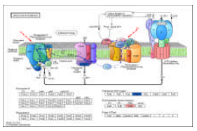
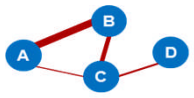
Comparative Transcriptomics to Study Water Deficit and Flooding

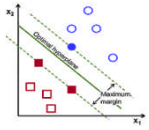
John J. Riascos & Hector. F. Espitia & Irene Tischer.
Presented by Jershon López

1 . Expression levels | 2. KEGG pathways | 3. **Co-expression networks** | 4. **Ranking** based on ***SVM-RFE***



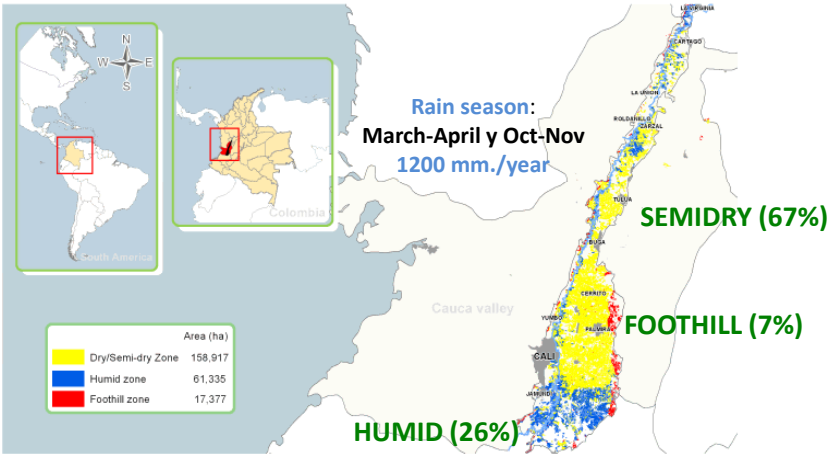






11th Germplasm& Breeding; 8th Molecular Biology ISSCT Workshop 1-5 June 2015

Sugarcane cultivation concentrates around the Cauca valley river



Rain season:
March-April y Oct-Nov
1200 mm./year

SEMIDRY (67%)

FOOTHILL (7%)

HUMID (26%)

Relevance

- Irrigation and drainage are integral part of Colombian sugarcane agriculture.
- Of the 230,000 ha of sugarcane cultivated land, approximately 91% (210,000 ha) uses irrigation; while almost 76% (175,000 ha) could need at some point a drainage system.
- While this favors high yields (Colombia's sugarcane industry ranks No 1. in production/ha), it also represents high costs (~50% of the total).

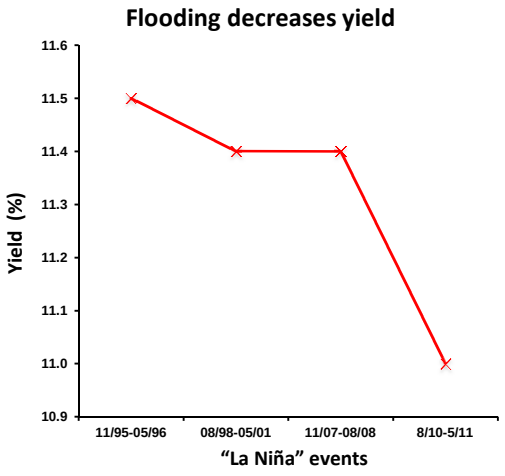
3

Colombian Sugarcane Research Center cenicaña

Relevance







4

Colombian Sugarcane Research Center cenicaña

Gene isolation and characterization

Approach 1

Isolation of **orthologous** genes associated with tolerance to drought and flooding

Selection of **specific** gene families

Approach 2

Gene expression studies to identify candidate genes

Selection of genes based on **differential expression** and **function**

Plant transformation and **molecular markers** development

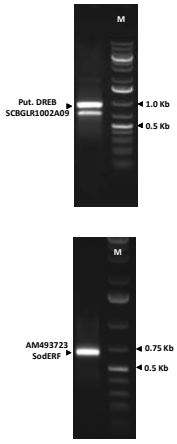
5

Colombian Sugarcane Research Center

cenicaña

Isolation of orthologs: DREB and ERF

Candidate target	Sugarcane orthologue ID	Isolated sequence (bp)	Full-length Yes	Full-length No	Most similar sequence in GenBank	% simlarity
Rice DREB2A	SCBGLR1002A09.g[1]	906	X		DRE binding factor 2 [*Oryza sativa Japonica* Group]	79
		915	X		Drought-responsive factor-like transcription factor DRFL1a [*Triticum aestivum*]	79
		710	X		Drought-responsive factor-like transcription factor DRFL1a [*Triticum aestivum*]	88
Members of the ERF3 superfamily	AM493723[1]	606	X		Ethylene Responsive transcription Factor [*Saccharum officinarum*]	97
		642	X		Ethylene Responsive transcription Factor [*Saccharum officinarum*]	80



