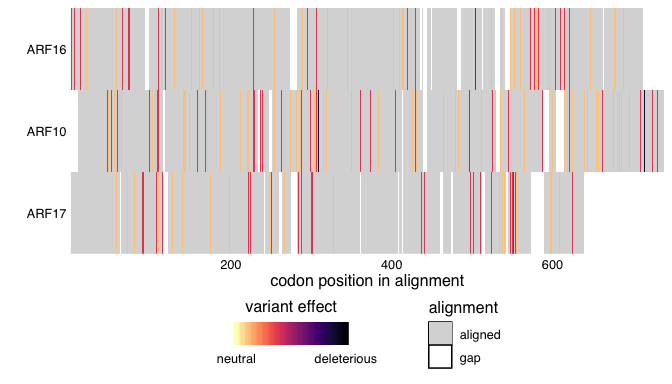
Figure 7 (ref:B-ARFalignment)

ARF14, ARF15, **ARF16 I don’t think this is correct ARF 16 is class C, perhaps this should be 12?**, ARF20, and ARF21 all do not have one or multiple of the conserved acidic residues in the PB1 domain, and all are within a subclade of the B ARFs. Interestingly, no non-synonymous variation is seen in these important residues for Aux/IAA and ARF interaction, suggesting that the amino acids that are in the regions may still be important for the function of these proteins. **Might cut this, it is interesting that perhaps homo/heterodimerization is not as important for this clade of ARFs but I think this conclusion mostly comes from looking at the alignments and not from the natural variation. Perhaps if it looks like these ARFs have higher than expected, or more deleterious variation across their PB1 domains**

### Class C *ARFs*

Class C ARFs show low nucleotide diversity scores, all values substantially lower than 1. ARF16 was the most conserved, whereas its clade members (ARF10, ARF17) had scores at least four times higher (Remington et al. [2004](#ref-remington_contrasting_2004)).

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 exhibit overall low diversity. Only the ARF10 DNA binding domain contains missense variants (S147T and A157S) (Fig. ARF10\_DBD Variants distribution). **Any insights into whether these may affect function? They are quite conservative.** Of the PB1 domain containing class C ARFs, ARF16 exhibits several missense variants which are sporadically distributed, in contrast to the conserved PB1 domain of ARF10 (Fig. Arf16\_PB\_variants distribution). **Interesting that the PB1 domain of ARF10 is conserved and the DBD of ARF16 is conserved. It this suggestive of subfunctionalization in this family, where one retains DNA and/or dimerization specificity and the other retains specificity for IAAs?**



Several members of class B (ARF2, 3 and 4) and C ARFs (ARF10, 16 and 17) are distinguished from other ARFs on the basis of their post transcriptional regulation by trans-acting small interfering RNAs (tasiRNAs) and microRNAs (miRNAs) (Rhoades et al. [2002](#ref-rhoades_prediction_2002); Allen et al. [2005](#ref-allen_micrornadirected_2005)). **(AMY check for conservation of Arf 6 and 8 targeted by miRNA 167).** The miRNA target of class C ARFs showed a high degree of conservation (Fig. ARF10\_16\_17\_Alignment). This is consistent with perturbation **what exactly does perturbation mean** studies that have shown drastic developmental phenotype for class C ARFs (J.-W. Wang et al. [2005](#ref-wang_control_2005); Mallory, Bartel, and Bartel [2005](#ref-mallory_micrornadirected_2005); P.-P. Liu et al. [2007](#ref-liu_repression_2007)). For the class B ARFs (2, 3 and 4) the tasiRNA binding motif is also highly conserved suggesting strong regulatory constraints on these genes (Fig ARF2\_3\_4\_Alignment). The DNA binding domain of ARF4 exhibits non-synonymous variation in 4 accession and would be interesting to carry out phenotypic characterization of these accessions for developmental growth defects (Fig ARF4\_DBD Variants map).

