

Prospecting for Green Revolution Genes

Learning Objectives:

- 1) Discover how changes in individual genes produce phenotypic change
- 2) Learn to apply bioinformatics tools to identify groups of related genes and investigate their evolutionary relationships
- 3) Understand that genes often are members of gene families that may arise through gene duplication
- 4) Be able to apply sequence analyses to identify mutations underlying specific phenotypes.
- 5) Understand how selection for specific phenotypes drove the Green Revolution

Pre-lab notes:

1. Please register at least 24 h in advance as a user for the DNA Subway that we will use in this lab. <http://dnasubway.iplantcollaborative.org/>
2. The following exercise was adapted from exercises generously provided by the developers of the DNA Subway: iPlant Genomics in Education Examples – GenomeProspecting <http://gfx.dnalc.org/files/evidence/Worksheets/>, accessed 11/16/11.

The “Green Revolution” that occurred during the 1960s and 1970s was based to a significant extent on the generation by breeders of semi-dwarf varieties of wheat, maize and rice that did not grow as tall as their predecessors, allowing them to divert more resources into building seeds while diminishing stalk breaking due to long stems and heavy seed loads. However, while highly desirable also for other crops, breeders were unable to routinely recreate the semi-dwarf phenotype in other plants – until biotechnology led the way.

The plant hormone gibberellic acid (GA) is known to stimulate plant growth by activating cell division and cell elongation. Consequently, the difference in stem length between short- and tall- growing crop varieties is frequently caused by genomic alterations that perturb plants’ GA signaling pathways. These changes often affect GA concentrations by altering its synthesis, degradation, or delivery to target cells. In other instances, these changes affect GA perception by changing GA receptor function or by altering proteins that are required to transduce the signal following arrival of GA at the receptor.

In 1985, Koorneef and colleagues isolated an X-ray caused, semi-dwarf, GA-insensitive mutant of *Arabidopsis thaliana* and, in 1997, Peng and colleagues isolated the gene underlying this phenotype. This gene was called *Gibberellic Acid Insensitive* or *GAI*; the mutant gene and protein were called *gai* and *gai*, respectively. The *GAI* protein was shown to function in transducing the signal after arrival of GA at its receptor.

Homologues of the *GAI* gene exist in many other plant species, which suggests that *GAI* may also play a similar role in other plants. *GAI* proteins belong to the highly conserved DELLA family of proteins which, in turn, belong to the larger family of GRAS proteins.

DELLA proteins act as transcriptional repressors. While they are not transcription factors themselves, they bind transcription factors to repress transcription. Upon arrival of GA these DELLA proteins are

Exercise III- Yellow Line

tagged for degradation by the process of ubiquitination and subsequently degraded. Once they are degraded the transcription factors they bind are free to regulate transcription—generally activating transcription. Mutations in this group of repressors can cause either a dwarf phenotype or may cause the plant to be taller than normal. How might different mutations in different parts of a repressor protein cause such different responses? We'll investigate this further in our exercise.

In the following exercises, we will search for and analyze members of the *GAI* gene family across a variety of different plant genomes using DNA Subway's Yellow Line which utilizes the TARGeT workflow. TARGeT stands for 'Tree Analysis of Related Genes and Transposons' and was developed in 2009 by Sue Wessler's lab at the University of Georgia to identify matching DNA sequences in different plant species utilizing sequence similarity searches and alignment tools.

Ok, let's get started!

Part I: Identifying Green Revolution Genes

First, we will use the *GAI* gene and *GAI* protein sequences provided in DNA Subway's Yellow Line to query plant genomes to identify *GAI* genes and other related genes and then examine their relatedness. In part II we'll look for mutations in these genes.

- 1) Log into DNA Subway (<http://dnasubway.iplantcollaborative.org/>). If you have not pre-registered, log in as a guest. As a guest you will have full functionality for this exercise—however your project will not be saved for future use.
- 2) You can enter the different Subway lines by clicking the boxes at the left. For this exercise we will use the Yellow Line—Prospect Genomes Using TARGeT.
- 3) So, to begin this exercise, click on the Yellow Line.
- 4) In order to begin our analysis we need to select a sequence source. We have a number of options here. We will all move through the first portion of this exercise together, so we'll all begin with the same sequence
- 5) Click on the Select a Sample Sequence button—indicated with the big red arrow below.

