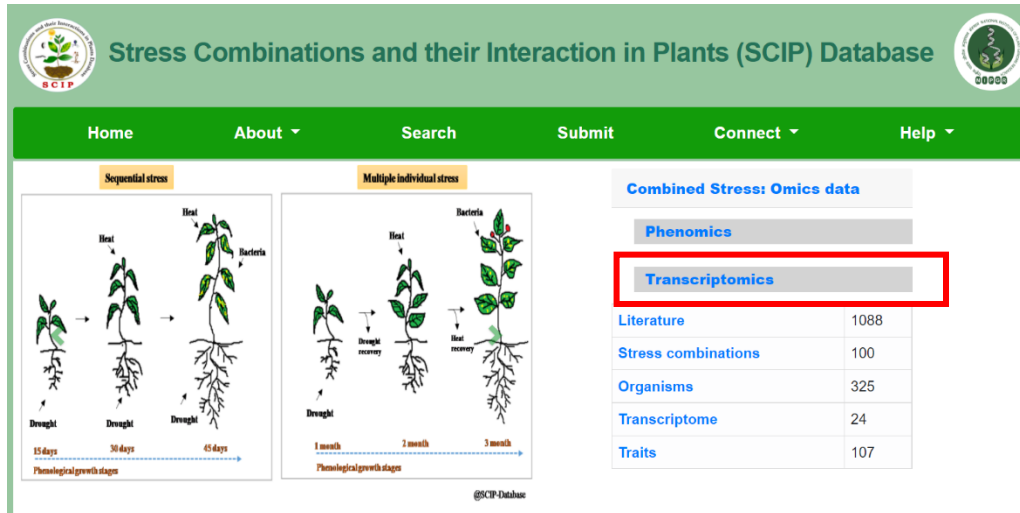**SCIPDb: Transcriptome navigation and data mining tutorial**

1) Assessing transcriptome data

a) User needs to click the Transcriptomics tab, from the home page of database.

b) On click, they will be redirected to the Transcriptomics section which is shown below.

c) User can get detailed explanation of the Combined stress transcriptome data from the highlighted section shown below

2) Mining transcriptome data:

Step 1: Select a plant from the drop-down menu highlighted.

Step 2: Corresponding combined stress transcriptome data available is automatically highlighted, in the second drop-down. Select a stress combination from the drop-down menu highlighted.

Combined stress: Transcriptome data

Search unique/common genes based on Plant

STEP I : Select a plant species:

Choose a Plant: [Thale cress ▼]

STEP II: Select a stress combination:

Choose a Stress combination:

[Please select Stress Combination ▼]

- Please select Stress Combination
- Drought & bacterial nonhost pathogen
- Drought & bacterial host pathogen
- Bacterial host pathogen & drought
- Drought recovery & bacterial host pathogen
- Dehydration & Nematode (Leaf)
- Dehydration & Nematode (Root)
- Salinity, mannitol & heat
- Cold & high light
- Heat & high light
- Salt & heat (Rasmussen et. al, 2013)
- Salt & high light
- Heat, Silwet & Flg22 peptide
- Cold, Silwet & Flg22 peptide
- Salt & heat (Suzuki N. et. al, 2016)
- Fungus, herbivory & drought
- Herbivory, fungus & drought
- Drought, herbivory & fungus
- Drought, pathogen & herbivory

[Reset] [Submit]

- SCIP database hosts a comprehensive collection of combined stress responsive differentially expressed genes (DEGs) identified in publically available transcriptomic data from various plant species.

- Transcriptomic data represented here includes a list of DEGs, log fold change and absolute fold change over control, obtained after in-house analysis of the combined stress transcriptome available in the public domain.

- This section provides an option for browsing transcriptomics data, based on drop down based selection of plants and stress combination.

- Result is presented in the form of interactive bootstrap table, that displays informations about list of differentially expressed genes (DEGSs) genes expressed uniquely and commonly under individual and combination of stresses along with gene annotation and fold change in expression.