# Discussion

Our analysis of the natural diversity in the gene families that make up the auxin nuclear signaling pathway has confirmed existing knowledge, supported existing hypotheses and generated new hypotheses. These results have been generated by a team of undergraduate and graduate students, postdoctoral researchers, and principal investigators as an alpha-test of the ViVa web application developed here. Our analysis has confirmed the importance of auxin-induced degradation of the Aux/IAA proteins as evidenced by the strong conservation of the *AFB* genes and the canonical degron region of the *Aux/IAA* genes. Despite this relative conservation, this analysis has also identified many accessions containing new alleles of interest in the *AFB* and *Aux/IAA* families. -interesting points from the AFBs -interesting points IAAs -TPLs -ARFs -DBD and PB1

Importance of -AFB -IAA degron -TPL -ARF DBD and PB1

# References

Allen, Edwards, Zhixin Xie, Adam M. Gustafson, and James C. Carrington. 2005. “microRNA-Directed Phasing During Trans-Acting siRNA Biogenesis in Plants.” *Cell* 121 (2): 207–21. doi:[10.1016/j.cell.2005.04.004](https://doi.org/10.1016/j.cell.2005.04.004).

Arase, Fumi, Hiroko Nishitani, Mayumi Egusa, Nami Nishimoto, Sumiko Sakurai, Naho Sakamoto, and Hironori Kaminaka. 2012. “IAA8 Involved in Lateral Root Formation Interacts with the Tir1 Auxin Receptor and ARF Transcription Factors in Arabidopsis.” *PloS One* 7 (8): e43414. doi:[10.1371/journal.pone.0043414](https://doi.org/10.1371/journal.pone.0043414).

Berleth, Thomas, Naden T. Krogan, and Enrico Scarpella. 2004. “Auxin Signals–Turning Genes on and Turning Cells Around.” *Current Opinion in Plant Biology* 7 (5): 553–63. doi:[10.1016/j.pbi.2004.07.016](https://doi.org/10.1016/j.pbi.2004.07.016).

Boer, D. Roeland, Alejandra Freire-Rios, Willy A. M. van den Berg, Terrens Saaki, Iain W. Manfield, Stefan Kepinski, Irene López-Vidrieo, et al. 2014. “Structural Basis for DNA Binding Specificity by the Auxin-Dependent ARF Transcription Factors.” *Cell* 156 (3): 577–89. doi:[10.1016/j.cell.2013.12.027](https://doi.org/10.1016/j.cell.2013.12.027).

Calderón Villalobos, Luz Irina A., Sarah Lee, Cesar De Oliveira, Anthony Ivetac, Wolfgang Brandt, Lynne Armitage, Laura B. Sheard, et al. 2012. “A Combinatorial Tir1/AFB-Aux/IAA Co-Receptor System for Differential Sensing of Auxin.” *Nat Chem Biol* 8 (5): 477–85. doi:[10.1038/nchembio.926](https://doi.org/10.1038/nchembio.926).

Coll, N. S., P. Epple, and J. L. Dangl. 2011. “Programmed Cell Death in the Plant Immune System.” *Cell Death and Differentiation* 18 (8): 1247–56. doi:[10.1038/cdd.2011.37](https://doi.org/10.1038/cdd.2011.37).

Covington, Michael F., and Stacey L. Harmer. 2007. “The Circadian Clock Regulates Auxin Signaling and Responses in Arabidopsis.” *PLoS Biology* 5 (8): e222. doi:[10.1371/journal.pbio.0050222](https://doi.org/10.1371/journal.pbio.0050222).

Delker, Carolin, Yvonne Pöschl, Anja Raschke, Kristian Ullrich, Stefan Ettingshausen, Valeska Hauptmann, Ivo Grosse, and Marcel Quint. 2010. “Natural Variation of Transcriptional Auxin Response Networks in Arabidopsis Thaliana.” *The Plant Cell* 22 (7): 2184–2200. doi:[10.1105/tpc.110.073957](https://doi.org/10.1105/tpc.110.073957).

Dezfulian, Mohammad H., Espanta Jalili, Don Karl A. Roberto, Britney L. Moss, Kerry Khoo, Jennifer L. Nemhauser, and William L. Crosby. 2016. “Oligomerization of SCF Tir1 Is Essential for Aux/IAA Degradation and Auxin Signaling in Arabidopsis.” *PLOS Genet* 12 (9): e1006301. doi:[10.1371/journal.pgen.1006301](https://doi.org/10.1371/journal.pgen.1006301).

Dharmasiri, Nihal, Sunethra Dharmasiri, Dolf Weijers, Esther Lechner, Masashi Yamada, Lawrence Hobbie, Jasmin S. Ehrismann, Gerd Jürgens, and Mark Estelle. 2005. “Plant Development Is Regulated by a Family of Auxin Receptor F Box Proteins.” *Developmental Cell* 9 (1): 109–19. doi:[10.1016/j.devcel.2005.05.014](https://doi.org/10.1016/j.devcel.2005.05.014).