Select Sequence Source *

- Enter a sequence in [FASTA format](#) (max 10kb):

- Select a sample sequence:

GAI Gene [ORF] (Zea mays/Corn)
GAI Gene Rht-1 [ORF] (Triticum aestivum/Wheat)
gai Gene rht-D1b, Mutant [ORF] (Triticum aestivum/Wheat)
GAI Protein (Zea mays/Corn)
GAI Protein (Arabidopsis thaliana/Mouse-ear cress)
GAI Protein (Oryza sativa/Rice)
gai Protein, Mutant (Arabidopsis thaliana/Mouse-ear cress)

Sequence type:

- DNA Protein



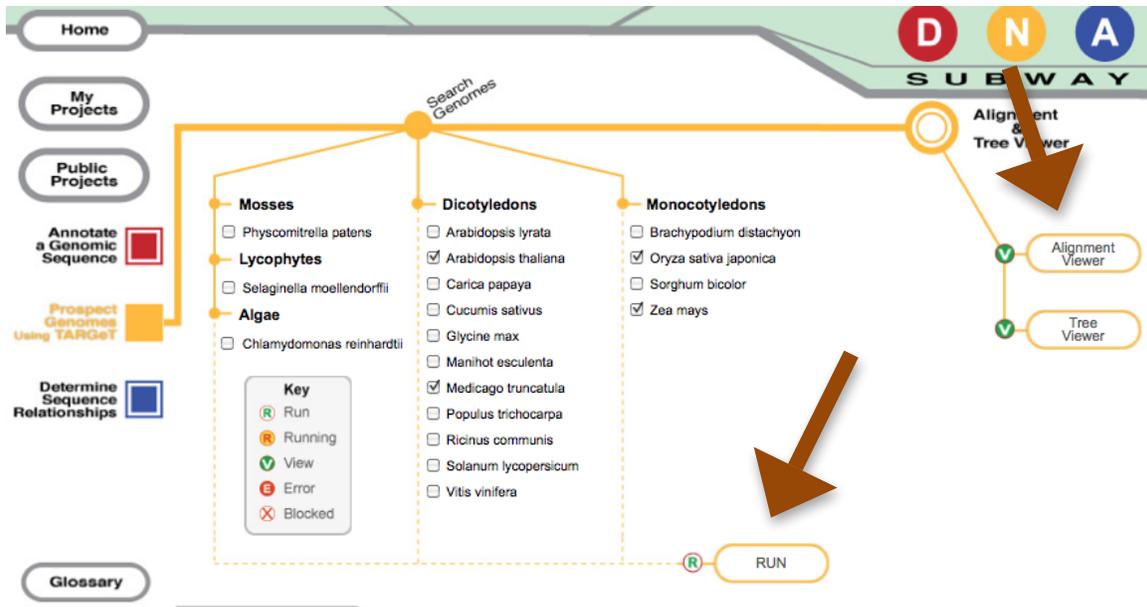
Exercise III- Yellow Line

- 6) Next, scroll down and select the GAI Protein (Arabidopsis thaliana/Mouse-ear cress). Make sure to select the GAI protein, not the GAI gene!
- 7) As you do this, you will see that project information in the right hand column is automatically filled in. You just need to complete the first box under “Name Your Project” by entering a Project title. Projects are saved for an indefinite time for registered users. This includes saving your progress through the project so that, if you stop midway through an analysis, it will be there when you return days later.
- 8) Next, click “Continue”
- 9) You will be taken to a page that shows the map of the full yellow line (shown below).

