Testing flextable

R Clay Wright

7/6/2018

Table of Contents

[List of working titles 2](#_Toc518739382)

[Abstract 2](#_Toc518739383)

[Intro 2](#_Toc518739384)

[Methods 3](#_Toc518739385)

[Nucleotide Diversity Calculation 3](#_Toc518739386)

[Detailed calculaiton simplification. 4](#_Toc518739387)

[Results 5](#_Toc518739388)

[*TIR1/AFB* genes 5](#_Toc518739389)

## knitr::opts\_chunk$set(warning = FALSE, message = FALSE, echo = FALSE, collapse = TRUE, comment = "#>", tidy = TRUE, cache = TRUE, cache.lazy = FALSE, out.width = "90%", fig.height=4, fig.width=7, out.extra='style="margin: auto; display: block; padding-top: 15px;"') #devtools::install\_github("wrightrc/r1001genomes", ref = "auxin-natural-variation") library(ape) library(r1001genomes) library(DT) library(tidyverse) library(magrittr) library(RColorBrewer) library(ggpmisc) library(ggthemes) library(ggseqlogo) library(ggtree) library(reticulate) library(scales) library(viridis) library(ips) library(knitr) #use\_python(python = "/usr/local/bin/python")

# List of working titles

* A comprehensive analysis of natural sequence variation within the Arabidopsis thaliana nuclear auxin signaling pathway
* Add your suggestions

# Abstract

* Motivating, high level statement
* Statement defining the cutting edge of your research area
* A statement of the form: “However, it remains unclear …” that defines what problem you are  
  addressing.
* A statement of the form:“Here, we …” that concisely states your contribution.
* A series of 3?5 statement about your results.
* A concluding statement.

# Intro

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 global 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 and is influencing the evolution of plant genomes. Outside of genome-wide studies however, this valuable 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 that may provide insight into the function and evolution of genes, families and networks. Here, we present a web application and R-package through which plant molecular biologists with little-to-no bioinformatics experience can make use of this rich dataset of genetic variation to formulate hypotheses as to the sequence/function/phenotype relationships determined by the gene, family, or network of their interest. We demonstrate the utility of this tool through comprehensive analysis and identification of potential functional variation in the nuclear auxin signaling pathway.

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, Young, Miller, & Fields, [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 and network evolution (**???**; Delker et al., [2010](#ref-delker_natural_2010)).

Here, we have utilized this tool to analyze the genetic variation in the nuclear auxin signaling pathway and formulate hypotheses regarding the functional implications of this variation and the evolution of the genes and gene families in this pathway. Our analysis also provided further confirmation of much of the existing knowledge of these genes demonstrating the validity of this approach.

# Methods

## 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, modifying it slightly, replacing with , being the number of sequences in the population of the form at location k: subscripts i and k are summed only over unique forms [need to clarify this]

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.

# Results

## *TIR1/AFB* genes

Auxin acts by binding to receptors that in turn target co-repressors for degradation. Auxin receptors (six in the model plant *Arabidopsis thaliana*) evolved through gene duplication and diversification early in the history of vascular plants (**???**). The rate of co-repressor degradation is determined by the identity of both the receptor and co-repressor (**???**), and this rate sets the pace of lateral root development (**???**).

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(Dharmasiri et al., [2005](#ref-dharmasiri_plant_2005); Parry et al., [2009](#ref-parry_complex_2009); Prigge et al., [2016](#ref-prigge_arabidopsis_2016)). Although TIR1/AFBs 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 (**???**). Additionally, the expression of AFB5 is stronly circadian-regulated (**???**) and AFB3 is more highly expressed in the roots in the presence of nitrate, allowing increased lateral root formation (**???**), suggesting more broad environmental regulation of this gene family may exist. Functionally, 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(**???**). This lack of complexation may allow observed high ubiquitous AFB1 accumulation (**???**). 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 (**???**). 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., n.d.). 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.

#> [1] "new genes:"  
#> character(0)

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

Calderón Villalobos, L. I. A., Lee, S., De Oliveira, C., Ivetac, A., Brandt, W., Armitage, L., … Estelle, M. (n.d.). A combinatorial TIR1/AFB-Aux/IAA co-receptor system for differential sensing of auxin. *Nat Chem Biol*, *8*(5), 477–85. <https://doi.org/10.1038/nchembio.926>

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>

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>

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 Genet*, *11*(2), e1004918. <https://doi.org/10.1371/journal.pgen.1004918>

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*, *106*(52), 22540–22545. <https://doi.org/10.1073/pnas.0911967106>

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>

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>