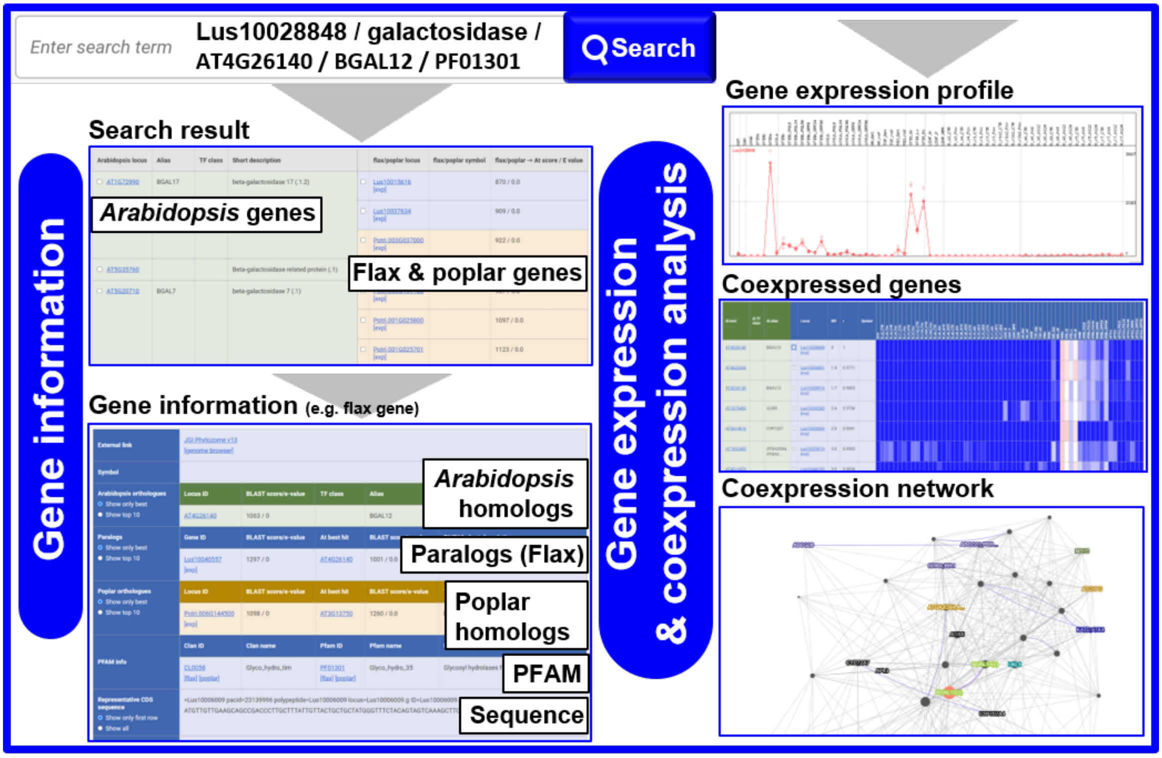
FIBexDB
Manual

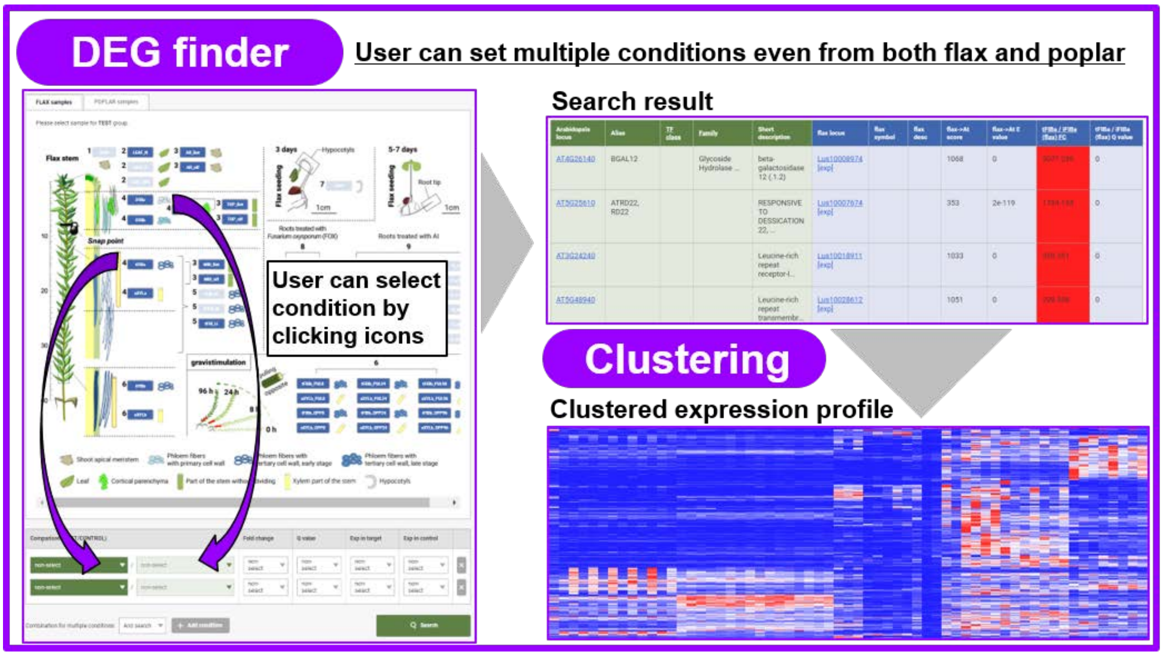
Quick start

There are two major ways to use FIBexDB: (i) browsing expression patterns of single or multiple genes and finding coexpressed genes by keyword (**Keyword search**), and (ii) finding up/downregulated genes in particular situations/tissues (**DEG finder**). Gene information contains information about gene symbol, Arabidopsis homologs, flax/poplar paralogs/homologs, PFAM, and sequence. Expression profile, coexpressed genes, and coexpression network are also shown. In DEG finder, users can set multiple search conditions even from both flax and poplar simultaneously.

Keyword search



DEG finder



Feel free to send a question, comment and/or bug report to Nobutaka Mitsuda (nobutaka.mitsuda@aist.go.jp).

To cite FIBexDB

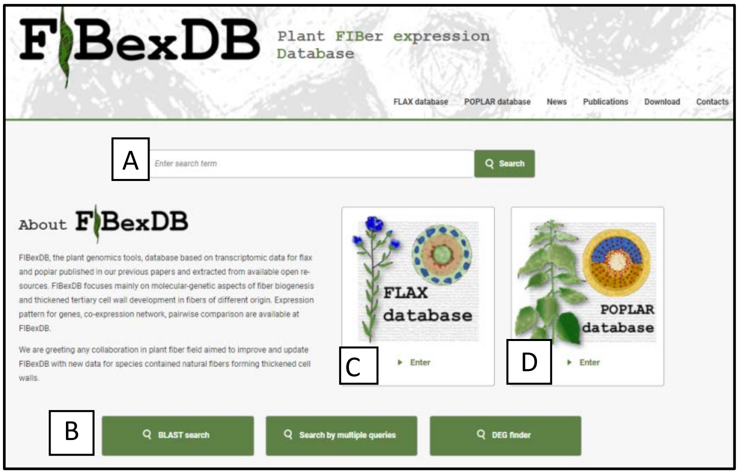
Mokshina N., Gorshkov O., Takasaki H., Onodera H., Sakamoto S., Gorshkova T., Mitsuda N. FIBexDB: A New Online Transcriptome Platform to Analyze Development of Plant Cellulosic Fibers. *Journal*, X: xxx-xxx (2021)

Introduction

FIBexDB is a database of gene expression profiles in different tissues of flax and poplar and at the different stages of plant development, with a major focus on samples bearing fibers with the thickened (tertiary) cell wall. For flax, we collected RNA-seq data for 61 samples from 9 different projects, comprising transcriptomes for 7 tissue types (shoot apical meristem, phloem fibers at different stages of development, core parenchyma, xylem tissue, hypocotyls, roots, and leaves), fibers from different flax cultivars and subspecies, as well as parts of mutant stems (reduced fibers); a complete list of samples with their detailed description can be found here. For poplar, we collected RNA-seq data for 88 samples from 7 projects, comprising transcriptomes of roots, leaves, normal wood, tension wood of wild type and transgenic poplars, and samples after treatments related to tension wood formation; a complete list of samples with their detailed description can be found here. To bridge the genes from two species, information regarding homologous relationships between flax, poplar and *Arabidopsis* is stored.