The screenshot shows the SUBWAY web interface. At the top, there are navigation buttons for Home, My Projects, and Public Projects. The main content area is titled "Search Genomes" and features a yellow line connecting the "Search Genomes" button to the "RUN" button, and then to the "Alignment & Tree Viewer" section. The "Search Genomes" section is divided into three columns: Mosses, Lycophytes, and Algae (all with checkboxes for Physcomitrella patens, Selaginella moellendorffii, and Chlamydomonas reinhardtii); Dicotyledons (with checkboxes for Arabidopsis lyrata, Arabidopsis thaliana, Carica papaya, Cucumis sativus, Glycine max, Manihot esculenta, Medicago truncatula, Populus trichocarpa, Ricinus communis, Solanum lycopersicum, and Vitis vinifera); and Monocotyledons (with checkboxes for Brachypodium distachyon, Oryza sativa japonica, Sorghum bicolor, and Zea mays). A "Key" section indicates that a red 'R' means "Run", a red 'R' with a dot means "Running", a green 'V' means "View", a red 'E' means "Error", and a red 'X' means "Blocked". The "RUN" button is currently blocked. The "Alignment & Tree Viewer" section has two buttons, "Alignment Viewer" and "Tree Viewer", both of which are currently blocked. At the bottom, the "Project Information" section shows details for a project named "module test", including the user "Nicholas Ewing", the organism "Arabidopsis thaliana", the protein name "GAI Protein", the class "GRAS Family Protein", and the function "Transcription Factor".

- 10) Next we will select the genomes to analyze. Select *Arabidopsis thaliana*, *Medicago truncatula* (a type of clover that is a model legume), *Oryza sativa* (rice), and *Zea mays* (corn) and then click “Run”. The figure below shows the location of the “Run” and “Alignment Viewer” buttons.
 - a. The “R” will blink until the analysis is complete.
 - b. In this analysis our sequence is being compared to each of the four genomes in entirety.
 - c. When the analysis is complete, the “R” button stops blinking and the buttons in front of the Alignment Viewer and Tree Viewer display a “V” indicating that they are ready for viewing.
 - d. Click on the “Alignment Viewer” button.

Exercise III- Yellow Line



- 11) After clicking on the “V” of Alignment Viewer, two windows will appear—one is a tree showing the similarity of sequences to one another and the other more colorful one is an alignment of the sequences.
- 12) Let's examine the sequence alignments first. Click on that window to bring it forward.
 - a. Note that identical amino acids are aligned across sequences with the single letter code used to designate amino acids.
 - b. Look at these alignments and notice that dashes are present in lots of positions. The program that does the alignment places these to optimize the positions of regions that match well. If these genes truly are related, these gaps could correspond to sites where deletions have occurred in some sequences or insertions have occurred in others.
 - c. Now look at the colored columns. These often include more than one amino acid. Do the amino acids designated with each color share biochemical characteristics? Go ahead and take a look at the amino acid abbreviations and structures anywhere on the web—such as <http://www.elmhurst.edu/~chm/vchembook/561aminostructure.html>.
 - d. If a single column includes amino acids that share biochemical characteristics are these amino acid substitutions less likely to cause a change in the structure or function of the protein compared to substitutions of very different amino acids?
- 13) Now let's look at the tree. Click on the tree from “Alignment Viewer”.
- 14) Do any patterns appear to you as you look at the tree?
- 15) How many major branches are there?
- 16) Click on an individual sequence in the tree and you will see that the sequence lights up in your alignment. Clicking on a node highlights the group that is connected to that node. Clicking elsewhere brings up a vertical line where you have clicked and highlights groups of sequences that are connected by nodes to the right of your click.

Exercise III- Yellow Line

- 17) Examine the sequences again. Do all of the monocot (corn and rice) sequences group together? Do all of the dicot (Arabidopsis and Medicago) sequences group together? Can you draw any conclusions about the evolutionary origins of the gene in our alignment based on these observations?
- 18) The tree that is shown was calculated from the alignment using the default technique which uses the percent of identical amino acids to generate the tree. By clicking on “calculate” on the tool bar of the alignment (shown below) it is possible to apply different statistics. Let’s try a second method.
- 19) We will calculate a new tree with “Neighbour Joining using BLOSUM62”. This is a more robust calculation that takes into account not only identical amino acids but also those that are similar since these substitutions are less likely to affect protein structure and function.
- 20) Click on “Calculate” to see the pull-down menu (shown below). From this pull-down menu select “Neighbour Joining using BLOSUM62”. A new tree will appear.

