J. Phytol. Res. **33**(1):41-69 2020 **ISSN 0970-5767**

IN SILICO IDENTIFICATION OF CROSS-TALKING ABIOTIC STRESS-TOLERANCE CONFERRING CANDIDATE GENE-ORTHOLOGS IN *ARABIDOPSIS AND POPULUS* USING GENE CO-EXPRESSION NETWORK ANALYSES AND COMPARATIVE GENOMICS

TARUN KANT
Genetics and Tree Improvement Division
Arid Forest Research Institute, Jodhpur
Corresponding Authors Email: tarunkant@icfre.org

The availability of high-quality gene expression microarray data for *Arabidopsis*, available in the public domain, provide a new opportunity for genome-wide exploration and discovery of genes associated with the response of a plant under abiotic stresses. Using this approach, a database of protein sequences and associated gene IDs involved in a plant's response to salinity stress has been created around the model plant *Arabidopsis thaliana*. This information can be used as a resource for the identification of orthologs in any other plant with sequence information available. A list of 140 *Populus* genes involved in salinity stress have been identified through the bioinformatics approach. This data can be used to clone out the genes for further characterization and testing and thereafter used for targeted salinity tolerance induction programmes under genetic improvement mandate of both agricultural and forestry species. Genes involved in a plant's response to osmotic stress, hyperosmotic salinity, cold and drought have also been identified in *Arabidopsis* and its counterpart ortholog has been identified in *Populus*. The information of tree orthologs for abiotic stress tolerance have become available for the first time in such a bigway and will lead to designing of better vectors for genetic engineering of plants in future.
Keywords: Bioinformatics, Osmotic shock, cold, drought, hyperosmotic salinity

Introduction

High salinity, drought and extremes of temperature are a few of the most well-known abiotic stresses that result in loss of productivity and even mortality in plants. Over 800 million ha of land is affected by the salinity of the soil[1]. New salt tolerance genes need to be identified for improvement work. Only a fraction of genes has known functions. Many new undiscovered genes are present particularly in tree genome[2], which have yet to be identified. These genes have the potential to be used for future tree/crop improvement programmes particularly those targeting abiotic stress tolerance. Genes from trees have rarely been identified, cloned and used against abiotic stress as a tree improvement strategy even though trees are excellent subjects to explore for genes in their highly evolved state for adaptation and survival under environmental extremes. The availability of high-quality gene expression microarray data for *Arabidopsis*, available in the public domain, provide a new opportunity for genome-wide exploration and discovery of genes associated with the response of a plant under abiotic stresses.

The native species of the abiotically-challenged region like saline and drought-prone regions have developed mechanisms with tweaked physiologies and morphologies such that they not just

survive, but also thrive in these extremes. This has taken hundreds of years. Trees are particularly important under such setups. Since trees live for many years, they supposedly have a better capability to survive the changes in seasonal cycles. A tree's genome is a store-house of such gene variants of ordinary genes that enable them to survive harsh environments, over many years, unlike seasonal plants. Trees, thus are living repositories of gene orthologs, which should be explored, identified and eventually cloned for use in effective genetic improvement programmes through the transgenic approach. Here, the gene ortholog mining of *Populustricocharpa*genes using co-expression network analysis based on *Arabidopsis thaliana* gene expression data from the public domain, coupled with gene ontology enrichment analysis is reported.

Material and Mehods

Gene co-expression analysis has emerged in the past decade as a powerful tool for gene function prediction. In essence, co-expression tries to look for the genes that show similar expression profiles across many experiments, for a gene of interest under consideration. Highly co-expressed genes may be involved in the biological process or processes of the query gene. And predicting orthologs that is, genes in another organism, that evolved from a common ancestral gene, and that tend to retain the same function, in the course of evolution, is a viable strategy for gene discovery. The following steps were deduced to carry out the said analysis:

Short-listing of bait genes (data-mining):

The first step was to prepare a curated list of genes of *Arabidopsis* that are implicated in abiotic stresses, particularly salinity stress. PubMed was queried for the purpose. PubMed is a free search engine accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics. The United States National Library of Medicine (NLM) at the National Institutes of Health (NIH) maintains the database as part of the Entrez system of information retrieval available at URL: www.ncbi.nlm.nih.gov/pubmed/

- To begin with, the PubMed database was queried to mark-out genes involved in abiotic stress response through documented experimental evidence in *Arabidopsis thaliana* using query terms Arabidopsis+Salt+Salinity+Gene+Proteins;
- The listed abstracts of the listed publications were downloaded and individually studied;
- The genes listed thereof were selected individually only if substantial experimental proof on the involvement of the gene was found in the abstract as per the search criteria;
- Curation of selected genes from the above database was carried out in a manually organized database.

Co-expression analysis:

The genes from the curated gene list acted as bait-genes to further query of *A. thaliana* co-expression networks using ATTED-II (http://atted.jp)[3]. ATTED-II is a co-expression database for plant species to aid in the discovery of relationships of unknown genes within a species. As an advanced co-expression analysis method, multispecies comparisons have the potential to detect alterations in gene relationships within an evolutionary context. However, determining the validity of comparative co-expression studies is difficult without quantitative assessments of the quality of co-expression data. ATTED-II was selected over CressExpress because it provided 16 co-expression platforms for nine plant species, including seven species supported by both microarray- and RNA sequencing (RNAseq)-based co-expression data. Two independent sources of co-expression data enable the assessment of the reproducibility of co-expression[3&4].

The co-expression networks were individually created for each of the bait gene using ATTED-II based on a correlational rank value called Mutual

Rank (MR). Correlation rank is asymmetric, that is, the rank of gene B from gene A is not the same as the rank of gene A from gene B. And thus, those two ranks are geometrically averaged, and is called the Mutual Rank (MR) given by the formula:

$$MR(AB) = \sqrt{(Rank(A \rightarrow B) \times Rank(B \rightarrow A))}.$$

The associated additional genes around the bait genes (network neighbours) were then pooled together with bait genes. This list of pooled genes was used as the main gene sub-set to look for orthologs in *Populus*. A similar approach was used by Yang et al (2011)[5].

Network Visualization and Cluster Analysis:

The list of primary genes (bait genes) and the newly identified genes found through co-expression networks were together processed through STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) Database (http://string-db.org/). STRING is a biological database and web resource of known and predicted protein-protein interactions[6-11] .

A co-expression network was reconstructed. Cluster analysis was carried out to divide the genes in the network into groups. The network visualization was carried out using Cytoscape environment vis-à-vis within STRING. Network visualization helped in the understanding of various pathways that were being used by the genes that were predicted.

SRING was a choice of the database at this stage because STRING imports data from experimentally derived protein-protein interactions through literature curation. Furthermore, STRING also store computationally predicted interactions from (i) text mining of scientific texts, (ii) interactions computed from genomic features, and (iii) interactions transferred from model organisms based on orthology[12]. Moreover, in STRING, all predicted or imported interactions are benchmarked against a common reference of functional partnership as annotated by KEGG (Kyoto Encyclopedia of Genes and Genomes).

Moreover, protein-protein interaction networks are an important ingredient for the system-level understanding of cellular processes. Such networks can be used for filtering and assessing functional genomics data and for providing an intuitive platform for annotating structural, functional and evolutionary properties of proteins. Exploring the predicted interaction networks can suggest new directions for future experimental research and provide cross-species predictions for efficient interaction mapping[13].

STRING Work-flow and parameters used:

i. Starting Point: At the STRING start page, Multiple sequenceswere selected. The pooled gene list (and vis a vis proteins list) was copied from excel sheet and pasted. Organism selected was *Arabidopdid thaliana* and searched.
ii. Network view appeared (Default view): The network view summarizes the network of predicted associations for a particular group of genes (and proteins they code for). The network nodes in the appearing diagrammatic representation are proteins. The edges represent the predicted functional associations. The edges are drawn according to the view settings. In confidence mode, which was selected in the present study, the thickness of the line indicated the degree of confidence prediction of the interaction.
iii. Data Settings (for fine-tuning and selecting parameters): In the data settings changes to the parameters influenced the output. when 'Update Settings' button was pressed after making the changes to the parameters. Under the active interaction sources type of evidence were selected manually which contributed to the prediction of the score. The minimum required interaction score put a threshold on the confidence score, such that only interaction above this score

were included in the predicted network. Lower score means more interaction, but also more false positives. The confidence score is the approximate probability that a predicted link exists between two enzymes in the same metabolic map in the KEGG database. Confidence limits were as follows:

- low confidence - 0.15,
- medium confidence - 0.4,
- high confidence - 0.7,
- highest confidence - 0.9.

In the present study, high confidence was used for analysis, to minimize false interactions.

iv. View Settings: Network-specific parameters are: 'edge scaling factor' - which reduced the length of high-scoring edges so that the images were drawn more compact, and low scoring hits spread out further. Lower values meant more compact images, higher values caused more spread. Options for selecting the meaning of network edges of the displayed network were:

- evidence - multiple lines where the colour indicates the type of interaction evidence
- confidence - line thickness indicates the strength of data support
- molecular action - line shape indicates the predicted mode of action

In the present study, confidence was used as the network edge parameter.

Finally, only the genes identifiers of *A. thaliana* were chosen through the network diagram, that were connected. A minimum of one connection was considered mandatory in the present analysis. As a result, all the orphan identifiers were screened out from the list of genes. This was an essential stem in the present analysis, which led to the finalization of only the genes that interacted and was in essence the true crux of a co-expression network.

Identification of *Populus*orthologs:

The steps involved in the identification of *Populusorthologs*used in the present investigation were as follows:

i. The list of all network-connected genes of *Arabidopsis* was tabulated.

ii. The protein sequences of each of the final selected gene from the co-expression network was individually obtained from The Arabidopsis Information Resource (TAIR) available at URL: https://www.arabidopsis.org/. TAIR maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*.

iii. Once all protein sequence data was curated in tables, individually each of the sequences was selected and blasted using NCBI Protein Blast (BlastP) available at URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp. The E-value cut-off was kept at $1X10^{-4}$.

iv. Top-hitting *Populus* ortholog was selected and the protein sequence was also obtained and saved.

5. Functional Analysis of *Populus* Genes

i. The accession of the genes selected were again processed through STRING database of protein interactions. This resulted in the development of the networks of the *Populus* genes.

ii. Gene Ontology (GO) enrichment analysis was performed on these networks within STRING Platform. This was also compared to the DAVID (Database for Annotation, Visualization and Integrated Gene Discovery) database. STRING Platform was found more suitable as the networks were already available on it as against DAVID, which is more enriched for non-plant species.

iii. GO Enrichment analysis led to the identification of the genes of the network which were also involved during Water deficits (drought) and Temperature (changes to temperature) besides Salinity which was primary dataset in the present analysis.

Thus, using the comparative genomic approach, querying of the *Populus* genome using known *Arabidopsis* genes lead to the

identification of *Populus*orthologs.These may be validated and characterized in wet-lab setup and similar candidate genes may be cloned from other tree species for use in genetic improvement programmes using a transgenic approach.

Results and discussion

Genereation of primary Bait Gene list:

In the present investigation, the first step was the identification of Bait Genes that would be used for the rest of the gene mining. The Bait gene here were the genes that would be used as bait inside a sea of gene data to attract the coexpressed genes and later to find out the orthologs.