1. Main (top/home) page:

Top page includes;
Query box for keyword search **(A)**. You can paste a gene symbol, keyword, gene identifier (Lus100XXXXX, PotriXXXGXXXXXX, ATXGXXXXX) or PFAM number (PFxxxx). In case of symbol, use abbreviated names with numbers (for example, like "CESA8" but not like "CESA") **(see 2. for the search result.)**. Links for **BLAST search**, **Search by multiple queries**, and integrated search for differentially expressed genes (**DEG finder**) in both flax and poplar **(B; see 4., 5., 7. for detail)**. Link to a flax database **(C)**, which only searches flax genes. Link to a poplar database **(D)**, which only searches poplar genes.
The search from the top page of FIBexDB only searches genes with Arabidopsis homologue while a search from the top page of each plant searches genes without Arabidopsis homolog, too.



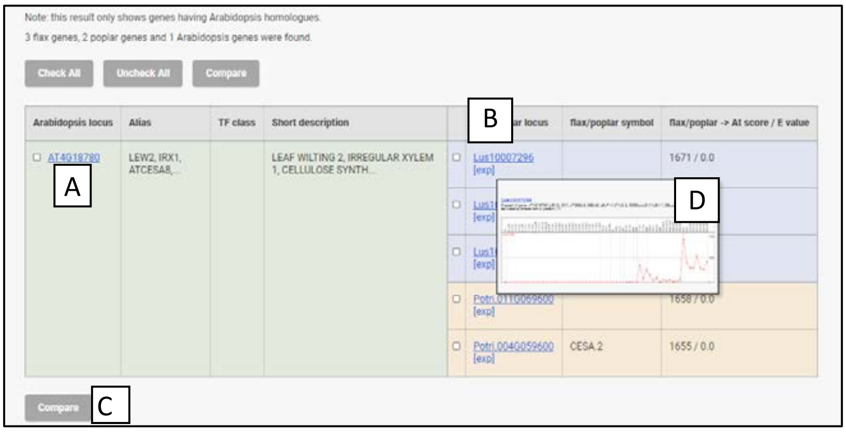
FIBexDB, the plant genomics tools, database based on transcriptomic data for flax and poplar published in our previous papers and extracted from available open resources. FIBexDB focuses mainly on molecular-genetic aspects of fiber biogenesis and thickened tertiary cell wall development in fibers of different origin. Expression pattern for genes, co-expression network, pairwise comparison are available at FIBexDB.

We are greeting any collaboration in plant fiber field aimed to improve and update FIBexDB with new data for species contained natural fibers forming thickened cell walls.

Fig.1 FIBexDB top page
(A) Query box for keyword search. **(B)** Links to BLAST search, search by multiple queries, and DEG finder. **(C)** Link to flax database, which only searches flax gene. **(D)** Link to poplar database, which only searches poplar gene.

2. Result of keyword search

The list of genes found in flax and poplar with their closest homologous *Arabidopsis* gene is given in the form of a table **(Fig. 2)**. Arabidopsis locus **(A)** is linked to corresponding TAIR page and flax/poplar locus **(B)** is linked to corresponding gene page in FIBexDB. You can select several genes and can compare gene expression profile by clicking **Compare** button **(C; see 5. for the result.)**. When you hover a mouse cursor on a "[exp]" hyperlink, you can preview the expression graph **(D)** without actually going to the corresponding page. When you click gene ID, the page for each gene in FIBexDB is shown (**see 3.**).



Note: this result only shows genes having Arabidopsis homologues.

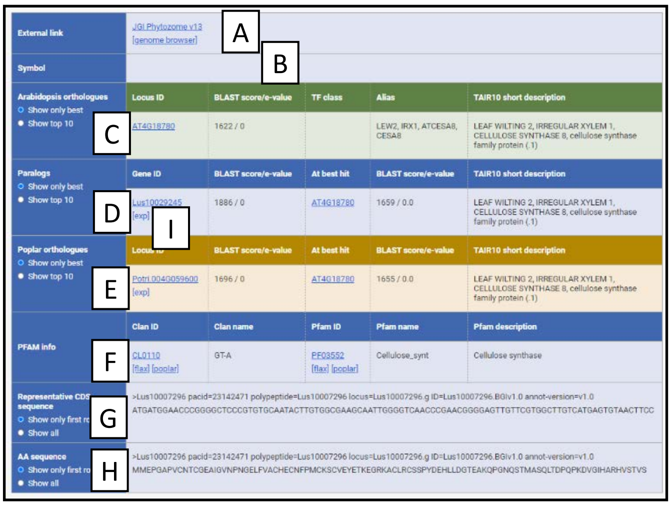
3 flax genes, 2 poplar genes and 1 Arabidopsis genes were found.

Arabidopsis locus	Alias	TF class	Short description	ar locus	flax/poplar symbol	flax/poplar -> At score / E value
☐ AT4G18780	LEW2, IRX1, ATCESA8,...		LEAF WILTING 2, IRREGULAR XYLEM 1, CELLULOSE SYNTH...	☐ Lus10007296 [exp]		1671 / 0.0
				☐ Lus1 [exp]		
				☐ Lus1 [exp]		
				☐ Potri.011G069600 [exp]		1658 / 0.0
				☐ Potri.004G059600 [exp]	CESA.2	1655 / 0.0

Fig. 2 Result of keywrord search
(A) Closest Arabidopsis gene of flax / poplar found genes. **(B)** Flax / poplar found genes. **(C)** "Compare" button to compare gene expression profiles of selected genes. **(D)** Preview of gene expression profile of each gene which appears when a mouse cursor is hovered on the "[exp]" hyperlink.

3. Information page for each gene

Information page for each gene is consisted of 4 consequent blocks (**Figs. 3-7**):
3-1. Basic information table (Fig. 3) contains;
External links (**A;** to corresponding page of JGI Phytozome v13 and [genome browser] in flax and poplar, as well as Popgenie and AspWood for poplar)
Gene symbol (**B;** if exists)
Arabidopsis homologues (**C;** show only best / show top 10) linked to TAIR DB
Paralogs within same species (**D;** show only best / show top 10) linked to FIBexDB corresponding gene page
Poplar / flax homologues (**E;** show only best/show top 10) linked to FIBexDB corresponding gene page
PFAM info including Clan/Pfam ID and name according to Pfam DB (**F**), linked to the search result of FIBexDB by the Clan/Pfam ID
Representative CDS sequence (**G;** Show only first row / Show all)
AA sequence (**H;** Show only first row / Show all).
The "[exp]" hyperlink in "Paralogs" and "Poplar / flax homologues" **(I)** shows a preview of gene expression profile when mouse cursor is hovered.



