## Introduction

## The genus Brassica

## riangle_of_U_sposato.pngThe genus Brassica consists of over thirty wild species and hybrids or morphotypes. Generally, species from the genus Brassica are used in foods like broccoli, cauliflower, cabbage, mustard and more. The Brassica genome has undergone more polyploidy events than Arabidopsis thaliana. Arabidopsis thaliana is notable for being a model organism because of its complexity paired with the fact that it has a relatively small genome. The Brassica genome has undergone two tetraploidy and two hexaploidy events, one more than Arabidopsis, since the eudicot paleohexaploidy event, which gave rise to Vitis, Prunus, Arabidopsis, and Brassica.

## The “Triangle of U” theory describes the genetic relationship between six species of Brassica: Brassica rapa, Brassica nigra, Brassica oleracea, Brassica juncea, Brassica carinata, and Brassica napus. B. juncea, B. carinata, and B. napus are all allotetraploids, hybrids with four times the chromosome set of haploids.

## UGT Gene Family

## UGT functions

## Uridine diphosphate (UDP) glycosyltransferases (UGTs) mediate transfer of glycosyl residues from activated nucleotide sugars to acceptor molecules ([Tang, Unleashing the Genome of the Brassica rapa](http://journal.frontiersin.org/article/10.3389/fpls.2012.00172/abstract)). They provide instructions for making enzymes that perform the process of glucuronidation, the addition of glucuronic acid to a substrate ([Genetics Home Reference, UGT gene family](http://ghr.nlm.nih.gov/geneFamily/ugt)). This pathway is particularly important in metabolism, and many regard the UGT enzyme as the most important enzyme in the pathway. In humans, these enzymes are responsible for the breakdown of several prescription drugs and pollutants.

## UGT Chemistry

## By mediating transfer of glycosyl residues from activated nucleotide sugars to acceptor molecules, UGTs regulate properties of those acceptors such as bioactivity, solubility and transport within cells and throughout organisms ([Ross, Higher plant glycosyltransferases](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC138907/)). The UGT enzymes in the metabolic process of glucuronidation, a very common process in Phase II metabolism.

## preliminary treeMethods

Figure . preliminary tree of Arabidopsis thaliana Glycosyltransferase Family 1

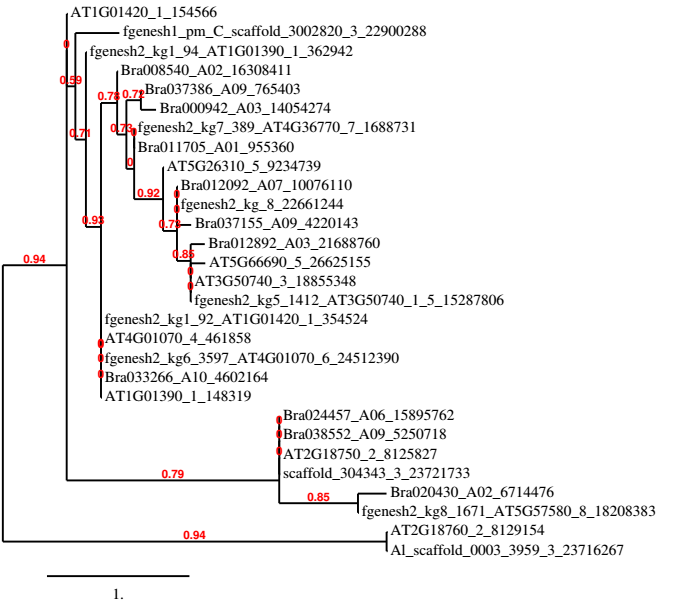
## Of the 28 Glycosyltransferase Families that The Arabidopsis Information Resource (TAIR) has on Arabidopsis thaliana, I chose to work with Family 1 due to the fact that most of the genes were similar in function as the TIGR Annotation suggested. A preliminary tree was constructed of 122 of these genes (those that CoGe had FASTA sequences for). With use of Keiko Yonekura-Sakakibara’s Functional genomics of family 1 glycosyltransferases in Arabidopsis, I began identifying several functions of the genes. We identified three clusters on the tree that we named Test groups. Test group 1 consists of AT2G36750, AT2G36760, AT2G36770, AT2G36780, AT2G36790, and AT2G36800. These genes exhibit functions of flavonol 7-*O*-glucosyltransferase and brassinosteroid *O*-glucosyltransferase. Test group 2 consists of AT3G21780, AT4G15720, AT4G15260, AT4G15280, AT3G21750, and AT3G21760. These genes exhibit the function of having ABA glucosyltransferase activity. Test group 3 consists of AT4G01070, AT1G01420, AT1G01390, AT3G50740, AT5G66690, AT5G26310, AT2G18570, AT2G18560, AT4G36670. These genes exhibit the function of monolignol 4-O-glycosyltransferase activity and having xeniobiotic glycosyltransferase activity. Each of these test groups genes’ FASTAs were placed into one file. Using GeVo, each *Arabidopsis thaliana* gene was visualized for syntenic regions in *Arabidopsis lyrata* and *Brassica rapa*. The FASTAs for those genes in were added to the file and the gevolinks were saved to a separate files for each test group. Using phylogeny.fr, new trees were constructed with *Arabidopsis thaliana, Arabidopsis lyrata,* and *Brassica rapa* for each test group.

## 

Tree 1. for test group 1

## 

Tree 2. for test group 2



Tree 3. for test group 3

**Links to Git Hub**

**Compose the formal ReadMe**

**Hypotheses**

According to how *Arabidopsis lyrata* and *Brassica rapa* diverged from *Arabidopsis thaliana*, and which duplication events occurred in that time, a ratio exists that would explain the UGT genes for each species if there were no losses of genes over time. That perfect ratio looks something like this.

|  |  |
| --- | --- |
| Arabidopsis thaliana: Arabidopsis lyrata | 1 : 1 |
| Arabidopsis thaliana: Brassica rapa | 1 : 3 |

However, simply looking from the data we have already retrieved from CoGe, Phytozome, and BRAD, those ratios are not what we see.