- The data also includes information about DEGs nature of change in expression i.e whether it is expressed uniquely or common or tailored in a particular stress combination.

- The DEGs has been mapped to KEGG pathway wherever possible

- In cases where pathway information wsa not available, DEGs has been mapped to KEGG genes

- KEGG Genes and pathway link can aid users to navigate through several other databases (E.g., TAIR, NCBI etc.) interlinked with SCIP database, and fetch other metadata like detailed gene annotation.

- User selection based output download of entire data is also provided

Step 3: Select desired category of genes: The differentially expressed genes (DEGs) for each transcriptome, has been divided into common (between individual and combined stress) and unique (unique in individual and combined stress) categories for better comprehension and presentation.

Case Study:

Plant selected: Thale Cress
Stress combination selected: Drought and bacterial non-host pathogen
Stress 1: Drought
Stress 2: Bacterial nonhost pathogen
Stress 3: Combined Drought and bacterial non-host pathogen
Selection made: Genes common to all categories
(Refer screenshots below: all depiction shown for above selections)

Choose a Stress combination:
Drought & bacterial nonhost pathogen

and combination of stresses along with gene annotation and fold change in expression.

- The data also includes information about DEGs nature of change in expression i.e whether it is expressed uniquely or common or tailored in a particular stress combination.

STEP III: Query SCIP database for desired genes:

- The DEGs has been mapped to KEGG pathway wherever possible

- In cases where pathway information wsa not available, DEGs has been mapped to KEGG genes

Query for desired genes:
Common genes : Genes common to all (D, NH and DNH)
Please select desired genes
Common genes : Genes common to all (D, NH and DNH)
Common genes : Genes common between drought (D) and nonhost pathogen (NH)
Common genes : Genes common between drought (D) and combined stress (DNH)
Common genes : Genes common between nonhost pathogen (NH) and combined stress (DNH)
Unique genes

- KEGG Genes and pathway link can aid users to navigate through databases (E.g., TAIR, NCBI etc.) interlinked with SCIP fetch other metadata like detailed gene annotation. based output download of entire data is also

Reset Submit

Step 4: On clicking the submit button, user will be redirected to the webpage containing the details of the gene which are common to all the three stress treatments for the species Thale cress.

STEP I : Select a plant species:

- SCII database hosts a comprehensive collection of combined stress responsive differentially expressed genes (DEGs) identified in publically available transcriptomic data from various plant species.

Choose a Plant: Thale cress

- Transcriptomic data represented here includes a list of DEGs, log fold change and absolute fold change over control, obtained after in-house analysis of the combined stress transcriptome available in the public domain.

- This section provides an option for browsing transcriptomics data, based on drop down based selection of plants and stress combination.

STEP II: Select a stress combination:

- Result is presented in the form of interactive bootstrap table, that displays informations about list of differentially expressed genes (DEGSs) genes expressed uniquely and commonly under individual and combination of stresses along with gene annotation and fold change in expression.

Choose a Stress combination:
Drought & bacterial nonhost pathogen

- The data also includes information about DEGs nature of change in expression i.e whether it is expressed uniquely or common or tailored in a particular stress combination.

STEP III: Query SCIP database for desired genes:

- The DEGs has been mapped to KEGG pathway wherever possible

- In cases where pathway information wsa not available, DEGs has been mapped to KEGG genes

Query for desired genes:
Common genes : Genes common to all (D, NH and DNH)

- KEGG Genes and pathway link can aid users to navigate through several other databases (E.g., TAIR, NCBI etc.) interlinked with SCIP database, and fetch other metadata like detailed gene annotation.
- User selection based output download of entire data is also provided

Click Here

Reset Submit

3) Understanding transcriptome data:

a) The final data depiction of transcriptome has been done in the form of Bootstrap table (see screenshot below).