External link: JGI Phytozome v13 [genome browser]

Symbol

Arabidopsis orthologues
○ Show only best
● Show top 10

Locus ID	BLAST score/e-value	TF class	Alias	TAIR10 short description
AT4G18780	1622 / 0		LEW2, IRX1, ATCESA8, CESA8	LEAF WILTING 2, IRREGULAR XYLEM 1, CELLULOSE SYNTHASE 8, cellulose synthase family protein (.1)

Paralogs
○ Show only best
● Show top 10

Gene ID	BLAST score/e-value	At best hit	BLAST score/e-value	TAIR10 short description
Lus10029245 [exp]	1886 / 0	AT4G18780	1659 / 0.0	LEAF WILTING 2, IRREGULAR XYLEM 1, CELLULOSE SYNTHASE 8, cellulose synthase family protein (.1)

Poplar orthologues
○ Show only best
● Show top 10

Locus ID	BLAST score/e-value	At best hit	BLAST score/e-value	TAIR10 short description
Potri.004G059600 [exp]	1696 / 0	AT4G18780	1655 / 0.0	LEAF WILTING 2, IRREGULAR XYLEM 1, CELLULOSE SYNTHASE 8, cellulose synthase family protein (.1)

PFAM info

Clan ID	Clan name	Pfam ID	Pfam name	Pfam description
CL0110 [flax] [poplar]	GT-A	PF03552 [flax] [poplar]	Cellulose_synt	Cellulose synthase

Representative CDS sequence
○ Show only first ro
● Show all

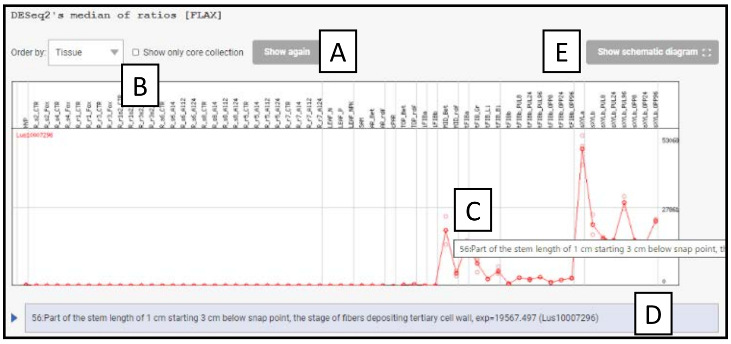
```
>Lus10007296 pacid=23142471 polypeptide=Lus10007296 locus=Lus10007296.g ID=Lus10007296.BGIv1.0 annot-version=v1.0
ATGATGGAACCCGGGGCTCCCGTGTGCAATACTTGTGGCGAAGCAATTGGGGTCAACCCGAACGGGGAGTTGTTCGTGGCTTGTCATGAGTGTAACTTCC
```

AA sequence
○ Show only first ro
● Show all

```
>Lus10007296 pacid=23142471 polypeptide=Lus10007296 locus=Lus10007296.g ID=Lus10007296.BGIv1.0 annot-version=v1.0
MMEPGAPVCNTCGEAIGVNPNGELFVACHECNFPMCKSCVEYETKEGRKACLRCSSPYDEHLLDGTEAKQPGNQSTMASQLTDPQPKDVGIHARHVSTVS
```

Fig. 3 Information page for each gene
(A) External links. **(B)** Gene symbol. **(C)** Arabidopsis homologues. **(D)** Paralogs within same species. **(E)** Poplar / flax homologues. **(F)** PFAM information. **(G)** Representative CDS sequence. **(H)** Amino-acid sequence. **(I)** Link for a preview of gene expression profile.

3-2. DESeq2's median of ratios (expression profile of each gene) are shown as a line graph **(Fig. 4)**. The order of the samples can be switched to "Tissue" or "Experiment" (**A;** select and click Show again to change). "Core collection" **(B)** set for flax and poplar includes the samples collected from stems of plants grown in normal conditions (without additional treatments) and samples from experiments performed in wild-type plants, respectively. A short sample description and expression value appear on the graph **(C)** and text box below **(D)** when you hover a mouse cursor on the corresponding point in the graph. Show schematic diagram **(E)** shows schematic illustration of sample collection.



56:Part of the stem length of 1 cm starting 3 cm below snap point, t

▸ 56:Part of the stem length of 1 cm starting 3 cm below snap point, the stage of fibers depositing tertiary cell wall, exp=19567.497 (Lus10007296)

Fig. 4 Graph of expression profile of each gene
(A) "Show again" button to switch the order of samples ("tissue" or "experiment"). **(B)** Checkbox to show only "core collection" samples. **(C and D)** Sample description and expression value which appear when a mouse cursor is hovered on each point.

3-3. List of coexpressed genes is shown in table format (**Fig. 5**). As a default, only top 10 genes are shown. Click Show all **(A)** to get a full list of coexpressed genes (r > 0.6 [Pearson's correlation coefficient] [max limit 300 genes]). Compare expression of checked genes **(B)** allows you to compare expression profiles of selected genes **(see 5.)**. Download tab-delimited text **(C)** allows you to download the resultant table as a tab-delimited text file. Arabidopsis gene **(D)** in the table is linked to TAIR DB. Locus **(E)** in the table is linked to the gene information page of FIBexDB. The "[exp]" hyperlink shows a preview of gene expression profile when mouse cursor is hovered. At TF class column **(F)** represents a belonging TF family based on Mitsuda et al. 2009, *Plant Cell Physiol.*, 50: 1232-1248. MR column **(G)** represents a mutual rank of coexpression between a pair of genes (proposed as an alternative to PCC measure of coexpression, Obayashi and Kinoshita, 2009, *DNA research*, 16: 249-260). r column **(G)** represents a Pearson's correlation coefficient of expression profiles between the query gene and each gene. Color of the heat map represents relative expression level among samples for each gene and therefore doesn't represent the absolute expression level. Sample name and expression value appear when a mouse cursor is hovered on each cell **(H)**.

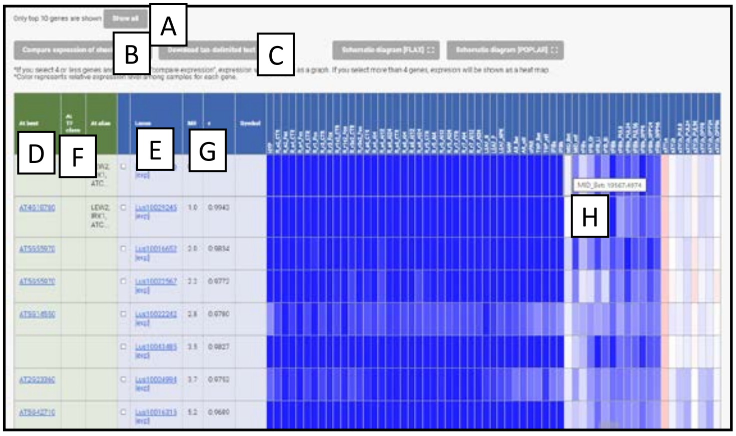
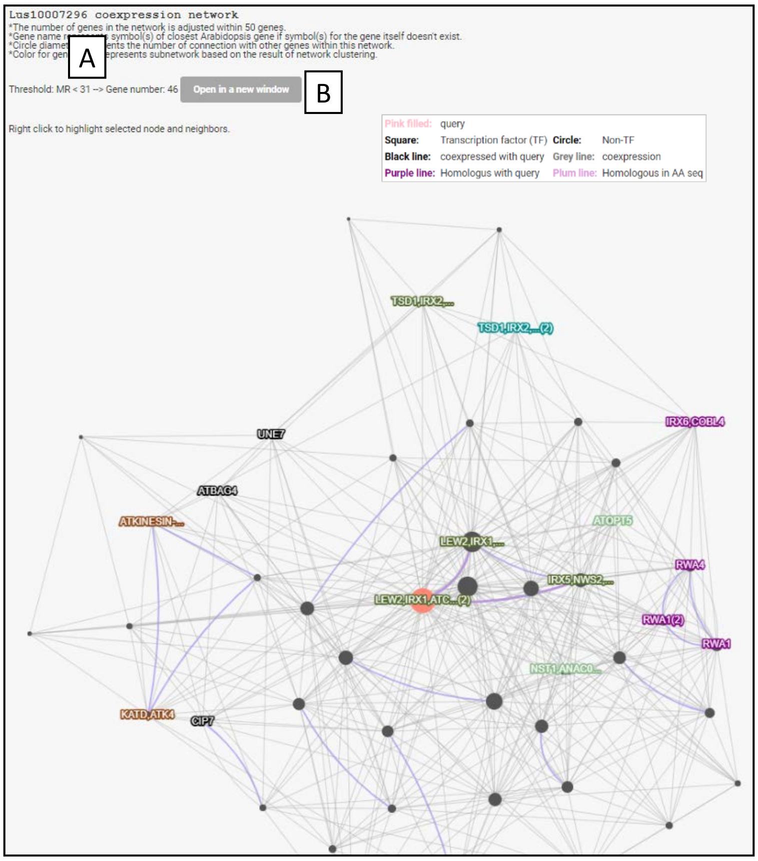


