**ARABIDOPSIS THALIANA**

**Arabidopsis Dataset Contained :**

* Samples : 1307
* SNPs : 214051
* Genotype Data Format : .ped
* Phenotypes format : .pheno

## **Phenotypes (Consider Number of Phenotypes : 4)**

|  |  |  |
| --- | --- | --- |
| **Abbreviation** | **Traits** | **Description** |
| **FLC** | Flowering Locus C | RNA was extracted from leaves after 4 wks of growth. FLC gene expression levels were determined by Northern hybridization quantified relative to TUBULIN expression |
| **FRI** | FRIGIDA | RNA was extracted from leaves after 4 wks of growth. FRI gene expression levels were determined by Northern hybridization quantified relative to TUBULIN expression |
| **RP\_GH** | Days to silk | Reproduction period: number of days between the appearance of the first flower and the plant complete senescence |
| **Anthocyanin\_22** | **Anthocyanin** | Results expressed as binary data, determined by the presence (1) or absence (0) of anthocyanin in all 4 plants / accession after 5wks of growth |

**Quantile P-vals:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Phenotype Name | 0% | 25% | 50% | 75% | 100% |
| Anthocyanin\_22 | 1.908469e-39 | 2.336407e-01 | 5.454623e-01 | 8.609825e-01 | 1.000000e+00 |
| FLC | 1.167325e-103 | 9.401442e-02 | 3.172489e-01 | 6.332525e-01 | 9.999924e-01 |
| FRI | 3.200037e-30 | 1.327477e-01 | 3.784224e-01 | 6.754587e-01 | 9.999944e-01 |
| RP\_GH | 6.965124e-17 | 2.198048e-01 | 4.771444e-01 | 7.392130e-01 | 9.999951e-01 |

**Significance Level**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Phenotype Name** | **0.10** | **0.05** | **0.01** | **0.001** | **0.0001** | **1e-5** | **1e-6** |
| **Anthocyanin\_22** | 26502 | 15589 | 4579 | 728 | **118** | - | - |
| FLC | 55401 | 37168 | 14894 | 4118 | 1190 | 367 | **121** |
| FRI | 44493 | 28863 | 11406 | 3499 | 1163 | 392 | **117** |
| RP\_GH | 27310 | 15368 | 4186 | 638 | **99** | - | - |

**Elapsed Time : (Highlighted is the total Execution time of whole Model To Complete)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Chain Number** | **Anthocyanin\_22** | **FLC** | **FRI** | **RP\_GH** |
| **1** | 4727.55 | 4949.49 | 4930.01 | 4220.4 |
| **2** | 4728.4 | **5031.15** | **4936.63** | 4160.25 |
| **3** | 3327.09 | 5027.14 | 4919.92 | **4247.02** |
| **4** | **4745.5** | 5018.28 | 4885.47 | 4237.56 |

**ANTHOCYANIN\_22**

|  |  |  |  |
| --- | --- | --- | --- |
| **SNP** | **Chromosome\_BP** | **Gene** | **Description** |
| SNP27317.1 | Chr1\_16924749 | AT1G44810 | DNA-binding storekeeper protein-related transcriptional regulator |
| SNP28067.1 | Chr1\_17338475 | AT1G47290 | 3beta-hydroxysteroid-dehydrogenase/decarboxylase isoform 1 |
| SNP66457.2 | Chr2\_9515219 | AT2G22425 | Microsomal signal peptidase 12 kDa subunit (SPC12) |
| SNP27381.2 | Chr1\_16976236 | AT1G44910 | pre-mRNA-processing protein 40A : Binds the carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II and functions as a scaffold for RNA processing machineries. |
| SNP25135.1 | Chr1\_14067286 | AT1G37037 | transposable\_element\_gene |
| SNP60540.1 | Chr2\_5558419 | No Gene | -- |
| SNP79487.1 | Chr2\_18986066 | No Gene | -- |
| SNP165493.1 | Chr5\_2602862 | AT5G08120 | Movement protein binding protein 2C : Microtubule-associated and viral movement protein binding protein. Negatively regulates KN1 association to plasmodesmata and  consequently  cell-to-cell transport. Involved in the alignment of cortical microtubules  the patterning of stomata and in restricting tobamoviral infections. |
| SNP165447.1 | Chr5\_2575329 | AT5G08030 | protein\_coding : Encodes a member of the glycerophosphodiester phosphodiesterase (GDPD) family. Phosphate inducible. Mutants show defects in root growth under phosphate limiting conditions. |
| SNP27368.1 | Chr1\_16970635 | AT1G44900 | minichromosome maintenance (MCM2/3/5) family protein : Encodes MCM2 (MINICHROMOSOME MAINTENANCE 2)  a protein essential to embryo development. Overexpression results in altered root meristem function. |

