A Practical Guide to
The Arabidopsis Information Resource

Part II

Ho-Ming Chen

Agricultural Biotechnology Research Center

Oct 8, 2019
Outline

Part I: Single gene information  (50 mins)

Break  (10 mins)

Part II: Tools and Genome-wide analysis  (50 mins)
Local alignment search tools @ TAIR

- search homologs
- map transcripts,
- examine the specificity of primers
- .....
BLAST program

Protein search is more sensitive than nucleotide search
Use BLASTP or BLASTX for searching homologs
BLAST datasets
Filter query: on or off

- Filtering masks off segments of the query sequence that have low compositional complexity such as “poly A” or “CT repeat”
- Speed up the program and eliminate statistically significant but biologically uninteresting reports

Query: 61 caaaaatttacttctgaaatctcttttgatctttcgggaaaaaaaaaaaaattcctaaagtaaggga 120

Subject: 62 caaaaatttacttctgaaatctcttttgatctttcgggaaaaaaaaaaaaattcctaaagtaaggga 121
Expectation value (E-value)

Increase E value for more hits (such as 100)
Decrease E value for highly similar hits (such as 0.0001)

Advanced BLAST™ Parameter Options

- Weight Matrix: blosum62
- Max Scores: 100
- Max Alignments: 50
- Gapped Alignments: ON
- Expectation: 10
- Nucleic Mismatch: -3
- Nucleic Match: 2
- Gap Opening Penalty: 0 (use default)
- Gap Extension Penalty: 0 (use default)
- Extension Threshold: 0 (use default)
- Word Size: 0 (use default)
- Query and Database Genetic Code: Standard or Universal (default)

Increase to 100 if query sequence is short (<20 nt) and expect imperfect hits
BLASTN

BLASTN: nucleotides query nucleotide database

Query= user-submitted sequence
(1662 letters)

Database: TAIR10 Transcripts (-introns, +UTRs) (DNA)
41,671 sequences; 64,867,051 total letters

Searching.................................done

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3295</td>
<td>0.0</td>
</tr>
<tr>
<td>359</td>
<td>7e-98</td>
</tr>
<tr>
<td>359</td>
<td>7e-98</td>
</tr>
<tr>
<td>359</td>
<td>7e-98</td>
</tr>
<tr>
<td>64</td>
<td>6e-09</td>
</tr>
<tr>
<td>64</td>
<td>6e-09</td>
</tr>
<tr>
<td>62</td>
<td>2e-08</td>
</tr>
<tr>
<td>60</td>
<td>9e-08</td>
</tr>
<tr>
<td>52</td>
<td>2e-05</td>
</tr>
<tr>
<td>52</td>
<td>2e-05</td>
</tr>
<tr>
<td>48</td>
<td>4e-04</td>
</tr>
<tr>
<td>48</td>
<td>4e-04</td>
</tr>
<tr>
<td>48</td>
<td>4e-04</td>
</tr>
<tr>
<td>48</td>
<td>4e-04</td>
</tr>
<tr>
<td>40</td>
<td>4e-04</td>
</tr>
<tr>
<td>44</td>
<td>0.006</td>
</tr>
<tr>
<td>44</td>
<td>0.006</td>
</tr>
<tr>
<td>44</td>
<td>0.006</td>
</tr>
<tr>
<td>44</td>
<td>0.006</td>
</tr>
</tbody>
</table>

15 hits

0.001
## BLASTX

BLASTX: nucleotides query amino acid database

### Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Score (bits)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5e-35</td>
<td></td>
</tr>
</tbody>
</table>

### Example Output:

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
<th>Score (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT3291441.1</td>
<td>Symbols: ATYB366, MYB65</td>
<td>1115 0.0</td>
</tr>
<tr>
<td>AT3291442.1</td>
<td>Symbols: MYB33, ATYB33</td>
<td>514 e-167</td>
</tr>
<tr>
<td>AT3291443.1</td>
<td>Symbols: MYB33, ATYB33</td>
<td>514 e-167</td>
</tr>
<tr>
<td>AT3291444.1</td>
<td>Symbols: MYB33, ATYB33</td>
<td>514 e-167</td>
</tr>
<tr>
<td>AT3291445.1</td>
<td>Symbols: MYB33, ATYB33</td>
<td>514 e-167</td>
</tr>
<tr>
<td>AT3291446.1</td>
<td>Symbols: MYB33, ATYB33</td>
<td>514 e-167</td>
</tr>
</tbody>
</table>

> 50 hits

---

**5e-35**
Drawbacks of sequence search @TAIR

For mapping sequence to the genome

• Need multiple searches if don’t know where the query sequence is produced

• Difficult for the comparison of gene structure
Graph View of Alignment to Genome @ Salk

http://signal.salk.edu/cgi-bin/atta

Increase E-value to 10 if query sequence is short

<table>
<thead>
<tr>
<th>Program</th>
<th>blastn, DNA</th>
<th>E-value: 1e-04</th>
<th>Hits: 5</th>
</tr>
</thead>
</table>

Cut and paste your sequence into here.