Fig. 5 List of coexpressed genes
(A) "Show all" button to show all coexpressed genes with r > 0.6 (max limit 300 genes). **(B)** "Compare expression of checked genes" button to show comparison of expression profiles of selected genes. **(C)** "Download tab-delimited text" button to allow you to download the expression profiles as a tab-delimited text file. **(D)** Closest Arabidopsis gene. **(E)** Flax/poplar coexpressed genes. **(F)** Transcription factor family of Arabidopsis gene. **(G)** Pearson's correlation coefficient (r) of expression profiles between a query gene and each gene and a mutual geometically averaged rank (MR) of them. **(H)** Sample name and expression value of the cell where a mouse cursor is hovered.

3-4. Coexpression network (Fig. 6)

The number of genes in the network is adjusted within 50 genes by changing cut-off value of MR **(A)**. Gene name represents symbol(s) of closest *Arabidopsis* gene if the symbol(s) for the gene itself doesn't exist. Size of each node represents the number of connection with other genes within this network. Color for gene name represents subnetwork based on the result of network clustering. Open in a new window **(B)** opens a new window only showing the network. Square node represents transcription factor (TF) gene. Circle represents non-TF. Pink-filled node represents the query gene. Black line represents coexpression with the query gene. Grey line represents coexpressoin. Purple line represents amino-acid sequence homology with the query gene. Plum line represents amino-acid sequence homology. You can move each node to anywhere you want by mouse dragging. Left click on each node moves you to the page for each gene. Right click on each node highlights the node, connected nodes, and connections. Right click on anywhere without node or edge cancels the highlight. A list of genes from the coexpressed network is presented in the table (Fig. 7).



Lus10007296 coexpression network

*The number of genes in the network is adjusted within 50 genes.
*Gene name r s symbol(s) of closest Arabidopsis gene if symbol(s) for the gene itself doesn't exist.
*Circle diamet ents the number of connection with other genes within this network.
*Color for gen epresents subnetwork based on the result of network clustering.

Threshold: MR < 31 --> Gene number: 46

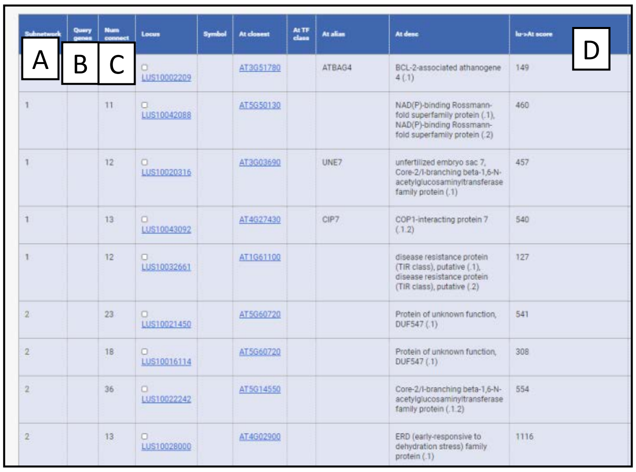
Right click to highlight selected node and neighbors.

Pink filled:	query		
Square:	Transcription factor (TF)	Circle:	Non-TF
Black line:	coexpressed with query	Grey line:	coexpression
Purple line:	Homologus with query	Plum line:	Homologous in AA seq

Fig. 6 Coexpression network
(A) Adopted Cut-off value of MR regarding coexpression to adjust the number of genes in the network within 50 genes **(B)** "Open in a new window" opens a new window only showing the network.

3-5. List of genes in the coexpression network (Fig. 7)

Genes appeared in the coexpression network (**Fig. 6**) is listed in the table below the network (**Fig. 7**). **Subnetwork** column **(A)** represents the subnetwork ID shown by the same color in the network. **Query genes** column **(B)** shows "Y" for the query gene of the coexpression network analysis. **Num connect** **(C)** indicates number of connections from each gene. **lu-> At score**(**pt-> At score**) and **lu-> At E val** (**pt-> At E value**) **(D)** represent score and E-value of BLAST search of flax/poplar gene against *Arabidopsis* closest gene, respectively.

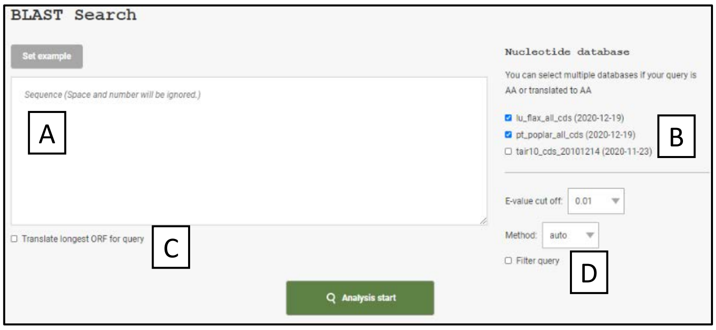


Subnetwork	Query genes	Num connect	Locus	Symbol	At closest	At TF class	At alias	At desc	lu->At score
			☐ LUS10002209		AT3G51780		ATBAG4	BCL-2-associated athanogene 4 (.1)	149
1		11	☐ LUS10042088		AT5G50130			NAD(P)-binding Rossmann-fold superfamily protein (.1), NAD(P)-binding Rossmann-fold superfamily protein (.2)	460
1		12	☐ LUS10020316		AT3G03690		UNE7	unfertilized embryo sac 7, Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein (.1)	457
1		13	☐ LUS10043092		AT4G27430		CIP7	COP1-interacting protein 7 (.1.2)	540
1		12	☐ LUS10032661		AT1G61100			disease resistance protein (TIR class), putative (.1), disease resistance protein (TIR class), putative (.2)	127
2		23	☐ LUS10021450		AT5G60720			Protein of unknown function, DUF547 (.1)	541
2		18	☐ LUS10016114		AT5G60720			Protein of unknown function, DUF547 (.1)	308
2		36	☐ LUS10022242		AT5G14550			Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein (.1.2)	554
2		13	☐ LUS10028000		AT4G02900			ERD (early-responsive to dehydration stress) family protein (.1)	1116

Fig. 7 List of genes in the coexpression network
(A) Subnetwork ID shown by the same color in the network. **(B)** "Y" is shown for the query gene. **(C)** The number of connections from each gene. **(D)** BLAST score and E-value from flax/poplar genes to Arabidopsis closest gene.