6

Colombian Sugarcane Research Center

cenicaña

Isolation of orthologs: IPT

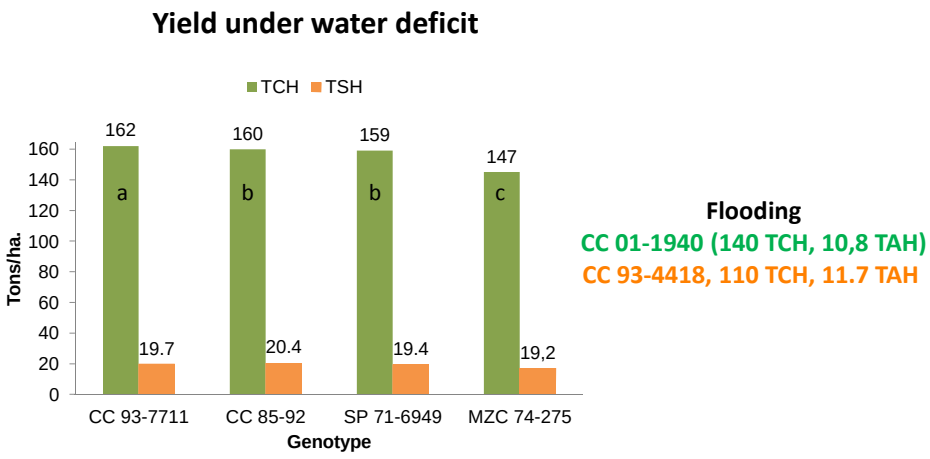
Candidate target	Sugarcane ortholog ID	Isolated sequence (bp)	Full-length		Most similar sequence in GenBank	% Similarity
			Yes	No		
Members of the IPT1-9 family	SCSGSB1005E06.g[1]	1030		3´RACE	Isopentenyl transferase [*Zea mays*]	76
	SCBGSB1029B02.g[1]	1392	X		Isopentenyl transferase IPT1 [*Zea mays*]	88
	SCSGFL4C06D01.g[1]	1001		3´RACE	*Zea mays* isopentenyl transferase IPT8 (IPT8) gene	80

7

Colombian Sugarcane Research Center

cenicaña

Genotypes of Interest



Flooding

CC 01-1940 (140 TCH, 10,8 TAH)