**FLC – Flowering locus C**

|  |  |  |  |
| --- | --- | --- | --- |
| **SNP** | **Chromosome\_BP** | **Gene** | **Description** |
| SNP125050 | Chr4\_519701 | AT4G01240 | S-adenosyl-L-methionine-dependent methyltransferases superfamily protein |
| SNP177934 | Chr5\_8781889 | AT5G25300 | F-box protein |
| SNP177086 | Chr5\_8433211 | AT5G24630 | double-stranded DNA binding: This gene is predicted to encode a protein that forms part of the topoisomerase VI complex. BIN4 is a nuclear-localized protein that can bind DNA. bin4 mutants are brassinolide-insensitive dwarves with severely reduced cell size in leaves roots and hypocotyls. Proper development of root hairs and trichomes is also disrupted in bin4 mutants and they have elevated levels of double strand breaks in their cotyledon cells. |
| SNP71819 | Chr2\_13683819 | AT2G32235 | protein\_coding |
| SNP210123 | Chr5\_24769497 | No Gene | -- |
| SNP95665 | Chr3\_8504155 | AT2G19690 | phospholipase A2-beta : Encodes one of the four Arabidopsis phospholipase PLA2 parologs: AT2G06925 (PLA2-ALPHA)  AT2G19690 (PLA2-BETA)  AT4G29460 (PLA2-GAMMA) and AT4G29470 (PLA2-DELTA). Involved in pollen development and germination and tube growth. Also involved in stomatal opening in response to light. |
| SNP174636 | Chr5\_7380507 | AT5G22300 | nitrilase 4 : encodes a nitrilase isomer. The purified enzyme shows a strong substrate specificity for beta-cyano-L-alanine  a intermediate product of the cyanide detoxification pathway. The mRNA is cell-to-cell mobile. |
| SNP153262 | Chr4\_13686358 | No Gene | -- |
| SNP177127 | Chr5\_8453078 | No Gene | -- |
| SNP97040 | Chr3\_9156792 | AT2G21390 | Coatomer subunit alpha-2. Part of endomembrane trafficking system. Interacts with SINAT1. |

**FRI - Frigida**

|  |  |  |  |
| --- | --- | --- | --- |
| **SNP** | **Chromosome\_BP** | **Gene** | **Description** |
| SNP9838 | Chr1\_5774855 | AT1G16880 / [SALKseq\_083319.1](https://www.arabidopsis.org/servlets/TairObject?type=polyallele&name=SALKseq_083319.1) | Encodes a ACT domain-containing protein. The ACT domain named after bacterial aspartate kinase chorismate mutase and TyrA (prephenate dehydrogenase)  is a regulatory domain that serves as an amino acid-binding site in feedback-regulated amino acid metabolic enzymes. The mRNA is cell-to-cell mobile. |
| SNP23374 | Chr1\_13065536 | No Gene | -- |
| SNP71492 | Chr2\_13417171 | AT2G31510 | IBR domain-containing protein |
| SNP92800 | Chr3\_7017951 | AT3G20090 | protein\_coding : Full\_name cytochrome P450 |
| SNP143372 | Chr4\_9041080 | AT4G15950 | Non-catalytic subunit common to Nuclear DNA-dependent RNA polymerases IV and V homologous to budding yeast RPB4. Role in gene silencing. Required for RNA-directed DNA methylation. |
| SNP23374 | Chr1\_13065536 | No Gene | -- |
| SNP86716 | Chr3\_3656161 | No Gene | -- |
| SNP85962 | Chr3\_3311425 | [SALKseq\_065316.1](https://www.arabidopsis.org/servlets/TairObject?type=polyallele&name=SALKseq_065316.1) | This genomic-terminal sequence of a TDNA insertion region lies within the Intergenic between AT3G10590.1 and AT3G10595.1=UTR3 (at chr 3 pos 3311410 (C/3311410-3311445) on the TAIR10). The sequences' 1-35 bps mapped on the genome |
| SNP39726 | Chr1\_23665978 | AT1G63780 | Ribosomal RNA processing Brix domain protein : Small nucleolar ribonucleoprotein protein involved in ribosomal RNA processing. Located in nucleolus and cajal bodies. |
| SNP166557 | Chr5\_3280283 | No Gene | -- |