4. BLAST search

The BLAST (Basic Local Alignment Search Tool) tool is implemented to find out flax and poplar genes by sequence homology (**Fig. 8**). You can simply paste the nucleotide (or amino acids [AA]) sequence (with or without a FASTA header) into the query sequence box (**A;** space and digit will be ignored). Select the desired dataset from the list of available databases **(B)**. You can select multiple databases if your search is intended as TBLASTN search. "Translate longest ORF for query" option **(C)** allows you to use the translated amino-acids sequence of the longest ORF as a query. "Filter query" option **(D)** allows you to mask the sequence with low complexity.



BLAST Search

Sequence (Space and number will be ignored.)

☐ Translate longest ORF for query

Nucleotide database

You can select multiple databases if your query is AA or translated to AA

☑ lu_flax_all_cds (2020-12-19)
☑ pt_poplar_all_cds (2020-12-19)
☐ tair10_cds_20101214 (2020-11-23)

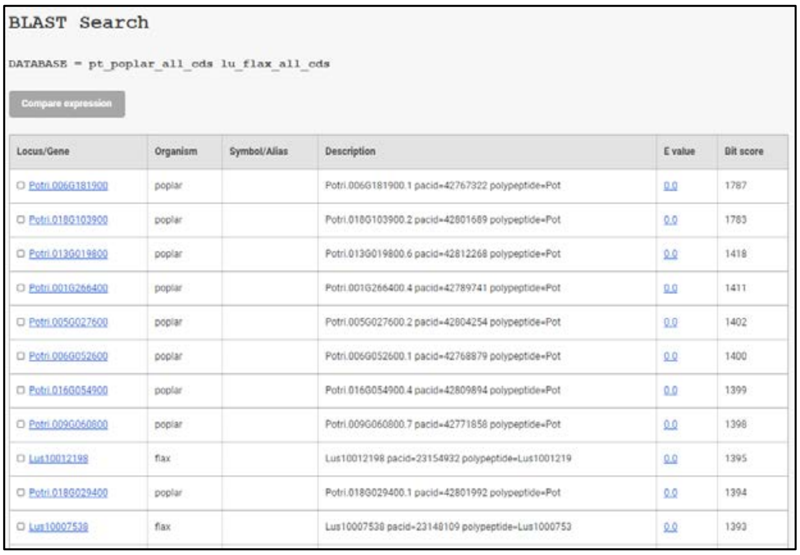
E-value cut off: 0.01

Method: auto

☐ Filter query

Fig. 8 BLAST search input
(A) Query sequence box. **(B)** Available database. **(C)** "Translate longest ORF for query" allows you to use the translated amino-acids sequence of the longest ORF as a query **(D)** "Filter query" option allows you to mask the sequence with low complexity.

BLAST search result is organized into a table format (**Fig. 9**) containing:
Locus/Gene
Organism (flax, poplar)
Symbol/Alias
Description
E value
Bit score
Compare expression allows you to compare expression profiles of selected genes.



BLAST Search

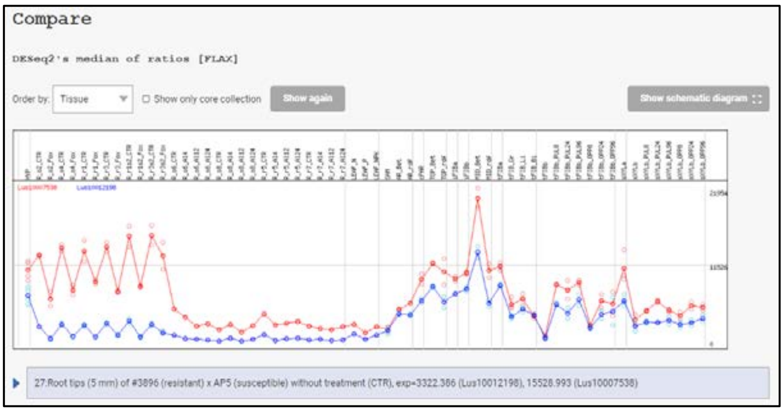
DATABASE = pt_poplar_all_cds lu_flax_all_cds

Locus/Gene	Organism	Symbol/Alias	Description	E value	Bit score
☐ Potri.006G181900	poplar		Potri.006G181900.1 pacid=42767322 polypeptide=Pot	0.0	1787
☐ Potri.018G103900	poplar		Potri.018G103900.2 pacid=42801689 polypeptide=Pot	0.0	1783
☐ Potri.013G019800	poplar		Potri.013G019800.6 pacid=42812268 polypeptide=Pot	0.0	1418
☐ Potri.001G266400	poplar		Potri.001G266400.4 pacid=42789741 polypeptide=Pot	0.0	1411
☐ Potri.005G027600	poplar		Potri.005G027600.2 pacid=42804254 polypeptide=Pot	0.0	1402
☐ Potri.006G052600	poplar		Potri.006G052600.1 pacid=42768879 polypeptide=Pot	0.0	1400
☐ Potri.016G054900	poplar		Potri.016G054900.4 pacid=42809894 polypeptide=Pot	0.0	1399
☐ Potri.009G060800	poplar		Potri.009G060800.7 pacid=42771858 polypeptide=Pot	0.0	1398
☐ Lus10012198	flax		Lus10012198 pacid=23154932 polypeptide=Lus1001219	0.0	1395
☐ Potri.018G029400	poplar		Potri.018G029400.1 pacid=42801992 polypeptide=Pot	0.0	1394
☐ Lus10007538	flax		Lus10007538 pacid=23148109 polypeptide=Lus1000753	0.0	1393

Fig. 9 Result of BLAST search

5. Comparison of multiple expression profiles

Expression profiles of multiple genes can be browsed in FIBexDB. If you select 4 or less genes from the result tables of **coexpressed genes**, **genes in the coexpression network**, **BLAST search result**, or **Cluster Analysis (see 6.)**, overlayed expression profiles of selected genes are shown in a graph (**Fig. 10**).



Compare

DESeq2's median of ratios [FLAX]

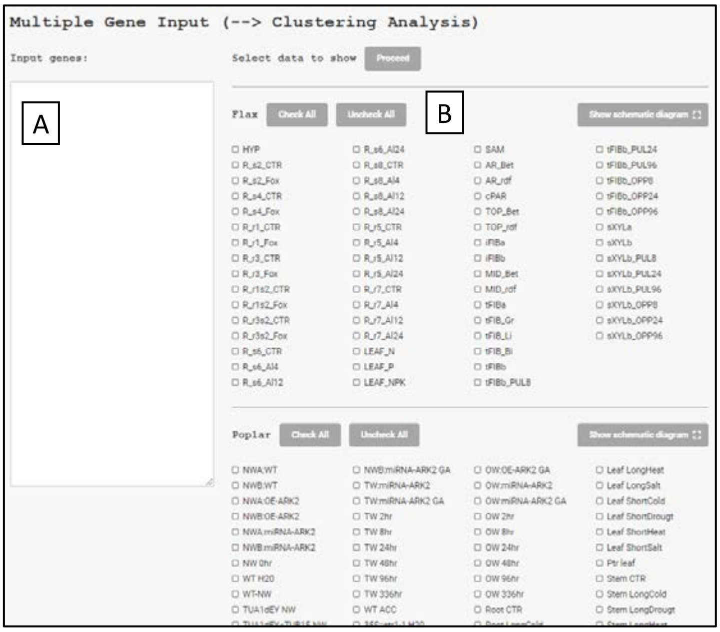
Order by: Tissue

☐ Show only core collection

27:Root tips (5 mm) of #3896 (resistant) x AP5 (susceptible) without treatment (CTR), exp=3322.386 (Lus10012198), 15528.993 (Lus10007538)

Fig. 10 Comparison of expression profiles of 4 or less genes

If you select more than 4 genes from above-mentioned result tables or use "Search by multiple queries" function (**Fig. 11**) from top pages of FIBexDB, heat-map table is shown as a result (**Fig. 12**). You can paste a list of flax/poplar genes **(A)** and select required samples from the list **(B)**.