This gene list was generated through a systematic key-word based text-mining of authentic peer-reviewed scientific literature based on sound experimentations from PubMed. The following table is the curated list of 74 Primary bait genes:

Table 1: *Arabidopsis thaliana* Gene IDs and the literature they have been identified from (reference)

#	***Gene ID***	***Citation details of publications refered*** [References]
1)	AT1G02730	Gu et al 2016; Yin et al 2011; Zhu et al 2010[14;15;16]
2)	AT1G03060	Steffens et al (2015)[17]
3)	AT1G05850	Sanchez-Rodriguez et al (2012)[18]
4)	AT1G06040	Datta et al (2007)[19]
5)	AT1G10940	Julkowaska et al (2015); McLoughlin et al (2012); Boudsocq et al (2004)[20;21;22]
6)	AT1G18890	Urao et al (1994); Harmon et al (2001); Cheng et al (2002) [23;24;25]
7)	AT1G24460	Kim &Bassham (2011); Roy &Bassham (2015)[26;27]
8)	AT1G29060	Tarte et al (2015)[28]
9)	AT1G35670	Urao et al (1994)[23]
10)	AT1G45688	Endler et al (2015)[29]
11)	AT1G57550	Medina et al (2007)[30]
12)	AT1G60940	McLoughlin et al (2012); Boudsocq et al (2004)[21;22]
13)	AT1G69270	Hong et al (1997); Osakabe et al (2010); Si et al (2014)[31;32;33]
14)	AT1G73660	Gao et al (2008)[34]
15)	AT1G78290	Boudsocq et al (2004); Kim et al (2012)[22;35]
16)	AT2G01450	Frick &Strader (2017)[36]
17)	AT2G01980	Wu et al (1996); Shi et al (2000); Qiu et al (2002); Shi et al (2002); Nah et al (2009); Oh et al (2010); Yue et al (2012)[37;38;39;40;41;42;43]
18)	AT2G03150	Guan et al (2013)[44]
19)	AT2G04240	Ko et al (2006)[45]
20)	AT2G17270	Zhu et al (2012)[46]
21)	AT2G26980	Kim et al (2003), Tang et al (2015)[47;48]
22)	AT2G38470	Jiang et al (2006)[49]
23)	AT2G39800	Strizhov et al (1997); Abraham et al (2003)[50;51]
24)	AT2G40950	Liu et al (2007), (2008)[52;53]
25)	AT2G41010	Perruc et al (2004)[54]
26)	AT2G41560	Geisler et al (200)[55]
27)	AT2G45640	Song & Galbraith (2006)[56]
28)	AT2G46400	Ding et al (2013)[57]
29)	AT2G47770	Balsemao-Pires et al (2011)[58]
30)	AT3G02140	Garcia et al (2008)[59]
31)	AT3G05880	Liu et al (2012)[60]
32)	AT3G12360	Sakamoto et al (2008)[61]
33)	AT3G16890	Zsigmond et al (2008)[62]
34)	AT3G26520	Schussler et al (2008)[63]
35)	AT3G45410	He et al (2004)[64]
36)	AT3G45700	Li et al (2016)[65]

37)	AT3G46550	Basu et al (2016)[66]
38)	AT3G47950	Vitrat et al (2001)[67]
39)	AT3G48850	Zhu et al (2012)[46]
40)	AT3G50500	Boudsocq et al (2004); Mogami et al (2015)[22;68]
41)	AT3G55530	Zhang et al (2015)[69]
42)	AT3G55610	Strizhov et al (1997); Szekely et al (2007)[50;70]
43)	AT4G01420	Saito et al (2018)[71]
44)	AT4G16830	Ambrosone et al (2015)[72]
45)	AT4G17615	Cheong et al (2003)[73]
46)	AT4G22330	Wu et al (2015)[74]
47)	AT4G22820	Adai et al (2005)[75]
48)	AT4G28088	Medina et al (2007)[30]
49)	AT4G30650	Medina et al (2007)[30]
50)	AT4G30660	Medina et al (2007)[30]
51)	AT4G30960	Guo et al (2001)[76]
52)	AT4G33000	Kim et al (2007); Quan et al (2007); Lin et al (2009)[77;78;79]
53)	AT4G33730	Chien et al (2015)[80]
54)	AT4G33950	Boudsocq et al (2004)[22]
55)	AT4G34890	Zarepour et al (2010)[81]
56)	AT4G35100	Pou et al (2016)[82]
57)	AT4G40010	Boudsocq et al (2004)[22]
58)	AT5G08590	Boudsocq et al (2004)[22]
59)	AT5G14040	Zhu et al (2012) [46]
60)	AT5G15970	Kai-Chau et al (2018)[83]
61)	AT5G17850	Cai& Lytton (2004)[84]
62)	AT5G19660	Liu et al (2008)[53]
63)	AT5G19690	Koiwa et al (2003)[85]
64)	AT5G24270	Ishitani et al (2000)[86]
65)	AT5G27150	Apse et al (1999)[87]
66)	AT5G35410	Liu et al (2000)[88]
67)	AT5G37850	Shi et al (2002), Gonzalez et al (2007)[89;90]
68)	AT5G42860	Endler et al (2015) [29]
69)	AT5G51110	Zhang et al (2015)[69]
70)	AT5G57630	Pandey et al (2015)[91]
71)	AT5G58580	Tian et al (2015)[92]
72)	AT5G63650	Boudsocq et al (2004)[22]
73)	AT5G63980	Quintero et al (1996)[93]
74)	AT5G66880	Boudsocq et al (2004)[22]

Gene-wise co-expressin Network :

For each of the 74 selected genes, co-expression networks were created using ATTED-II. Genes directly connected with the gene on the network based on Microarray results from public domain database were enlisted based on Mutual Rank (MR) values and Network Map was created through Cytoscape web.

Creation of Pooled Gane :

The co-expression networks (as shown in the previous section) resulted in enlisting of additional genes that were directly connected to the main bait gene in each of the networks created using MR values from microarray data through ATTED-II database.

76 genes were originally used as bait genes. Co-expression networks enabled identification of co-expressed genes under salt stress condition. The original bait genes were clubbed together with the identified co-expressed genes. As a result a new pooled gene list was created

containing a total of 397 genes (Appendix 1). This new list was used to create a new more complex co-expression network using the STRING database.

Co-expression Network Analysis Based on Pooled Genes:

The co-expression network was created using the STRING database at High Confidence level. High Confidence level resulted in the elimination of false-positive interactions

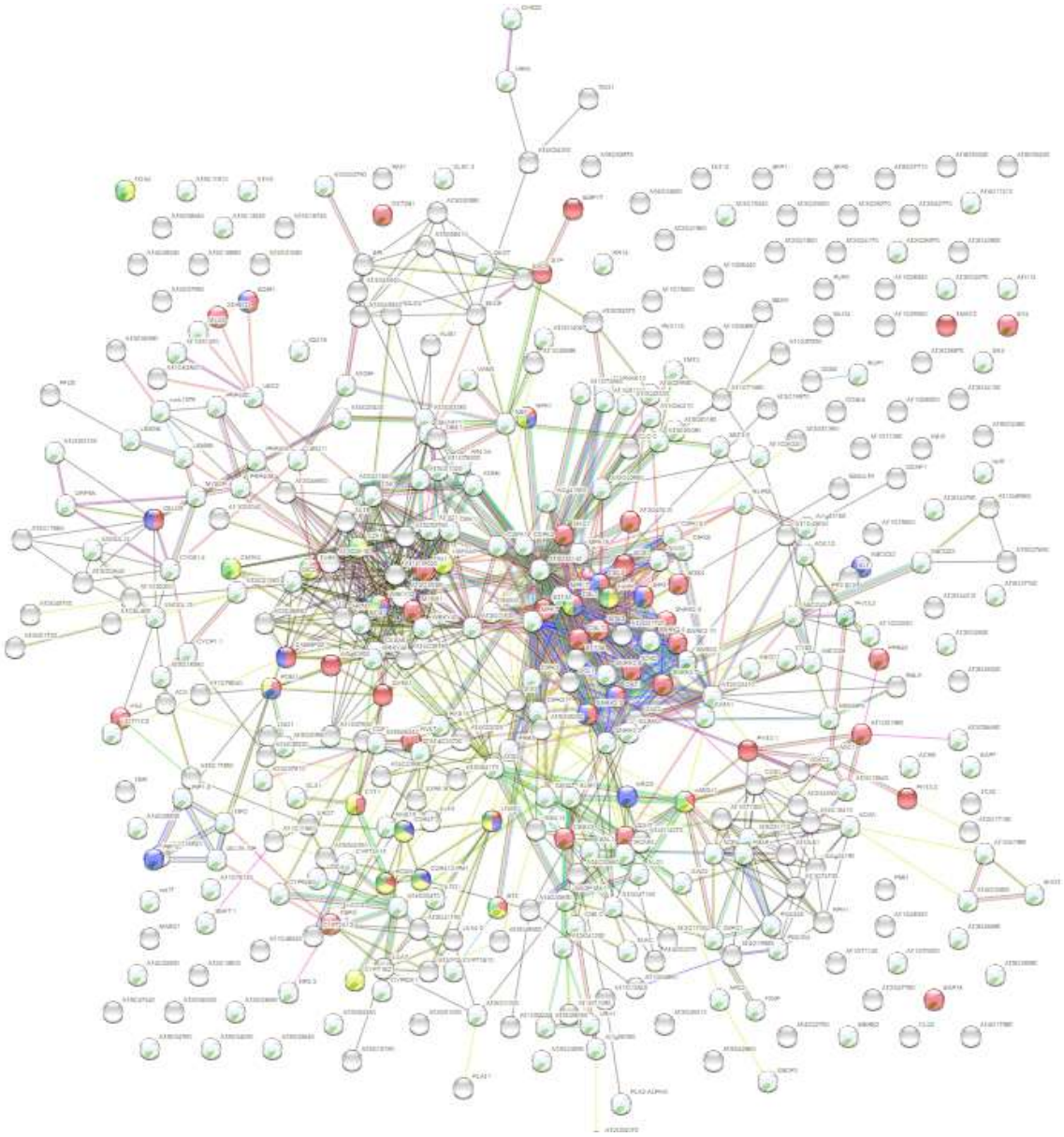

Fig.1: Co-expression network of genes with along with un-connected genes.

The genes not connected through any other gene in the network were eliminated. The resulting genes were now the final selected genes indicating clear clustering within the network. The resulting co-expression network, which is cleaner and without the noise of non-correlated genes, is presented below:

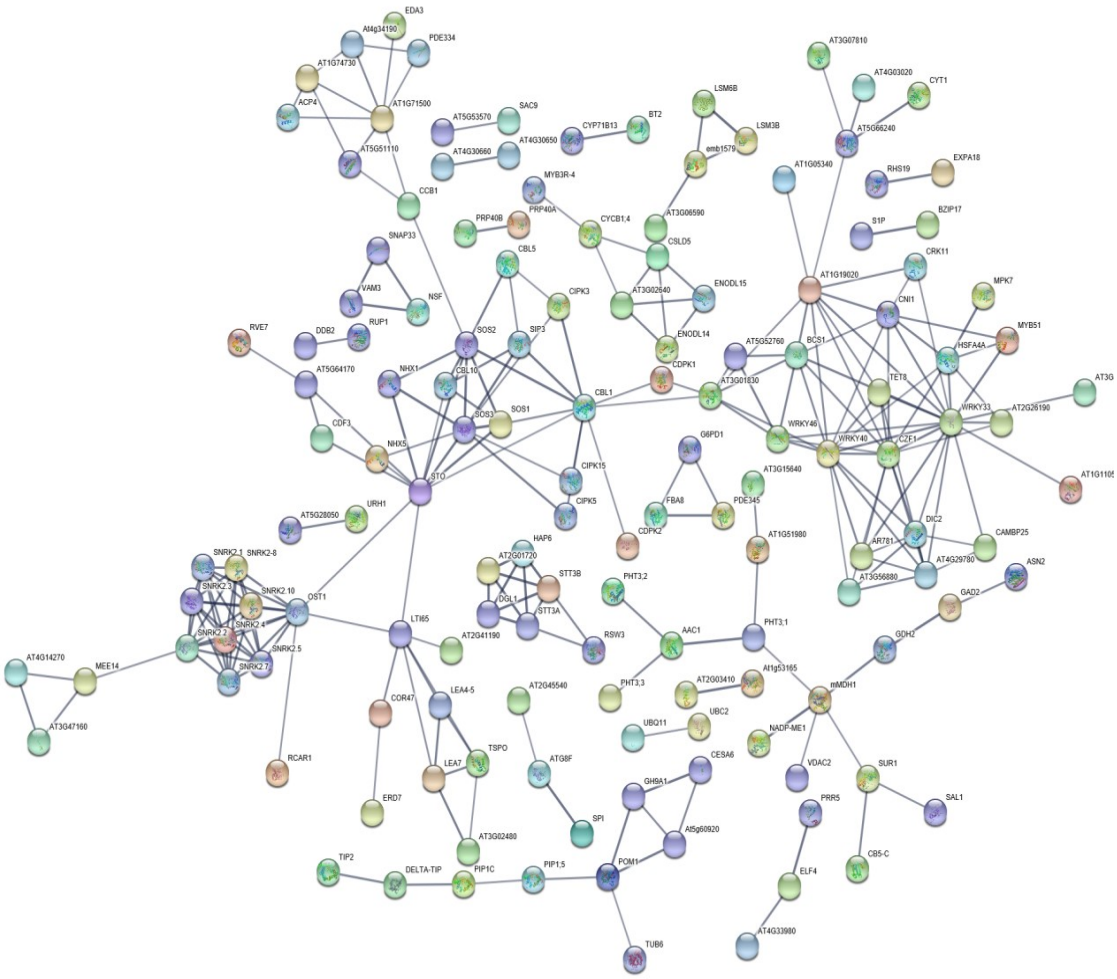

Fig.2: Final Co-expression network showing the interacting genes only along with clear sub-network clusters

