



Brassica Oleracea Database



Aaron Trautman ; Luz Hernandez



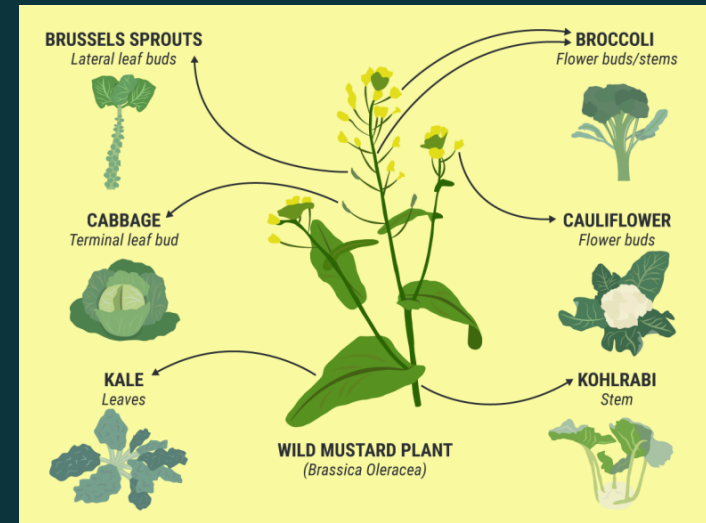
Background

Brassica oleracea comes from the mustard family

Many cultivated plants were developed from *B. oleracea*

- Cabbage, Kale, Collard Greens, Brussel Sprouts, Broccoli, Kohlrabi ...

High in many compounds beneficial to human health



Database Contents

Gene Expression Omnibus (GEO)

- Tissue expression data
- 7 tissue types

Kyoto Encyclopedia of Genes and Genomes (KEGG)

- Pathway data
- Gene data

Connect the two datasets through BLAST

- Blast GEO read data against Arabidopsis thaliana genome from KEGG

kegg_pathway_description	
kegg_path_id	varchar
path_description	varchar

entrez_to_pathway	
kegg_path_id	varchar
entrez_gene_id	varchar

ncbi_to_entrez	
ncbi_gene_id	numeric
entrez_gene_id	varchar

parsed_gene_ids	
genomic_id	varchar
ncbi_gene_id	numeric

arabidopsis_genes_to_brassica_read_counts	
genomic_id	Varchar
description	Varchar
bud_count	numeric
callus_count	numeric
flower_count	numeric
leaf_count	numeric
root_count	numeric
siliqua_count	numeric
stem_count	numeric
avg_read_count	numeric
sd_read_count	numeric

What can this database tell us?



Q1. Biotin is needed for cell development, the production of fatty acids, and the break down of amino acids and fats. Biotin deficiency is very rare but possible. Supplements are used to treat nerve pain associated with diabetes to hair loss.

What genes are associated with biotin metabolism? What tissues have the highest expression?

Q1 Results

38 genes associated with biotin metabolism

Expression total as measured by read counts:

21 genes > 1 standard deviation above the mean

7 genes > 2 standard deviations above the mean

The above reads came from the bud, leaf, root, silique, and the stem. 18 of the highly expressed genes came from the bud.

What can this database tell us?

Q2. Leaf tissues are known to be high in beneficial compounds, and have been researched for potential health benefits. Some Native American tribes even use them to treat several conditions.

What genes in *B. oleracea* are more highly expressed within the leaf tissue than any others and which pathways do these genes affect?

Q2 results

1074 genes more highly expressed in the leaf tissues than any other tissue

83 pathways (20 pathways involved in compound biosynthesis)

- Pantothenate biosynthesis (Vitamin B5)
- Glucosinolate biosynthesis (anti-cancerous compounds)
- Diterpenoid biosynthesis (anti-inflammatory, anti-microbial)
- Stilbenoid, diarylheptanoid and gingerol biosynthesis (anti-oxidants)

Difficulties and Limitations

- The ID's that NCBI BLAST tool gives each result is not easily linked back to a gene id.
 - Used a simple python script to pull NCBI gene id's from the website
- KEGG database can be difficult to access
- KEGG is not comprehensive, NCBI contains more genes
- Not all BLAST matches were linked to the pathways (13000/21000)
- BLAST *B. oleracea* reads against *A. thaliana* genome

What would we do different?

- Compare *Brassica oleracea* reads against *B. oleracea* genome
 - Would those results differ from our results
- Completely map all genes to all pathways
- Use a different database to compare results, and identify other genes/pathways

Thank you!

