

Tutorial: Hayai-Annotation Plants

Reference:

**Hayai-Annotation Plants: an ultra-fast and comprehensive gene annotation system
in plants**

Andrea Ghelfi*, Kenta Shirasawa, Hideki Hirakawa, Sachiko Isobe

Laboratory of Plant Genetics and Genomics, Department of Frontier Research and
Development, Kazusa DNA Research Institute, Kisarazu, 292-0818, Japan.

* andreaghelfi@kazusa.or.jp

127.0.0.1

Hayai-Annotation Plants

Kazusa DNA Research Institute

Type of Alignment
 Local
 Global

Type of Algorithm
 Protein Existence Level
 Alignment Score

Upload FASTA File
Browse... No file selected

Max hits per query 1 Minimum Sequence Identity (%) 90

Evalue 1e-6 Minimum Query Coverage (%) 80

Submit

Download

Show 10 entries

Search:

query	uniprot	seqID	length	mis	gaps	startquery	endquery	starttarget	endtarget	evalue	score	gene_name	Protein_Evidence	EC	evidence
No data available in table															

No data available in table

Showing 0 to 0 of 0 entries Previous Next

Since database is split in 4 files this number is multiplied by 4 for the final file. Thus, if desired maximum hits per query is 20, divide by four and type 5 in this window.

After selecting the parameters above, browse your FASTA file (cDNA or protein).

Default values:
Type of alignment = local
Max hits per query = 1
E-value = 1e-6
Minimum Sequence Identity = 80
Minimum Query Coverage = 80
Type of algorithm = Protein Existence Level

127.0.0.1

Hayai-Annotation Plants

Kazusa DNA Research Institute

Type of Alignment

Local
 Global

Max hits per query: 1

Minimum Sequence Identity (%): 90

Type of Algorithm

Protein Existence Level
 Alignment Score

Evalue 1e-: 6

Minimum Query Coverage (%): 80

Upload FASTA File

Browse... Araport11_genes.201606.pep.
Upload complete

Submit

Download

Show 10 entries

Search:

query	uniprot	seqID	length	mis	gaps	startquery	endquery	starttarget	endtarget	evalue	score	gene_name	Protein_Evidence	EC	evidence
No data available in table															

Showing 0 to 0 of 0 entries

Previous Next

Hayai-Annotation Plants uploaded the FASTA file.

After uploading your FASTA file, click “Submit” once. Please wait a few minutes for the results, the page will automatically update.

127.0.0.1

Type of Algorithm

- Protein Existence Level
- Alignment Score

Evalue 1e-
6

Minimum Query Coverage (%)
80

Upload FASTA File

Browse... Araport11_genes.201606.pep.
Upload complete

Submit

Download

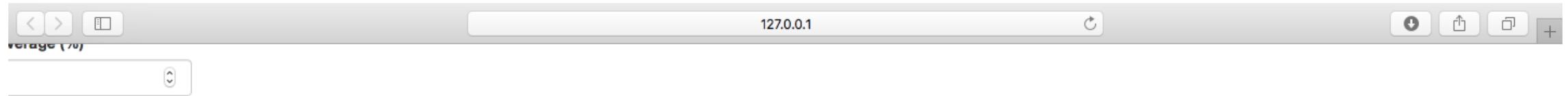
When Hayai-Annotation Plants finish its analysis, click here to download all results.

Use this search engine to find your targets.

Show 10 entries

Search:

	query	uniprot	seqID	length	mis	gaps	startquery	endquery	starttarget	endtarget	evalue	score	gene_name	Protein_Evidence
1	AT1G01010.1	NAC1_ARATH	100	429	0	0	1	429	1	429	1.7e-257	890.6	NAC domain-containing protein 1	PE 2
2	AT1G01020.1	Q5MK24_ARATH	100	245	0	0	1	245	1	245	1.3e-134	481.9	ARV1	PE 2
3	AT1G01030.1	NGA3_ARATH	100	358	0	0	1	358	1	358	8.9e-215	748.4	B3 domain-containing transcription factor NGA3	PE 2
4	AT1G01040.2	DCL1_ARATH	99.8	1910	2	1	1	1910	1	1909	0	3809.2	Endoribonuclease Dicer homolog 1	PE 1
5	AT1G01050.1	IPYR1_ARATH	100	212	0	0	1	212	1	212	1.1e-119	431.8	Soluble inorganic pyrophosphatase 1	PE 1