<table>
<thead>
<tr>
<th>Summary:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>chr5</td>
</tr>
<tr>
<td>chr5</td>
</tr>
<tr>
<td>chr3</td>
</tr>
<tr>
<td>chr4</td>
</tr>
</tbody>
</table>
Graph View of Alignment to Genome @ Salk

http://signal.salk.edu/cgi-bin/atta

Change the view with Edit Preference

CDS only

query
Bulk data retrieval
## Bulk Data Retrieval and Analysis

### Gene Descriptions
Download gene descriptions using locus identifiers (AGI codes) or gene names.

### Gene Ontology (GO) Annotations
Search and download GO annotations using locus identifiers. Functionally classify Arabidopsis genes and view the results as tables or pie charts.

### Plant Ontology (PO) Annotations
Search and download PO annotations using locus identifiers.

### Sequences
Download nucleotide and protein sequences using locus identifiers. Send FASTA formatted sequences to ClustalW multiple sequence alignment program at the European Bioinformatics Institute.

### Proteins
Search proteins according to predicted protein properties. Now includes SCOP's structural class prediction.

### Locus History
Find out how a locus has been modified during different releases of the genome by entering a locus name.

### Microarray Elements
Search array elements on the Affymetrix 8K, 25K, and all the AFGC arrays.
Bulk data retrieval for GO

Home > Tools > Bulk Data Retrieval > GO Annotations

GO annotation search, functional categorization and download
Gene Ontology at TAIR

Paste locus identifiers (such as At1g01030) into the textbox and press one of the submit buttons below. The identifiers have to be separated by tabs, commas, carriage returns or spaces. Alternatively, you can upload a file, same formatting as for the textbox. Clicking on Get all GO annotations will display in detail all the GO annotations done to your set of genes. Clicking on Functional categorization will group the genes into broad functional categories based on the high level terms in GO hierarchy.

You may download the whole genome GO annotations from TAIR Downloads Site.

Do you want to look for over-represented terms in your data set?
Try GO Term Enrichment: A GO tool for statistical comparisons of annotations between two data sets

Whole Genome Categorization

Get:

Whole Genome Categorization

GO Annotations

Locus Identifiers:

Upload file:
Output of bulk data retrieval for GO

Displaying 45 records.

<table>
<thead>
<tr>
<th>Keyword Category</th>
<th>Functional Category</th>
<th>Annotation Count</th>
<th>Gene Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO Cellular Component</td>
<td>nucleus</td>
<td>9949</td>
<td>9332</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>other intracellular components</td>
<td>8755</td>
<td>5208</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>other cytoplasmic components</td>
<td>8721</td>
<td>5189</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>chloroplast</td>
<td>5680</td>
<td>3695</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>other membranes</td>
<td>5168</td>
<td>3775</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>mitochondria</td>
<td>3456</td>
<td>3009</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>plasma membrane</td>
<td>3382</td>
<td>3238</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>extracellular</td>
<td>2864</td>
<td>2521</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>plastid</td>
<td>2401</td>
<td>1380</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>cytosol</td>
<td>1944</td>
<td>1702</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>unknown cellular components</td>
<td>1914</td>
<td>1914</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>other cellular components</td>
<td>1111</td>
<td>1098</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>ribosome</td>
<td>977</td>
<td>480</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>Golgi apparatus</td>
<td>975</td>
<td>922</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>cell wall</td>
<td>745</td>
<td>018</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>ER</td>
<td>649</td>
<td>588</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>other binding</td>
<td>6201</td>
<td>6789</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>unknown molecular functions</td>
<td>7697</td>
<td>7697</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>transferase activity</td>
<td>5618</td>
<td>2948</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>other enzyme activity</td>
<td>5017</td>
<td>3650</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>hydrolase activity</td>
<td>4890</td>
<td>3087</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>nucleotide binding</td>
<td>3990</td>
<td>3414</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>DNA or RNA binding</td>
<td>3975</td>
<td>3395</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>kinase activity</td>
<td>3659</td>
<td>1351</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>protein binding</td>
<td>3129</td>
<td>2599</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>transporter activity</td>
<td>2553</td>
<td>1288</td>
</tr>
</tbody>
</table>