Multiple Gene Input (--> Clustering Analysis)

Input genes:

Select data to show

Flax

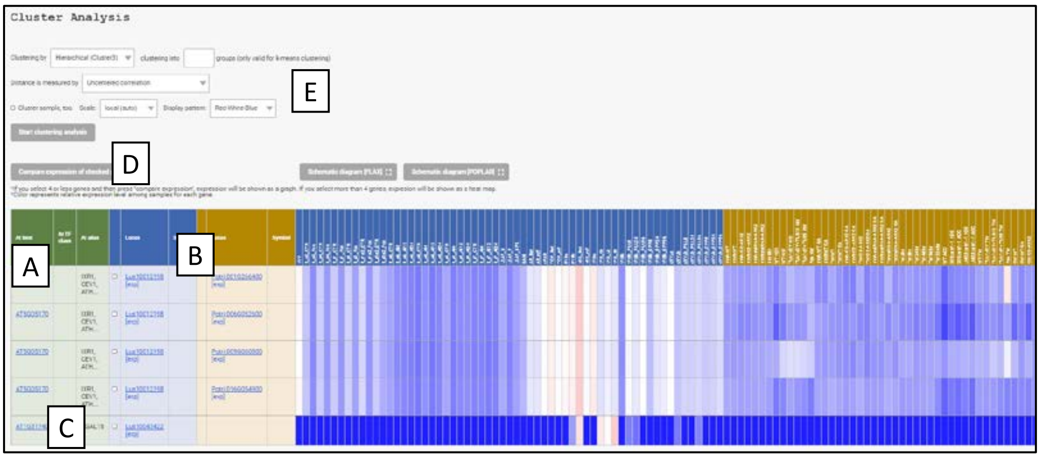
☐ HYP
☐ R_s2_CTR
☐ R_s2_Fox
☐ R_s4_CTR
☐ R_s4_Fox
☐ R_r1_CTR
☐ R_r1_Fox
☐ R_r3_CTR
☐ R_r3_Fox
☐ R_r1s2_CTR
☐ R_r1s2_Fox
☐ R_r3s2_CTR
☐ R_r3s2_Fox
☐ R_s6_CTR
☐ R_s6_Al4
☐ R_s6_Al12
☐ R_s6_Al24
☐ R_s8_CTR
☐ R_s8_Al4
☐ R_s8_Al12
☐ R_s8_Al24
☐ R_r5_CTR
☐ R_r5_Al4
☐ R_r5_Al12
☐ R_r5_Al24
☐ R_r7_CTR
☐ R_r7_Al4
☐ R_r7_Al12
☐ R_r7_Al24
☐ LEAF_N
☐ LEAF_P
☐ LEAF_NPK
☐ SAM
☐ AR_Bet
☐ AR_rdf
☐ cPAR
☐ TOP_Bet
☐ TOP_rdf
☐ iFIBa
☐ iFIBb
☐ MID_Bet
☐ MID_rdf
☐ tFIBa
☐ tFIB_Gr
☐ tFIB_Li
☐ tFIB_Bi
☐ tFIBb
☐ tFIBb_PUL8
☐ tFIBb_PUL24
☐ tFIBb_PUL96
☐ tFIBb_OPP8
☐ tFIBb_OPP24
☐ tFIBb_OPP96
☐ sXYLa
☐ sXYLb
☐ sXYLb_PUL8
☐ sXYLb_PUL24
☐ sXYLb_PUL96
☐ sXYLb_OPP8
☐ sXYLb_OPP24
☐ sXYLb_OPP96

Poplar

☐ NWA:WT
☐ NWB:WT
☐ NWA:OE-ARK2
☐ NWB:OE-ARK2
☐ NWA:miRNA-ARK2
☐ NWB:miRNA-ARK2
☐ NW 0hr
☐ WT H20
☐ WT-NW
☐ TUA1dEY NW
☐ NWB:miRNA-ARK2 GA
☐ TW:miRNA-ARK2
☐ TW:miRNA-ARK2 GA
☐ TW 2hr
☐ TW 8hr
☐ TW 24hr
☐ TW 48hr
☐ TW 96hr
☐ TW 336hr
☐ WT ACC
☐ OW:OE-ARK2 GA
☐ OW:miRNA-ARK2
☐ OW:miRNA-ARK2 GA
☐ OW 2hr
☐ OW 8hr
☐ OW 24hr
☐ OW 48hr
☐ OW 96hr
☐ OW 336hr
☐ Root CTR
☐ Leaf LongHeat
☐ Leaf LongSalt
☐ Leaf ShortCold
☐ Leaf ShortDrougt
☐ Leaf ShortHeat
☐ Leaf ShortSalt
☐ Ptr leaf
☐ Stem CTR
☐ Stem LongCold
☐ Stem LongDrougt

Fig. 11 Search by multiple queries
(A) You can paste a list of flax/poplar genes. **(B)** You can select required samples.

Result is shown as a heat-map table (**Fig. 12**) but for up to 10 genes as a default and the entire retrieved data can be sent to cluster analysis **(see 6.)**. Color of the heat map represents relative expression level among samples for each gene and therefore doesn't represent the absolute expression level. If you enter this function from the top page of FIBexDB, only genes with Arabidopsis homologue **(A)** are shown with the data of flax/poplar homologues (**B;** if exist). If the found particular Arabidopsis gene doesn't have a homolog in either species, then expression value for missing species is shown as 0 **(C)**. You can send the retrieved data to clustering analysis **(D)** after setting parameters **(E)**. Distance between genes / samples can be calculated by one of following methods; "**Uncentered correlation (default)", "Pearson correlation", "Uncentered correlation, absolute value", "Pearson correlation, absolute value",** "**Spearman's rank correlation", "Kendall's tau", "Euclidean distance", "City-block distance".** Coloration of the heat map in the result page of clustering analysis can be selected from "local (auto)" and "global (auto)". "local (auto)" determines cell color dependent on expression level among all samples of each gene. "global (auto)" determines cell color dependent on expression level among all cells and therefore reflects absolute expression level.


Cluster Analysis

Clustering by: Hierachical (Cluster3) clustering into ☐ groups (only valid for k-means clustering)

Distance is measured by: Uncentered correlation

☐ Cluster sample, too Scale: local (auto) Display pattern: Red-White-Blue

*If you select 4 or less genes and then press "compare expression", expression will be shown as a graph. If you select more than 4 genes, expression will be shown as a heat map.
*Color represents relative expression level among samples for each gene.

Fig. 12 The result of multiple gene input
(A) Homologous Arabidopsis gene is shown. **(B)** flax/poplar gene is shown. **(C)** A gene with a homolog only in either species shows 0 expression value in the missing species. **(D)** You can send the entire retrieved data to clustering analysis. **(E)** You can select two modes of clustering namely, hieralchical clustering and k-means clustering. The number of groups should be designated for k-means clustering. If you check "cluster sample, too", then the order of samples is changed according to the expression similarity of samples.