[SALKseq\_083319.1](https://www.arabidopsis.org/servlets/TairObject?type=polyallele&name=SALKseq_083319.1) :

This genomic-terminal sequence of a TDNA insertion region lies within the CDS of AT1G16880.1 (at chr 1 pos 5774857 (C/5774857-5774919) on the TAIR10). The sequences' 1-63 bps mapped on the genome 61-78 bps onto the T-DNA. Pools: P 73

R 77

C 72 N 80. The insertion site is at the 3' end of the sequence mapping.

**RP\_GH -** Days to silk

|  |  |  |  |
| --- | --- | --- | --- |
| **SNP** | **Chromosome\_BP** | **Gene** | **Description** |
| SNP162169 | Chr5\_612228 | AT5G02710/ SALKseq\_088774.8 | zinc/iron-chelating domain protein : |
| SNP117645 | Chr3\_19826356 | AT3G53480 | pleiotropic drug resistance : Negative regulator of auxin polar transport inhibitors. ABCG37 regulates auxin distribution and homeostasis in roots by excluding IBA from the root apex but does not act directly in basipetal transport. ABCG37 and ABCG36 act redundantly at outermost root plasma membranes and  transport IBA out of the cells. Also involved in root transmembrane secretion of fluorescent phenolics involved in Fe uptake. The mRNA is cell-to-cell mobile. |
| SNP198334 | Chr5\_18576148 | AT5G45800 | Leucine-rich repeat protein kinase family protein |
| SNP35181 | Chr1\_21696736 | AT1G58400 | Disease resistance protein (CC-NBS-LRR class) family |
| SNP119613 | Chr3\_21010312 | No Gene | -- |
| SNP168399 | Chr5\_4205302 | AT5G13190 | GSH-induced LITAF domain protein |
| SNP98775 | Chr3\_9847527 | No Gene | -- |
| SNP128113 | Chr4\_1891993 | No Gene | -- |
| SNP45627 | Chr1\_26406589 | AT1G70110 | Concanavalin A-like lectin protein kinase family protein |
| SNP115096 | Chr3\_18356294 | AT1G49590 | C2H2 and C2HC zinc fingers superfamily protein |

SALKseq\_088774.8 : transposable\_element\_insertion\_site , This genomic-terminal sequence of a TDNA insertion region lies within the Intergenic between AT5G02700.1 and AT5G02710.1=P355 (at chr 5 pos 612212 (C/612212-612277) on the TAIR10). The sequences' 1-66 bps mapped on the genome75-95 bps onto the T-DNA. Pools: P 76R 80C 80 N 78. The insertion site is at the 3' end of the sequence mapping.