The network indicates the interaction within the selected pooled genes and has more relevance as it gives a bigger picture of various pathways that are involved at the time a plant experiences salinity stress. The interaction data was exported and was analyzed using CYTOSCAPE, which is an open-source software platform for visualizing complex molecular interaction networks and integrating with gene expression profile data.

Using Cytoscape, cluster analysis was done on network interactions using ClusterViz. The entire network was subjected to cluster analysis using the maximal clique-based EAGLE algorithm which helps in identification of even the overlapped clusters. Default EAGLE algorithm parameters (Clique Size Threshold: 3; Output Threshold: 2) were selected. A total of 11 sub-network clusters are formed as shown in the table below:

Table 2: Sub-network cluster details

Cluster No	***Nodes No***	***Edges No***	***Node ID (Genes)***
1	28	51	LEA7, CDPK2, AT3G02480, CIPK5, STO, CDF3, NHX1, AT5G64170, RVE7, OST1, CBL10, CBL5, SOS1, ERD7, NHX5, CDPK1, CIPK15, SOS3, SOS2, COR47, CIPK3, LTI65, CBL1, AT2G41190, AT3G01830, TSPO, SIP3, LEA4-5
2	23	67	CAMBP25, AR781, CRK11, MPK7, HSFA4A, BCS1, AT1G05340, MYB51, WRKY33, CZF1, TET8, AT3G46600, AT3G56880, AT2G26190, AT1G11050, DIC2, CNI1, AT4G29780, AT1G19020, WRKY40, AT3G01830, WRKY46, AT5G52760
3	17	13	UBC2, AT3G07810, UBQ11, SPI, CYT1, MEE14, ATG8F, AT2G45540, AT5G66240, AT4G03020, AT3G47160, AT4G14270, SAC9, AT5G53570, VAM3, SNAP33, NSF
4	15	8	ELF4, PRR5, RUP1, DDB2, PRP40A, PRP40B, AT4G33980, EXPA18, BT2, RHS19, CYP71B13, AT4G30650, AT4G30660, URH1, AT5G28050
5	15	14	PHT3;2, PHT3;3, AAC1, PHT3;1, SAL1, mMDH1, NADP-ME1, VDAC2, AT1G51980, GAD2, AT3G15640, GDH2, ASN2, SUR1, CB5-C
6	11	9	emb1579, AT3G06590, LSM3B, LSM6B, PDE345, BZIP17, FBA8, S1P, At1g53165, AT2G03410, G6PD1
7	10	37	SNRK2.3, RCAR1, SNRK2.1, SNRK2.10, SNRK2.7, SNRK2.4, OST1, SNRK2.5, SNRK2-8, SNRK2.2
8	9	10	TIP2, PIP1;5, At5g60920, PIP1C, DELTA-TIP, TUB6, POM1, GH9A1, CESA6
10	6	12	AT2G01720, STT3B, HAP6, STT3A, DGL1, RSW3
11	6	9	MYB3R-4, AT3G02640, ENODL14, ENODL15, CYCB1;4, CSLD5

The cluster analysis resulted in the identification of a set of 140 network connected genes. These were the final list of genes of *Arabidopsis* that were considered relevant for a plants response to salinity stress.

Retrival of Protein Sequences of Network Associated Genes, Protein Blast and Identification of Populus Orththolog:

The protein sequences for each of the 140 salt stress-related genes identified through co-expression network analysis were retrieved from The Arabidopsis Information Resource (TAIR) database. The protein sequences of *Arabidopsis* of the network linked genes were Blasted against the background of non-redundant protein sequence database (nr) of *Populustrichocarpa* (taxid:3694) using blastp (protein-protein BLAST) algorithm of NCBI Blast suite. The following information was retrieved:

Local Extreme Metrics: These measures treat each aligned segment independently. Where there are multiple matches to the same subject (database) sequence, only the metric for the best match is considered. The E(xpect) Value is the traditional BLAST statistic used to sort the output by significance.

1. *E(xpect) Value*: the number of alignments expected by chance with a particular score or better. The expect value is the default sorting metric and normally gives the same sorting order as Max Score.
2. *Max(imum) Score*: the highest alignment score of a set of aligned segments from the same subject (database) sequence. The score is calculated from the sum of the match

rewards and the mismatch, gap open and extend penalties independently for each segment. This normally gives the same sorting order as the E Value.

Total Metrics : These metrics are summed over or include all aligned segments for the same subject sequence. These are most useful for analyzingBLASTmatches to genomic sequences.

1. *Tot(al) Score*: the sum of alignment scores of all segments from the same subject sequence. This sorting order may help promote the position of mRNA matches to genomic sequences where there are multiple exons. The Total Score is useful for distinguishing hits to functional multi-exon genes from those to the corresponding intronlessretrotransposed pseudogenes.
2. *Query Coverage*: the percent of the query length that is included in the aligned segments. This is calculated over all segments as with the Tot Score.

From the data analysis of shortlisted network highlighted *Arabidopsis* genes and the *Populus* gene data retrieved after BlastP, it was found that several Genes encode proteins of unknown function or with a function similar but not established to another protein or uncharacterized protein and are **Domains of Unknown Functions (DUFs)** indicating Genes of putative nature. Moreover, several sequences were of **Transcription Factors (TFs)** or their activators or suppressors. This information is as follows:

Table 3: DUFs and TFs identified during the analysis (Arabidopsis thaliana)

Species	***DUFs***	***TFs***
Arabidopsis thaliana	1. AT1G05340 (Uncharacterized protein) 2. AT1G19020 (Uncharacterized protein) 3. AT1G74730 (Uncharacterized protein) 4. AT2G03410 (MO25-like protein) 5. AT2G26530 (Unknown function) 6. AT3G02640 (Uncharacterized protein) 7. AT4G23190 (Cysteine-rich receptor-like protein kinase 11) 8. AT4G29780 (Uncharacterized protein) 9. AT4G33980 (Uncharacterized protein) 10. AT5G28050 (Cytidine/deoxycytidylate deaminase-like protein)	1. AT1G80840 (WRKY DNA-binding protein 40; Transcription factor) 2. AT2G38470 (WRKY DNA-binding protein 33; Transcription factor) 3. AT2G46400 (WRKY DNA-binding protein 46; Transcription factor) 4. AT3G06590 (Transcription factor bHLH148) 5. AT3G46600 (Scarecrow-like protein 30; Probable transcription factor involved in plant development) 6. AT3G47500 (Cycling DOF factor 3; Transcription factor) 7. AT4G18880 (Heat shock transcription factor A4A; Transcriptional activator) 8. AT5G11510.1 (Myb domain protein 3r-4) 9. AT5G24470 (Pseudo-response regulator 5; Transcriptional repressor of CCA1 and LHY) 10. AT5G64170 (Dentin sialophosphoprotein-related; Transcriptional coactivator)

Table 4: DUFs and TFs identified during the analysis (Populustricocharpa)

Species	***DUFs***	***TFs***
Populustrichocarpa	1. ABK92801.1 (unknown protein) 2. XP_011009946.1 (PREDICTED: chitinase-like protein 2) 3. POPTR_017G133000 (probable receptor-like protein kinase) 4. POPTR_015G070700 (uncharacterized protein 5. POPTR_019G073900 (uncharacterized protein 6. POPTR_012G070600 (uncharacterized protein LOC7484752) 7. POPTR_002G099700 (serine/threonine-protein kinase SAPK2) 8. POPTR_001G044500 (probable WRKY transcription factor 40) 9. POPTR_016G011100 (putative MO25-like protein) 10. POPTR_017G014100 (probable aminotransferase TAT2) 11. POPTR_002G128500 (uncharacterized protein) 12. POPTR_010G179300 (uncharacterized protein) 13. NT39234.1 (hypothetical protein POPTR_004G026100) 14. XP_006372028.1 (protein ALP1-like) 15. POPTR_006G182500 (UPF0057 membrane protein) 16. POPTR_006G182500 (UPF0057 membrane protein) 17. POPTR_002G125900 (uncharacterized protein) 18. POPTR_010G003500 (uncharacterized protein 19. POPTR_012G112400 (probable pterin-4-alpha-carbinolamine dehydratase, chloroplastic) 20. POPTR_007G100000 (probable glucan 1,3-alpha-glucosidase)	1. POPTR_002G096800 (transcription factor MYB93) 2. POPTR_016G128300 (probable WRKY transcription factor 33) 3. POPTR_002G168700 (probable WRKY transcription factor 46 isoform X1) 4. POPTR_008G103500 (transcription factor bHLH148) 5. POPTR_009G033300 (scarecrow-like protein 14) 6. POPTR_009G045400 (cyclic dof factor 2) 7. POPTR_018G038000 8. (transcription factor MYB3R-1 isoform X1) 9. POPTR_015G002300 (two-component response regulator-like APRR5 isoform X1) 10. POPTR_001G205800 (protein LNK1)

The findings of DUFs or Genes of putative nature are important for future research aimed at their characterization because these are reflected genes that have a role in abiotic stress management in some way but are not yet characterized or known. This is a very important set of findings as this will lead to the identification of new gene information that is implicated in a plant's response to abiotic stresses.

1. Gene ontology (GO) enrichment:

GO enrichment analysis was carried out using plantgenei.org GO Enrichment tool that calculates gene function enrichment for a selected gene set. All the implicated genes of *Populus* identified were used in this enrichment analysis. The genes for which the gene ID was not yet available were not included in this analysis. A total of 124 genes were analyzed. The tables below are colour coded for level of significance:

Correlation Colour code:

Least Significant←----------→ Highly Significant

A. Go Enrichment usisng Populus Tearms:

i. GO - Biological process: Upon GO Enrichment using the Enrichment tool of plantgenei.org, it was discovered that a total of 10 biological processes are correlated with genes that have a role in salt stress response of a plant. These indicate that the biological process for single-organism carbohydrate metabolic process, macromolecule glycosylation, protein glycosylation, carbohydrate metabolic process, glycosylation are significantly upregulated followed by processes for cellular protein modification, macromolecule modification, peptidyl-asparagine modification and protein N-linked glycosylation via asparagine.

