Comparative transcription profiles of *Solanum* wild species under drought conditions: Preliminary results

M.F. Lucca
Instituto de Biotecnología
CICVyA - INTA
Buenos Aires, Argentina
e-mail: flucca@balcarce.inta.gov.ar

E. Hopp
Instituto de Biotecnología
CICVyA - INTA
Buenos Aires, Argentina
e-mail: ehopp@cnia.inta.gov.ar

R. Romero-Zaliz
DECSAI - UGR
Granada, España
e-mail: rocio@decsai.ugr.es

***Abstract*—Potato (*Solanum tuberosum*) is the third most important crop and the most relevance dicot in human feeding. Water stress is the greatest limitation of crop production. In response to drought, the plant triggers strategies that involve morphological, physiological and biochemical changes, polygenic traits controlled by additive genes and often synergistic. We focused on the analysis of transcriptional profiling on genotypes of *Solanum* wild species: *S. venturii* and *S. cardiophyllum* that present contrasting behavior under water stress using cDNA microarray hybridizations. *Solanum* wild species are a rich source of novel genes for introduction into cultivated species to enhance drought tolerance. Bioinformatic analysis of expression data of *Solanum* wild species showed significant differences in transcriptional profiles. Differentially expressed genes vary not only in number but also in their temporal expression profiles. These differences between wild potato species suggest new ways to exploit the genetic variability of wild germplasm for crop breeding.**

Microarrays; clustering; bioinformatics; germoplasm; potato

I. BACKGROUND

The potato is a key member of the family Solanaceae (Juss.) and it is also the most important dicot in the human feeding, constituting an ideal food in developing countries and especially in areas with altitudes above 2000 meters, in which there are limited number crops and where the man and animals require a higher caloric intake.

The genus *Solanum* is constituted by numerous species, differing about 5500 varieties [1]. The habitat of these species is highly variable growing at different altitudes from sea level to over 4000 meters and do so on an extensive range of temperature, photoperiod and water supplies. Also, the wild genotypes adapted to different ecological conditions ranging from highland, tropics to desert regions [2].

The potato breeders are focused on the identification of new genes of agronomic importance traits from the worldwide potato germplasm banks. These studies aim to incorporate new traits to select cultivars to be used in the production and creating new varieties through a combination of traditional breeding techniques and biotechnological methods.

The genetic base of cultivated potatoes out of the Andes is very narrow, making it difficult to develop new higher yielding varieties, or more resistance to major biotic and abiotic stresses in this crop. The growth and development of plants depends on genetic factors and environmental [3]. We use the concept of "stress" when in excess or deficiency of one o more of these factors; the development and plant growth is affected. Abiotic stresses are the main cause of yield loss worldwide, reporting losses of over 50% in major crops [4][5]. The adaptation of plants to stress involves the activation of various physiological, metabolic and developmental responses and they are polygenic traits controlled by additive genes and often synergistic [6].

The mechanisms in which genes act in response to water deficit stress can involve different pathways inside the cell; those undergo changes in their gene expression. These changes may be directly regulated by water stress conditions, secondary stresses and/or injury responses. Moreover, different subsets of mRNA transcripts are expressed at different times during the water stress period. In spite of the advances in molecular biology, changes in gene expression are induced by complex transduction signals that are still not been clearly established [7].

In order to analyze gene expression we use potato cDNA microarrays. This technology allows the simultaneous measurement of expression levels of thousands of genes represented on the array [8]. Unlike traditional molecular biology, the successful use of microarrays requires the application of statistics and informatics to design the arrays and experiments, and also to analyze and manage the data. To navigate this large amount of information we need bioinformatics tools. Bioinformatics is a rapidly developing field whose multi-discipline nature serves to enhance the knowledge of biological processes, as it involves the use of different techniques including applied mathematics, informatics, statistics, computer science, artificial intelligence, chemistry, and biochemistry to solve biological problems usually on the molecular level.

A bioinformatic well-known methodology called clustering, often used for gene expression data analysis, allows the partition of a pool of genes contained on an array into subsets, or clusters, so that the data in each subset participates in some biological process of interest, like water stress.