Charts for Functional Categorization

Functional Categorization by annotation for: GO Cellular Component

- nucleus: 16.969% (raw value = 9949)
- other intracellular components: 14.933% (raw value = 8755)
- other cytoplasmic components: 14.875% (raw value = 872)
- chloroplast: 9.686% (raw value = 5680)
- other membranes: 8.811% (raw value = 5166)
- mitochondria: 5.895% (raw value = 3456)
- plasma membrane: 5.768% (raw value = 3382)
- extracellular: 4.783% (raw value = 2864)
- plastid: 4.095% (raw value = 2401)
- cytosol: 3.316% (raw value = 1914)
- unknown cellular components: 3.265% (raw value = 1914)
- other cellular components: 1.895% (raw value = 1111)
- ribosome: 1.666% (raw value = 977)
- Golgi apparatus: 1.663% (raw value = 975)
- cell wall: 1.271% (raw value = 745)
- ER: 1.107% (raw value = 649)
Enrichment of GO term

**GO annotation search, functional categorization and download**  Gene Ontology at TAIR

Paste locus identifiers (such as At1g01030) into the textbox and press one of the submit buttons below. The identifiers have to be separated by tabs, commas, carriage returns or spaces. Alternatively, you can upload a file, same formatting as for the textbox. Clicking on Get all GO annotations will display in detail all the GO annotations done to your set of genes. Clicking on Functional categorization will group the genes into broad functional categories based on the high level terms in GO hierarchy.

You may download the whole genome GO annotations from TAIR Downloads Site.

**Do you want to look for over-represented terms in your data set?**
Try **GO Term Enrichment**: A GO tool for statistical comparisons of annotations between two data sets

---

**Whole Genome Categorization**

Get:  
Whole Genome Categorization

**GO Annotations**

**Locus Identifiers:**

**Upload file:**  
選擇檔案 未選擇任何檔案
Tool for GO Term Enrichment for Plants

Home > Tools > GO Term Enrichment

GO Term Enrichment for Plants
Statistical Over/Under Representation (powered by PANTHER)

Use this tool to identify Gene Ontology terms that are over or under-represented in a set of genes (for example from co-expression or RNAseq data). The data are sent to the PANTHER Classification System which contains up to date GO annotation data for Arabidopsis and other plant species. Choose the advanced setting if you want to change parameters or explore PANTHER’s other tools for analyzing sets of genes. [Help]

Enter a list of valid identifiers, separated by newline.  Try a sample gene list

Choose Organism
Arabidopsis thaliana

Choose GO aspect
biological process
Result of GO term enrichment

<table>
<thead>
<tr>
<th>GO biological process complete</th>
<th>Arabidopsis thaliana (REF)</th>
<th>upload_1</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>#</td>
<td>expected</td>
</tr>
<tr>
<td>cell-cell junction assembly</td>
<td>5</td>
<td>5</td>
<td>.45</td>
</tr>
<tr>
<td>cell junction assembly</td>
<td>6</td>
<td>5</td>
<td>.54</td>
</tr>
<tr>
<td>cellular component organization or biogenesis</td>
<td>12555</td>
<td>1298</td>
<td>1129.41</td>
</tr>
<tr>
<td>cell junction assembly</td>
<td>3161</td>
<td>219</td>
<td>284.35</td>
</tr>
<tr>
<td>cellular component assembly</td>
<td>788</td>
<td>41</td>
<td>70.89</td>
</tr>
<tr>
<td>cellular component biogenesis</td>
<td>1309</td>
<td>71</td>
<td>117.75</td>
</tr>
<tr>
<td>indole-containing compound catabolic process</td>
<td>12</td>
<td>7</td>
<td>1.08</td>
</tr>
<tr>
<td>indole-containing compound metabolic process</td>
<td>100</td>
<td>30</td>
<td>9.00</td>
</tr>
<tr>
<td>cellular metabolic process</td>
<td>9524</td>
<td>970</td>
<td>856.75</td>
</tr>
<tr>
<td>metabolic process</td>
<td>11121</td>
<td>1153</td>
<td>1000.41</td>
</tr>
<tr>
<td>organonitrogen compound metabolic process</td>
<td>4900</td>
<td>503</td>
<td>440.79</td>
</tr>
<tr>
<td>organic substance metabolic process</td>
<td>9935</td>
<td>1004</td>
<td>893.72</td>
</tr>
</tbody>
</table>
Identification of *cis*-elements from co-expressed genes