REFERENCES : <https://gbrowse.arabidopsis.org/cgi-bin/gb2/gbrowse/arabidopsis/>

**TASSEL MLM MODEL**

Anthocyanin\_22

|  |  |  |
| --- | --- | --- |
| **Chromosome\_BP** | **Gene** | **Description** |
| Chr1\_12663142 | AT1G34580 | Major facilitator superfamily protein |
| Chr1\_16930622 | No Gene | -- |
| Chr3\_6414091 | AT3G18640 | Zinc finger C-x8-C-x5-C-x3-H type family protein |
| Chr3\_18694551 | AT3G50380 | vacuolar protein sorting-associated protein : putative (DUF1162) |
| Chr5\_22340810 | No Gene | -- |
| Chr1\_14028301 | AT1G36990 | C-jun-amino-terminal kinase-interacting protein |
| Chr1\_14053491 | AT1G37030 | transposable\_element\_genenon-LTR retrotransposon family (LINE)  has a 5.1e-17 P-value blast match to GB:NP\_038605 L1 repeat  Tf subfamily  member 30 (LINE-element) (Mus musculus)(source:TAIR10) |
| Chr1\_14142053 | No Gene | -- |
| Chr1\_19430888 | AT5G08030 | protein\_coding : Encodes a member of the glycerophosphodiester phosphodiesterase (GDPD) family. Phosphate inducible. Mutants show defects in root growth under phosphate limiting conditions. |
| Chr3\_11738155 | AT3G29830 | protein\_coding - F-box/RNI-like superfamily protein |

FLC

|  |  |  |
| --- | --- | --- |
| **Chromosome\_BP** | **Gene** | **Description** |
| "Chr5\_10172992" | No Gene | -- |
| "Chr1\_6148689" | AT1G32460 | hypothetical protein |
| "Chr1\_4128051" | AT1G17850 | Rhodanese/Cell cycle control phosphatase superfamily protein |
| "Chr1\_1493521" | AT1G05170 | Galactosyltransferase family protein : 3-galactosyltransferase that acts redundantly with CAGE2 during primary and secondary cell wall biosynthesis. |
| "Chr1\_1896873" | AT1G06200 | Peptidase S24/S26A/S26B/S26C family protein |
| "Chr2\_221352" | AT2G01490 | phytanoyl-CoA dioxygenase (PhyH) family protein : Encodes a phytanoyl-CoA 2-hydroxylase (PAHX). The mRNA is cell-to-cell mobile. |
| "Chr2\_5084182" | No Gene | -- |
| "Chr1\_11266302" | AT1G31480 | shoot gravitropism 2 (SGR2) : encodes a novel protein that may be part of a gene family represented by bovine phosphatidic acid-preferring phospholipase A1 (PA-PLA1)containing a putative transmembrane domain. SGR2 is involved in the formation and function of the vacuole. |
| "Chr3\_10660281" | AT3G28430 | CLEC16A-like protein : Encodes a peripheral membrane protein localized at the Golgi apparatus that is involved in membrane trafficking  vacuole development and in flavonoid accumulation in the seed coat. Mutant seed color is pale brown. |
| "Chr1\_9752585" | AT1G27980 | dihydrosphingosine phosphate lyase : Encodes an ER-localized sphingoid long-chain base-1-phosphate lyase involved in the dehydration stress response. |

FRI

|  |  |  |
| --- | --- | --- |
| **Chromosome\_BP** | **Gene** | **Description** |
| "Chr4\_268990" | AT4G00650 | FRIGIDA-like protein : Encodes a major determinant of natural variation in Arabidopsis flowering time. Dominant alleles of FRI confer a vernalization requirement causing plants to overwinter vegetatively. Many early flowering accessions carry loss-of-function fri alleles .Twenty distinct haplotypes that contain non-functional FRI alleles have been identified and the distribution analyzed in over 190 accessions. The common lab strains- Col and Ler each carry loss of function mutations in FRI. |
| "Chr4\_276143" | AT4G00660 | RNAhelicase-like 8 : Immunoprecipitation with mRNA decapping complex members DCP1 and DCP2. |
| "Chr4\_275349" | AT4G00660 | Immunoprecipitation with mRNA decapping complex members DCP1 and DCP2. |
| "Chr4\_244182" | AT4G00570 | NAD-dependent malic enzyme 2 : Encodes an NAD-dependent malic enzyme (NAD-ME) that does not act on oxaloacetate indicating that it belongs to EC 1.1.1.39. It is a member of the beta family of NAD-MEs in plants. It appears to function as a homodimer or as a heterodimer with the alpha-type NAD-ME2 (At2g13560). NAD-ME2 transcript and protein levels are higher during the night than during the day. |
| "Chr4\_280202" | AT4G00680 | actin depolymerizing factor 8 : Actin-depolymerizing factor (ADF) expressed specifically in root hairs high thermostability. |
| "Chr4\_269260" | AT4G00650 | FRIGIDA-like protein : Encodes a major determinant of natural variation in Arabidopsis flowering time. Dominant alleles of FRI confer a vernalization requirement causing plants to overwinter vegetatively. Many early flowering accessions carry loss-of-function fri alleles .Twenty distinct haplotypes that contain non-functional FRI alleles have been identified and the distribution analyzed in over 190 accessions. The common lab strains- Col and Ler each carry loss of function mutations in FRI. |
| "Chr4\_248566" | AT4G00580 | COP1-interacting protein-related |
| "Chr4\_244954" | AT4G00570 | Encodes an NAD-dependent malic enzyme (NAD-ME) that does not act on oxaloacetate  indicating that it belongs to EC 1.1.1.39. It is a member of the beta family of NAD-MEs in plants. It appears to function as a homodimer or as a heterodimer with the alpha-type NAD-ME2 (At2g13560). NAD-ME2 transcript and protein levels are higher during the night than during the day. |
| "Chr3\_6479859" | AT3G18800 | transmembrane protein |
| "Chr4\_308466" | AT4G00740 | S-adenosyl-L-methionine-dependent methyltransferases superfamily protein : Encodes a Golgi-localized type II membrane pectin methyltransferase regulating cell wall biosynthesis in suspension cells. |