In this work, we apply clustering over a set of gene expression data together with statistical analysis of microarrays to uncover novel knowledge on different *Solanum* wild species. This information can be used to maintain the organism metabolism and limit the harm under prolonged water stress conditions, since wild genotypes of

Solanum are a rich source of novel genes for introduction into cultivated species to enhance stress tolerance.

II. Materials and Methods

A. Drought treatment

To induce water stress, five-week-old *Solanum cardiophyllum* (CPH18325) and *S. venturii* (VNT8239) plants were randomly sorted in two groups (20-25 plants per group). One group of plants was submitted to non-irrigated conditions (withholding water) and the second group was normally watered. All the plants were irrigated 72 hs before starting the treatment. Plant water status was monitored by a Scholander pressure chamber and with the determination of the relative water content (RWC) [9]. Leaflets tissue from three-four plants per group was collected at four time-points: 0, 5, 7, and 10 days after initiation drought treatment. The leaflets at each time-point were pooled, frozen in liquid nitrogen and stored at –80°C.

B. RNA extraction

RNA was extracted from leaf tissue following TRIZOL method (Invitrogen). RNA integrity was verified on agarose gel and the concentration was adjusted to 1µg/ml by ethanol precipitation and re-suspension.

C. Microarray

We designed microarrays assays using 10K potato cDNA array, version 4 provided by The Institute for Genomic Research (TIGR). A total of 15,264 cDNAs were spotted in duplicate on the slides (22,824 useable elements –spots- from 32,448 total spots).

We followed MIAME (Minimum Information About a Microarray Experiment) standard to design microarray experiments. With n species and t time-points and r repeats, we used 32 slides (2 species; 4 time-points; 4 repeats (dye-swap)).

D. Preparation of labelled cDNA probes

Synthesis of cDNA was carried out by a direct labeling reaction of reverse transcription following TIGR protocols [10].

E. Microarrays Hybridization

Hybridization of glass potato microarrays was performed according to TIGR protocols [11].

F. Data analysis

Scanning of the microarrays was performed in a GenPix scanner (Axon Instruments, USA). Limma bioconductor package was used to read and analyze all microarrays. Minimum method for background correction was used along with default normalization within arrays (printtiploess) and aquantile normalization between arrays. Duplicate spots were averaged. A linear model was calculated and all p-values were adjusted using an empirical Bayes method to minimize the false discovery rate. Results with an adjusted p-value lower than 0.003 were kept.

G. Fuzzy c-means clustering

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunn in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition. It is based on minimization of the following objective function, with respect to U, a fuzzy c-partition of the data set, and to V, a set of K prototypes:

$$J_m(U,V) = \sum_{j=1}^{n}\sum_{i=1}^{c} u_{ij}^{m} \left\| X_j - V_i \right\|^2, \quad 1 \le m \le \infty$$

where m is any real number greater than 1, u_{ij} is the degree of membership of X_j in the cluster i, X_j is the jth of d-dimensional measured data, V_i is the d-dimension center of the cluster and ||*|| is any norm expressed the similarity between any measured data and the center [12].

H. Clustering of expression data

Differentially expressed genes were clustered using fuzzy c-means methods using Euclidean distance measure [12][13]. Instead of clustering the expression values by themselves we calculated the difference between contiguous time points and use these values to cluster. Using these new values changes in expression with the same sign and relatively similar strength are clustered together. With this technique we simulate a simple correlation study. Fuzzy clustering was used to provide the possibility of a gene belonging to more than one cluster, thus allowing the expert to make the final decision of which cluster each gene should be assigned.

III. Analysis of Results

We analyzed all microarrays from both species and extracted the list of differential expressed genes. The ten first differentially expressed genes of each species are shown in Tables I and II.

Fuzzy c-means clustering was applied to the list of differential expressed genes to obtain 10 clusters for each species, where all genes were mapped to a single cluster by selecting the cluster with highest fuzzy membership (Figures 1 and 2) (we also tried with different number of clusters without any important improvement).