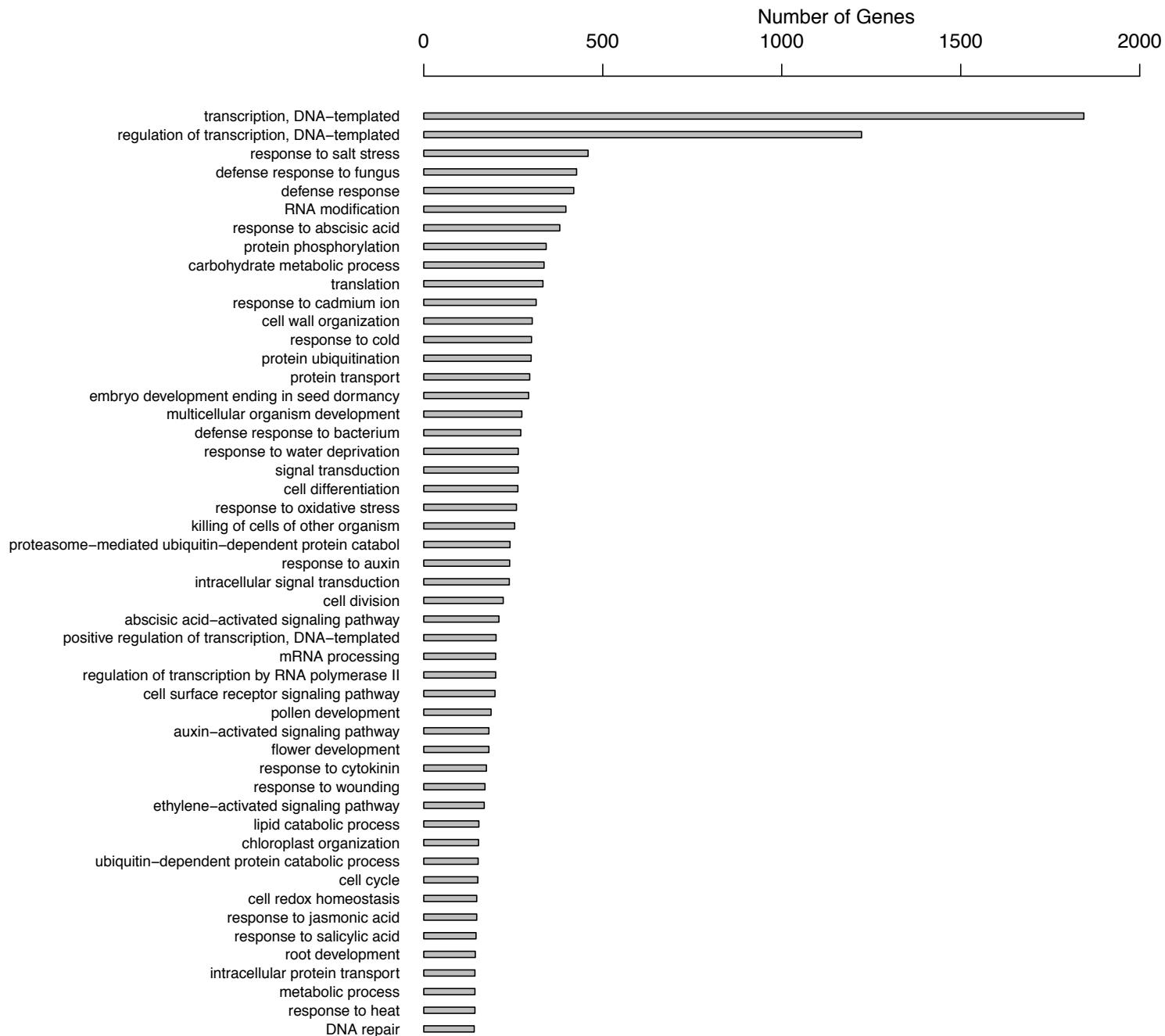


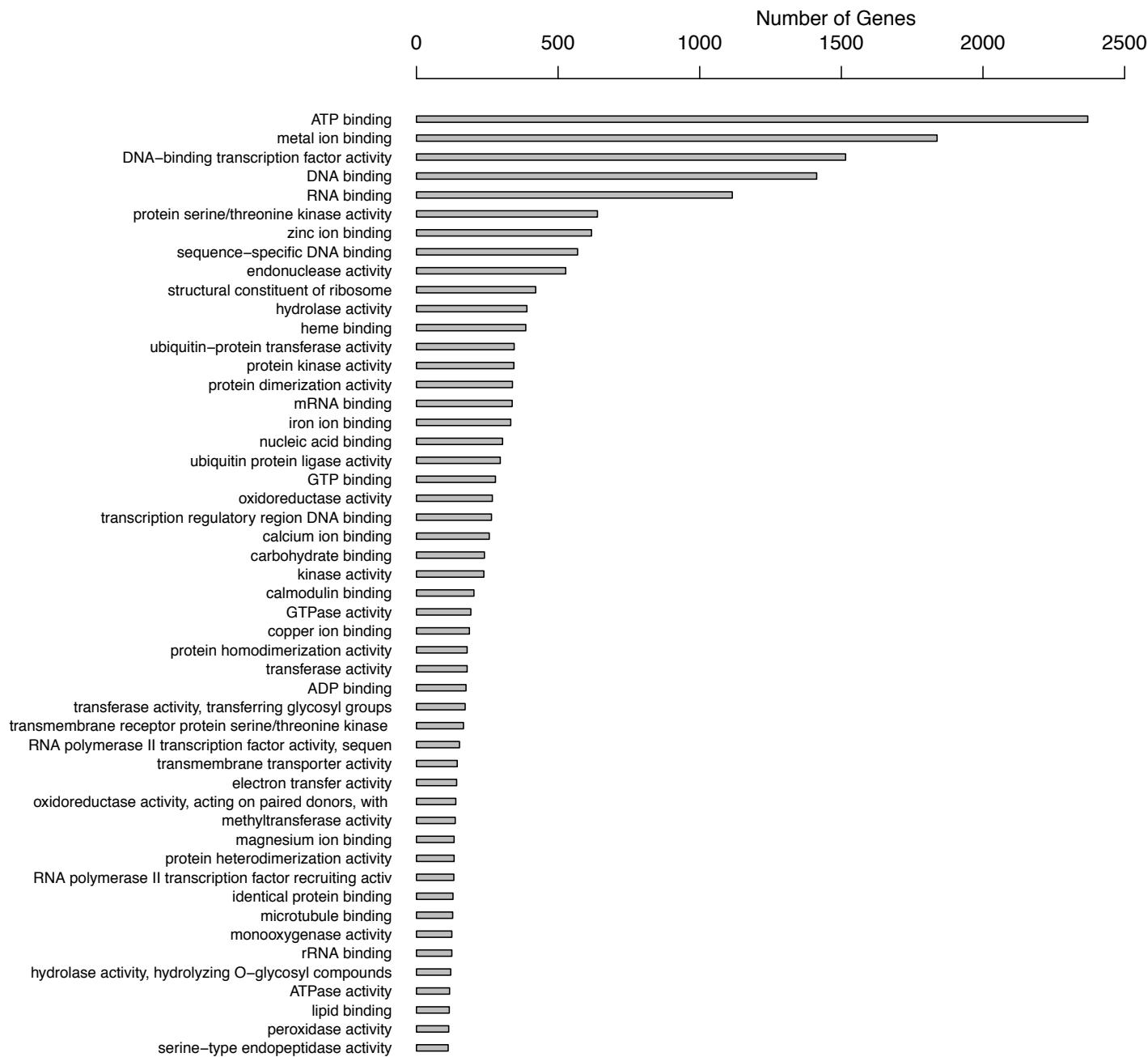
This is the remaining columns of Hayai-Annotation Plants output table.