b) The header of the output (shown in arrows below) is based on the users selection of plant species, stress combination and corresponding category (**Drought and bacterial non-host pathogen, in Thale cress and gene common to all three stress treatments).**

c) The interactive bootstrap table, lists all the common DEGs between the three stress treatments (D: Drought, NH: Nonhost pathogen, and Drought and non-host pathogen).

d) The 6 columns (highlighted below) correspond
1.   Sl.no.
2.   Gene ID's (DEGs)
3.   Gene Name
4.   Transcript fold change under each treatment
5.   Nature of Change in Expression (similar expression/response under all stress [positive/negative in all three treatments] or tailored response [not same expression trend in all three treatments])
6.   Pathway and Metadata (All the DEGs has been mapped to KEGG pathways wherever possible, in cases where pathway information was not available link to KEGG Genes has been provided). Clicking on the pathway link will take users to KEGG (See box highlighted below in Pathway and Metadata column).

Home	Phenome	Transcriptome	Meta-analysis

Drought and nonhost pathogen stress in Thale cress ←
Genes common to all stresses (D, NH and DNH) ←
Search results matching your query

Show 10 entries   Search:

Sl.no.	Gene ID	Gene name	Transcript Fold change (FC)	Nature of change in expression	Pathway & metadata
1	AT1G62300	WRKY family transcription factor	NH:8.7; D:1.8 ; DNH: -2.1	Tailored response	KEGG Genes Link
2	AT5G39520	Protein of unknown function (DUF1997)	DNH: 3.1; NH: 1.50; D: 8.11	Similar expression/response under all stress	KEGG Genes Link
3	AT5G38210	Protein kinase family protein	DNH: 2.40; NH: 2.30; D: 6.68	Similar expression/response under all stress	KEGG Genes Link
4	AT1G51860	Leucine-rich repeat protein kinase family protein	DNH: 6.70; NH: 2.90; D: -1.90	Tailored response	KEGG Genes Link
5	AT1G47890	Receptor like protein 7	DNH: 6.60; NH: 1.80; D: -1.50	Tailored response	KEGG Genes Link
6	AT1G58225	NA	DNH: 6.50; NH: 1.50; D: -1.50	Tailored response	KEGG Genes Link
7	AT5G26340	Major facilitator superfamily protein	DNH: 2.70; NH: 1.50; D: 6.40	Similar expression/response under all stress	KEGG Genes Link
8	AT1G76970	Target of Myb protein 1	DNH: 6.10; NH: 2.40; D: -2.30	Tailored response	KEGG Genes Link
9	AT1G66880	Protein kinase superfamily protein	DNH: 6.00; NH: 2.50; D: -1.60	Tailored response	KEGG Genes Link
10	AT4G33150	Lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional enzyme	DNH: 5.60; NH: 3.00; D: -1.40	Tailored response	ath00310~Lysine degradation; ath01100~Metabolic pathways; ath01110~Biosynthesis of secondary metabolites

Showing 1 to 10 of 20 entries   Previous 1 2 Next

e) The user can get the details of the platform for transcriptome study, along with the plant ecotype, corresponding GEO or SRA ID and publication link, from the table highlighted in below screenshot.