CC 93-4418, 110 TCH, 11.7 TAH

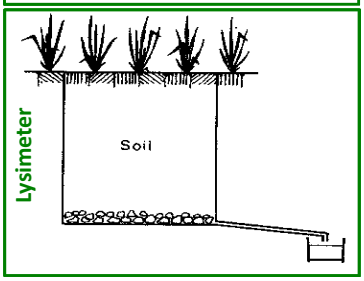
Colombian Sugarcane Research Center

cenicaña

Flooding experiment layout







9

Colombian Sugarcane Research Center

cenicaña

Flooding experiment layout

CC 01-1940 | CC 93-4418

Control

Stress



2 DAF (Mild) | 10 DAF (Severe) | 2 DAF (Mild) | 10 DAF (Severe)

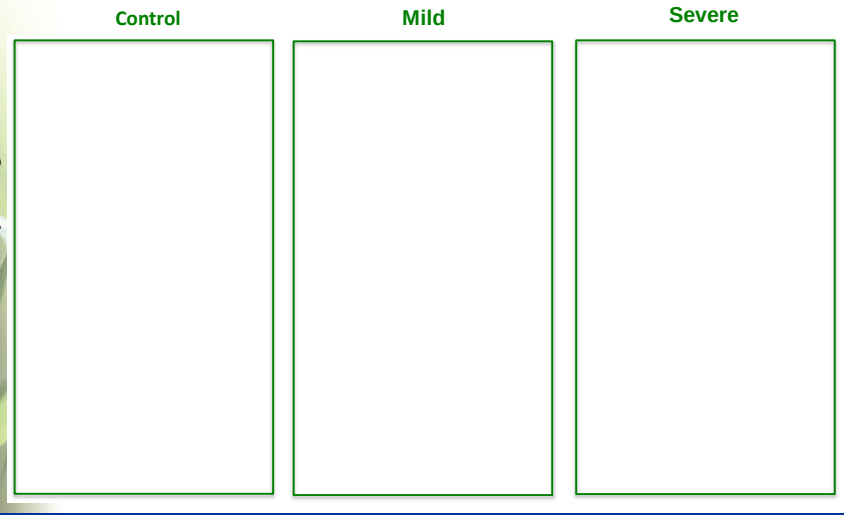
10

Signal transduction changes and metabolic adaptation occur within hours (Zou et al., 2010)

Colombian Sugarcane Research Center

cenicaña

Drought experiment layout



11

Colombian Sugarcane Research Center cenicaña

Sequencing summary

- ✓ **Roots and leaves** were harvested
- ✓ **4 treatments** (Medium Stress and Control) (Severe Stress and Control)
- ✓ **3 biological replicates** (2 RNASeq and 1 qPCR validation)

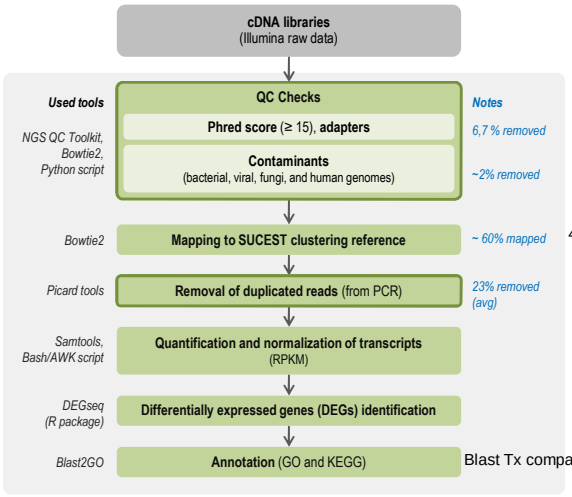
Experiment	Genotypes	# RNASeq libraires	# sequences (100 bp)	# sequences used for DEG
Water deficit	CC 93-7711, SP 71-6949, MZC 74-275	42	2.139.070.976	1.754.412.963
Flooding	CC 93-4418, CC 01-1940	30	1.117.434.112	936.171.634
Total			**4.373.939.200**	**3.256.505.808**

12

Colombian Sugarcane Research Center cenicaña

1

DGE Pipeline



40 % no in SUCEST

Blast Tx comparison

13

Colombian Sugarcane Research Center cenicaña

Water deficit

Leaves

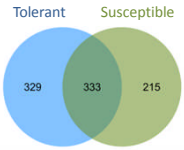


Root

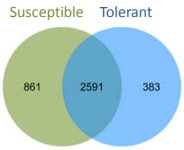


Flooding

Leaves



Root



6,411 **Total DEG** 4712

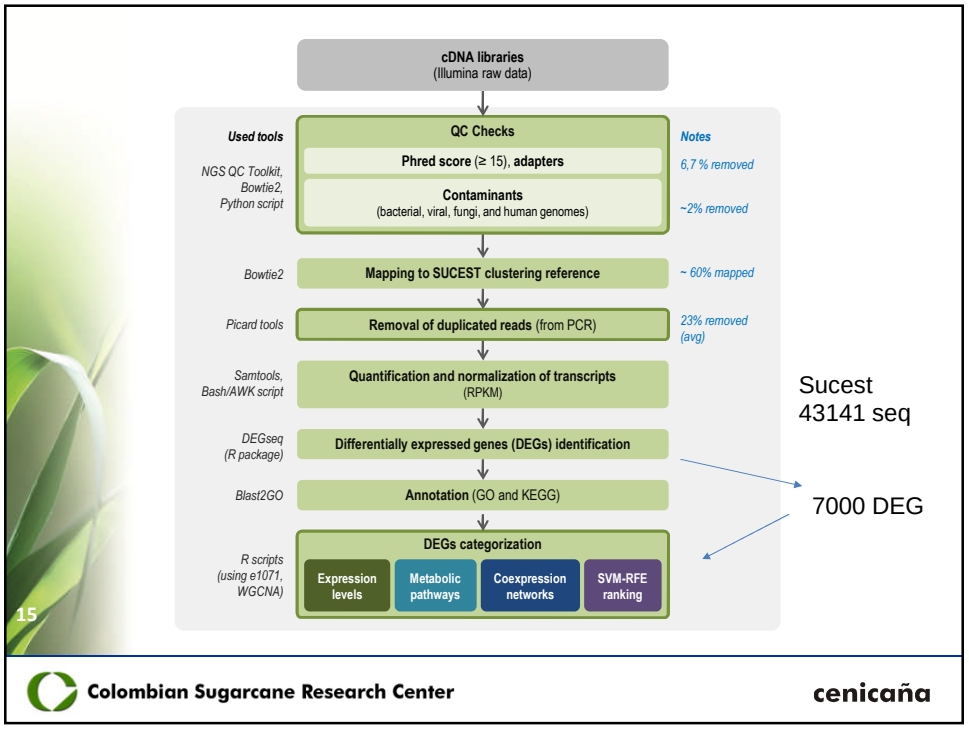
187 + 183 - (6) = **364** **Only Tolerants** 329 + 383 – (11) = **710**