Table 5:GO - Biological process

GO ID	***P Value (corrected)***	***Statistics***	***Description***
GO:0044723	3.072e-05	12/59 | 439/14903	single-organism carbohydrate metabolic process
GO:0043413	7.617e-04	5/59 | 71/14903	macromolecule glycosylation
GO:0006486	7.617e-04	5/59 | 71/14903	protein glycosylation
GO:0005975	1.353e-03	14/59 | 998/14903	carbohydrate metabolic process
GO:0070085	1.633e-03	5/59 | 88/14903	glycosylation
GO:0006464	1.090e-02	20/59 | 2227/14903	cellular protein modification process
GO:0036211	1.090e-02	20/59 | 2227/14903	protein modification process
GO:0043412	1.221e-02	20/59 | 2288/14903	macromolecule modification
GO:0018196	2.170e-02	2/59 | 11/14903	peptidyl-asparagine modification
GO:0018279	2.170e-02	2/59 | 11/14903	protein N-linked glycosylation via asparagine

ii. GO- Molecular Functions: Upon GO Enrichment using the Enrichment tool of plantgenei.org, it was discovered that a total of 8 Molecular Functions are correlated with genes that have a role in salt stress response of a plant. These

indicate that oligosaccharyltransferase activity is having high significance followed by dolichyl-diphosphooligosaccharide-protein glycotransferase activity. Activities of moderate significance include hydrogen ion transmembrane transporter activity, monovalent inorganic cation transmembrane transporter activity, solute:hydrogen antiporter activity.

Table 6:GO –Molecular Functions

GO ID	***P Value (corrected)***	***Statistics***	***Description***
GO:0004576	1.459e-06	5/79 | 19/19622	oligosaccharyl transferase activity
GO:0004579	1.946e-03	3/79 | 15/19622	dolichyl-diphosphooligosaccharide-protein glycotransferase activity
GO:0015078	1.154e-02	5/79 | 139/19622	hydrogen ion transmembrane transporter activity
GO:0015077	1.797e-02	5/79 | 190/19622	monovalent inorganic cation transmembrane transporter activity
GO:0015299	1.921e-02	3/79 | 48/19622	solute:hydrogen antiporter activity
GO:0015298	1.921e-02	3/79 | 48/19622	solute:cation antiporter activity
GO:0004332	1.994e-02	2/79 | 9/19622	fructose-bisphosphate aldolase activity
GO:0016832	2.425e-02	2/79 | 11/19622	aldehyde-lyase activity

iii. KEGG: According to the GO enrichment for Kyoto Encyclopedia for Genes and Genomes (KEGG), **a** total of 6 KEGG IDs are correlated with the genes of Populus identified. The number of physiological pathways involved are manifold.

Table 7:GO – KEGG Identifiers

KEGG ID	***P Value (corrected)***	***Statistics***	***Description***
K14498	2.901e-04	3/48 | 9/5582	serine/threonine-protein kinase SRK2 [EC:2.7.11.1]
K07151	6.482e-04	2/48 | 3/5582	dolichyl-diphosphooligosaccharide--protein glycosyltransferase [EC:2.4.1.119]
K01623	5.017e-03	2/48 | 9/5582	fructose-bisphosphate aldolase, class I [EC:4.1.2.13]
K09873	1.062e-02	2/48 | 15/5582	aquaporin TIP
K13412	1.644e-02	2/48 | 21/5582	calcium-dependent protein kinase [EC:2.7.11.1]
K00517	4.801e-02	2/48 | 41/5582	beta-carotene 15,15'-monooxygenase [EC:1.14.99.36]

Table 8: Further bifurcation of KEGG pathways per KEGG ID

KEGG ID	***KEGG Pathways involved***
K14498	ko04016 - MAPK signaling pathway - plant ko04075 - Plant hormone signal transduction
K07151	ko00510 - N-Glycan biosynthesis ko00513 - Various types of N-glycan biosynthesis ko01100 - Metabolic pathways ko04141 - Protein processing in endoplasmic reticulum
K01623	ko00010 - Glycolysis / Gluconeogenesis ko00030 - Pentose phosphate pathway ko00051 - Fructose and mannose metabolism ko00680 - Methane metabolism ko00710 - Carbon fixation in photosynthetic organisms ko01100 - Metabolic pathways ko01110 - Biosynthesis of secondary metabolites ko01120 - Microbial metabolism in diverse environments ko01130 - Biosynthesis of antibiotics ko01200 - Carbon metabolism ko01230- Biosynthesis of amino acids
K09873	-
K13412	ko04626 - Plant-pathogen interaction ko05145 - Toxoplasmosis
K00517	-

USING AT SYNONYMOUS GO TERMS

Table 9:GO - Biological process (ATI based)

GO id	***P value ©***	***Statistics***	***Description***
GO:0009414	3.490e-09	11/81 | 189/24222	response to water deprivation
GO:0009409	4.330e-09	12/81 | 243/24222	response to cold
GO:0009651	4.898e-09	18/81 | 441/24222	response to salt stress
GO:0006970	2.400e-06	7/81 | 96/24222	response to osmotic stress
GO:0030244	2.589e-05	5/81 | 46/24222	cellulose biosynthetic process
GO:0042538	2.978e-05	5/81 | 49/24222	hyperosmotic salinity response
GO:0009738	4.584e-05	5/81 | 55/24222	abscisic acid mediated signaling pathway
GO:0010200	1.546e-04	6/81 | 127/24222	response to chitin
GO:0009644	2.487e-04	4/81 | 38/24222	response to high light intensity
GO:0018279	2.623e-04	2/81 | 2/24222	protein N-linked glycosylation via asparagine
GO:0006885	2.623e-04	2/81 | 2/24222	regulation of pH
GO:0050832	2.846e-04	5/81 | 90/24222	defense response to fungus
GO:0009628	2.866e-04	3/81 | 13/24222	response to abiotic stimulus
GO:0042742	5.685e-04	6/81 | 176/24222	defense response to bacterium
GO:0010118	8.717e-04	3/81 | 21/24222	stomatal movement
GO:0006814	1.083e-03	3/81 | 23/24222	sodium ion transport
GO:0015840	1.106e-03	2/81 | 4/24222	urea transport
GO:0048015	3.633e-03	2/81 | 7/24222	phosphatidylinositol-mediated signaling
GO:0006833	4.579e-03	2/81 | 8/24222	water transport

GO id	*P value ©*	*Statistics*	*Description*
GO:0031347	6.961e-03	2/81 | 10/24222	regulation of defense response
GO:0009737	7.105e-03	6/81 | 304/24222	response to abscisic acid stimulus
GO:0006810	7.120e-03	7/81 | 421/24222	transport
GO:0006468	8.925e-03	11/81 | 1049/24222	protein phosphorylation
GO:0007033	9.589e-03	2/81 | 13/24222	vacuole organization
GO:0007165	9.838e-03	7/81 | 455/24222	signal transduction
GO:0043622	1.073e-02	2/81 | 14/24222	cortical microtubule organization
GO:0007623	1.190e-02	3/81 | 61/24222	circadian rhythm
GO:0006108	1.309e-02	2/81 | 16/24222	malate metabolic process
GO:0009832	3.258e-02	2/81 | 29/24222	plant-type cell wall biogenesis
GO:0048573	3.258e-02	2/81 | 29/24222	photoperiodism, flowering
GO:0009826	3.271e-02	3/81 | 98/24222	unidimensional cell growth
GO:0046686	3.664e-02	5/81 | 330/24222	response to cadmium ion
GO:0006812	4.072e-02	3/81 | 108/24222	cation transport
GO:0042542	4.493e-02	2/81 | 36/24222	response to hydrogen peroxide

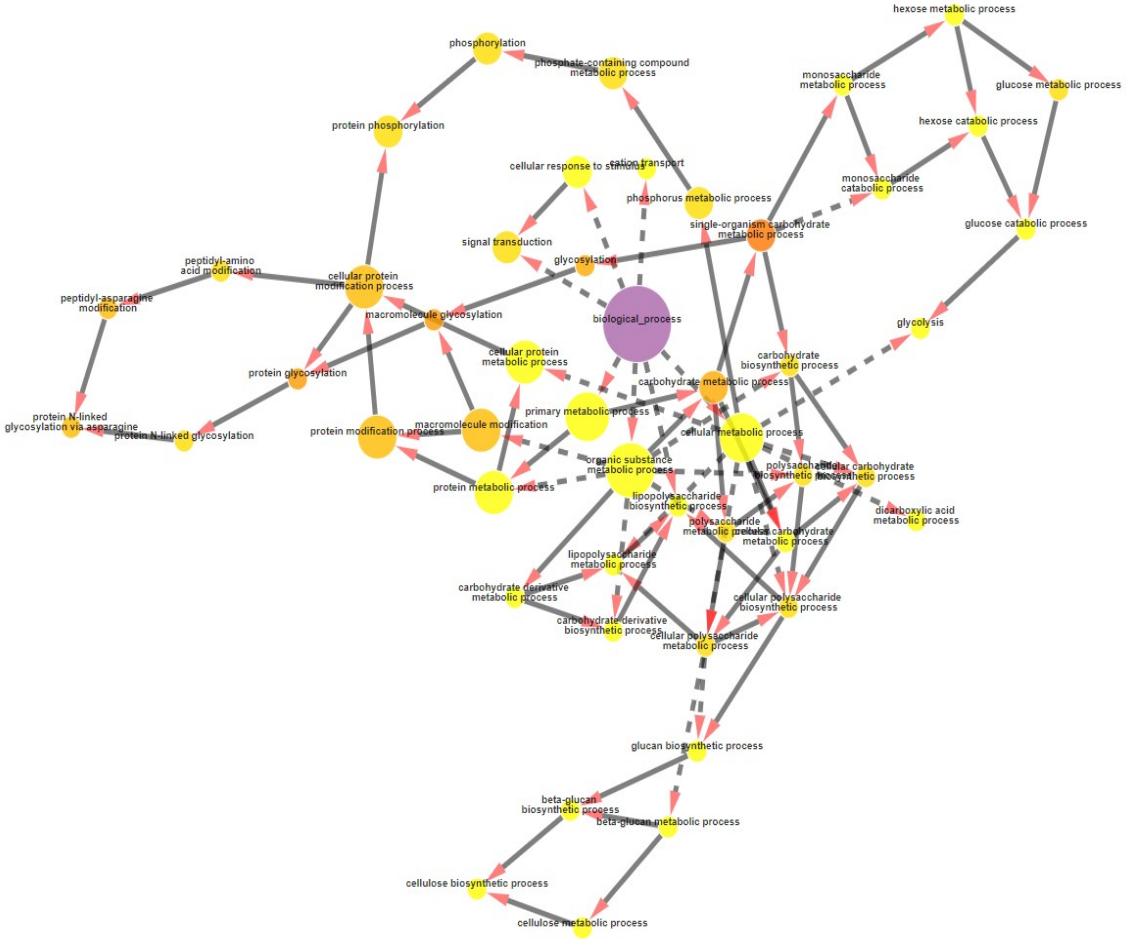

Fig. 3: GO Biological Process (Diagramatic representation – Force Directed Tree Layout)