Sl.no.	Gene ID	Gene name	Transcript Fold change (FC)	change in expression	Pathway & metadata
1	AT1G62300	WRKY family transcription factor	NH:8.7; D:1.8 ; DNH: -2.1	Tailored response	KEGG Genes Link
2	AT5G39520	Protein of unknown function (DUF1997)	DNH: 3.1; NH: 1.50; D: 8.11	Similar expression/response under all stress	KEGG Genes Link
3	AT5G38210	Protein kinase family protein	DNH: 2.40; NH: 2.30; D: 6.68	Similar expression/response under all stress	KEGG Genes Link
4	AT1G51860	Leucine-rich repeat protein kinase family protein	DNH: 6.70; NH: 2.90; D: -1.90	Tailored response	KEGG Genes Link
5	AT1G47890	Receptor like protein 7	DNH: 6.60; NH: 1.80; D: -1.50	Tailored response	KEGG Genes Link
6	AT1G58225	NA	DNH: 6.50; NH: 1.50; D: -1.50	Tailored response	KEGG Genes Link
7	AT5G26340	Major facilitator superfamily protein	DNH: 2.70; NH: 1.50; D: 6.40	Similar expression/response under all stress	KEGG Genes Link
8	AT1G76970	Target of Myb protein 1	DNH: 6.10; NH: 2.40; D: -2.30	Tailored response	KEGG Genes Link
9	AT1G66880	Protein kinase superfamily protein	DNH: 6.00; NH: 2.50; D: -1.60	Tailored response	KEGG Genes Link
10	AT4G33150	Lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional enzyme	DNH: 5.60; NH: 3.00; D: -1.40	Tailored response	ath00310~Lysine degradation; ath01100~Metabolic pathways; ath01110~Biosynthesis of secondary metabolites

Showing 1 to 10 of 20 entries   Previous 1 2 Next

Platform	Plant	Stress	GEO/SRA IDS	Publication
Gene Chip Gene 1.0ST, Cat#901915, Affymetrix, California, USA	Arabidopsis thaliana ecotype Col-0 (accession number#CS70000)	Drought - 40% FC & pathogen- Pseudomonas syringae pv. tabaci (Pta)- 3x10000000 CFU/mL, 5x10000000 CFU/mL and 2x100000000 CFU/mL	GSE79681	Choudhary et al., 2017

4) Transcriptome: Interactive Visualizations

Users are presented with various kinds of interactive visualization, in the transcriptomics section of SCIPDb.

I) Visualizations for common category of DEGs

1) Interactive Heat Map

a) The user can fetch the corresponding heatmap for the top 20 differentially expressed genes by clicking on the highlighted section shown below.

| 10 | AT4G33150 | Lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional enzyme | DNH: 5.60; NH: 3.00; D: -1.40 | Tailored response | ath00310~Lysine degradation; ath01100~Metabolic pathways; ath01110~Biosynthesis of secondary metabolites |

Showing 1 to 10 of 20 entries   Previous **1** 2 Next

Platform	Plant	Stress	GEO/SRA IDS	Publication
Gene Chip Gene 1.0ST, Cat#901915, Affymetrix, California, USA	Arabidopsis thaliana ecotype Col-0 (accession number#CS70000)	Drought - 40% FC & pathogen- Pseudomonas syringae pv. tabaci (Pta)- 3x10000000 CFU/mL, 5x10000000 CFU/mL and 2x100000000 CFU/mL	GSE79681	Choudhary et al., 2017

Click here ➡ **View Heat Map**   View Venn diagram   Download complete results

b) After clicking the "View Heat Map", the user will be directed to the page containing an interactive heat map as illustrated below.

c) Each column corresponds to the stress treatments (**Here Drought, nonhost pathogen and Drought and nonhost**)

d) Each row corresponds to Gene Ids.

