A catalogue of 5700 expressed genes in cassava: identification of genes implicated in cassava bacterial blight resistance and starch biosynthesis

Lopez, C, Jorge, V, Piégu, B, Mba, C, Cortes, D, Restrepo, S, Soto, M, Cooke, R, Delseny, M, Tohme, J and Verdier V.

Biotechnology Research Unit, Centro Internacional de Agricultura Tropical (CIAT) AA 6713, Cali, Colombia

UMR5096, IRD-CNRS-Université de Perpignan,52 Paul Alduy, 66860 Perpignan Cedex, France.

INTRODUCTION

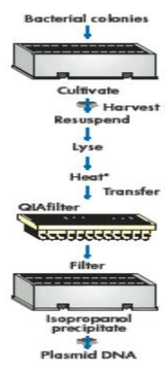
Cassava (*Manihot esculenta* subsp. *esculenta* Crantz), is the fourth more important basic product after rice, wheat and maize. In addition to being a basic food in the human diet, cassava constitutes one of the more common raw materials for the starch production. The yield, the dry matter content (indirectly starch content) and planting materials are affected by Cassava Bacterial Blight (CBB). One tool which holds a lot of promise in unraveling the complexities in gene expression is Expressed Sequence Tags (ESTs). In this study, we have targeted 2 economically important characters, starch content and CBB resistance to generate several types of libraries including subtracted. A high collection of EST was generated. We report here the single pass sequencing of 18202 cDNA clones from the 5'end

MATERIALS AND METHODS

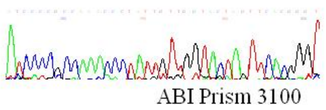
cDNA libraries construction

Characters	Variety	Characteristics	N° of librairies	Tissue	Type of library	DAI	EST sequenced	EST analyzed
CBB	MBRA685	resistant	6	stem	sustractive	pool	768	438
					non sustractive	pool	384	258
					Stratagene	/	2304	1560
					cDNA AFLP	pool	250	241
	MCOL1522	susceptible	2	stem	sustractive	pool	384	258
					Stratagene	48h	2304	1721
	SG107-35	highly resistant	4	leaf/st	sustractive	pool	1152	592
					non sustractive	pool	288	128
Starch	MPER183	low dry	1	root	Sratagene	6mon	4992	3391
	CM523-7	high dry	1	root	Stratagene	6mon	5376	3608
							18202	**12195**
					Genbank			848
								13043

Minipreps, sequence and sequence analyses



REAL96-QIAGEN



18202 sequences generated
848 cassava ESTs from Genbank

Development of a pipeline

Chromatogram → Trim end sequences and vector → Assemblage contigs and singletons

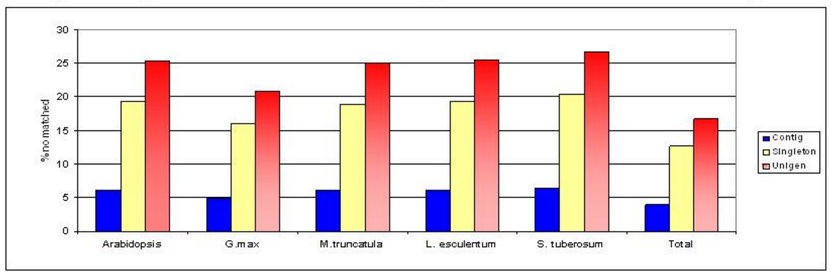
PHRED VECSCREEN StackPACK

RESULTS

Characterization of a unigene set

After trimming a total of 13043 sequences was obtained. These were assembled in a unigene set of 5700 sequences, including 1875 Tentative Contigs (TC) and 3825 singletons

Comparing the cassava unigene set with other plant species



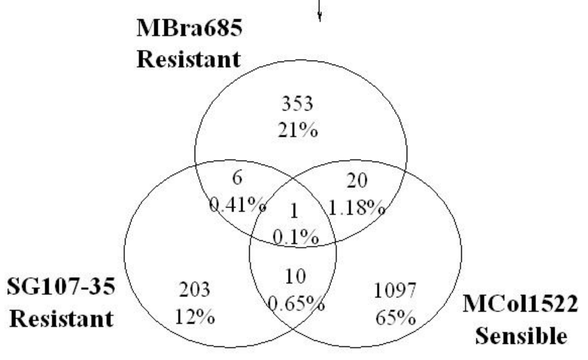
Conceptual translation of 16% of the sequences did not show any similarity of available EST of other plant species. They might represent new cassava-specific genes, although some of them might be 5' or 3' untranslated sequences.