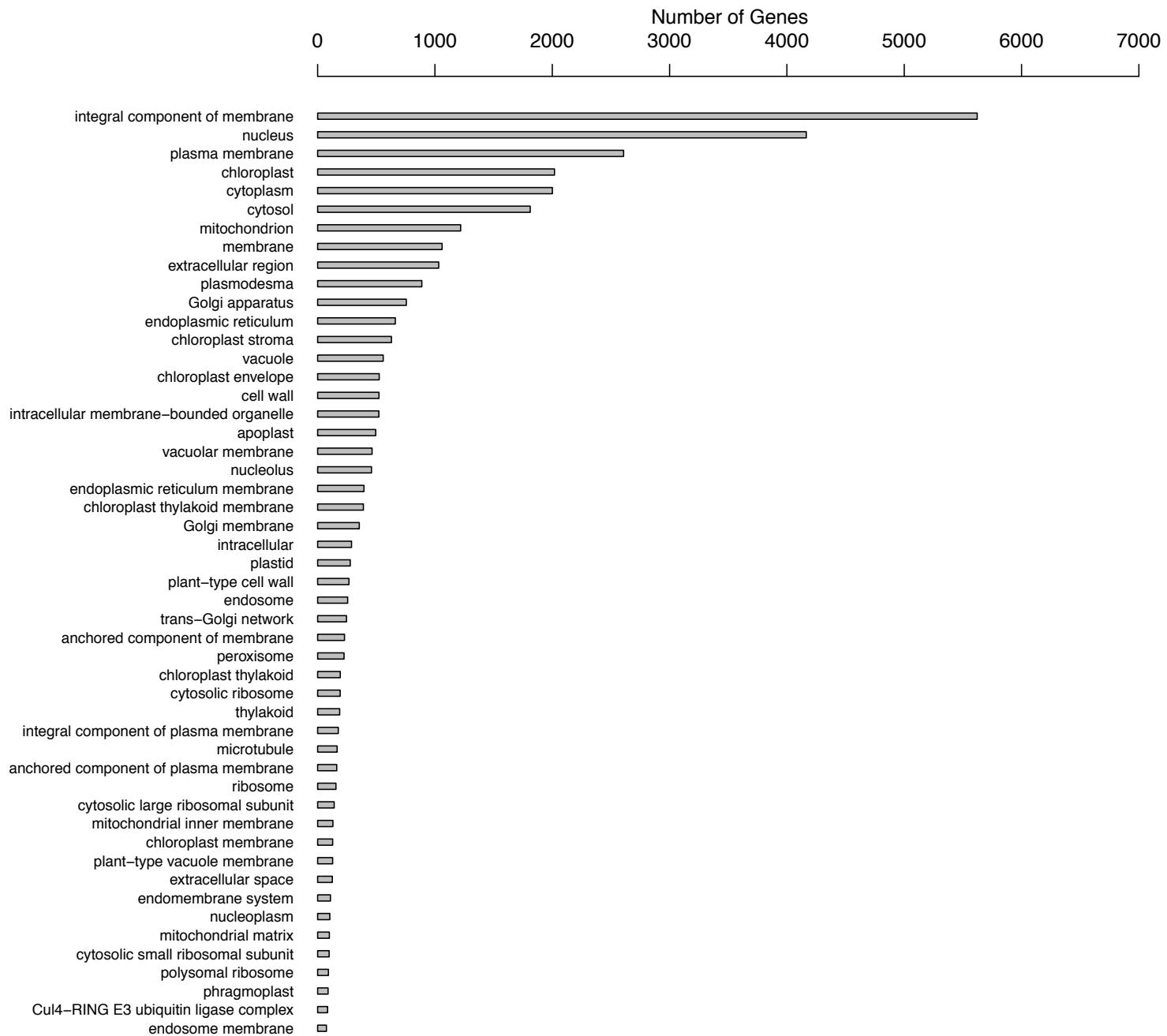
endtarget	evalue	score	gene_name	Protein_Evidence	EC	evidence_type	GO_BP	GO_BP_name	GO_MF	GO_MF_name	GO_CC	GO_CC_name
429	1.7e-257	890.6	NAC domain-containing protein 1	PE 2		ISS	GO:0006351	transcription, DNA-templated	GO:0003700	DNA-binding transcription factor activity	GO:0016021	integral component of membrane
245	1.3e-134	481.9	ARV1	PE 2		IBA	GO:0097036	regulation of plasma membrane sterol distribution			GO:0005783	endoplasmic reticulum
358	8.9e-215	748.4	B3 domain-containing transcription factor NGA3	PE 2		IGI	GO:0006351	transcription, DNA-templated	GO:0003700	DNA-binding transcription factor activity	GO:0005634	nucleus
1909	0	3809.2	Endoribonuclease Dicer homolog 1	PE 1	3.1.26.-	IBA	GO:0006396	RNA processing	GO:0003723	RNA binding	GO:0016442	RISC complex
212	1.1e-119	431.8	Soluble inorganic pyrophosphatase 1	PE 1	3.6.1.1	IDA	GO:0006796	phosphate-containing compound metabolic process	GO:0000287	magnesium ion binding	GO:0005829	cytosol