TABLE I. TEN BEST DIFFERENTIAL EXPRESSED GENES FOR SPECIES I (CPH18325)

Name	Clone name	Time 5	Time 7	Time 10	adj.P.Val
peptidase M20/M25/M40 family protein similar to acetylornithine deacetylase	STMHR62	0.20	0.16	-0.96	1.90e-06
Metallothionein-like protein type 2 B	STMGS26	0.32	0.55	0.401	3.30e-06
Avr9/Cf-9 rapidly elicited protein 284	STMGP26	-1,31	-0.43	-0.21	1.41e-05
GDSL-motif lipase/hydrolase family protein similar to family II lipase EXL3	STMCV95	-0.09	-0.55	-1,78	2.62e-05
1-deoxyxylulose 5-phosphate synthase	STMCY65	-0.07	0.11	-0.86	2.62e-05
GDSL-motif lipase/hydrolase family protein similar to family II lipase EXL3	STMDB20	-0.14	-0.54	-1,70	2.62e-05
WIZZ	STMEP07	-1,04	-0.32	-0.74	2.62e-05
expressed protein	STMGJ17	-1,16	-0.10	-0.15	2.62e-05
null	STMEP27	0.45	-0.03	-0.60	2.62e-05
Ferritin 1 (Fragment)	STMGY57	1.34	0.06	-0.30	2.71e-05

TABLE II. TEN BEST DIFFERENTIAL EXPRESSED GENES FOR SPECIES II (VNT8239)

Name	Clone name	Time 5	Time 7	Time 10	adj.P.Val
Jasmonic acid 2	STMIY82	0.17	1.57	0.56	1,47E-02
dehydrodolichyl diphosphate synthase putative / DEDOL-PP synthase putative similar	STMIR06	0.11	0.41	1.30	3,90E-03
expressed protein	STMDO28	0.19	0.88	0.52	7,70E-03
2-oxoglutarate-dependent dioxygenase	STMGO18	-0.47	0.73	-0.88	9,09E-02
hypothetical protein	STMCZ51	0.08	0.73	-0.62	1,39E-01
Abscisic acid and environmental stress inducible protein TAS14 (Dehydrin TAS14)	STMHQ27	0.30	2.10	1.47	1,63E-01
24K germin like protein precursor	STMCX35	-0.03	-1.46	0.37	1,71E-01
17.6 kD class I small heat shock protein (Type I small heat shock protein 17.6 kDa isoform)	STMGB34	0.30	0.96	0.72	2,25E-01
TATA-binding protein-associated factor TAFII55 family protein contains Pfam profile	STMHN79	0.29	0.92	0.64	3,08E-01
Putative non-specific lipid transfer protein StnsLTP	STMGQ20	0.19	1.87	0.57	3,39E-01

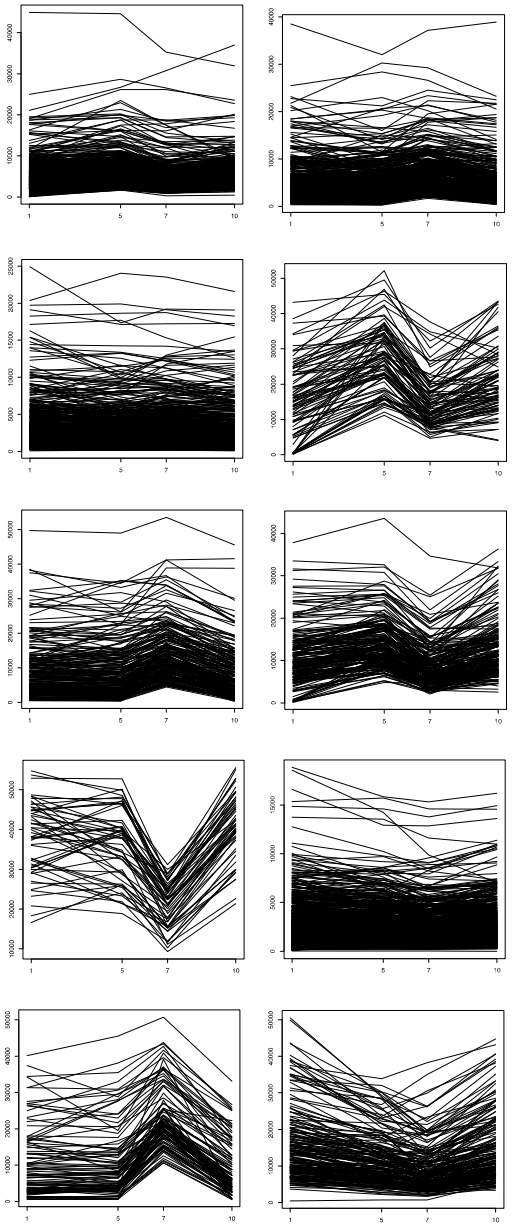

Figure 1. Species I (CPH18325) fuzzy c-means clustering of differential expressed genes. The x-axis correspond to time points while the y-axis correspond to expression level.