Genes involved in defense

Some ESTs showed similarity to known genes related to defense.

- 39 Protein kinases
  - 5 Xa21-like protein
- 13 similarity to R-gene
- 10 Cytochrome P450
- 3 peroxidases
- Transcription factor WRKY
- EDS1
- Chitinase
- 56% Non similarity

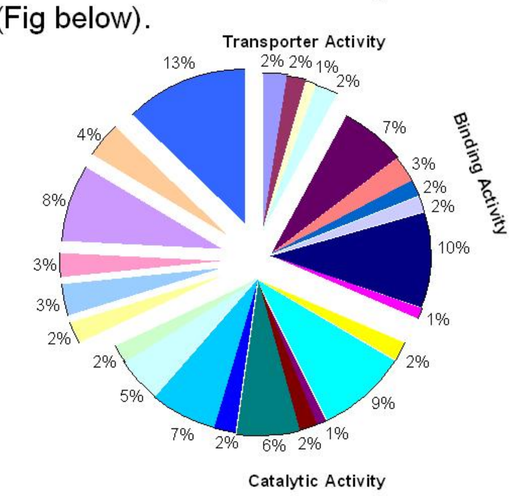
We identified sequences that were only present in the libraries corresponding to inoculated plants .



Comparing the specific or common sequences among the different cassava cultivars we observed that most of them were specific to each cultivar-*Xam* interaction (either susceptible or resistant reactions)

Functional categories

Functional categories were defined using the Gene Ontology (GO) classification scheme (http://www.geneontology.org). A total of 63% of unigenes did not have a significant match. Thus, 37% of the unigene set was assigned a putative function using this method (Fig below).



Biological process

Transporter Activity (7.84%)
- unknown transporter
- carrier transporter
- ion transporter
- others transporter

Binding Activity (23.70%)
- unknown binding
- nucleotide binding
- protein binding
- metal ion binding
- nucleic acid binding
- others binding

Catalytic Activity (36.30%)
- unknown catalytic
- hydrolase activity
- isomerase activity
- lyase activity
- transferase activity
- ligase activity
- oxidoreductase activity
- kinase activity
- others catalytic

- signal transducer activity
- transcription regulator activity
- translation regulator activity
- structural molecule
- others molecular activity
- unknown molecular function

Characterization of EST associated to starch biosynthesis

Putative genes involved in the starch biosynthesis were identified in the cassava EST collection

- 3 Isoformes alpha-1,4 glucan phosphorylase
- 2 Tuber-specific and sucrose-responsive element binding factor
- 2 Isoformes Starch phosphorylase
- Glycogen [starch] synthase, chloroplast precursor (GBSSII)
- Granule bound starch synthase II precursor
- Granule-bound glycogen [starch] synthase
- Soluble glycogen [starch] synthase, chloroplast precursor (SS III)
- Protein induced upon tuberization

Comparing in silico gene expression

We performed a differential analysis using the method reported by Stekel et al. (2000) to study differences in gene expression between cultivars CM523-7 and MPer183. Results are presented below.

Similarity	No of sequences for CM523-7	No of sequences for Mper183
Putative xyloglucan endotransglucanase	21	59
WRKY family transcription factor	14	1
Putative ADP-ribosylation factor	16	1
Shaggy-related protein kinase	12	1
Fumarase -related	12	1

CONCLUSIONS AND PERSPECTIVES

- The cassava EST data presented here is the first effort in the large scale sequencing of the cassava expressed genome and also in cataloguing cassava genes.
- A unigene set of 5700 sequences was identified and a putative function assigned to 37% of unigenes.
- A number of ESTs were found to be present only in the *Xam* challenged libraries.
- The EST resource will increase the density of gene markers on the cassava genetic map.
- Information obtained here will be used to develop microarray technology for further cassava gene expression studies.
- Based on this, developing new cassava varieties having high dry matter content and durable resistance to CBB will be the next challenge.

This research was supported by grants from Agropolis and Colciencias, Colombia, (contract No 344-98, code 2236-12-051-98)