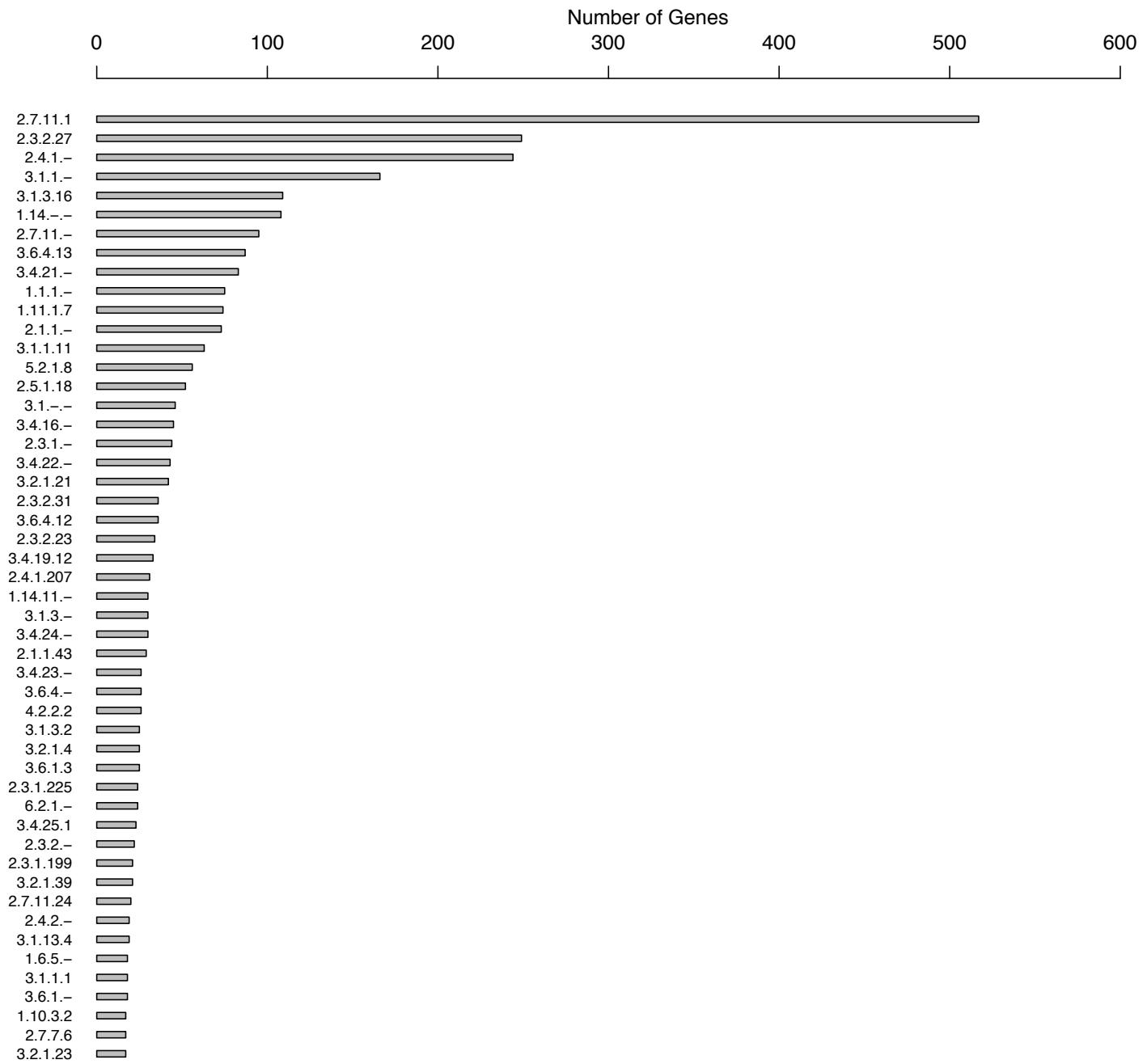
Hayai-Annotation Plants

Downloaded Output Files

GO Biological Process – Top 50 – Gene Level

GO Molecular Function – Top 50 – Gene Level

GO Cellular Component – Top 50 – Gene Level

Enzyme Commission – Top 50

GO_BP_table

Home Insert Page Layout Formulas Data Review View

Cut Copy Paste Format

Calibri (Body) 12 A A = Wrap Text General Conditional Formatting

Merge & Center %

A1 fx uniprot

	A	B	C	D
1	uniprot	query	GO_BP	GO_BP_name
2	14310_ARATH	AT1G22300.1	GO:0009742	brassinosteroid mediated signaling pathway
3	14310_ARATH	AT1G22300.1	GO:0009737	response to abscisic acid
4	14332_ARATH	AT1G78300.1	GO:0009742	brassinosteroid mediated signaling pathway
5	14333_ARATH	AT5G38480.1	GO:0009873	ethylene-activated signaling pathway
6	14333_ARATH	AT5G38480.1	GO:0009409	response to cold
7	14333_ARATH	AT5G38480.1	GO:0006995	cellular response to nitrogen starvation
8	14333_ARATH	AT5G38480.1	GO:0016036	cellular response to phosphate starvation
9	14333_ARATH	AT5G38480.1	GO:0050826	response to freezing
10	14333_ARATH	AT5G38480.1	GO:0051365	cellular response to potassium ion starvation
11	14333_ARATH	AT5G38480.1	GO:0019222	regulation of metabolic process
12	14333_ARATH	AT5G38480.1	GO:0009631	cold acclimation
13	14335_ARATH	AT5G16050.1	GO:0046686	response to cadmium ion
14	14336_ARATH	AT5G10450.4	GO:0050826	response to freezing
15	14336_ARATH	AT5G10450.4	GO:0042742	defense response to bacterium
16	14336_ARATH	AT5G10450.4	GO:0046686	response to cadmium ion
17	14336_ARATH	AT5G10450.4	GO:0045732	positive regulation of protein catabolic process