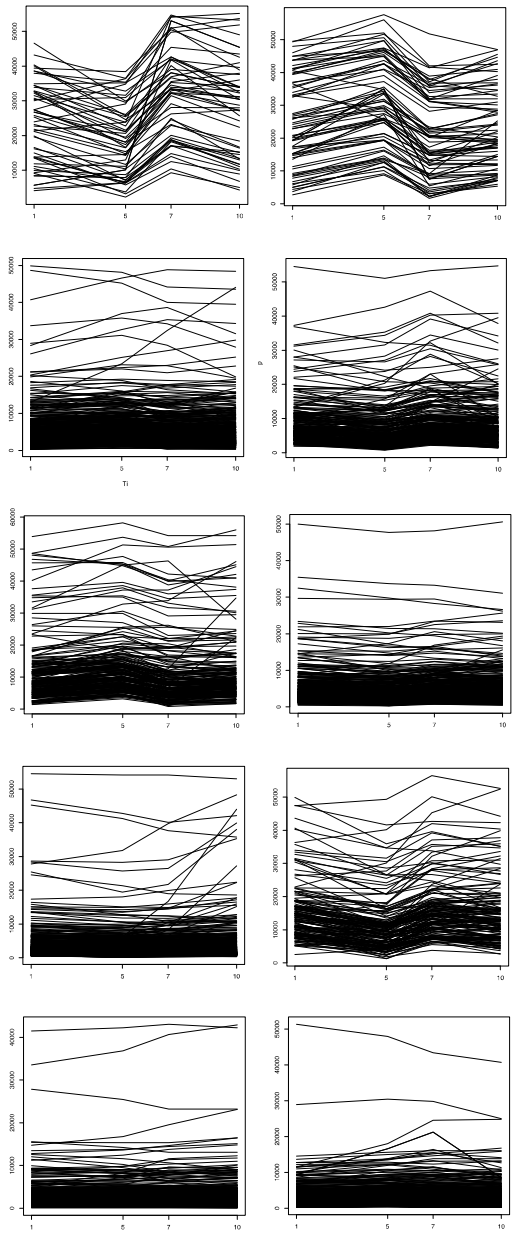

Figure 2. Species II (VNT8239) fuzzy c-means clustering of differential expressed genes. The x-axis correspond to time points while the y-axis correspond to expression level.

From these clusters we can see that there are several genes with a very similar expression in both species. Then, we study which genes behave similarly in both species and which have substantial differences. In the latter group we found the most interesting results.

In the next figures we show some genes with different behavior between species I and II.

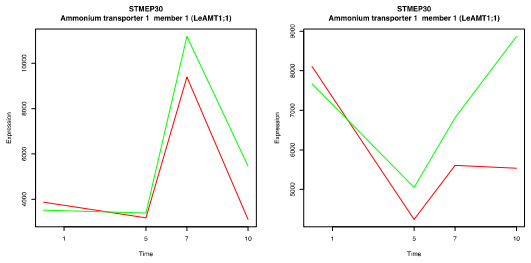

Figure 3. Left side: STMEP30 for species I (CPH18325). Right side: STMEP30 for species II (VNT8239). Green lines represent control and red ones treatment (drought).

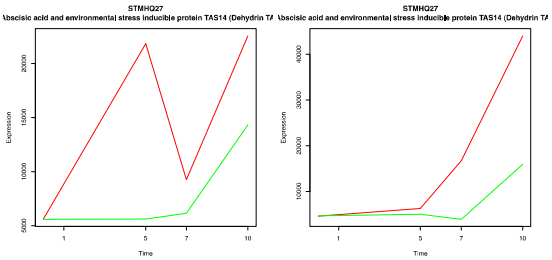

Figure 4. Left side: STMHQ27 for species I (CPH). Right side: STMHQ27 for species II (VNT8239). Green lines represent control and red ones treatment (drought).