Dreher, Kate A., Jessica Brown, Robert E. Saw, and Judy Callis. 2006. “The Arabidopsis Aux/IAA Protein Family Has Diversified in Degradation and Auxin Responsiveness.” *The Plant Cell* 18 (3): 699–714. doi:[10.1105/tpc.105.039172](https://doi.org/10.1105/tpc.105.039172).

Flores-Sandoval, Eduardo, D. Magnus Eklund, and John L. Bowman. 2015. “A Simple Auxin Transcriptional Response System Regulates Multiple Morphogenetic Processes in the Liverwort Marchantia Polymorpha.” *PLOS Genetics* 11 (5): e1005207. doi:[10.1371/journal.pgen.1005207](https://doi.org/10.1371/journal.pgen.1005207).

Guilfoyle, Tom J., and Gretchen Hagen. 2007. “Auxin Response Factors.” *Current Opinion in Plant Biology*, Cell signalling and gene regulationedited by jian-kang zhu and ko shimamoto, 10 (5): 453–60. doi:[10.1016/j.pbi.2007.08.014](https://doi.org/10.1016/j.pbi.2007.08.014).

Guseman, Jessica M., Antje Hellmuth, Amy Lanctot, Tamar P. Feldman, Britney L. Moss, Eric Klavins, Luz Irina A. Calderón Villalobos, and Jennifer L. Nemhauser. 2015. “Auxin-Induced Degradation Dynamics Set the Pace for Lateral Root Development.” *Development* 142 (5): 905–9. doi:[10.1242/dev.117234](https://doi.org/10.1242/dev.117234).

Havens, Kyle A., Jessica M. Guseman, Seunghee S. Jang, Edith Pierre-Jerome, Nick Bolten, Eric Klavins, and Jennifer L. Nemhauser. 2012. “A Synthetic Approach Reveals Extensive Tunability of Auxin Signaling.” *Plant Physiol* 160 (1): 135–42. doi:[10.1104/pp.112.202184](https://doi.org/10.1104/pp.112.202184).

Ito, Jun, Hidehiro Fukaki, Makoto Onoda, Lin Li, Chuanyou Li, Masao Tasaka, and Masahiko Furutani. 2016. “Auxin-Dependent Compositional Change in Mediator in Arf7- and Arf19-Mediated Transcription.” *Proceedings of the National Academy of Sciences*, May, 201600739. doi:[10.1073/pnas.1600739113](https://doi.org/10.1073/pnas.1600739113).

Kagale, Sateesh, and Kevin Rozwadowski. 2011. “EAR Motif-Mediated Transcriptional Repression in Plants: An Underlying Mechanism for Epigenetic Regulation of Gene Expression.” *Epigenetics* 6 (2): 141–46.

Ke, Jiyuan, Honglei Ma, Xin Gu, Adam Thelen, Joseph S. Brunzelle, Jiayang Li, H. Eric Xu, and Karsten Melcher. 2015. “Structural Basis for Recognition of Diverse Transcriptional Repressors by the TOPLESS Family of Corepressors.” *Science Advances* 1 (6): e1500107. doi:[10.1126/sciadv.1500107](https://doi.org/10.1126/sciadv.1500107).

Kim, Byung Chul, Moon Soo Soh, Bong Joong Kang, Masaki Furuya, and Hong Gil Nam. 1996. “Two Dominant Photomorphogenic Mutations of Arabidopsis Thaliana Identified as Suppressor Mutations of Hy2.” *The Plant Journal* 9 (4): 441–56. doi:[10.1046/j.1365-313X.1996.09040441.x](https://doi.org/10.1046/j.1365-313X.1996.09040441.x).

Kliebenstein, Daniel J. 2008. “A Role for Gene Duplication and Natural Variation of Gene Expression in the Evolution of Metabolism.” *PLOS ONE* 3 (3): e1838. doi:[10.1371/journal.pone.0001838](https://doi.org/10.1371/journal.pone.0001838).

Koenig, Daniel, Emmanuelle Bayer, Julie Kang, Cris Kuhlemeier, and Neelima Sinha. 2009. “Auxin Patterns Solanum Lycopersicum Leaf Morphogenesis.” *Development (Cambridge, England)* 136 (17): 2997–3006. doi:[10.1242/dev.033811](https://doi.org/10.1242/dev.033811).