d) Each of the cells are hoverable and displays the row (Gene ID) and column (treatment, here Drought) information
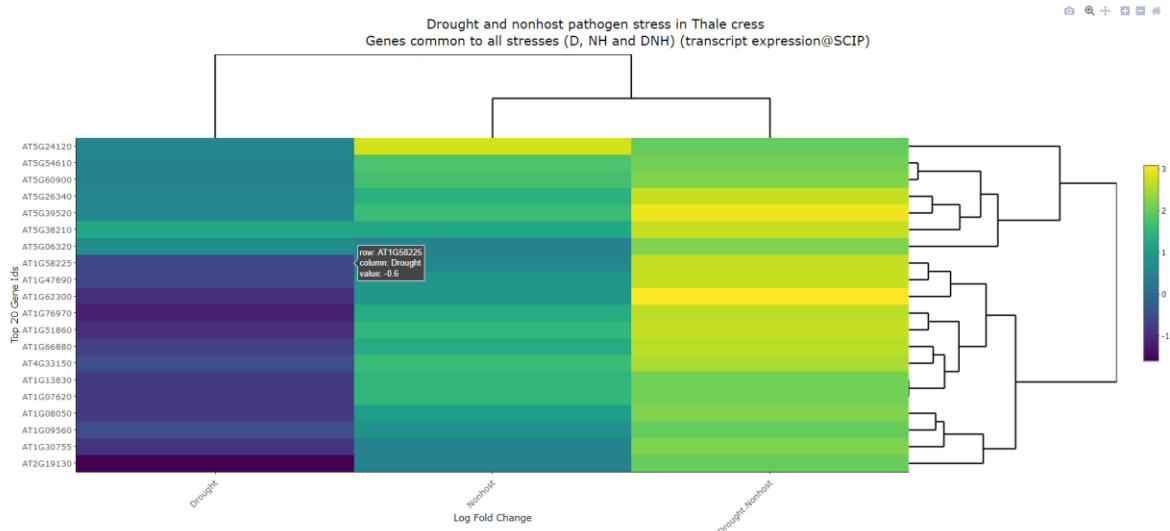
2) Venn diagram

a) The user can fetch the corresponding Venn diagram for the categories selected by clicking on the highlighted section shown below.

6	AT1G58225	NA	DNH: 6.50; NH: 1.50; D: -1.50	Tailored response	**KEGG Genes Link**
7	AT5G26340	Major facilitator superfamily protein	DNH: 2.70; NH: 1.50; D: 6.40	Similar expression/response under all stress	**KEGG Genes Link**
8	AT1G76970	Target of Myb protein 1	DNH: 6.10; NH: 2.40; D: -2.30	Tailored response	**KEGG Genes Link**
9	AT1G66880	Protein kinase superfamily protein	DNH: 6.00; NH: 2.50; D: -1.60	Tailored response	**KEGG Genes Link**
10	AT4G33150	Lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional enzyme	DNH: 5.60; NH: 3.00; D: -1.40	Tailored response	ath00310~Lysine degradation; ath01100~Metabolic pathways; ath01110~Biosynthesis of secondary metabolites

Showing 1 to 10 of 20 entries

Previous **1** 2 Next

| Platform	Plant	Stress	GEO/SRA IDS	Publication
Gene Chip Gene 1.0ST, Cat#901915, Affymetrix, California, USA	Arabidopsis thaliana ecotype Col-0 (accession number#CS70000)	Drought - 40% FC & pathogen- Pseudomonas syringae pv. tabaci (Pta)- 3x10000000 CFU/mL, 5x10000000 CFU/mL and 2x100000000 CFU/mL	GSE79681	Choudhary et al., 2017

Click here

View Heat Map   View Venn diagram   Download complete results

b) After clicking the "View Venn diagram", the user will be directed to the page containing the Venn diagram like the one below.

c) It depicts common and unique DEGs between the three stress treatments (**here Drought, nonhost pathogen and drought and nonhost)**.

Drought and nonhost pathogen stress in Thale cress
Venn diagram depicting common and unique genes between Drought (D), Nonhost (NH) and combined stress (DNH) @SCIP

II) Visualizations for unique category of DEGs

1) Interactive Manhattan plot

a) If the user has selected the Unique Genes category (highlighted below) then separate set of visualizations are presented to them.

expression i.e whether it is expressed uniquely or common or tailored in a particular stress combination.