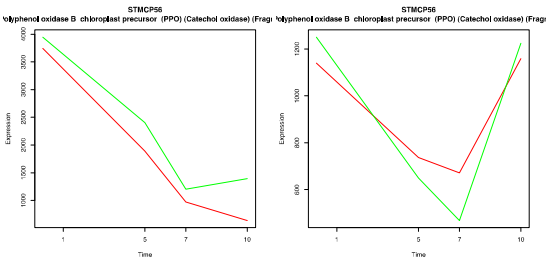

Figure 5. Left side: STMCP56 for species I (CPH18325). Right side: STMCP56 for species II (VNT8239). Green lines represent control and red ones treatment (drought).

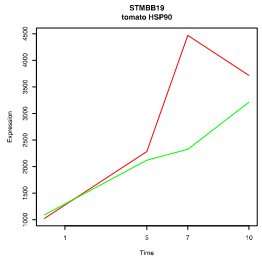

Figure 6. STMBB19 for species I (CPH18325). Green lines represent control and red ones treatment (drought).

The comparative transcriptome analysis of CPH18325 and VNT8239 species showed that they implemented different strategies to cope the stress, with differences in the number of genes as in the kinetics of them during the drought treatment.

In both species there is differential expression of stress response genes that validate the trial (Tables I and II) such Metallothioneins 1, Ferritin and Dehydrins that act to keep cellular homeostasis. Genes behaved in groups in both species in response to stress, but with differential temporal profiles of expression (Figures 1 and 2) between species. In VNT8239, analyzing the differential expression of some genes, we observe contrasting expressions patterns, suggesting a common synchronous regulation, as exemplified by the gene non-specific putative lipid transfer protein StnsLTP that contrast to Photosystem I reaction center subunit X psak and also Ethylene-responsive late embryogenesis-like protein in contrast to methyl esterase Pectin, confirming the genetic diversity of responses in *Solanum* wild species germplasm to drought. While CPH18325 show little differential gene expression (235 genes) associated with cell wall metabolism, transcription, protein metabolism, signaling and response genes to biotic and abiotic, VNT8239 displayed at all times a massive differential expression comprising 4133 genes. Of the total differentially expressed genes, species share the expression of 151 genes.

Some of these genes shown to be differentially down regulated by stress in both species, as Ammonium transporter 1 (STMEP30), even though with different dynamics in each species (Figure 3). Others genes showed to be induced in both species as Dehydrin Tas14, with a strong and early induction in CPH18325 (Figure 4). The Polyphenol oxidase B precursor (Figure 5) is an example of contrasting expression between genes in common genes group.

Although analyzing the common genes and their dynamics helps us to understand the differential response to drought, the most interesting are the unique genes associated with most resistant species, *S. cardiophyllum*, such as Heat Shock Protein 90, which appears as a candidate gene associated with water stress in potato (Figure 6).

These differences in expression between wild potato species suggest new ways to exploit the genetic variability of American germplasm for potato breeding.

DISCUSSION

S. cardiophyllum 18325 shown to be more resistant to drought than *S. venturii* 8239, with RWC of 87 and 67% respectively. Native environments of CPH18325 and VNT8239 species are very different. *S. cardiophyllum* was reported to be from Mexico, developed with dry shrubs vegetation, on the edge of old fields and crops [14] and *S. venturii* was described by Hawkes and Hjerting [15] and Spooner and Clausen [16] in NW of Argentina, particularly in wet slopes, including *Alnus* and *Polylepis* forests, or along cultivated fields, roads and river terraces at an altitude between 1900-3000 meters.

Watkinson et al [17] suggest that adaptation and acclimation to drought involves the coordinated participation of regulatory networks and, in particular, genes associated with signaling, gene transcription and metabolic pathways involved in the distribution of resources, stress defense and protection.