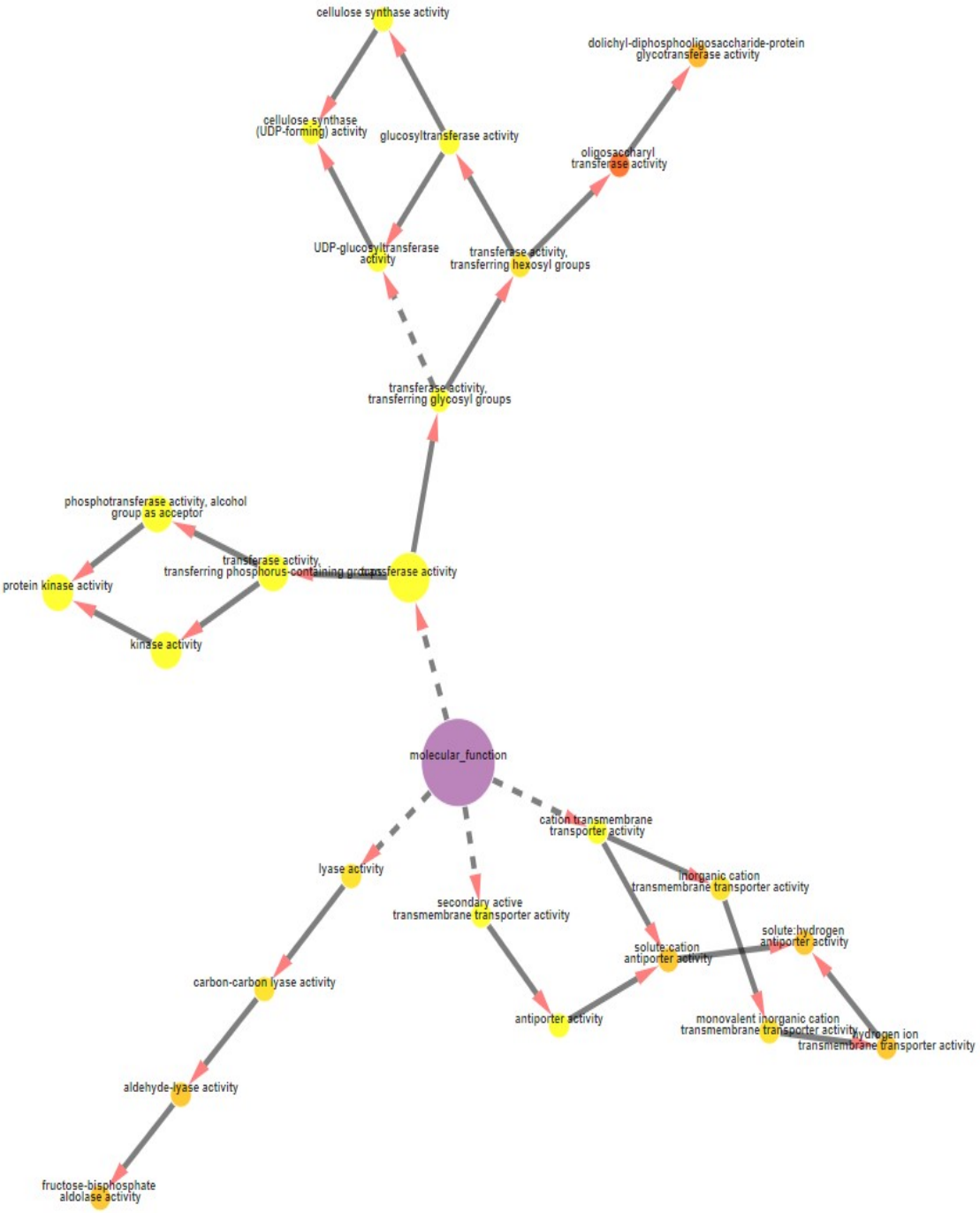

Fig. 4: GO Molecular Function (Diagramatic representation – Force Directed Tree Layout)

Identification of Genes Having Role in Management of other Abotic Stresses

A total of 74 genes were identified using text-mining that had experimental pieces of evidence for their involvement in a plants response to salinity stress, directly or indirectly. With the help of co-expression network analysis for each of the 74 shortlisted genes, 320 more genes were identified. These together with 74 original bait genes were clubbed together and a complex gene expression network was created between these 394 genes. At a high confidence level, only 140 genes were found to be strongly connected. These 140 genes formed the base for the identification of orthologs in *Populus.*

It is known that there is a cross-talk between gene functions and molecular pathways for various abiotic stresses. Hence to Functional enrichments for Biological Processes using Gene Ontology data for following 4 abiotic stresses was performed:

1. ***Osmotic shock***: Osmotic shock or osmotic stress is a sudden change in the solute concentration around a cell, causing a rapid change in the movement of water across its cell membrane. Under conditions of high concentrations of either salt, substrates or any solute in the supernatant water is drawn out of the cells through osmosis. Osmotic shock is an initial response of the plant under the state of high salinity or drought.
2. ***Cold***: Low temperature, usually below 4-degree Celcius is the starting point of frost, which precipitates at freezing point. Cold and frost attacks also force the plant to initiate a series of metabolic processes to counteract the stress.
3. ***Drought***: Deprivation of water is a major stress that threatens the integrity of a cell and can onset a chain reaction leading to the disruption of major metabolic processes. Salinity and drought are quite inter-connected because salinity induces a state of physiological stress.
4. ***Hyperosmotic salinity response***: Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of detection of, or exposure to, an increase in the concentration of salt (particularly but not exclusively sodium and chloride ions) in the environment. Hyper salinity is also considered as the later stage of the salinity response because at this stage the toxicity effects of ion accumulation start to become pronounced.

Form the above discussion, it is clear that all the above responses are very well connected and there is a crosstalk between these processes. Functional enrichment analysis for the above 4 biological processes was performed and colour coding was done for each response – RED for OSMOTIC STRESS, BLUE for RESPONSE TO COLD, GREEN for WATER DEPRIVATION (DROUGHT) and YELLOW for HYPEROSMOTIC SALINITY. The following co-expression network shows the colour coded genes:

The Arabidopsis gene list for each of the 4 responses along with their Populus ortholog counterpart is tabulated below and is only suggestive, until proven in a wet-lab format separately.

Conclusion

Trees are excellent subjects to explore genes in their highly evolved state for adaptation and survival under environmental extremes. Tree genes have not yet been fully exploited for genetic improvement using a targeted transgenic approach, due to lack of availability of information on gene function from trees. This work has been an effort to identify tree genes that can later be used to study gene expression and carry out genetic engineering work to produce GM plants that can better withstand abiotic stresses

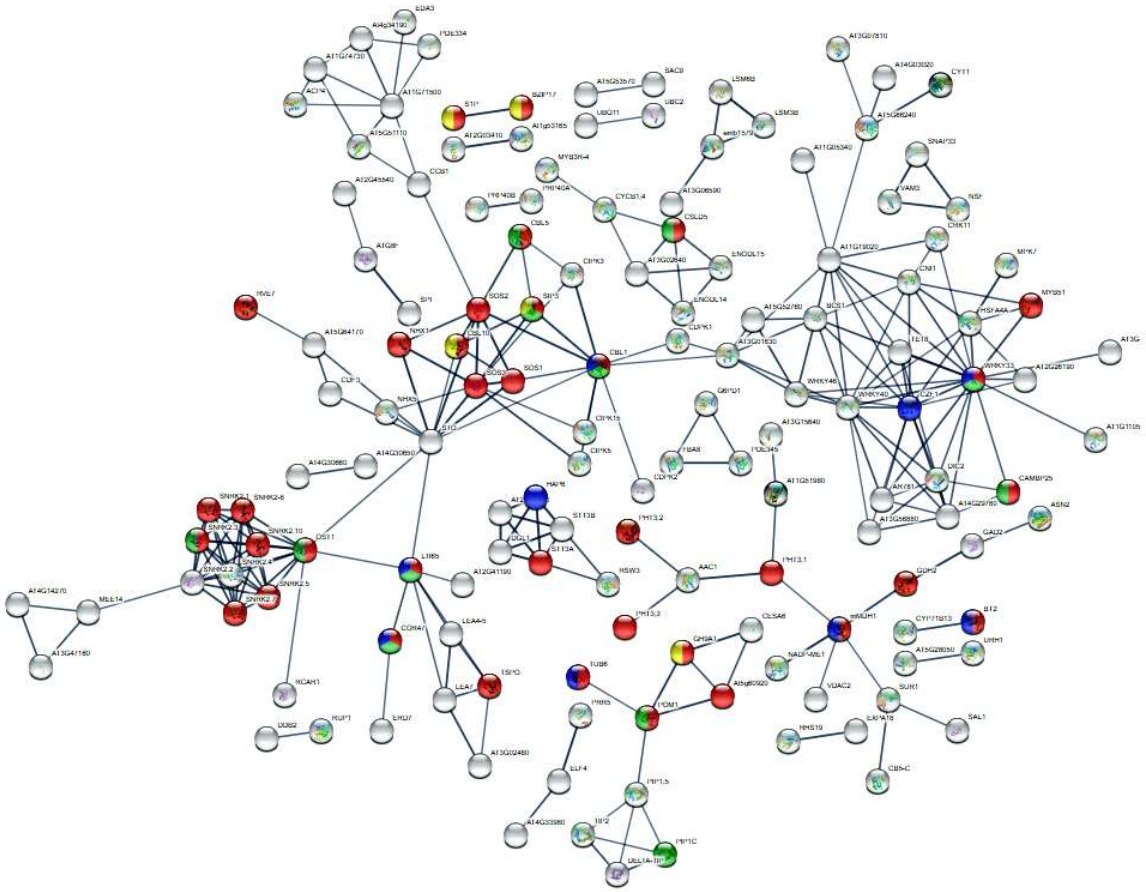

Fig.5: Co-Expression Network indicating Salinity responsive genes highlighted for their involvement in other abiotic stresses (Red=Osmotic shock responsive genes; Blue = Cold responsive genes; Green = Drought responsive genes; Yellow = Hyperosmotic salinity responsive genes)

Table 10: Genes indicated to be involved in OSMOTIC STRESS/SHOCK (RED in Netw

ork diagram)

Arabidopsis thaliana Gene	***PopulustrichocarpaOrtholog***
AT5G19660 **(SITE-1 protease)**	POPTR_018G081400 (subtilisin-like protease SBT6.1)
AT2G40950 **(Basic helix-loop-helix domain-containing protein)**	POPTR_016G032400 (bZIP transcription factor 17)
AT2G39770 **(CYTOKINESIS DEFECTIVE 1)**	POPTR_010G198800 (mannose-1-phosphate guanylyltransferase 1)
AT4G01420 **(Calcineurin B-like protein 5; Acts as a calcium sensor)**	POPTR_012G015100 (calcineurin B-like protein 4 isoform X2)
AT1G02730 **(1,4-beta-D-xylan synthase)**	POPTR_014G125100 (cellulose synthase-like protein D5)
AT1G18330 **(EARLY-PHYTOCHROME-RESPONSIVE1)**	POPTR_015G030400 (protein REVEILLE 7)

AT5G27150 **(Na+/H+ exchanger 1)**	POPTR_013G031700 (sodium/hydrogen exchanger 1)
AT5G35410 **(SALT OVERLY SENSITIVE 2)**	POPTR_018G130500 (CBL-interacting serine/threonine-protein kinase 24 isoform X1)
AT4G30960 **(SOS3-interacting protein 3; CIPK serine-threonine protein kinases interact with CBL proteins)**	POPTR_006G186200 (CBL-interacting protein kinase 9)
AT4G33000 **(Calcineurin B-like protein 10; Acts as a calcium sensor)**	POPTR_006G230200 (calcineurin B-like protein 10)
AT2G01980 **(SALT OVERLY SENSITIVE 1)**	POPTR_010G100900 (sodium/hydrogen exchanger 8 isoform X2)
AT5G24270 **(SALT OVERLY SENSITIVE 3)**	POPTR_015G013100 (calcineurin B-like protein 4 isoform X2)
AT4G17615 **(Calcineurin B-like protein 1)**	POPTR_003G084200 (calcineurin B-like protein 9 isoform X1)
AT1G18570	POPTR_002G096800 (transcription factor MYB93)
AT2G38470 **(WRKY DNA-binding protein 33; Transcription factor)**	POPTR_016G128300 (probable WRKY transcription factor 33)
AT5G08590.1 **(SNF1-related protein kinase 2.1)**	POPTR_003G015400 (serine/threonine-protein kinase SRK2A)
AT1G78290 **(SNF1-RELATED PROTEIN KINASE 2-8; Involved in gene regulation and confers tolerance to drought and osmotic stress)**	POPTR_002G099700 (serine/threonine-protein kinase SAPK2)
AT5G66880 **(Serine/threonine-protein kinase SRK2)**	POPTR_005G134400 (serine/threonine-protein kinase SRK2I)
AT5G63650 **(SNF1-related protein kinase 2.5)**	POPTR_003G015400 (serine/threonine-protein kinase SRK2A)
AT4G40010 **(SNF1-related protein kinase 2.7)**	POPTR_007G096400 (Serine/threonine-protein kinase SAPK2 isoform X1)
AT5G52300 **(CAP160 protein)**	POPTR_012G141300 (low-temperature-induced 65 kDa protein)
AT1G20440 **(Dehydrin COR47)**	POPTR_005G248100 (phosphoprotein ECPP44)
AT2G47770.1 **(Tryptophan-rich sensory protein-like protein)**	POPTR_002G206100 (translocator protein homolog)
AT5G19690 **(Staurosporin and temperature sensitive 3-like A)**	POPTR_018G086000 (dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A)
AT3G48850 **(Phosphate transporter 3;2)**	POPTR_015G104400 (mitochondrial phosphate carrier protein 3, mitochondrial)
AT1G51980 **(Insulinase (Peptidase family M16) protein)**	POPTR_010G036700 (mitochondrial-processing peptidase subunit