STEP III: Query SCIP database for desired genes:

- The DEGs has been mapped to KEGG pathway wherever possible

- In cases where pathway information wsa not available, DEGs has been mapped to KEGG genes

Query for desired genes:

| Unique genes |
| Please select desired genes |
| Common genes : Genes common to all (D, NH and DNH) |
| Common genes : Genes common between drought (D) and nonhost pathogen (NH) |
| Common genes : Genes common between drought (D) and combined stress (DNH) |
| Common genes : Genes common between nonhost pathogen (NH) and combined stress (DNH) |
| Unique genes |

- KEGG Genes and pathway link can aid users to navigate through databases (E.g., TAIR, NCBI etc.) interlinked with SCIP etch other metadata like detailed gene annotation.

based output download of entire data is also

Reset Submit

b) The user can click on the highlighted section (shown below) to get the interactive gene ontology (GO) enrichment, Manhattan plot for this unique combined stress gene categories.

| 9 | AT5G24110 | WRKY DNA-binding protein 30 | 5.90 | | Uniquely expressed under drought and nonhost pathogen stress | KEGG Genes Link |
| 10 | AT4G39830 | Cupredoxin superfamily protein | 5.80 | | Uniquely expressed under drought and nonhost pathogen stress | KEGG Genes Link |

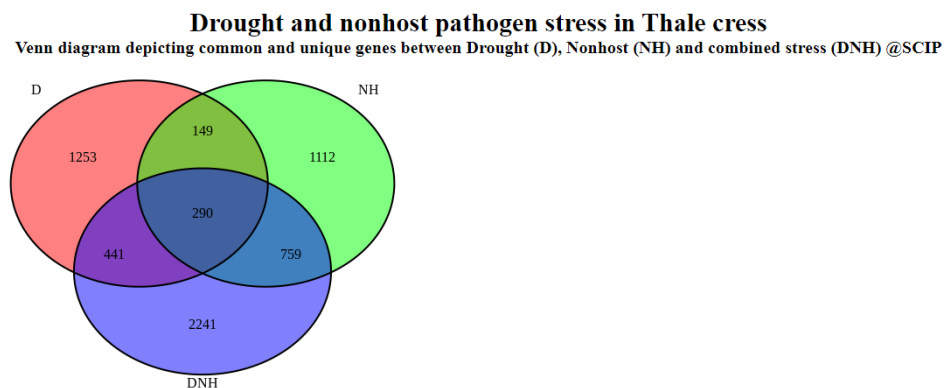Showing 1 to 10 of 20 entries   Previous **1** 2 Next

Platform	Plant	Stress	GEO/SRA IDS	Publication
Gene Chip Gene 1.0ST, Cat#901915, Affymetrix, California, USA	Arabidopsis thaliana ecotype Col-0 (accession number#CS70000)	Drought - 40% FC & pathogen- Pseudomonas syringae pv. tabaci (Pta)- 3x10000000 CFU/mL, 5x10000000 CFU/mL and 2x100000000 CFU/mL	GSE79681	Choudhary et al., 2017

Click Here ⟶ View GO Enrichment   View Co-functional network   Download complete results

c) After clicking the "View GO Enrichment", the user will be directed to the page containing the interactive Manhattan plot like the one below.

d) The Y-axis is specific to three GO functional categories, GO: Molecular function, GO: Cellular compartment and GO: Biological process. KEGG and other pathway databases are also shown here.

e) On X-axis enriched GO ids are shown in negative logarithm scale.

f) Each of the circle shown in plot below are interactive and on hover show other details of GO Ids and name.

g) The colour here depicts and corresponds to three GO functional categories, GO: Molecular function, GO: Cellular compartment and GO: Biological process. If pathway enrichment is detected that is also shown in different colour.

Manhattan plot depicting functional enrichment analysis of differentially expressed unique genes under Drought and non-host pathogen

h) The user can download the complete GO ontology enrichment results from the link highlighted in above screenshot.

II) Interactive Co-expression network

a) The user can also fetch the corresponding Co-Expression Network from the highlighted section shown below.