Promoter sequences
Statistical motif analysis of promoter sequences

Home > Tools > Motif Analysis  (6-mer words)

Statistical Motif Analysis in Promoter or Upstream Gene Sequences

The program compares the frequencies of 6-mer "words" in your query set of sequences (on both strands) with the frequencies of the words in the current genomic sequence set of 33516 sequences, each containing 500 (or 1000) bp upstream of the start codon of each gene. You can type in sets of AGI locus identifiers (e.g. At1g01030) or sets of fasta sequences. Make sure each fasta header is formatted as such, fasta symbol (>), immediately followed by a unique ID, a space, then all other descriptions (e.g. >ABCD1.1 my gene). Ensure that there are no sequences appearing more than once in your query set.

Locus ID

Upload file:  選擇檔案  未選擇檔案

Dataset:
- 500 bp upstream
- 1000 bp upstream
- 3000 bp upstream
Bulk sequence retrieval for sequences

Home > Tools > Bulk Data Retrieval > Sequences

Sequence Bulk Download and Analysis

This tool can be used to download a variety of sequences from the Arabidopsis Genome Initiative (AGI) in FASTA or tab-delimited format. Individual or sets of AGI locus identifiers (e.g., At1g01030) can be typed into the textbox below or uploaded from your desktop. FASTA formatted sequences can also be sent to the multiple sequence alignment program Clustal W. More information on the tool can be found by following the link to the Help document.

Please use the ftp server if you want to download entire datasets. Click HERE to obtain details about the sequence datasets.

For Intron Sequences ONLY: A special format is required for obtaining intron sequences using the Intron Sequences database. Each entry must include the locus identifier followed by the gene model suffix. To specify which intron, the model id.suffix is followed by the intron number and the number of the intron you wish to retrieve. For example, to obtain the sequence of the first intron on the AGAMOUS gene would be AT4G18960.1-1.

For Upstream/Downstream/Intergenic sequences: For these data types we ONLY provide the sequence relative to the (based on the maximum extent of all gene models for that locus) regardless of whether a locus name or a specific gene model is used for the search.

### Araport11 transcripts
- Araport11 coding sequences
- Araport11 genomic locus sequences
- Araport11 protein sequences
- Araport11 Loci Upstream Seq - 500bp
- Araport11 Loci Upstream Seq - 1000bp
- Araport11 Loci Upstream Seq - 3000bp
- Araport11 Loci Downstream Seq - 500bp
- Araport11 Loci Downstream Seq - 1000bp
- Araport11 Loci Downstream Seq - 3000bp
- Intergenic Sequences
- Intron Sequences
- Araport11 3' UTRs
- Araport11 5' UTRs
Motif analysis with MEME

DNA motif in Promoter, RNA transcripts or Protein motif

Sequence Bulk Download and Analysis

http://meme-suite.org/tools/meme
Identify genes with specific motifs

BLAST

Patmatch
Patmatch: short sequences search

Home > Tools > Patmatch

Search *cis*-elements in promoters or genes
Search protein motifs in proteins

- short (<20 nt)
- ambiguous/degenerate patterns
- sequence >20 bp with no degenerate positions, use BLAST or FASTA, which are much faster
Patmatch: pattern syntax

Supported Pattern Syntax and Examples:

<table>
<thead>
<tr>
<th>Search Type</th>
<th>Character</th>
<th>Meaning</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peptide</strong></td>
<td>IFVLWMAGCYP</td>
<td>Exact match</td>
<td>DQGT</td>
</tr>
<tr>
<td></td>
<td>TSHEDQNKR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>Any hydrophobic residue (IFVLWMA...</td>
<td>AAAAAAJJ</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>Any hydrophilic residue (TS...</td>
<td>TTTTTTOO</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>D or N</td>
<td>FLGB</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>E or Q</td>
<td>GLFGZ</td>
</tr>
<tr>
<td></td>
<td>X or .</td>
<td>Any amino acid</td>
<td>DXXXNW...VSK</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Nucleotide searches</strong></th>
<th>ACTGU</th>
<th>Exact match</th>
<th>ACCGCGTGAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>Any purine base (AG)</td>
<td>AAGGCGGGRRRR</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Any pyrimidine base (CT)</td>
<td>CCCATAYGGYY</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>G or C</td>
<td></td>
<td>YGGTWCAMWTGTYY</td>
</tr>
<tr>
<td>W</td>
<td>A or T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>A or C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>G or T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>A or C or G</td>
<td></td>
<td>CCGG...WHW...CCGG</td>
</tr>
<tr>
<td>H</td>
<td>A or C or T</td>
<td></td>
<td>{3,5}HWH...CCGG</td>
</tr>
<tr>
<td>D</td>
<td>A or G or T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>C or G or T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N or X or .</td>
<td>Any base</td>
<td></td>
<td>ATGCTNNNATCG</td>
</tr>
</tbody>
</table>
## Patmatch: pattern syntax

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
|  | A subset of elements [TC] = T or C  
[WFY]XXXDN[RK][ST] |
|  | An excluded subset of elements  
[^TA] = not T or A,  
(matches nucleotides C or G)  
NDBB…[VILM]Z[DE]…[^PG] |
|  | Specifies a sub-pattern  
(YPT) = YPT  
(YDXXX){2,} |
|  | {m} = exactly m times  
{m,} = at least m times  
{,m} = 0 to m times  
{m,n} = between m and n times  
L{3,5}X{5}DGZ |
|  | Constrains pattern to N-terminus or 5' end  
<MNTD (pep)  
<ATGX{6,10}RTTRTT (nuc) |
|  | Constrains pattern to C-terminus or 3' end  
sbgz> (pep)  
yatrtga> (nuc) |
Identify promoter sequences meeting following criteria
1. With DRE element **RCCGAC** (ACCGAC or GCCGAC)
2. With ABRE element **GACACGT**
3. The distance between DRE and ABRE: 10-200 nt

**RCCGACN\{10,200\}GACACGT**
Patmatch: output example

Your PatMatch request has been queued and will be processed. Please do not hit "Stop" and resubmit your job. Such action will add to the queue of jobs already in progress and have the effect of placing your request to the end of the line.

PatMatch Search is now Starting...

Queued
Running

<table>
<thead>
<tr>
<th>Hits found</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequences with hits</td>
<td>28</td>
</tr>
<tr>
<td>Sequences searched</td>
<td>33602</td>
</tr>
<tr>
<td>Bytes searched</td>
<td>16800250</td>
</tr>
<tr>
<td>Pattern</td>
<td>RCCGACN[10,200]GACACGT</td>
</tr>
<tr>
<td>Dataset searched</td>
<td>TAIR10 Loci Upstream Sequences -- 500 bp (DNA)</td>
</tr>
<tr>
<td>Download all matches as a text file</td>
<td>download</td>
</tr>
</tbody>
</table>

Next Results >

<table>
<thead>
<tr>
<th>Hit#</th>
<th>Sequence name</th>
<th># of hits</th>
<th>Hit pattern</th>
<th>Matching Positions</th>
<th>Hit sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AT1G04080</td>
<td>1</td>
<td>ACCGACCGTTGGGTTCCGGAGAAQA CGGCTGCTCATAAATCTATCTGCG CTAATCGATGATCTTGCTCTCAACTC TAAGATCAAGGACACGGCAAGGATG AGAAGACGTCTACCAGCGATCCACGC CCGGCTTCGCCGGTGCCTTTAACAAG TTCTTCCGACACGT</td>
<td>33 196</td>
<td>sequence</td>
</tr>
<tr>
<td>2</td>
<td>AT1G05100</td>
<td>1</td>
<td>GCCGACACGTGGCTCGGAGACAAATA CGGAACCTCCCGGTGGAAACAGAAGAC AATTACTTTTTTGGACTACCTGAAAAA TTCGACACGT</td>
<td>11 95</td>
<td>sequence</td>
</tr>
</tbody>
</table>
## TAIR FTP download

Home > Download > Download Overview

### Genes
Contains TAIR’s genome release files, gene family data, and lists of gene names, aliases and descriptions for all TAIR genes.

### GO and PO Annotations
Contains files of all Gene Ontology (biological process, cellular component and molecular function) and Plant Ontology (plant growth and developmental stage, plant structure) annotations in TAIR.

### Maps
Contains files of map coordinates for Arabidopsis genes and their features, ESTs, cDNAs, BAC clones, polymorphisms, and markers on the sequence map (SeqViewer data subdirectory). Also contains map data for older maps, including physical and genetic maps.