	alpha)
AT2G41010 **(Calmodulin binding protein 25)**	POPTR_016G029600 (calmodulin-binding protein 25)
AT5G07440 **(Glutamate dehydrogenase 2)**	POPTR_015G111000 (glutamate dehydrogenase 2)
AT1G53240 **(Malate dehydrogenase 1)**	POPTR_011G096300 (malate dehydrogenase, mitochondrial)
AT3G48360 **(BTB and TAZ domain protein 2)**	POPTR_012G091200 (BTB/POZ and TAZ domain-containing protein 1 isoform X1)
AT5G49720 **(Endoglucanase 25)**	POPTR_003G151700 (endoglucanase 25)
AT5G60920 **COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family**	POPTR_0015s07100 similar to probable phytochelatinsynthetase
AT1G05850 **(POM-POM1; No chitinase activity)**	XP_011009946.1 (PREDICTED: chitinase-like protein 2)
AT5G12250 **(Beta-6 tubulin; Tubulin is the major constituent of microtubules)**	POPTR_016G033200 (tubulin beta-5 chain)

Table 11: Genes indicated to be involved in COLD RESPONSE (BLUE in Network diagram)

Arabidopsis thaliana Gene	***PopulustrichocarpaOrtholog***
AT4G17615 **(Calcineurin B-like protein 1)**	POPTR_003G084200 (calcineurin B-like protein 9 isoform X1)
At2g40140 **zinc finger (CCCH-type) family protein**	POPTR_0010s19520 zinc finger (CCCH-type) family protein
AT4G21150 **(HAPLESS 6; Essential subunit of the N-oligosaccharyl transferase (OST) complex)**	POPTR_002G036600 (dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2 isoform X1)
AT5G52300 **(CAP160 protein)**	POPTR_012G141300 (low-temperature-induced 65 kDa protein)
AT1G20440 **(Dehydrin COR47)**	POPTR_005G248100 (phosphoprotein ECPP44)
AT5G12250 **(Beta-6 tubulin; Tubulin is the major constituent of microtubules)**	POPTR_016G033200 (tubulin beta-5 chain)
AT1G53240 **(Malate dehydrogenase 1)**	POPTR_011G096300 (malate dehydrogenase, mitochondrial)
AT3G48360 **(BTB and TAZ domain protein 2)**	POPTR_012G091200 (BTB/POZ and TAZ domain-containing protein 1 isoform X1)

Table 12. Genes indicated to be involved during DROUGHT RESPONSE (GREEN in Network diagram)

Arabidopsis thaliana Gene	***PopulustrichocarpaOrtholog***
AT1G02730 **(1,4-beta-D-xylan synthase)**	POPTR_014G125100 (cellulose synthase-like protein D5)
AT4G01420 **calcineurin B-like protein 5**	None found
AT4G30960 **(SOS3-interacting protein 3; CIPK serine-threonine protein kinases interact with CBL proteins)**	POPTR 006G186200 (CBL-interacting protein kinase 9)
AT4G33000 **calcineurin B-like protein 10**	None found
AT2G41010 **CALMODULIN (CAM)-BINDING PROTEIN OF 25 KDA**	POPTR_0006s24630
AT5G66880 **(Serine/threonine-protein kinase SRK2)**	POPTR_005G134400 (serine/threonine-protein kinase SRK2I)
AT4G33950 **OPEN STOMATA 1, OST1**	POPTR_0004s15270 (OTS1)
AT1G20440 **(Dehydrin COR47)**	POPTR_005G248100 (phosphoprotein ECPP44)
AT1G05850 **(POM-POM1; No chitinase activity)**	XP_011009946.1 (PREDICTED: chitinase-like protein 2)

Table 12. Genes indicated to be involved during HYPEROSMOTIC SALINITY RESPONSE (YELLOW in Network diagram)

Arabidopsis thaliana Gene	***PopulustrichocarpaOrtholog***
AT5G19660 **SITE-1 protease**	POPTR_0018s08810
AT2G40950 **BZIP17**	POPTR_0016s03220 bZIP transcription factor family protein
AT4G30960 **SIP3**	POPTR_0006s20030 similar to CBL-INTERACTING PROTEIN KINASE 6
AT4G33000 **CALCINEURIN B-LIKE PROTEIN 10**	POPTR_0006s24630

like salinity, drought and temperature perturbations, and at the same time compensate for diminishing productivity. This is highly significant in the face of fast-changing climatic conditions of the globe.

A total of 74 genes known to be involved in a plant's response to salinity stress (and associated physiological drought stress) have been identified and shortlisted backed by high-quality experimental evidence from the global scientific literature.

Gene co-expression networks were created around each of the 74 genes individually. This exercise enabled the identification of 321 more genes that are involved during a plant's response to

salinity stress. Together, a database of 395 genes, their protein sequences and their known function in *Arabidopsis* has been created as a battery of genes that can be utilized as a readily assembled package any time in many ways for accessing and improving abiotic stress response of a plant, through tools of bioinformatics and molecular genetics.

The battery of 395 genes was used to construct a gene co-expression network. The Network was able to bring out the specific correlation between all the genes. The network was constructed using STRING database and analyzed using Cytoscape. Clustering of gene Sub-networks was done using ClusterViz within Cytoscape. This resulted in the delineation of 11 sub-networks (clusters) based on maximal clique-based EAGLE algorithm.

The Cluster analysis resulted in the identification of a set of 140 network connected genes. These were the final list of genes of Arabidopsis that were considered relevant for a plants response to salinity stress.

For all the final 140 salinity implicated genes, the protein sequences were obtained from the TAIR database. The protein sequences were blasted against the *Populustrichocarpa* protein database at NCBI. The reciprocal blast was carried out for the top-ranking best hits of the identified ortholog against the *Arabidopsis* background. Thus a total of 140 Poplar genes were finally identified through the cross-species comparative genomics approach. This list of Poplar genes is the main product of the analysis and the project. It is a unique set of information which together forms the major leads for future gene cloning and characterization work for isolation of genes for salt tolerance. Of course, this would need wet-lab validation.

Another important finding was the identification of uncharacterized proteins that are the indicators of Domains of Unknown Functions (DUFs) pointing towards Genes of putative nature; as well as Transcription Factors (TFs) or their activators or suppressors. In *Arabidopsis,* a total of 10 DUFs and 10 TFs were identified, while in *Populustrichocarpa* 20 DUFs and 10 TFs were identified. The findings of DUFs or Genes of putative nature are important for future research aimed at their characterization because these indicate genes that have a role in abiotic stress management in some way but are not yet characterized or known. This is a very important set of findings as this is the first step that will lead to the discovery of new genes that are implicated in a plant's response to abiotic stresses.

Gene Ontology (GO) enrichment analysis was also performed using both String database as well as PlantGenei. It was found that a total of several biological processes (10 processes with *P. trichocarpa*database background and 34 with *Arabidopsis* background) are correlated with genes that have a role in salt stress response of a plant. The topmost biological processes at a corrected P-value of 2.490e-09 was in which 11 genes were involved followed by a response to cold at P-value of 4.330e-09 with 12 genes and closely with the response to salt stress with 18 genes at P-value of 4.898e-09. This was followed by the osmotic stress response, cellulose biosynthesis and hyperosmotic salinity response.

Based on the GO enrichment analysis, further analysis was done for pinpointing out the other Abiotic stress (Osmotic shock, cold, drought and hyperosmotic salinity response) responsive genes, from within the generated network as well as additionally (if any). This resulted in the identification of 34 *Arabidopsis* genes and counterpart *Populus* orthologous genes for osmotic shock response, 8 for the cold response for both species, 9 Arabidopsis and 8 *Populus* genes for drought response and 4 each of *Arabidopsis* and populous for Hyperosmotic salinity response. Many of the identified *Populus* orthologous genes

were uncharacterized and unmarked for a functional significance before this analysis. This analysis has thus brought to light the new *Populus* genes.

Acknowledgements:
The PI (author) acknowledges the Indian Council of Forestry Research & Education (ICFRE), Dehradun, an autonomous council under Ministry of Environment, Forest & Climate Change, Govt. of India - for funding the research project and Director AFRI, Jodhpur for administrative and research support, which enabled its execution and successful conclusion.

References

1. FAO (2008). Handbook for saline soil management (Ed: R. Vargas, E.I. Pankova, S.A. Balyuk, P.V. Krasilnikov and G.M. Khasankhanova). Published by the Food and Agriculture Organization of the United Nations and Lomonosov Moscow State University. P.144
2. Neale DB, Langley CH, Salzberg SL and Wegrzyn JL 2013, Open access to tree genomes: The path to a better forest. Genome Biology. Vol 14. 120. 10.1186/gb-2013-14-6-120.
3. Obayashi T, Aoki Y, Tadaka S and Kagaya Y 2018, ATTED-II in 2018: A Plant Co-expression Database Based on Investigation of Statistical Property of the Mutual Rank Index. Plant Cell Physiology, 59, e3
4. Aoki Y, Okamura Y, Tadaka S, Kinoshita K and Obayashi T 2016, ATTED-II in 2016: a plant co-expression database towards lineage-specific co-expression. Plant Cell Physiology. 57 e5
5. Yang X, Yea C-Y, Bisaria A, Tuskana GA and Kalluri UC 2011, Identification of candidate genes in Arabidopsis and Populus cell wall biosynthesis using text-mining, co-expression network analysis and comparative genomics. Plant Science 181: 675– 687
6. Snel B, Lehmann G, Bork P and Huynen MA 2000, STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 28 (18) 3442–3444. doi:10.1093/nar/28.18.3442. PMC 110752. PMID 10982861
7. von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B and Bork P 2007, STRING 7—recent developments in the integration and prediction of protein interactions. Nucleic Acids Res. 35 (Database issue): D358–62. doi:10.1093/nar/gkl825. PMC 1669762. PMID 17098935
8. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P and von Mering C 2009, STRING 8—a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 37 (Database issue): D412–6. doi:10.1093/nar/gkn760. PMC 2686466. PMID 18940858
9. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ and von Mering C 2011, The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 39 (Database issue): D561–8. doi:10.1093/nar/gkq973. PMC 3013807. PMID 21045058
10. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ and von Mering C 2015, STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43 (Database issue): D447–52. doi:10.1093/nar/gku1003. PMC 4383874. PMID 25352553
11. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P,

von Mering C and Jensen LJ 2013, STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41 (Database issue): D808–15. doi:10.1093/nar/gks1094. PMC 3531103. PMID 23203871