7	AT5G16900	Leucine-rich repeat protein kinase family protein	6.30		Uniquely expressed under drought	**KEGG Genes Link**
8	AT5G05300	Unknown	6.30		Uniquely expressed under drought and nonhost pathogen stress	**KEGG Genes Link**
9	AT5G24110	WRKY DNA-binding protein 30	5.90		Uniquely expressed under drought and nonhost pathogen stress	**KEGG Genes Link**
10	AT4G39830	Cupredoxin superfamily protein	5.80		Uniquely expressed under drought and nonhost pathogen stress	**KEGG Genes Link**

Showing 1 to 10 of 20 entries

Previous **1** 2 Next

| Platform	Plant	Stress	GEO/SRA IDS	Publication
Gene Chip Gene 1.0ST, Cat#901915, Affymetrix, California, USA	Arabidopsis thaliana ecotype Col-0 (accession number#CS70000)	Drought - 40% FC & pathogen- Pseudomonas syringae pv. tabaci (Pta)-3x10000000 CFU/mL, 5x10000000 CFU/mL and 2x100000000 CFU/mL	GSE79681	**Choudhary et al., 2017**

Click Here

View GO Enrichment   **View Co-functional network**   **Download complete results**

b) After clicking the "View Co-functional Network", the user will be directed to the page containing the interactive Co-Expression Network as shown below.
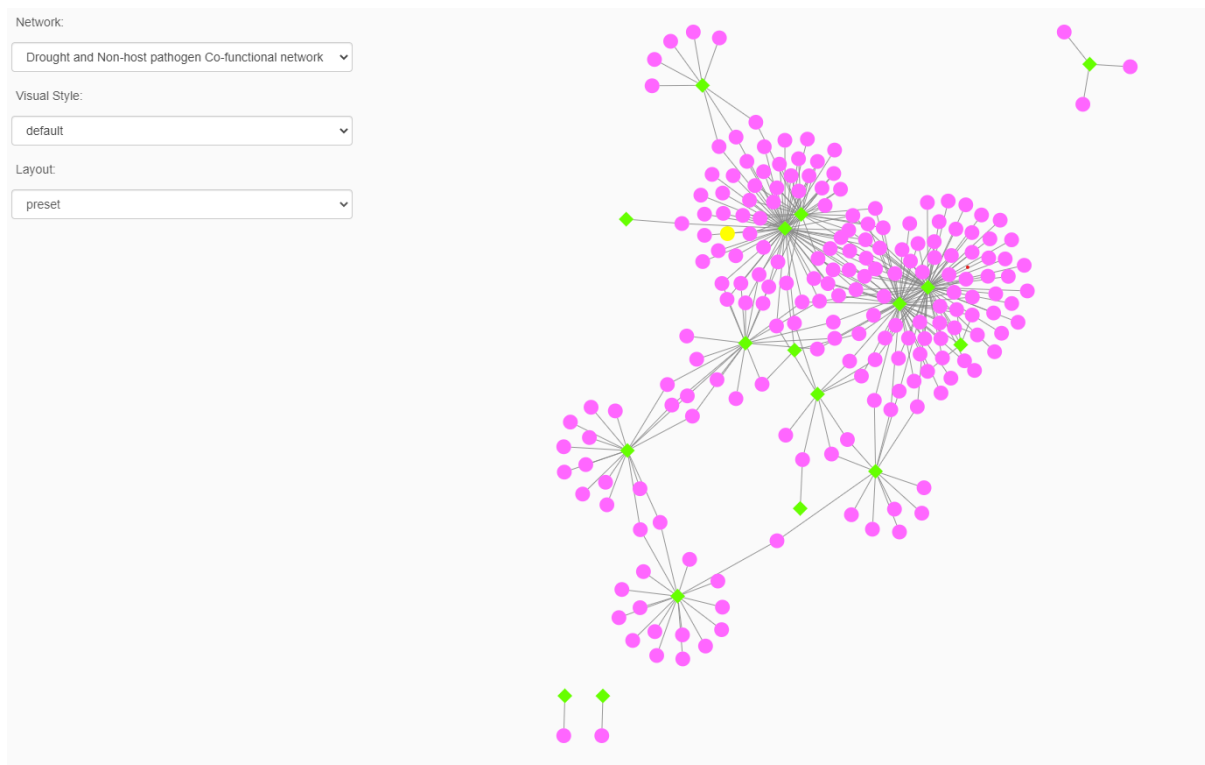
c) Nodes in green are the top 20 combined genes (sorted on Log FC) for the selected transcriptome.