### Microarray Data
Contains Affymetrix, AFGC, Agilent and CATMA mapping files matching array element identifiers to AGI loci. Also contains original and reanalyzed microarray datasets for some experiments including AtGenExpress.

### Pathways
TAIR FTP download: genes

Download - Genes
- Arabidopsis genome release
- gene_aliases_20130831.txt 876
- Gene families
- TAIR10 genome release
- TAIR6 genome release
- TAIR7 genome release
- TAIR8 genome release
- TAIR9 genome release
- TAIR gene confidence ranking
- TIGR release versions

Download - TAIR10 genome release
- gene_description_20131231.txt.gz 4,474 KB 2013-12-31
- README_TAIR10.txt 5 KB 2012-08-22
- TAIR10-Subcellular_Predictions.xlsx 3,964 KB
- TAIR10 blastsets
- TAIR10 chromosome files
- TAIR10_domain_architectures.tab.10 2,608 KB
- TAIR10_functional_descriptions 25,396 KB 2013-01-01
- TAIR10_functional_descriptions.bk 25,396 KB
- TAIR10_functional_descriptions_20130831.txt 25,394 KB
- TAIR10 gene confidence ranking
- TAIR10 gene lists
- TAIR10 gene transcript associations
- TAIR10 gff3
- TAIR10_locushistory.txt 2,053 KB 2010-11-23
- TAIR10 NCBI mapping files
- TAIR10_sequence_edits.txt 0 KB 2010-12-07
- TAIR10_TAIRAccessionID_AGI_mapping.txt 792
- TAIR10_TAIRlocusaccessionID_AGI_mapping.txt
- TAIR10 transposable elements
- Tair10 XML
TAIR FTP download: sequences

Download - Sequences

- Araport11 blastsets
- TAIR10 blastsets

Download - TAIR10 blastsets

- downstream sequences
  - Readme_blastdatasets_TAIR10.txt 4 KB 2012-04-16
  - TAIR10_3_utr_20101028 8,718 KB 2010-11-10
  - TAIR10_5_utr_20101028 6,027 KB 2010-11-10
  - TAIR10_bac_con_20101028 137,399 KB 2010-11-10
  - TAIR10_cdna_20101214_updated 71,717 KB 2012-04-16
  - TAIR10_cdna_20110103_representative_gene_model_updated 56,587 KB 2012-04-16
  - TAIR10_cds_20101214_updated 49,453 KB 2012-04-16
  - TAIR10_cds_20110103_representative_gene_model_updated 38,019 KB 2012-04-16
  - TAIR10_exon_20101028 81,803 KB 2010-11-10
  - TAIR10_intergenic_20101028 51,663 KB 2010-11-10
  - TAIR10_intron_20101028 41,668 KB 2010-11-10
  - TAIR10_pep_20101214_updated 20,006 KB 2012-04-16
  - TAIR10_pep_20110103_representative_gene_model_updated 15,437 KB 2012-04-16
  - TAIR10_seq_20101214_updated 101,193 KB 2012-05-07
  - TAIR10_seq_20110103_representative_gene_model_updated 76,879 KB 2012-04-16
- upstream sequences
Database/Tool Help

This section of TAIR's website contains links to documents to help you maximally utilize TAIR's database and tools.

Use the Definitions help. Wherever you see a symbol next to a word, click on the symbol to see a definition for the word or phrase.

If you have a specific topic you need help with, you can search ONLY within the help section of TAIR using the search box below.

Tutorials
- AraCyc Tutorial
- TAIR Gene Ontology Tools
- TAIR Microarray Tools
- Getting Started Guide
- Database/Tool Help Pages
- FAQ
- TAIR Glossary
- TAIR User Guide

Home > Help

GO annotation search, functional categorization and download

Paste locus identifiers (such as At1g01030) into the textbox and press one of the submit buttons below. The identifiers have to be separated by tabs, commas, carriage returns or spaces. Alternatively, you can upload a file, same formatting as for the textbox. Clicking on Get all GO annotations will display in detail all the GO annotations done to your set of genes. Clicking on Functional categorization will group the genes into broad functional categories based on the high level terms in GO hierarchy.
Thank you!

For questions, you can contact

Ho-Ming Chen 陳荷明

Office: Agricultural Technology Building A638
Email: homing@gate.sinica.edu.tw
Phone: 2787-2073