12. Wodak SJ, Pu S, Vlasblom J and Séraphin B 2009, Challenges and rewards of interaction proteomics. Mol Cell Proteomics. 8 (1) 3–18. doi:10.1074/mcp.R800014-MCP200
13. Schwartz AS, Yu J, Gardenour KR, Finley Jr RL and Ideker T 2008, Cost-effective strategies for completing the interactome. Nature Methods. 6 (1) 55–61. doi:10.1038/nmeth.1283. PMC 2613168
14. Gu F, Bringmann M, Combs J, Yang J, Bergmann D and Nielsen E 2016, The Arabidopsis CSLD5 functions in cell plate formation in a cell cycle dependent manner. Plant Cell. 28(7) 1722-1737. Epub 2016 Jun 27
15. Yin L, Verhertbruggen Y, Oikawa A, Manisseri C, Knierim B, Prak L, Jensen JK, Knox JP, Auer M, Willats WG and Scheller HV 2011, The Cooperative Activities of CSLD2, CSLD3, and CSLD5 Are Required for Normal Arabidopsis Development. Mol Plant. 4(6) 1024-1037. Epub 2011 Apr 6
16. Zhu J, Lee BH, Dellinger M, Cui X, Zhang C, Wu S, Nothnagel EA and Zhu JK 2010, A cellulose synthase-like protein is required for osmotic stress tolerance in Arabidopsis. Plant J. 63(1) 128-140
17. Steffens A, Bräutigam A, Jakoby M, and Hülskamp M 2015, The BEACH Domain Protein SPIRRIG Is Essential for Arabidopsis Salt Stress Tolerance and Functions as a Regulator of Transcript Stabilization and Localization. PLoS biology. 13(7). e1002188. doi:10.1371/journal.pbio.1002188
18. Sanchez-Rodriguez C, Bauer S, Hematy K, Saxe F, Ibanez AB, Vodermaier V, Konlechner C, Sampathkumar A, Ruggeberg M, Aichinger E, Neumetzler L, Burgert I, Somerville C, Hauser MT and Persson S 2012, Chitinase-like1/pom-pom1 and its homolog CTL2 are glucan-interacting proteins important for cellulose biosynthesis in Arabidopsis. Plant Cell. 24(2) 589-607. Epub 2012 Feb 10
19. Datta S, Hettiarachchi C, Johansson H and Holm M 2007, SALT TOLERANCE HOMOLOG2, a B-box protein in Arabidopsis that activates transcription and positively regulates light-mediated development. Plant Cell 19 3242–3255
20. Julkowska MM, McLoughlin F, Galvan-Ampudia CS, Rankenberg JM, Kawa D, Klimecka M, Haring MA, Munnik T, Kooijman EE and Testerink C 2015, Identification and functional characterization of the Arabidopsis Snf1-related protein kinase SnRK2.4 phosphatidic acid-binding domain. Plant Cell Environ. 38(3) 614-624
21. McLoughlin F, Galvan-Ampudia CS, Julkowska MM, Caarls L, van der Does D, Lauriere C, Munnik T, Haring MA and Testerink C 2012, The Snf1-related protein kinases SnRK2.4 and SnRK2.10 are involved in maintenance of root system architecture during salt stress. Plant J. 72(3) 436-449
22. Boudsocq M, Barbier-Brygoo H and Lauriere C 2004, Identification of nine sucrose nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in Arabidopsis thaliana. J Biol Chem. 279(40) 41758-41766. Epub 2004 Jul 29
23. Urao T, Katagiri T, Mizoguchi T, Yamaguchi-Shinozaki K, Hayashida N and Shinozaki K 1994, Two genes that encode Ca(2+)-dependent protein kinases are induced by drought and high-salt stresses in Arabidopsis

thaliana. Mol Gen Genet. 244(4) 331-40
24. Harmon AC, Gribskov M, Gubrium E and Harper JF 2001, The CDPK superfamily of protein kinases. New Phytologist. 151 175-183.
25. Cheng SH, Willmann MR, Chen HC and Sheen J 2002, Calcium signaling through protein kinases. The Arabidopsis calcium-dependent protein kinase gene family. Plant Physiol. 129(2) 469-85
26. Kim SJ and Bassham DC 2011, TNO1 is involved in salt tolerance and vacuolar trafficking in Arabidopsis thaliana. Plant Physiol. 156(2) 514-526. Epub 2011 Apr 26.
27. Roy R and Bassham DC 2015, Gravitropism and Lateral Root Emergence are Dependent on the Trans-Golgi Network Protein TNO1. Front Plant Sci. 12 6 969. eCollection 2015
28. Tarte VN, Seok HY, Woo DH, Le DH, Tran HT, Baik JW, Kang IS, Lee SY, Chung T and Moon YH 2015, Arabidopsis Qc-SNARE gene AtSFT12 is involved in salt and osmotic stress responses and Na(+) accumulation in vacuoles. Plant Cell Rep. 34(7) 1127-1138. Epub 2015 Feb 18
29. Endler A, Kesten C, Schneider R, Zhang Y, Ivakov A, Froehlich A, Funke N and Persson S 2015, A Mechanism for Sustained Cellulose Synthesis during Salt Stress. Cell. 162(6) 1353-1364. Epub 2015 Sep 3
30. Medina J, Ballesteros ML and Salinas J 2007, Phylogenetic and functional analysis of Arabidopsis RCI2 genes. J Exp Bot. 58(15-16) 4333-4346
31. Hong SW, Jon JH, Kwak JM and Nam HG 1997, Identification of a receptor-like protein kinase gene rapidly induced by abscisic acid, dehydration, high salt, and cold treatments in Arabidopsis thaliana. Plant Physiol. 113(4) 1203-1212
32. Osakabe Y, Mizuno S, Tanaka H, Maruyama K, Osakabe K, Todaka D, Fujita Y, Kobayashi M, Shinozaki K and Yamaguchi-Shinozaki K 2010, Overproduction of the membrane-bound receptor-like protein kinase1, RPK1, enhances abiotic stress tolerance in Arabidopsis. J Biol Chem. 285(12) 9190-9201. Epub 2010 Jan 20
33. Shi CC, Feng CC, Yang MM, Li JL, Li XX, Zhao BC, Huang ZJ and Ge RC 2014, Overexpression of the receptor-like protein kinase genes AtRPK1 and OsRPK1 reduces the salt tolerance of Arabidopsis thaliana. Plant Sci. 217-218 63-70. Epub 2013
34. Gao L and Xiang CB 2008, The genetic locus At1g73660 encodes a putative MAPKKK and negatively regulates salt tolerance in Arabidopsis. Plant MolBiol 67(1-2) 125-34. Epub 2008 Feb 26
35. Kim MJ, Park MJ, Seo PJ, Song JS, Kim HJ and Park CM 2012, Controlled nuclear import of NTL6 transcription factor reveals a cytoplasmic role of SnRK2.8 in drought stress response. Biochem J. 15 448(3) 353-363.
36. Frick EM and Strader LC 2017, Kinase MPK17 and the peroxisome division factor PMD1 influence salt-induced peroxisome proliferation. Plant Physiol. 2017 Sep 20
37. Wu SJ, Ding L and Zhu JK 1996, SOS1, a Genetic Locus Essential for Salt Tolerance and Potassium Acquisition. Plant Cell. 8(4) 617-627
38. Shi H, Ishitani M, Kim C and Zhu JK 2000, The Arabidopsis thaliana salt tolerance gene SOS1 encodes a putative Na+/H+ antiporter. Proc Natl AcadSci U S A. 97(12) 6896-6901
39. Qiu QS, Guo Y, Dietrich MA, Schumaker KS and Zhu JK 2002, Regulation of SOS1, a plasma membrane Na+/H+ exchanger in Arabidopsis thaliana, by SOS2 and SOS3. Proc Natl AcadSci U S A. 99(12) 8436-8441

40. Shi H, Quintero FJ, Pardo JM and Zhu JK 2002, The putative plasma membrane Na(+)/H(+) antiporter SOS1 controls long-distance Na(+) transport in plants. Plant Cell. 14(2) 465-477
41. Nah G, Pagliarulo CL, Mohr PG, Luo M, Sisneros N, Yu Y, Collura K, Currie J, Goicoechea JL, Wing RA and Schumaker KS 2009, Comparative sequence analysis of the SALT OVERLY SENSITIVE1 orthologous region in Thellungiellahalophila and Arabidopsis thaliana. Genomics. 94(3) 196-203. Epub 2009 May 28
42. Oh DH, Lee SY, Bressan RA, Yun DJ and Bohnert HJ 2010, Intracellular consequences of SOS1 deficiency during salt stress. J Exp Bot. 61(4) 1205-1213. Epub 2010 Jan 6.
43. Yue Y, Zhang M, Zhang J, Duan L and Li Z 2012, SOS1 gene overexpression increased salt tolerance in transgenic tobacco by maintaining a higher K(+)/Na(+) ratio. J Plant Physiol. 169(3) 255-61. Epub 2011 Nov 23
44. Guan Q, Wu J, Yue X, Zhang Y and Zhu J 2013, A nuclear calcium-sensing pathway is critical for gene regulation and salt stress tolerance in Arabidopsis. PLoS Genet. 9(8) e1003755
45. Ko JH, Yang SH and Han KH 2006, Upregulation of an Arabidopsis RING-H2 gene, XERICO, confers drought tolerance through increased abscisic acid biosynthesis. Plant J. 47(3) 343-355. Epub 2006 Jun 22.
46. Zhu W, Miao Q, Sun D, Yang G, Wu C, Huang J and Zheng C 2012, The Mitochondrial Phosphate Transporters Modulate Plant Responses to Salt Stress via Affecting ATP and Gibberellin Metabolism in Arabidopsis thaliana. PLoS One. 7(8) e43530. Epub 2012 Aug 24
47. Kim KN, Cheong YH, Grant JJ, Pandey GK and Luan S 2003, CIPK3, a calcium sensor-associated protein kinase that regulates abscisic acid and cold signal transduction in Arabidopsis. Plant Cell. 15(2) 411-423.
48. Tang RJ, Zhao FG, Garcia VJ, Kleist TJ, Yang L, Zhang HX and Luan S 2015, Tonoplast CBL-CIPK calcium signaling network regulates magnesium homeostasis in Arabidopsis. Proc Natl AcadSci U S A. 112(10) 3134-3139. Epub 2015 Feb 2
49. Jiang Y and Deyholos MK 2006, BMC Plant Biology. 6(1) p.25. Available at: http://dx.doi.org/10.1186/1471-2229-6-25.
50. Strizhov N, Abraham E, Okresz L, Blickling S, Zilberstein A, Schell J, Koncz C and Szabados L 1997, Differential expression of two P5CS genes controlling proline accumulation during salt-stress requires ABA and is regulated by ABA1, ABI1 and AXR2 in Arabidopsis. Plant J. 12(3) 557-569
51. Abraham E, Rigo G, Szekely G, Nagy R, Koncz C and Szabados L 2003, Light-dependent induction of proline biosynthesis by abscisic acid and salt stress is inhibited by brassinosteroid in Arabidopsis. Plant Mol Biol. Feb 51(3) 363-372
52. Liu JX, Srivastava R, Che P and Howell SH 2007, Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. Plant J. 51(5) 897-909. Epub 2007 Jul 28
53. Liu JX, Srivastava R, Che P and Howell SH 2008, Salt stress signaling in Arabidopsis thaliana involves a membrane-bound transcription factor AtbZIP17 as a signal transducer. Plant Signal Behav. 3(1) 56-57
54. Perruc E, Charpenteau M, Ramirez BC, Jauneau A, Galaud JP, Ranjeva R and Ranty B 2004, A novel

calmodulin-binding protein functions as a negative regulator of osmotic stress tolerance in Arabidopsis thaliana seedlings. Plant J. 38(3) 410-420