18	14336_ARATH	AT5G10450.4	GO:0009409	response to cold
19	14336_ARATH	AT5G10450.4	GO:0009742	brassinosteroid mediated signaling pathway
20	14338_ARATH	AT5G65430.3	GO:0019222	regulation of metabolic process
21	14338_ARATH	AT5G65430.3	GO:0050826	response to freezing
22	1A110_ARATH	AT1G62960.1	GO:0042218	1-aminocyclopropane-1-carboxylate biosynthetic process
23	1A111_ARATH	AT4G08040.1	GO:0009693	ethylene biosynthetic process
24	1A111_ARATH	AT4G08040.1	GO:0009835	fruit ripening

GO_BP_table

GO_MF_table

GO_CC_table

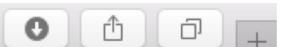
Contains information of all GO terms and codes for each gene, including UniProt ids.

GO_BP_counts

	GO_BP_name	Counts
1	transcription, DNA-templated	1894
2	regulation of transcription, DNA-templated	1242
3	response to salt stress	457
4	defense response to fungus	433
5	defense response	432
6	RNA modification	400
7	response to abscisic acid	380
8	protein phosphorylation	353
9	carbohydrate metabolic process	337
10	translation	331
11	response to cadmium ion	315
12	protein transport	312
13	response to cold	308
14	cell wall organization	306
15	protein ubiquitination	306
16	embryo development ending in seed dormancy	299
17	multicellular organism development	280
18	defense response to bacterium	275
19	cell differentiation	272
20	response to water deprivation	262
21	response to oxidative stress	261
22	signal transduction	261
23	killing of cells of other organism	258
24	proteasome-mediated ubiquitin-dependent protein catabol	246
25	intracellular signal transduction	245
26	response to auxin	245

GO_BP_counts
GO_MF_counts
GO_CC_counts

Contains the sum of all GO terms.