The screenshot shows the FAST TRACK TO GENE ANNOTATION AND GENOME ANALYSIS software interface. The 'Calculate' menu is open, showing options like 'Average Distance Using % Identity', 'Neighbour Joining Using % Identity', 'Average Distance Using BLOSUM62', and 'Neighbour Joining Using BLOSUM62'. A red arrow points to the 'Neighbour Joining Using BLOSUM62' option. The main window displays a sequence alignment of GAI protein from various species, including Zea mays, Arabidopsis thaliana, and Medicago truncatula. Below the alignment are conservation and quality plots. At the bottom, there is a 'module test' section with user information and a description of the protein as a GRAS Family Protein and Transcription Factor.

- 21) As you examine your new tree do you see similar overall patterns? Again, how many major branches do you see?
- 22) Refer back to the sequence alignments and examine the first 80 amino acids. Do you see why these proteins have been named DELLA proteins?
- 23) Does the similarity in the first 80 amino acids correspond to major differences in the tree? Go to the tree and click on nodes, between nodes and on individual sequences to investigate this.

Exercise III- Yellow Line

- 24) Where does the sequence we used in our original search fall among these?
- 25) The DELLA and closely related non-DELLA proteins in our alignment are members of a gene family of which each species have multiple members.
- 26) How many members of this gene family does each species appear to have based on our alignment here?
- 27) What processes could allow gene families to come about through evolution?
- 28) The major groups of flowering plants are the dicots (includes *Arabidopsis* and *Medicago* in our tree) and the monocots (includes corn and rice in our tree).
- 29) Considering the ancient divergence of monocots and dicots, what can you conclude about the timing of the evolutionary steps in the generation of the members of this gene family?

Part II. Identifying Mutations in DELLA proteins that affect plant growth and yield

- 1) Return to the start of the DNA Subway Yellow Line to begin a new project.
- 2) Under Select Sequence source select the gai Protein, Mutant (*Arabidopsis thaliana*/Mouse-ear cress).
- 3) Name this new project and continue on.
- 4) Under search genomes—select *Arabidopsis thaliana* only and run the analysis.
- 5) Locate the mutant (gai) protein in the alignment.
- 6) How does the *Arabidopsis* mutant gai protein differ from the other DELLA proteins?
- 7) To visualize the mutation even more clearly, let's try one of the other tools.
- 8) Select the mutant gai protein and one of the other *Arabidopsis* genes encodes a protein that is most similar to gai at its N-terminus.
- 9) From the calculate menu, select pairwise alignment.
- 10) The alignment will pop up—and the nature of the mutation is obvious if you haven't spotted it already.
- 11) How many amino acids long is the deletion in the DELLA region?
- 12) Can we determine the exact length of the deletion in nucleotides based on this information? If so, how long is it?
- 13) The figure below shows the key domains of the DELLA proteins. DELLA proteins are transcriptional repressors. When the hormone GA is absent the GRAS domain binds transcription factors, inactivating them. When GA is present the DELLA domain binds the protein GID1. This binding causes the DELLA protein to be tagged for degradation (using ubiquitination). With DELLA proteins degraded the transcription factors are able to bind promoters and turn on gene expression leading to increased growth of stems.