d) Nodes in pink are their corresponding co-relates.

e) On selection nodes are highlighted in yellow, and other meta-data like GO information, paralog information and ranking of nodes selected are presented in the form of table, as shown below.

Copyright SCIPDb 2021
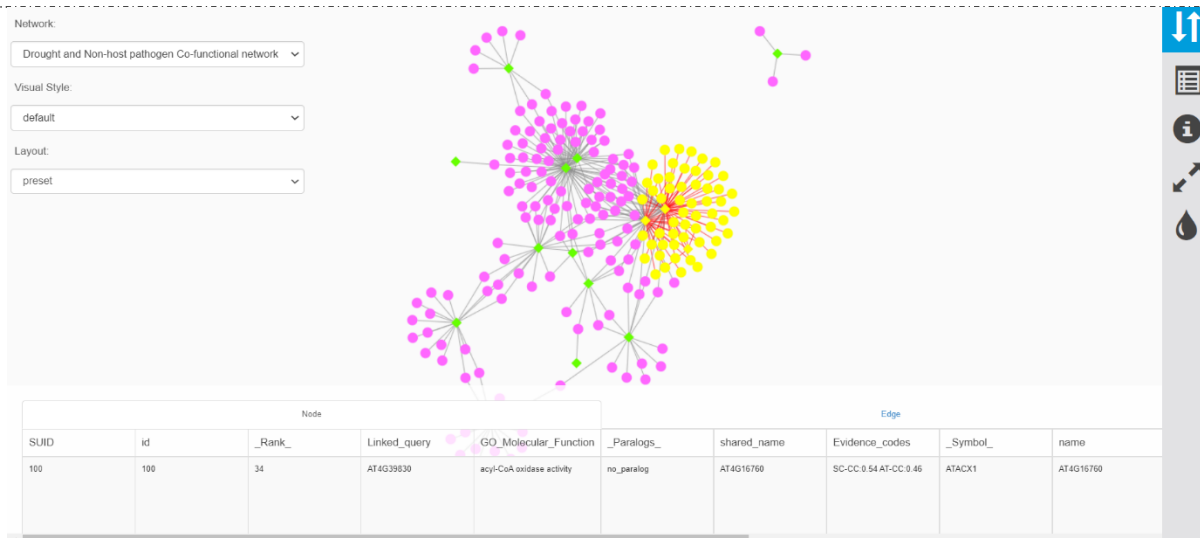
5) Transcriptome: Downloading data

a) The User can download the complete dataset from the "Download complete results" section which is highlighted below, both for common and unique genes categories.

				stress	
9	AT5G24110	WRKY DNA-binding protein 30	5.90	Uniquely expressed under drought and nonhost pathogen stress	KEGG Genes Link
10	AT4G39830	Cupredoxin superfamily protein	5.80	Uniquely expressed under drought and nonhost pathogen stress	KEGG Genes Link

Showing 1 to 10 of 20 entries   Previous **1** 2 Next

Platform	Plant	Stress	GEO/SRA IDS	Publication
Gene Chip Gene 1.0ST, Cat#901915, Affymetrix, California, USA	Arabidopsis thaliana ecotype Col-0 (accession number#CS70000)	Drought - 40% FC & pathogen- Pseudomonas syringae pv. tabaci (Pta)- 3x10000000 CFU/mL, 5x10000000 CFU/mL and 2x100000000 CFU/mL	GSE79681	Choudhary et al., 2017

View GO Enrichment   View Co-functional network   **Download complete results** ← Click Here