KEGG Mapper

A suite of KEGG mapping tools

► About KEGG Mapper

[Search Pathway](#)
[Search&Color Pathway](#)
[Color Pathway](#)

[Search Brite](#)
[Search&Color Brite](#)
[Join Brite](#)
[Join Brite Table](#)

[Search Module](#)
[Search&Color Module](#)

[Search Disease](#)

[Reconstruct Pathway](#)
[Reconstruct Brite](#)
[Reconstruct Module](#)
[Map Taxonomy](#)

About KEGG Mapper

KEGG Mapper is a collection of tools for [KEGG mapping](#): KEGG pathway mapping, BRITE mapping, and MODULE mapping. Two pathway mapping tools, "Search Pathway" and "Color Pathway", were made available from the beginning of the KEGG project. The naming of these tools was somewhat misleading since both involved searching and coloring procedures. The latter is now called "Search&Color Pathway" and a new "Color Pathway" is introduced for coloring of a selected pathway map (without searching).

Pathway mapping tools

[Search Pathway](#) - basic pathway mapping tool

- Search against all pathway maps in one category (ko, org, etc.)
- Multiple search objects may be specified in one line
- Mapped objects are marked in red

[Search&Color Pathway](#) - pathway mapping tool with coloring options

- Search against all pathway maps in one category (ko, org, etc.)
- Each search object is specified in one line together with color attributes (background color and foreground color)
- Mapped objects can be marked in any color
- NCBI and UniProt identifiers may be used



KEGG Mapper – Search Pathway

1. Type ec

Search against:

Enter: map, ko, ec, rn, hsadd, or

Enter objects:

Examples:

Select

2. Choose file: unique_EC.csv (this file is generated by Hayai-Annotation Plants and can be found in the Downloaded files)

Alternatively, enter the file name containing the data:

Reconstruct Pathway
Reconstruct Brite
Reconstruct Module
Map Taxonomy

Convert ID
Draw Phylogram

Annotate Sequence
BlastKOALA

3. Click Exec

Include aliases

Display objects not found in the search

Search pathways containing all the objects (AND search)

Sort by the pathway list

Show all objects

- [ec01100 Metabolic pathways \(566\)](#)
- [ec01110 Biosynthesis of secondary metabolites \(325\)](#)
- [ec01130 Biosynthesis of antibiotics \(157\)](#)
- [ec01120 Microbial metabolism in diverse environments \(122\)](#)
- [ec00230 Purine metabolism \(40\)](#)
- [ec00270 Cysteine and methionine metabolism \(35\)](#)
- [ec00520 Amino sugar and nucleotide sugar metabolism \(33\)](#)
- [ec00860 Porphyrin and chlorophyll metabolism \(29\)](#)
- [ec00500 Starch and sucrose metabolism \(28\)](#)
- [ec00260 Glycine, serine and threonine metabolism \(28\)](#)
- [ec00240 Pyrimidine metabolism \(26\)](#)
- [ec00564 Glycerophospholipid metabolism \(25\)](#)
- [ec00010 Glycolysis / Gluconeogenesis \(25\)](#)
- [ec00900 Terpenoid backbone biosynthesis \(24\)](#)
- [ec00620 Pyruvate metabolism \(23\)](#)
- [ec00710 Carbon fixation in photosynthetic organisms \(21\)](#)
- [ec00970 Aminoacyl-tRNA biosynthesis \(21\)](#)
- [ec00330 Arginine and proline metabolism \(21\)](#)
- [ec00250 Alanine, aspartate and glutamate metabolism \(21\)](#)
- [ec00051 Fructose and mannose metabolism \(20\)](#)
- [ec00630 Glyoxylate and dicarboxylate metabolism \(19\)](#)

This is the output of KEGG Mapper, each line represents one pathway and in parenthesis is the number of genes found in that pathway.