

Rapid Transcriptome Sequencing of sweetpotato

Sweetpotato (*Ipomoea batatas* L. (Lam)) is an important staple crop in developing countries. It is a rustic crop with high tolerance to abiotic stresses such as drought, but is susceptible to a range of virus diseases and insect pests. Its hexaploidy and its outcrossing behaviour cause high genetic variability that can be fixed by vegetative propagation.

The availability of genomic resources for sweetpotato improvement is in strong contrast with the importance of this crop. Until recently available sweetpotato sequence data were restricted to around 22 000 ESTs and to ~1500 gene sequences. In the scope of a *Generation Challenge Program* – funded Project we had the opportunity to apply Next Generation Sequencing tools to characterize the Sweetpotato transcriptome under drought stress with the aim to produce functional genomics resources such as gene sequences and markers.

Sequencing runs used two normalized cDNA collections from stems and leaves from drought stressed plants and yielded 523.914 expressed sequence tags, which were assembled together with 22.094 publicly available ESTs (GeneBank) into 31.165 contigs (sets of overlapping DNA segments) and 29.080 singletons. This sequence information was used to design 247 new SSR primer pairs and 200 SNP markers for stress genes. BLASTx of the assembly with the manually annotated and curated *A. thaliana* UniProtKB database yielded 35.251 hits and indicated the presence of 14.145 unique genes, along with 24.994 sequences with no match to protein sequences.

The sweetpotato gene index is available at http://gpcpr.grinfor.net/index.php?app=databases&inc=dataset_details&dataset_id=712.

Methods

Pot-grown plants of the sweetpotato variety "Tanzania" (CIP accession number 440160) were exposed to water stress in the greenhouse for eight weeks. Then leaf and stem tissue was sampled separately for the production of two normalized cDNA libraries (www.evrogen.com). 7 µg cDNA of each library was submitted separately to a quarter 454 sequencing run at the School of Biological Sciences at the University of Liverpool.

Sequence cleaning and assembly was performed on the CIP High Performance Computer (<http://hpc.cip.cgiar.org/>). Adaptor primer and SMART oligonucleotide sequences as well as low complexity regions present in the 454 raw reads were masked using open source software (www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html, www.repeatmasker.org/RMDownload.html). De novo hybrid assembly of the 454 reads together with 20 094 publicly available sweetpotato EST sequences (<http://www.ncbi.nlm.nih.gov/sites/entrez>) was performed with the NGen software (Lasergene) using the following parameters: Match Size: 25, Gap Penalty: 7, Mismatch Penalty: 12, Match Score: 10, Minimal Match Percentage: 75, Match Spacing: 40. Assembly quality was assessed by self-blast and blastn analysis with mRNA sequences of *Solanales*, *A. thaliana* as well as with *I. batatas* and *I. nil* EST assemblies (www.plantgdb.org/download/download.php?dir=/Sequence/ESTcontig/ipomoea_batatas/current_version, http://plantta.jcvi.org/cgi-bin/plantta_release.pl, http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=morning_glory). 400 representative contigs were checked manually for the correct alignment of the fragments.

Contigs and singletons were annotated through comparison with protein sequences deposited at UNIPROT, TAIR and NCBI (<http://www.uniprot.org/uniprot>, <http://www.arabidopsis.org>, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein>) using BLAST 2.2.20 on the CIP High Performance Computer (<http://hpc.cip.cgiar.org/>). GO-annotation was performed using Blast2Go (Conesa A et al., 2005).

Acknowledgements

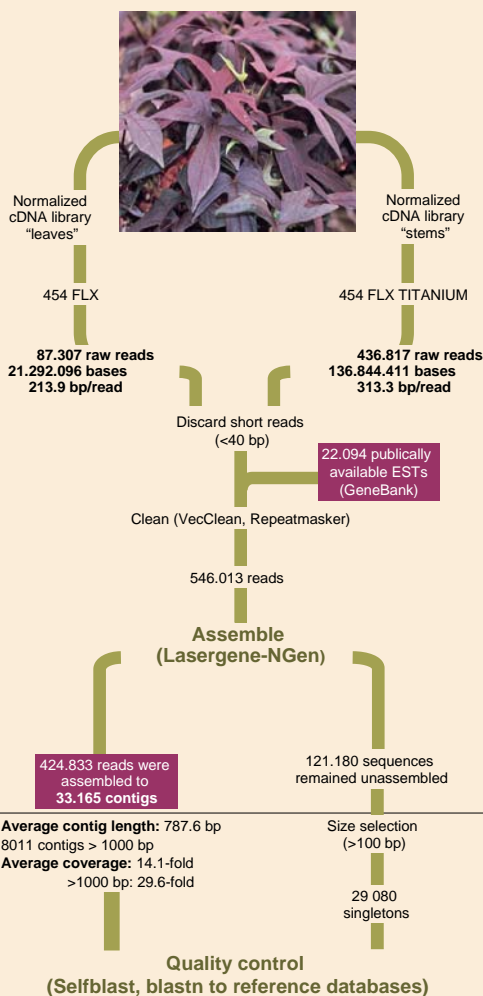
The works were financed by the Generation Challenge Program. We thank Raymundo Gutierrez for growing the plant material.

References

Conesa A et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18): 3674-3676

Roland Schafleitner
Luz Rosalina Tincopa,
Omar Palomino,
Anthony Collins and
Reinhard Simon

International Potato Center,
Av La Molina 1895,
Apartado 1558, Lima 12, Peru
r.schafleitner@cgiar.org



Quality control
(Selfblast, blastn to reference databases)

**Sweetpotato
Gene Index**
60.245 sequences

247 new
SSR Marker

Annotation
Blast2GO

SNP Detection
In 2000 stress
response genes

GO-Location	#	GO-Function	#	#			
Intracellular	332	ER	339	Transcription factor	421	Blinding	769
Cytoplasm	1186	Codg	225	Translation Factor	31	Nucleotide binding	362
Cytosol	591	Membrane	1193	Enzyme regulator	24	DNA binding	454
Nucleus	981	Plasma membrane	1119	Catalytic activity	379	RNA binding	239
Vacuole	477			Dehydrogenase	275	ATP binding	1201
Ribosome	358	Mitochondrion	946	Oxidase	130	Protein binding	1450
Plastid	628	Peroxisome	188	Hydrolase	520	Carbohydrate binding	30
Chloroplast	1807			Transferase	649	Lipid binding	15
Thylakoid	199	Apoplast	238	Kinase	135	Ion binding	3079
O-evolve. comp.	11	Cell wall	234	Synthase	357		
				Electron carrier	222	Struct. Molecule	143
				Transport	186	Nutrient reservoir	190
				Struct. Ribosome	564		

GO-Process	#	#	#		
Reproduction	8	Gen. o. precursors and energy	80	Biosynthetic processes	190
Transcription	251	Photosynthesis	161	C-metabolism	243
Transcription regulation	793	Photosynth. e-transport	1314	Gal metabolism	26
Translation	1038	e-transport chain	98	Amino acid metabol.	34
Translation initiation	109	Mitochondrial e-transport	38	Lipid metabolism	89
Signal transduction	238	Transport	1766	Nucleic acid metabolism	56
Cell-cell signaling	11	Intracell. Protein transport	209	Secondary metabolism	9
Cell cycle	58	Amino acid transport	47	Terpene biosynthesis	14
Redox	92				
Protein metabolism	90	Response to stress	277		
Protein folding	203	Defense response	201		
Protein phosphorylation	393	Cell death	81		
Proteolysis	248	Response to stimulus	47		