Korasick, David A., Corey S. Westfall, Soon Goo Lee, Max H. Nanao, Renaud Dumas, Gretchen Hagen, Thomas J. Guilfoyle, Joseph M. Jez, and Lucia C. Strader. 2014. “Molecular Basis for AUXIN RESPONSE FACTOR Protein Interaction and the Control of Auxin Response Repression.” *Proceedings of the National Academy of Sciences* 111 (14): 5427–32. doi:[10.1073/pnas.1400074111](https://doi.org/10.1073/pnas.1400074111).

Kück, Patrick, Karen Meusemann, Johannes Dambach, Birthe Thormann, Björn M. von Reumont, Johann W. Wägele, and Bernhard Misof. 2010. “Parametric and Non-Parametric Masking of Randomness in Sequence Alignments Can Be Improved and Leads to Better Resolved Trees.” *Frontiers in Zoology* 7 (March): 10. doi:[10.1186/1742-9994-7-10](https://doi.org/10.1186/1742-9994-7-10).

Lee, Min-Soo, Ji-Hyun An, and Hyung-Taeg Cho. 2016. “Biological and Molecular Functions of Two EAR Motifs of Arabidopsis Iaa7.” *Journal of Plant Biology* 59 (1): 24–32. doi:[10.1007/s12374-016-0453-1](https://doi.org/10.1007/s12374-016-0453-1).

Liu, Po-Pu, Taiowa A. Montgomery, Noah Fahlgren, Kristin D. Kasschau, Hiroyuki Nonogaki, and James C. Carrington. 2007. “Repression of AUXIN RESPONSE Factor10 by microRNA160 Is Critical for Seed Germination and Post-Germination Stages.” *The Plant Journal: For Cell and Molecular Biology* 52 (1): 133–46. doi:[10.1111/j.1365-313X.2007.03218.x](https://doi.org/10.1111/j.1365-313X.2007.03218.x).

Long, Jeff A., Carolyn Ohno, Zachery R. Smith, and Elliot M. Meyerowitz. 2006. “TOPLESS Regulates Apical Embryonic Fate in Arabidopsis.” *Science* 312 (5779): 1520–3. doi:[10.1126/science.1123841](https://doi.org/10.1126/science.1123841).

Ma, Honglei, Jingbo Duan, Jiyuan Ke, Yuanzheng He, Xin Gu, Ting-Hai Xu, Hong Yu, et al. 2017. “A D53 Repression Motif Induces Oligomerization of TOPLESS Corepressors and Promotes Assembly of a Corepressor-Nucleosome Complex.” *Science Advances* 3 (6). doi:[10.1126/sciadv.1601217](https://doi.org/10.1126/sciadv.1601217).

Mallory, Allison C., David P. Bartel, and Bonnie Bartel. 2005. “MicroRNA-Directed Regulation of Arabidopsis AUXIN RESPONSE Factor17 Is Essential for Proper Development and Modulates Expression of Early Auxin Response Genes.” *The Plant Cell* 17 (5): 1360–75. doi:[10.1105/tpc.105.031716](https://doi.org/10.1105/tpc.105.031716).

Martin-Arevalillo, Raquel, Max H. Nanao, Antoine Larrieu, Thomas Vinos-Poyo, David Mast, Carlos Galvan-Ampudia, Géraldine Brunoud, Teva Vernoux, Renaud Dumas, and François Parcy. 2017. “Structure of the Arabidopsis TOPLESS Corepressor Provides Insight into the Evolution of Transcriptional Repression.” *Proceedings of the National Academy of Sciences*, July, 201703054. doi:[10.1073/pnas.1703054114](https://doi.org/10.1073/pnas.1703054114).

Melamed, Daniel, David L. Young, Christina R. Miller, and Stanley Fields. 2015. “Combining Natural Sequence Variation with High Throughput Mutational Data to Reveal Protein Interaction Sites.” *PLoS Genet* 11 (2): e1004918. doi:[10.1371/journal.pgen.1004918](https://doi.org/10.1371/journal.pgen.1004918).

Moss, Britney L., Haibin Mao, Jessica M. Guseman, Thomas R. Hinds, Antje Hellmuth, Marlies Kovenock, Anisa Noorassa, et al. 2015. “Rate Motifs Tune Auxin/Indole-3-Acetic Acid Degradation Dynamics.” *Plant Physiology* 169 (1): 803–13. doi:[10.1104/pp.15.00587](https://doi.org/10.1104/pp.15.00587).