55. Geisler M, Frangne N, Gomes E, Martinoia E and Palmgren MG 2000, The ACA4 gene of Arabidopsis encodes a vacuolar membrane calcium pump that improves salt tolerance in yeast. Plant Physiol. 124(4) 1814-1827
56. Song CP and Galbraith DW 2006, AtSAP18, an orthologue of human SAP18, is involved in the regulation of salt stress and mediates transcriptional repression in Arabidopsis. Plant Mol Biol. 60(2) 241-257
57. Ding ZJ, Yan JY, Xu XY, Li GX and Zheng SJ 2013, WRKY46 functions as a transcriptional repressor of ALMT1 regulating Al-induced malate secretion in Arabidopsis. Plant J. 76(5) 825-835. Epub 2013 Nov 5
58. Balsemao-Pires E, Jaillais Y, Olson BJ, Andrade LR, Umen JG, Chory J and Sachetto-Martins G 2011, The Arabidopsis translocator protein (AtTSPO) is regulated at multiple levels in response to salt stress and perturbations in tetrapyrrole metabolism. BMC Plant Biol. 11 108
59. Garcia ME, Lynch T, Peeters J, Snowden C and Finkelstein R 2008, A small plant-specific protein family of ABI five binding proteins (AFPs) regulates stress response in germinating Arabidopsis seeds and seedlings. Plant Mol Biol. 67(6) 643-658.
60. Liu B, Feng D, Zhang B, Mu P, Zhang Y, He Y, Qi K, Wang J and Wang H 2012, Musa paradisica RCI complements AtRCI and confers Na(+) tolerance and K(+) sensitivity in Arabidopsis. Plant Sci. 184 102-111. Epub 2011 Dec 13
61. Sakamoto H, Matsuda O and Iba K 2008, ITN1, a novel gene encoding an ankyrin-repeat protein that affects the ABA-mediated production of reactive oxygen species and is involved in salt-stress tolerance in Arabidopsis thaliana. Plant J. 56(3) 411-422. Epub 2008 Aug 6
62. Zsigmond L, Rigo G, Szarka A, Szekely G, Otvos K, Darula Z, Medzihradszky KF, Koncz C, Koncz Z, and Szabados L 2008, Arabidopsis PPR40 connects abiotic stress responses to mitochondrial electron transport. Plant Physiol. 146(4) 1721-1737. Epub 2008 Feb 27
63. Schussler MD, Alexandersson E, Bienert GP, Kichey T, Laursen KH, Johanson U, Kjellbom P, Schjoerring JK and Jahn TP 2008, The effects of the loss of TIP1;1 and TIP1;2 aquaporins in Arabidopsis thaliana. Plant J. 56(5) 756-767. Epub 2008 Sep 19
64. He XJ, Zhang ZG, Yan DQ, Zhang JS and Chen SY 2004, A salt-responsive receptor-like kinase gene regulated by the ethylene signaling pathway encodes a plasma membrane serine/threonine kinase. TheorAppl Genet. 109(2) 377-383
65. Li B, Byrt CS, Qiu J, Baumann U, Hrmova M, Evrard A, Johnson AA, Birnbaum KD, Mayo GM, Jha D, Henderson SW, Tester M, Gilliham M and Roy SJ 2015, Identification of a stelar-localised transport protein that facilitates root-to-shoot transfer of chloride in Arabidopsis. Plant Physiol. 170(2) 1014-1029. Epub 2015 Dec 11
66. Basu D, Tian L, Debrosse T, Poirier E, Emch K, Herock H, Travers A and Showalter AM 2016, Glycosylation of a Fasciclin-Like Arabinogalactan-Protein (SOS5) Mediates Root Growth and Seed Mucilage Adherence via a Cell Wall Receptor-Like Kinase (FEI1/FEI2) Pathway in Arabidopsis. PLoS One. 11(1) e0145092
67. Vitart V, Baxter I, Doerner P and Harper JF 2001, Evidence for a role in growth and salt resistance of a plasma

membrane H+-ATPase in the root endodermis. Plant J. 27(3) 191-201
68. Mogami J, Fujita Y, Yoshida T, Tsukiori Y, Nakagami H, Nomura Y, Fujiwara T, Nishida S, Yanagisawa S, Ishida T, Takahashi F, Morimoto K, Kidokoro S, Mizoi J, Shinozaki K and Yamaguchi-Shinozaki K 2015, Two Distinct Families of Protein Kinases Are Required for Plant Growth under High External Mg2+ Concentrations in Arabidopsis. Plant Physiol. 167(3) 1039-1057
69. Zhang H, Cui F, Wu Y, Lou L, Liu L, Tian M, Ning Y, Shu K, Tang S and Xie Q 2015, The RING Finger Ubiquitin E3 Ligase SDIR1 Targets SDIR1-INTERACTING PROTEIN1 for Degradation to Modulate the Salt Stress Response and ABA Signaling in Arabidopsis. Plant Cell. 27(1) 214-227. Epub 2015 Jan 23
70. Szekely G, Abraham E, Cseplo A, Rigo G, Zsigmond L, Csiszar J, Ayaydin F, Strizhov N, Jasik J, Schmelzer E, Koncz C and Szabados L 2008, Duplicated P5CS genes of Arabidopsis play distinct roles in stress regulation and developmental control of proline biosynthesis. Plant J. 53(1) 11-28. Epub 2007 Oct 27
71. Saito S, Hamamoto S, Moriya K, Matsuura A, Sato Y, Muto J, Noguchi H, Yamauchi S, Tozawa Y, Ueda M, Hashimoto K, Koster P, Dong Q, Held K, Kudla J, Utsumi T and Uozumi N 2018, N-myristoylation and S-acylation are common modifications of Ca(2+) -regulated Arabidopsis kinases and are required for activation of the SLAC1 anion channel. New Phytol. 2018 Mar 2
72. Ambrosone A, Batelli G, Nurcato R, Aurilia V, Punzo P and Bangarusamy D K 2015, The Arabidopsis RNA-binding protein AtRGGA regulates tolerance to salt and drought stress. Plant Physiol. 168 292–306
73. Cheong YH, Kim KN, Pandey GK, Gupta R, Grant JJ and Luan S 2003, CBL1, a calcium sensor that differentially regulates salt, drought, and cold responses in Arabidopsis. Plant Cell. 15(8) 1833-1845
74. Wu JX, Li J, Liu Z, Yin J, Chang ZY, Rong C, Wu JL, Bi FC and Yao N 2015, The Arabidopsis ceramidase AtACER functions in disease resistance and salt tolerance. Plant J. 81(5) 767-780
75. Adai A, Johnson C, Mlotshwa S, Archer-Evans S, Manocha V and Vance V 2005, Computational prediction of miRNAs in Arabidopsis thaliana. Genome Res. 15 78–91
76. Guo Y, Halfter U, Ishitani M and Zhu JK 2001, Molecular characterization of functional domains in the protein kinase SOS2 that is required for plant salt tolerance. Plant Cell. 13(6) 1383-1400
77. Kim BG, Waadt R, Cheong YH, Pandey GK, Dominguez-Solis JR, Schultke S, Lee SC, Kudla J and Luan S 2007, The calcium sensor CBL10 mediates salt tolerance by regulating ion homeostasis in Arabidopsis. Plant J. 52(3) 473-484. Epub 2007 Sep 6
78. Quan R, Lin H, Mendoza I, Zhang Y, Cao W, Yang Y, Shang M, Chen S, Pardo JM and Guo Y 2007, SCABP8/CBL10, a Putative Calcium Sensor, Interacts with the Protein Kinase SOS2 to Protect Arabidopsis Shoots from Salt Stress. Plant Cell. 19(4) 1415-31. Epub 2007 Apr 20
79. Lin H, Yang Y, Quan R, Mendoza I, Wu Y, Du W, Zhao S, Schumaker KS, Pardo JM and Guo Y 2009, Phosphorylation of SOS3-LIKE CALCIUM BINDING PROTEIN8 by SOS2 Protein Kinase Stabilizes Their Protein Complex and Regulates Salt Tolerance in Arabidopsis. Plant Cell. 21(5) 1607-1619. Epub 2009 May 15
80. Chien PS, Nam HG and Chen YR 2015, A salt-regulated peptide derived from the CAP superfamily protein negatively regulates salt-stress tolerance in Arabidopsis. J Exp Bot.

66(17) 5301-5313. doi: 10.1093/jxb/erv263. Epub 2015 Jun 20
81. Zarepour M, Kaspari K, Stagge S, Rethmeier R, Mendel RR and Bittner F 2010, Xanthine dehydrogenase AtXDH1 from Arabidopsis thaliana is a potent producer of superoxide anions via its NADH oxidase activity. Plant Mol Biol. 72(3) 301-310. Epub 2009 Nov 14
82. Pou A, Jeanguenin L, Milhiet T, Batoko H, Chaumont F and Hachez C 2016, Salinity-mediated transcriptional and post-translational regulation of the Arabidopsis aquaporin PIP2;7. Plant Mol Biol. 92(6) 731-744. Epub 2016 Sep 26
83. Kai-Chau Huang, Wei-Chih Lin and Wan-Hsing Cheng 2018, Salt hypersensitive mutant 9, a nucleolar APUM23 protein, is essential for salt sensitivity in association with the ABA signaling pathway in Arabidopsis. BMC Plant Biol. 18 40
84. Cai X and Lytton J 2004, The cation/Ca(2+) exchanger superfamily: phylogenetic analysis and structural implications. MolBiolEvol. 21(9) 1692-1703. Epub 2004 May 26
85. Koiwa H, Li F, McCully MG, Mendoza I, Koizumi N, Manabe Y, Nakagawa Y, Zhu J, Rus A, Pardo JM, Bressan RA and Hasegawa PM 2003, The STT3a subunit isoform of the Arabidopsis oligosaccharyltransferase controls adaptive responses to salt/osmotic stress. Plant Cell. 15(10) 2273-2284. Epub 2003 Sep 5
86. Ishitani M, Liu J, Halfter U, Kim CS, Shi W and Zhu JK 2000, SOS3 function in plant salt tolerance requires N-myristoylation and calcium binding. Plant Cell. 12(9) 1667-1678.
87. Apse MP, Aharon GS, Snedden WA and Blumwald E 1999, Salt tolerance conferred by overexpression of a vacuolar Na+/H+ antiport in Arabidopsis. Science. 285(5431) 1256-1258
88. Liu J, Ishitani M, Halfter U, Kim CS and Zhu JK 2000, The Arabidopsis thaliana SOS2 gene encodes a protein kinase that is required for salt tolerance. Proc Natl AcadSci U S A. 97(7) 3730-3734
89. Shi H, Xiong L, Stevenson B, Lu T and Zhu JK 2002, The Arabidopsis salt overly sensitive 4 mutants uncover a critical role for vitamin B6 in plant salt tolerance. Plant Cell. 14(3) 575-588.
90. Gonzalez E, Danehower D and Daub ME 2007, Vitamer Levels, Stress Response, Enzyme Activity and Gene Regulation of Arabidopsis Lines Mutant in the Pyridoxine/Pyridoxamine 5-Phosphate Oxidase (PDX3) and the Pyridoxal Kinase (SOS4) Genes Involved in the Vitamin B6 Salvage Pathway. Plant Physiol. 145(3) 985-996
91. Pandey GK, Kanwar P, Singh A, Steinhorst L, Pandey A, Yadav AK, Tokas I, Sanyal S, Kim BG, Lee SC, Cheong YH, Kudla J and Luan S 2015, CBL-interacting protein kinase, CIPK21, regulates osmotic and salt stress responses in Arabidopsis. Plant Physiol. 169(1) 780-792. Epub 2015 Jul 21
92. Tian M, Lou L, Liu L, Yu F, Zhao Q, Zhang H, Wu Y, Tang S, Xia R, Zhu B, Serino G and Xie Q 2015, The RING finger E3 ligase STRF1 is involved in membrane trafficking and modulates salt-stress response in Arabidopsis thaliana. Plant J. 82(1) 81-92. Epub 2015 Mar 7
93. Quintero FJ, Garciadeblas B and Rodriguez-Navarro A 1996, The SAL1 gene of Arabidopsis, encoding an enzyme with 3'(2'),5'-bisphosphate nucleotidase and inositol polyphosphate 1-phosphatase activities, increases salt tolerance in yeast. Plant Cell. 8(3) 529-537