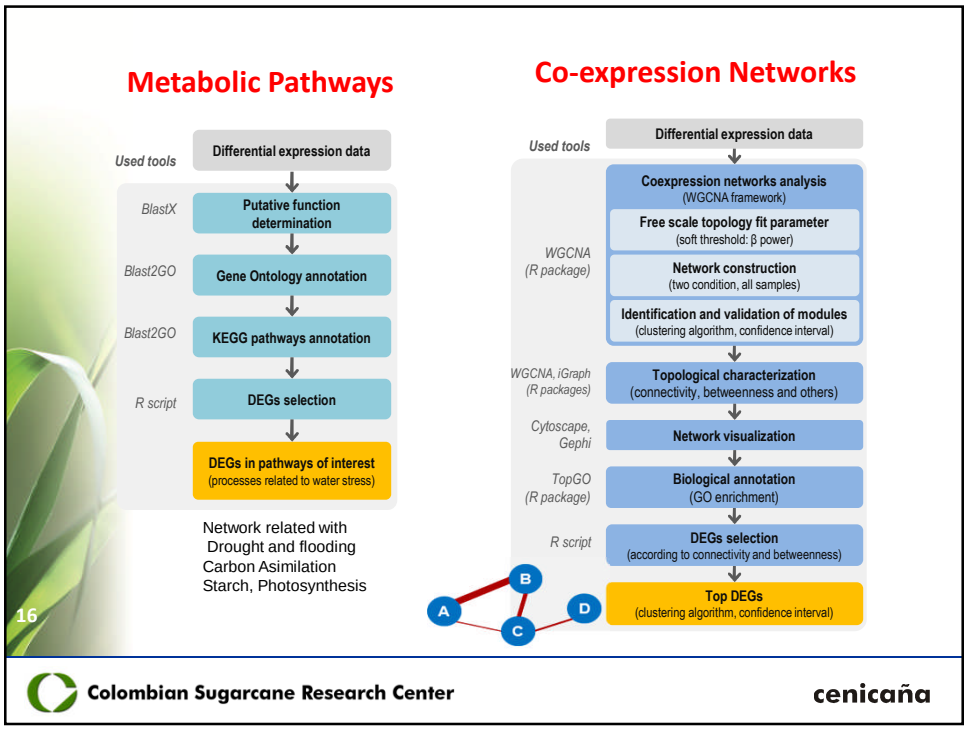
364 + 710 –(48) = **1017**

Colombian Sugarcane Research Center cenicaña

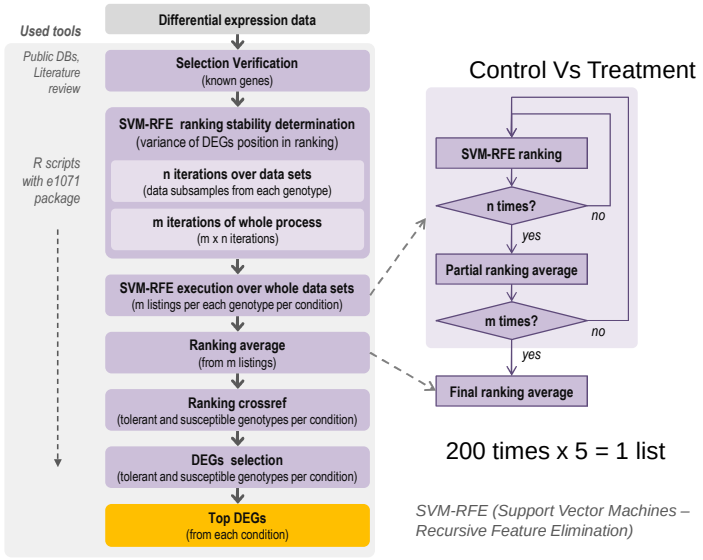
Slide 13

1 biotecnologia, 5/19/2015


cDNA libraries
(Illumina raw data)
Used tools
Notes
QC Checks
Phred score (≥ 15), adapters
6,7 % removed
NGS QC Toolkit, Bowtie2, Python script
Contaminants
(bacterial, viral, fungi, and human genomes)
~2% removed
Bowtie2
Mapping to SUCEST clustering reference
~ 60% mapped
Picard tools
Removal of duplicated reads (from PCR)
23% removed (avg)
Samtools, Bash/AWK script
Quantification and normalization of transcripts (RPKM)
DEGseq (R package)
Differentially expressed genes (DEGs) identification
Blast2GO
Annotation (GO and KEGG)
R scripts (using e1071, WGCNA)
DEGs categorization
Expression levels
Metabolic pathways
Coexpression networks
SVM-RFE ranking
Sucest
43141 seq
7000 DEG
15
Colombian Sugarcane Research Center
cenicaña


Metabolic Pathways
Used tools
Differential expression data
BlastX
Putative function determination
Blast2GO
Gene Ontology annotation
Blast2GO
KEGG pathways annotation
R script
DEGs selection
DEGs in pathways of interest
(processes related to water stress)
Network related with
Drought and flooding
Carbon Asimilation
Starch, Photosynthesis
Co-expression Networks
Used tools
Differential expression data
Coexpression networks analysis
(WGCNA framework)
Free scale topology fit parameter
(soft threshold: β power)
WGCNA
(R package)
Network construction
(two condition, all samples)
Identification and validation of modules
(clustering algorithm, confidence interval)
WGCNA, iGraph
(R packages)
Topological characterization
(connectivity, betweenness and others)
Cytoscape, Gephi
Network visualization
TopGO
(R package)
Biological annotation
(GO enrichment)