RP\_GH

|  |  |  |
| --- | --- | --- |
| **Chromosome\_BP** | **Gene** | **Description** |
| **"Chr1\_2535598"** | AT1G08110 | DNA-binding storekeeper protein-related transcriptional regulator |
| "Chr1\_2546868" | AT1G08130 | Encodes the Arabidopsis DNA ligase 1 that provides the major DNA ligase activity in cells and plays a key role in both DNA replication and excision repair pathways. In addition  it is an important component of the active DNA demethylation machinery and is indispensable for cell viability. AtLIG1 expresses one major and two minor mRNA transcripts differing only in the length of the 5' untranslated leader sequences preceding a common ORF. Translation from the first in-frame start codon produces an AtLIG1 isoform that is targeted exclusively to the mitochondria. Translation initiation from the second in-frame start codon produces an AtLIG1 isoform targeted only to th e nucleus. |
| "Chr1\_2630539" | No Gene | -- |
| "Chr1\_2776433" | AT1G44910 | Protein kinase superfamily protein : enhanced disease resistance 1 (EDR1) confers resistance to powdery mildew disease caused by the fungus Erysiphe cichoracearum The mRNA is cell-to-cell mobile. |
| "Chr1\_2829344" | AT1G08840 | DNA replication helicase : Encodes a homolog of human and yeast DNA2. Mutants have increased sensitivity to DNA damage stress. |
| "Chr1\_2836038" | AT1G08840 | DNA replication helicase : Encodes a homolog of human and yeast DNA2. Mutants have increased sensitivity to DNA damage stress. |
| "Chr1\_2840166" | AT1G08845 | Ribosomal L18p/L5e family protein : Member of the uL18 RNA-binding protein family. uL18 proteins share a short structurally conserved domain that binds the 5S rRNA and allow its incorporation into ribosomes. |
| "Chr1\_2881735" | AT1G08960 | cation exchanger 11 : Encodes a member of the Potassium-dependent sodium-calcium exchanger like-family that localizes to the plasma membrane and nuclear periphery and has a role in mediating high-afﬁnity K+ uptake and Na+ transport in yeast. |
| "Chr1\_3190241" | AT1G09820 | Pentatricopeptide repeat (PPR-like) superfamily protein |
| "Chr1\_9481923" | AT1G27290 | transmembrane protein |

Error Rates

[1] "\nRMSE results: 4.85730351476308"

[1] "\nR-squared results: 0.292841384663739"

Error rate for: FRI

Error rate for fit nuts:[1] "\nMSE results: 0.261615792103355"

[1] "\nRMSE results: 0.511483911871483"

[1] "\nR-squared results: 0.358578733947313"

Error rate for: FLC

Error rate for fit nuts:[1] "\nMSE results: 0.312801938817856"

[1] "\nRMSE results: 0.559286991461321"

[1] "\nR-squared results: 0.643040447867527"