Mutte, Sumanth K., Hirotaka Kato, Carl Rothfels, Michael Melkonian, Gane Ka-Shu Wong, and Dolf Weijers. 2018. “Origin and Evolution of the Nuclear Auxin Response System.” *eLife* 7 (March): e33399. doi:[10.7554/eLife.33399](https://doi.org/10.7554/eLife.33399).

Nanao, Max H., Thomas Vinos-Poyo, Géraldine Brunoud, Emmanuel Thévenon, Meryl Mazzoleni, David Mast, Stéphanie Lainé, et al. 2014. “Structural Basis for Oligomerization of Auxin Transcriptional Regulators.” *Nature Communications* 5: 3617. doi:[10.1038/ncomms4617](https://doi.org/10.1038/ncomms4617).

Overvoorde, Paul J., Yoko Okushima, José M. Alonso, April Chan, Charlie Chang, Joseph R. Ecker, Beth Hughes, et al. 2005. “Functional Genomic Analysis of the AUXIN/INDOLE-3-ACETIC ACID Gene Family Members in Arabidopsis Thaliana.” *The Plant Cell* 17 (12): 3282–3300. doi:[10.1105/tpc.105.036723](https://doi.org/10.1105/tpc.105.036723).

Park, Briton, Matthew T. Rutter, Charles B. Fenster, V. Vaughan Symonds, Mark C. Ungerer, and Jeffrey P. Townsend. 2017. “Distributions of Mutational Effects and the Estimation of Directional Selection in Divergent Lineages of Arabidopsis Thaliana.” *Genetics* 206 (4): 2105–17. doi:[10.1534/genetics.116.199190](https://doi.org/10.1534/genetics.116.199190).

Parry, G., L. I. Calderon-Villalobos, M. Prigge, B. Peret, S. Dharmasiri, H. Itoh, E. Lechner, W. M. Gray, M. Bennett, and M. Estelle. 2009. “Complex Regulation of the Tir1/AFB Family of Auxin Receptors.” *Proceedings of the National Academy of Sciences* 106 (52): 22540–5. doi:[10.1073/pnas.0911967106](https://doi.org/10.1073/pnas.0911967106).

Pierre-Jerome, Edith, Britney L. Moss, Amy Lanctot, Amber Hageman, and Jennifer L. Nemhauser. 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): 11354–9. doi:[10.1073/pnas.1604379113](https://doi.org/10.1073/pnas.1604379113).

Press, Maximilian O., Keisha D. Carlson, and Christine Queitsch. 2014. “The Overdue Promise of Short Tandem Repeat Variation for Heritability.” *Trends in Genetics: TIG* 30 (11): 504–12. doi:[10.1016/j.tig.2014.07.008](https://doi.org/10.1016/j.tig.2014.07.008).

Prigge, Michael J., Kathleen Greenham, Yi Zhang, Aaron Santner, Cristina Castillejo, Andrew M. Mutka, Ronan C. O’Malley, Joseph R. Ecker, Barbara N. Kunkel, and Mark Estelle. 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): 1383–90. doi:[10.1534/g3.115.025585](https://doi.org/10.1534/g3.115.025585).

Remington, David L., Todd J. Vision, Thomas J. Guilfoyle, and Jason W. Reed. 2004. “Contrasting Modes of Diversification in the Aux/IAA and ARF Gene Families.” *Plant Physiology* 135 (3): 1738–52. doi:[10.1104/pp.104.039669](https://doi.org/10.1104/pp.104.039669).

Rhoades, Matthew W., Brenda J. Reinhart, Lee P. Lim, Christopher B. Burge, Bonnie Bartel, and David P. Bartel. 2002. “Prediction of Plant microRNA Targets.” *Cell* 110 (4): 513–20.

Ronquist, F., and J. P. Huelsenbeck. 2003. “MrBayes 3: Bayesian Phylogenetic Inference Under Mixed Models.” *Bioinformatics* 19 (12): 1572–4. doi:[10.1093/bioinformatics/btg180](https://doi.org/10.1093/bioinformatics/btg180).