R script
DEGs selection
(according to connectivity and betweenness)
Top DEGs
(clustering algorithm, confidence interval)
A
B
C
D
16
Colombian Sugarcane Research Center
cenicaña

SVM-RFE Ranking



SVM-RFE (Support Vector Machines – Recursive Feature Elimination)

17

Colombian Sugarcane Research Center cenicaña

Selection criteria

1 . Expression levels | 2. KEGG pathways | 3. **Co-expression networks** | 4. **Ranking** based on ***SVM-RFE***

- Most overexpressed gene in the study
- Most overexpressed gene in Tol. Gen.
- Highest expression difference (Tolerant-Susceptible)
- Most overexpressed gene in leaves
- Most overexpressed gene in leaves of Tol. Gen.
- Highest expression difference (Tolerant-Susceptible) in leaves
- Most overexpressed gene in roots
- Most overexpressed gen in roots of Tol. Gen.
- Highest expression difference (Tolerant-Susceptible) in roots
- DEG in KEGG pathways associated with water deficit
- Highest betweness/module
- Highest connectivity/module
- Highest betweenesss + connectivity
- Best ranked gen in Tol. Gen.
- Highest ranking difference (Tolerant–Susceptible)
- Highest ranked gen in Tol. Gen. Expressed only in leaves.
- Highest ranking difference (Tolerant–Susceptible) expressed only in leaves.
- Highest ranked gen in Tol. Gen. Expressed only in roots.
- Highest ranking difference (Tolerant–Susceptible) expressed only in roots.

18

Colombian Sugarcane Research Center cenicaña

10 Genes have been initially selected for function testing

Genetic transformation

***A. thaliana* and Rice** → **Sugarcane**

✓ #1 y #2 in evaluation

19

Colombian Sugarcane Research Center

cenicaña

Thank you!

John J. Riascos, Ph.D.
jjriascos@cenicana.org

Jershon López Ph.D.
Jlopez@cenicana.org

cenicaña
Colombian Sugarcane Research Center