Wild *Solanum* species CPH18325 and VNT8239 showed different responses to drought. These species come from very different geographical areas, which suggest that differences in stress responses reflect adaptation to native environment through specific differential gene expression strategies. The information obtained on these adaptive strategies could be used in potato breeding programs to take advantage of genetic diversity and develop varieties resistant to drought. Since *S. cardiophyllum* 18325 varied in expression of only 235 genes and 57 genes were unique, they are interesting candidate genes to be evaluated as new genes for drought resistance.

From the computational point of view, the clustering of genes with similar behavior provided an interesting insight and helped in the study of wild *Solanum* species. Alternative clustering methods could also have been used to group genes with similar behavior and we intend to test them in future work to improve our preliminary results.

ACKNOWLEDGMENT

We thank Dr. Enrique Ritter and Neiker-Tecnalia Institute for wild *Solanum* germplasm used in this study and the facilities where they conducted drought treatments.

R. Romero-Zaliz is supported by The Spanish Ministerio de Educación y Ciencia (TIN2009-13950), Junta de Andalucía (TIC-02788) and GENIL program (PYR-2010-28).

REFERENCES

[1] T. F. and A. O. of the U. Nations, "FAO Statistics 2008." [Online]. Available: http://www.potato2008.org/es/index.html.

[2] P. Rousselle, Y. Robert, and J. C. Crosnier, *La Patata*. Mundi-Prensa Libros, 1999, p. 601.

[3] J. Levitt, *Responses of Plants to Environmental Stresses (Physiological Ecology): Chilling, freezing, and high temperature stresses*. Academic Pr, 1980, p. 497.

[4] E. Bray, E. A., Bailey-Serres, J., and Weretilnyk, "Responses to abiotic stresses," in *Biochemistry and Molecular Biology of Plants*, B. B. Buchanan, W. Gruissem, and R. L. Jones, Eds. American Society of Plant Physiologists, 2000, pp. 1158-1203.

[5] J. S. Boyer, "Plant Productivity and Environment.," *Science (New York, N.Y.)*, vol. 218, no. 4571, pp. 443-448, Oct. 1982.

[6] H. J. Bohnert, D. E. Nelson, and R. G. Jensen, "Adaptations to Environmental Stresses.," *The Plant cell*, vol. 7, no. 7, pp. 1099-1111, Jul. 1995.

[7] E. A. Bray, "Classification of Genes Differentially Expressed during Water-deficit Stress in Arabidopsis thaliana: an Analysis using Microarray and Differential Expression Data," *Annals of Botany*, vol. 89, no. 7, pp. 803-811, Jun. 2002.

[8] G. J. Bryan and I. Hein, "Genomic resources and tools for gene function analysis in potato.," *International journal of plant genomics*, vol. 2008, p. 216513, Jan. 2008.

[9] M. J. Reigosa Roger, L. González, and M. González-Vilar, *Handbook of Plant Ecophysiology Techniques*. Dordrecht: Kluwer Academic Publishers, 2003, pp. 207-212-212.

[10] TIGR Protocol, "http://jcvi.org/potato/images/SGED_SOP_5.1.1.pdf." [Online]. Available: http://jcvi.org/potato/images/SGED_SOP_5.1.1.pdf.

[11] TIGR Protocol, "http://jcvi.org/potato/images/SGED_SOP_6.1.1.pdf."

[12] J. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press, 1981.

[13] R. Nock and F. Nielsen, "On weighting clustering.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 8, pp. 1223-35, Aug. 2006.

[14] J. Hawkes, *The potato: evolution, biodiversity and genetic resources*. London: Belhaven, 1990.

[15] J. G. Hawkes and J. P. Hjerting, "Some wild potato species from Argentina.," *Phyton, Vicente Lopez*, vol. 9, pp. 140-46, 1960.

[16] D. M. Spooner and A. M. Clausen, "Wild potato (Solanum sect. Petota) germplasm collecting expedition to Argentina in 1990, and status of Argentinian potato germplasm resources."

[17] J. Watkinson et al., "Accessions of Solanum tuberosum ssp. andigena Show Differences in Photosynthetic Recovery After Drought Stress as Reflected in Gene Expression Profiles," *Plant Science*, vol. 171, 2006.