Starita, Lea M., Nadav Ahituv, Maitreya J. Dunham, Jacob O. Kitzman, Frederick P. Roth, Georg Seelig, Jay Shendure, and Douglas M. Fowler. 2017. “Variant Interpretation: Functional Assays to the Rescue.” *The American Journal of Human Genetics* 101 (3): 315–25. doi:[10.1016/j.ajhg.2017.07.014](https://doi.org/10.1016/j.ajhg.2017.07.014).

Szemenyei, Heidi, Mike Hannon, and Jeff A. Long. 2008. “TOPLESS Mediates Auxin-Dependent Transcriptional Repression During Arabidopsis Embryogenesis.” *Science (New York, N.Y.)* 319 (5868): 1384–6. doi:[10.1126/science.1151461](https://doi.org/10.1126/science.1151461).

Tan, Xu, Luz Irina A. Calderon-Villalobos, Michal Sharon, Changxue Zheng, Carol V. Robinson, Mark Estelle, and Ning Zheng. 2007. “Mechanism of Auxin Perception by the Tir1 Ubiquitin Ligase.” *Nature* 446 (7136): 640–5. doi:[10.1038/nature05731](https://doi.org/10.1038/nature05731).

Vidal, Elena A, José M Álvarez, and Rodrigo A Gutiérrez. 2014. “Nitrate Regulation of Afb3 and Nac4 Gene Expression in Arabidopsis Roots Depends on Nrt1.1 Nitrate Transport Function.” *Plant Signaling & Behavior* 9 (6): e28501. doi:[10.4161/psb.28501](https://doi.org/10.4161/psb.28501).

Wang, Jia-Wei, Ling-Jian Wang, Ying-Bo Mao, Wen-Juan Cai, Hong-Wei Xue, and Xiao-Ya Chen. 2005. “Control of Root Cap Formation by MicroRNA-Targeted Auxin Response Factors in Arabidopsis.” *The Plant Cell* 17 (8): 2204–16. doi:[10.1105/tpc.105.033076](https://doi.org/10.1105/tpc.105.033076).

Winkler, Martin, Michael Niemeyer, Antje Hellmuth, Philipp Janitza, Gideon Christ, Sophia L. Samodelov, Verona Wilde, et al. 2017. “Variation in Auxin Sensing Guides AUX/IAA Transcriptional Repressor Ubiquitylation and Destruction.” *Nature Communications* 8 (June): 15706. doi:[10.1038/ncomms15706](https://doi.org/10.1038/ncomms15706).

Wright, Erik S. 2015. “DECIPHER: Harnessing Local Sequence Context to Improve Protein Multiple Sequence Alignment.” *BMC Bioinformatics* 16 (October): 322. doi:[10.1186/s12859-015-0749-z](https://doi.org/10.1186/s12859-015-0749-z).

Wright, R. Clay, Mollye L. Zahler, Stacey R. Gerben, and Jennifer L. Nemhauser. 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): 583–91. doi:[10.1534/genetics.117.300092](https://doi.org/10.1534/genetics.117.300092).

Wu, Miin-Feng, Nobutoshi Yamaguchi, Jun Xiao, Bastiaan Bargmann, Mark Estelle, Yi Sang, and Doris Wagner. 2015. “Auxin-Regulated Chromatin Switch Directs Acquisition of Flower Primordium Founder Fate.” *eLife* 4 (October): e09269. doi:[10.7554/eLife.09269](https://doi.org/10.7554/eLife.09269).

Yang, Xiaoqing, Sungsu Lee, Jai-Hyun So, Suni Dharmasiri, Nihal Dharmasiri, Lei Ge, Carolyn Jensen, Roger Hangarter, Lawrence Hobbie, and Mark Estelle. 2004. “The Iaa1 Protein Is Encoded by Axr5 and Is a Substrate of SCF(TIR1).” *The Plant Journal: For Cell and Molecular Biology* 40 (5): 772–82. doi:[10.1111/j.1365-313X.2004.02254.x](https://doi.org/10.1111/j.1365-313X.2004.02254.x).

Yu, Hong, Yi Zhang, Britney L. Moss, Bastiaan O. R. Bargmann, Renhou Wang, Michael Prigge, Jennifer L. Nemhauser, and Mark Estelle. 2015. “Untethering the Tir1 Auxin Receptor from the SCF Complex Increases Its Stability and Inhibits Auxin Response.” *Nature Plants* 1 (3): 14030. doi:[10.1038/nplants.2014.30](https://doi.org/10.1038/nplants.2014.30).