Exercise III- Yellow Line

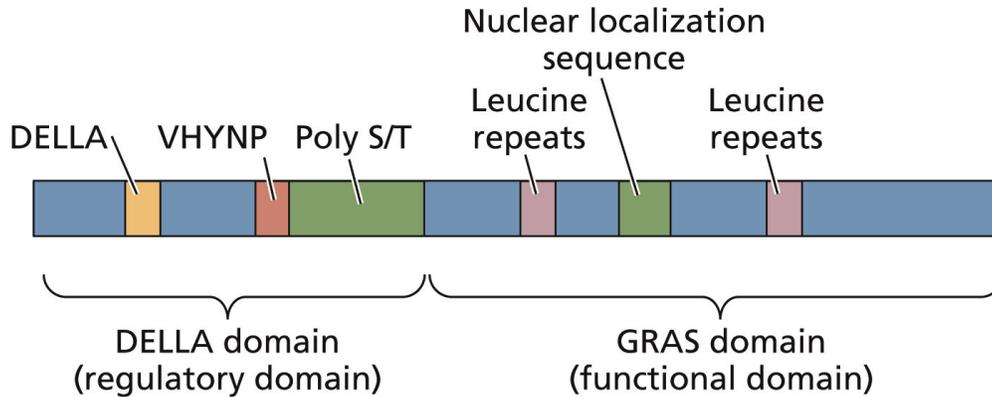


Figure from: Taiz and Zeiger, Plant Physiology, 5e.

- 14) What phenotypic effect would you predict the mutation you have discovered to cause?
- 15) Prepare a diagram that shows how the signal transduction pathway might be affected as a result of this mutation.
- 16) Other mutations in DELLA proteins are known to cause plants to be taller than normal. Propose a mechanism by which changes elsewhere in this protein could cause plants to be taller than normal. Expand the diagram you prepared in the previous question to illustrate this.
- 17) Now, run your own analysis.
- 18) Return to the Yellow Line and see if you can identify the mutation in the *Zea mays* D8-1 gene.
- 19) What phenotype could be expected as a result of this mutation?
- 20) Prepare a diagram that illustrates how this mutation can lead to the predicted phenotype.
- 21) Now, prepare a diagram (or use the one above) to illustrate how this mutation could lead to the opposite mutant phenotype

Readings:

Han, Y., Burnette III, J.M. and Wessler, S.R. (2009). TARGeT: a web-based pipeline for retrieving and characterizing gene and transposable element families from genomic sequences. *Nucl. Acids. Res.* 37:e78

Hedden, P. (2003) The genes of the Green Revolution. *Trends In Genetics* 19:5-9.

Koornneef, M., Elgerma, A., Hanhart, C.J., van Loenen-Martinet, E.P., van Rijn, F. and Zeevaart, J.A.D. (1985). A gibberellin insensitive mutant of *Arabidopsis thaliana*. *Physiol. Plant.* 65:33-39.

Pearce, S, Saville R, Vaughan, S.P., Chandler, P.M., Wilhelm, E.P. Sparks, C.A., Al-Kaff, N., Korolev, A., Boulton, M.I., Phillips, A.L., Hedden, P., Nicholson, P., Thomas S.G. (2011) Molecular Characterization of Rht-1 Dwarfing Genes in Hexaploid Wheat *Plant Physiology* 157: 1820-1831.

Exercise III- Yellow Line

Peng, J. and Harberd, N.P. (1993). Derivative Alleles of the Arabidopsis Gibberellin-Insensitive (*gai*) Mutation Confer a Wild-Type Phenotype. *The Plant Cell* 5:351-360.

Peng, J., Richards, D., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., Sudhakar, D., Christou, P., Snape, J.W., Gale, M.D. and Harberd, N.P. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400:256-261.

Peng, J., Carol, P., Richards, D.E., King, K.E., Cowling, R.J., Murphy, P.M. and Harberd, N.P. (1997). The *Arabidopsis GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. *Gene & Development* 11:3193-3205.

Sánchez-Fernández, R., Ardiles-Díaz, W., Van Montagu, M., Inzé, D. and May, M.J. (1998). Cloning of a novel *Arabidopsis thaliana* RGA-like gene, a putative member of the VIID-domain transcription factor family. *J. Exp. Bot.* 49:1609-1610.

Wu, J., Kong, X., Wan, J., Liu, X., Zhang, X., Guo, X., Zhou, R., Zhao, G., Jing, R., Fu, X., Jia, J. (2011) Dominant and pleiotropic effects of a *GAI* gene in wheat results from a lack of interaction between *DELLA* and *GID1*. *Plant Physiol.* 157: 2120–2130,