6. Clustering Analysis

FIBexDB can perform clustering analysis after showing expression profiles of multiple genes to sort out the listed genes based on expression profile. Hierarchical clustering changes order of genes based on expression profile of each gene (**Fig. 13**). The phylogeny of the genes is shown in left **(A)** but this may be omitted if the number of genes is too many. The phylogeny of samples is shown in top **(B)** if the "cluster sample, too" option is selected. If You enter "Search by multiple queries" from the top page of FIBexDB and select samples from both flax and poplar, expression in both plants is shown. If you set conditions for both flax and poplar in the "DEG finder" and send the resultant DEGs to the clustering analysis, it's same. You can extract the particular clade of genes by selecting cluster ID shown in left **(A)** and press "Extract checked genes as queries for multiple search" **(C)**. This sends extracted genes to multiple gene input (**Fig. 11**).

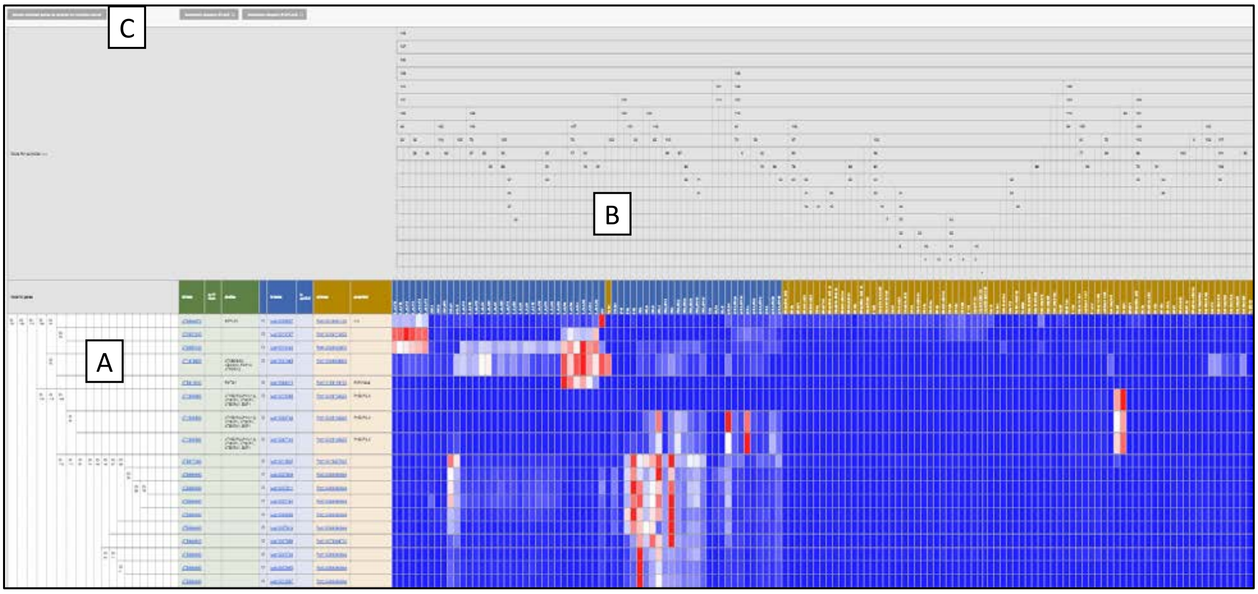


Fig. 13 Result of hieralchical clustering analysis
(A) Phylogeney of the genes is shown with cluster ID. Genes in a particular cluster can be extracted by selecting the cluster ID and press "Extract checked genes as queries for multiple search" button **(C)**. **(B)** Phylogeny of the samples is shown if the "cluster sample, too" option was selected.
Entire heat-map image is shown at the bottom of the result page (**Fig. 14**).

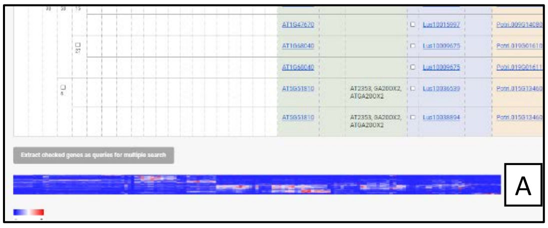


Fig. 14 Entire heat-map image **(A)** is shown at the bottom of the result page

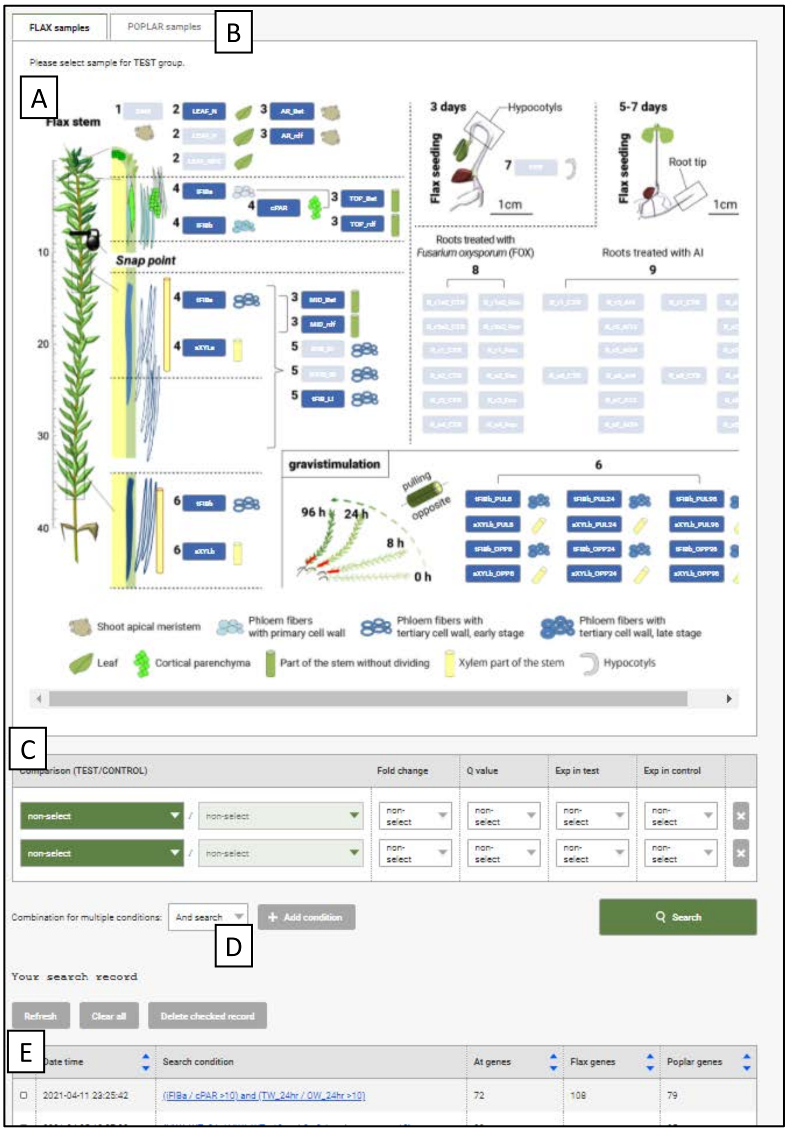
K-means clustering classifies genes into pre-determined number of groups based on gene expression profile (**Fig. 15**).

Fig. 15 Result of k-means clustering (without clustering samples)
Genes are classified into pre-determined number of groups. Group (cluster) ID is shown in left **(A)**.

7. DEG finder

FIBexDB can be used to list genes up/downregulated in a particular tissue or condition via DEG finder (**Fig. 16**). This function is available from the top page of FIBexDB as well as from each flax and poplar top pages. To select pairs for comparison, click graphical icon in schematic pictures of flax (and/or poplar) **(A)**. You can switch species by clicking each tab **(B)**. Once selecting a sample for a target (TEST), samples that can be selected as a CONTROL are shown as active state. By clicking the graphical icons for test and control samples, condition is automatically set in the list of comparisons below picture **(C)**. Multiple combined comparisons up to 9 pairs of samples can be done simultaneously both from flax and poplar. Differential expression analysis was performed using DESeq2 package for R. For each pair of comparison (TEST/CONTROL), following conditions can be set;

Fold change (>10/ >5/ >3/ >2/ >1.5/ >0/ <0.67/ <0.5/ <0.33/ <0.2/ <0.1)
Q value (<0.1/<0.05/<0.03/<0.01)
Exp in target (expression value in test sample; >0.1/>10)
Exp in control (expression value in control sample; >0.1/>10)
Select "And search" or "Or search" (**D;** for multiple conditions). Click **+ Add condition** to add pair for comparison (up to 9 pairs simultaneously). History of your search conditions is shown below **(E)**. **Delete checked record** or **clear all** to delete your search records. This search history is stored in your browser and not stored in our server. Mixed search of flax and poplar identifies common Arabidopsis homologues whose counterparts in flax and poplar meet search conditions (**Fig. 17**).



	Date time	Search condition	At genes	Flax genes	Poplar genes
☐	2021-04-11 23:25:42	(FIBa / cPAR >10) and (TW_24hr / OW_24hr >10)	72	108	79

Fig. 16 DEG finder input
(A) You can select samples to set as a test and control sample for comparison by clicking icons. **(B)** You can switch flax/poplar by clicking each tab. **(C)** Set conditions by click of above picture are indicated. Search parameters should be set. **(D)** Your search history is shown.

DEGs finder results are shown as a table (**Fig. 17**) with the following columns;

Arabidopsis locus is linked to TAIR DB

Alias according to TAIR DB

At TF class showing TF family name. By clicking the header, genes are sorted by alphabetical order of the TF family name.

Family showing more general name of gene family. By clicking the header, genes are sorted by alphabetical order of the family name.

Short description according to TAIR DB

flax locus (**poplar locus**)

flax symbol (**poplar symbol**) Gene symbol according to Phytozome DB

flax desc (**poplar desc**) description of flax (poplar) gene according to Phytozome DB

flax ->At score (**poplar-> At score**) score of BLAST search result against closest *Arabidopsis* homologue

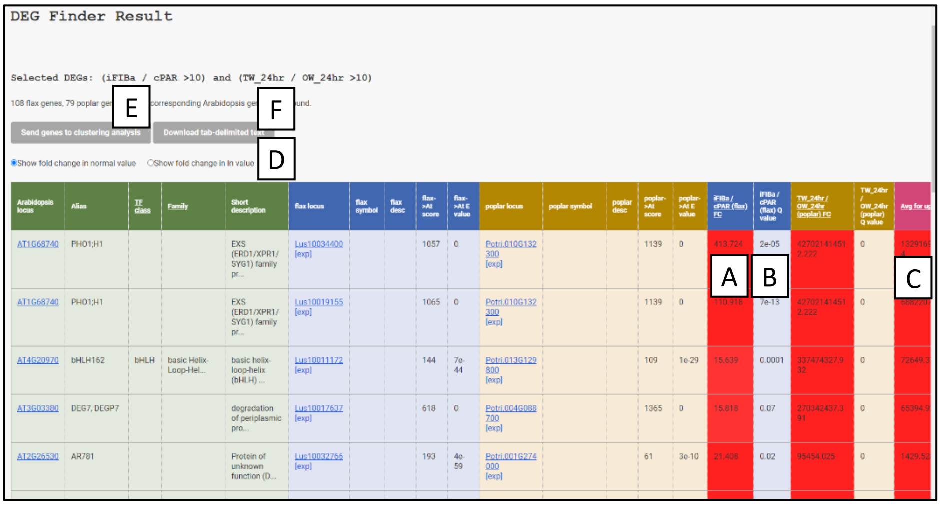
flax ->At E value (**poplar-> At E value**) E-value of BLAST search result against closest *Arabidopsis* homologue

TEST/CONTROL (flax) FC (**TEST/CONTROL (poplar) FC**) Fold change (FC) value of the comparison. By clicking the header, genes are sorted by the value in this column **(A)**.

TEST/CONTROL (flax) Q value (**TEST/CONTROL (poplar) Q value**) Q value of the comparison. By clicking the header, genes are sorted by the value in this column **(B)**.

Avg for up/down Average FC of all comparisons if you commonly select either of “upregulation” or “downregulation” for all comparisons. By clicking the header, genes are sorted by the value in this column **(C)**.

You can switch "Show fold change in normal value" or "Show fold change in ln (natural logarithm) value" by selecting radio button **(D)**. You can send genes to clustering analysis through the “multiple gene input” (**Fig. 11**) by clicking **Send genes to clustering analysis** button **(E)**. You can download the result as a tab-delimited text by clicking **Download tab-delimited text** button **(F)**.



DEG Finder Result

Selected DEGs: (iFIBa / cPAR >10) and (TW_24hr / OW_24hr >10)

108 flax genes, 79 poplar ge corresponding Arabidopsis ge und.

● Show fold change in normal value ○ Show fold change in ln value

Arabidopsis locus	Alias	TF class	Family	Short description	flax locus	flax symbol	flax desc	flax->At score	flax->At E value	poplar locus	poplar symbol	poplar desc	poplar->At score	poplar->At E value	iFIBa / cPAR (flax) FC	iFIBa / cPAR (flax) Q value	TW_24hr / OW_24hr (poplar) FC	TW_24hr / OW_24hr (poplar) Q value	Avg for up
AT1G68740	PHO1;H1			EXS (ERD1/XPR1/SYG1) family pr...	Lus10034400 [exp]			1057	0	Potri.010G132300 [exp]			1139	0	413.724	2e-05	427021414512.222	0	1329169 4
AT1G68740	PHO1;H1			EXS (ERD1/XPR1/SYG1) family pr...	Lus10019155 [exp]			1065	0	Potri.010G132300 [exp]			1139	0	110.918	7e-13	427021414512.222	0	688220
AT4G20970	bHLH162	bHLH	basic Helix-Loop-Hel...	basic helix-loop-helix (bHLH) ...	Lus10011172 [exp]			144	7e-44	Potri.013G129800 [exp]			109	1e-29	15.639	0.0001	337474327.932	0	72649.3
AT3G03380	DEG7, DEGP7			degradation of periplasmic pro...	Lus10017637 [exp]			618	0	Potri.004G088700 [exp]			1365	0	15.818	0.07	270342437.391	0	65394.9
AT2G26530	AR781			Protein of unknown function (D...	Lus10032766 [exp]			193	4e-59	Potri.001G274000 [exp]			61	3e-10	21.408	0.02	95454.025	0	1429.52

Fig. 17 DEG finder result
(A) Fold change (FC) value of the comparison. **(B)** Q value of the comparison is shown. **(C)** Average FC of all comparisons if you commonly select either of “upregulation” or “downregulation” for all comparisons. **(D)** You can switch "Show fold change in normal value" or "Show fold change in ln (natural logarithm) value" by selecting radio button. **(E)** You can send genes to clustering analysis through the “multiple gene input” by clicking this button. **(F)** You can download the result as a tab-delimited text by clicking this button.
For **(A-C)**, by clicking the header, genes are sorted by the value in this column.