



CitEST libraries

Maria Luísa P. Natividade Targon¹, Marco Aurélio Takita^{1,2}, Alexandre M. do Amaral^{1,3},
Alessandra A. de Souza¹, Eliane Cristina Locali-Fabris¹, Sílvia de Oliveira Dorta¹, Kleber Martins Borges¹,
Juliana Mendonça de Souza¹, Carolina Munari Rodrigues¹, Adriano Reis Lucheta¹,
Juliana Freitas-Astúa^{1,4} and Marcos Antonio Machado¹

¹Centro APTA Citros Sylvio Moreira, Instituto Agronômico de Campinas, Cordeirópolis, SP, Brazil.

²Centro de Recursos Genéticos Vegetais, Instituto Agronômico de Campinas, Campinas, SP, Brazil.

³EMBRAPA Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil.

⁴EMBRAPA Mandioca e Fruticultura Tropical, Cruz das Almas, BA, Brazil.

Abstract

In order to obtain a better understanding of what is citrus, 33 cDNA libraries were constructed from different citrus species and genera. Total RNA was extracted from fruits, leaves, flowers, bark, seeds and roots, and subjected or not to different biotic and abiotic stresses (pathogens and drought) and at several developmental stages. To identify putative promoter sequences, as well as molecular markers that could be useful for breeding programs, one shotgun library was prepared from sweet orange (*Citrus sinensis* var. *Olimpia*). In addition, EST libraries were also constructed for a citrus pathogen, the oomycete *Phytophthora parasitica* in either virulent or avirulent form. A total of 286,559 cDNA clones from citrus were sequenced from their 5' end, generating 242,790 valid reads of citrus. A total of 9,504 sequences were produced in the shotgun library and the valid reads were assembled using CAP3. In this procedure, we obtained 1,131 contigs and 4,083 singletons. A total of 19,200 cDNA clones from *P. parasitica* were sequenced, resulting in 16,400 valid reads. The number of ESTs generated in this project is, to our knowledge, the largest citrus sequence database in the world.

Key words: ESTs, shotgun, citrus, phytopathogen.

Received: July 24, 2006; Accepted: April 17, 2007.

Introduction

Citrus is an important crop worldwide, with an annual production estimated at over 105 million tons in the period of 2000-2004 (FAO, 2005). Brazil is one of the main citrus fruit producing countries, together with the Mediterranean countries, the United States and China. More than two thirds of global citrus fruit production comes from these countries. The processing of citrus fruits represents approximately one third of the fruit production, more than 80 percent in frozen concentrated orange juice production. São Paulo State in Brazil, and Florida State in the USA, are the two main orange juice producers. Brazil exports approximately 99 percent of its production while 90 percent of Florida's production is consumed domestically and only 10 percent is exported.

Cultivated citrus species are susceptible to various pathogens including bacteria, fungi, nematodes, viruses

and viroids; these are responsible for severe losses worldwide (Deng *et al.*, 2001). Citrus trees usually grow as a combination of productive scion variety bud-grafted onto a rootstock variety adapted to soil and environmental conditions (Forment *et al.*, 2005). Citrus breeding is an expensive, long-term process, and citrus breeders have always faced many problems due to the complex genetic background of this crop.

The Citrus genome size (*n*) is approximately 367,000,000 bp, making it too expensive to be completely sequenced. On the other hand, expressed sequence tags (ESTs) are, today, a fast and inexpensive way of identifying new genes for obtaining data on gene expression and regulation, and for reconstructing metabolic maps based on genome data. ESTs of many important crops have been generated, including sugarcane (Vettore *et al.*, 2001, 2003), *Populus* (Sterky *et al.*, 2004), *Vitis* species (da Silva *et al.*, 2005). For citrus, a genomic shotgun library was also constructed in this work, but the focus was the cDNA libraries which were constructed from different tissues of citrus spe-

cies and genera, either in the presence of a pathogen/abiotic stress, or not.

Pathogens are also a focus of CitEST. One of the major concerns of the citrus industry is the phytosanitary problem caused by the different pathogens that attack the crop. Two of these pathogens, *Xylella fastidiosa* and *Xanthomonas axonopodis* pv. citri, had their genomes completely sequenced (Simpson *et al.*, 2000; da Silva *et al.*, 2002). In CitEST, EST libraries were constructed for the oomycete *Phytophthora parasitica* in either virulent or avirulent form.

In this paper, we report the construction and general data obtained from the CitEST libraries.

Materials and Methods

Biological material

The citrus species and genera as well as all the tissue sources used to construct the cDNA libraries are listed in Table 1. The material was collected at the Centro APTA Citros Sylvio Moreira - IAC (Cordeiropolis, SP, Brazil)

Table 1 - RNA sources used to construct the cDNA libraries.

Library	Species	Tissue
CS00-C1-100	<i>C. sinensis</i> var. Pera	Leaves from trees grown in greenhouse
CS00-C1-101	<i>C. sinensis</i> var. Pera	Leaves from trees infected with <i>Xylella fastidiosa</i> , grown in greenhouse
CS00-C1-650	<i>C. sinensis</i> var. Pera	Second pair of leaves from seedling grown in greenhouse
CS00-C1-102	<i>C. sinensis</i> var. Pera	Leaves from 8 months old plants, after 30 days of <i>X. fastidiosa</i> inoculation
CS00-C2-003	<i>C. sinensis</i> var. Pera	Bark from plants grown in greenhouse
CS00-C3-700	<i>C. sinensis</i> var. Pera	Peel of fruits, 1 cm in diameter, from plant grown in greenhouse
CS00-C3-701	<i>C. sinensis</i> var. Pera	Peel of fruits, 2.5 cm in diameter, from plant grown in greenhouse
CS00-C3-702	<i>C. sinensis</i> var. Pera	Peel of fruits, 5 cm in diameter, from plant grown in greenhouse
CS00-C3-703	<i>C. sinensis</i> var. Pera	Peel of fruits, 7 cm in diameter, from plant grown in greenhouse
CS00-C3-704	<i>C. sinensis</i> var. Pera	Peel of fruits, 8 cm in diameter, from plant grown in greenhouse
CS00-C3-705	<i>C. sinensis</i> var. Pera	Peel of fruits, 9 cm in diameter, from plant grown in greenhouse
CS00-C5-003	<i>C. sinensis</i> var. Pera	Flowers at different developmental stages
CS00-C1-401	<i>C. sinensis</i> var. Pera	Leaves from plants inoculated with CiLV, grown in greenhouse
CS12-C1-001	<i>C. sinensis</i> var. Olimpia	Leaves from trees grown in the field
CS13-C1-001	<i>C. sinensis</i> var. Pera	Leaves from trees grown in the field
CR05-C1-100	<i>C. reticulata</i> var. Ponkan	Leaves from healthy plant, grown in phytotron
CR05-C1-102	<i>C. reticulata</i> var. Ponkan	Leaves collected from plant 30 dai with <i>X. fastidiosa</i> , grown in phytotron
CR05-C1-103	<i>C. reticulata</i> var. Ponkan	Leaves collected from plant 60 dai with <i>X. fastidiosa</i> , grown in phytotron
CR05-C3-700	<i>C. reticulata</i> var. Ponkan	Peel of fruits, 1 cm in diameter, from plant grown in the field
CR05-C3-701	<i>C. reticulata</i> var. Ponkan	Peel of fruits, 2.5 cm in diameter, from plant grown in the field
CR05-C3-702	<i>C. reticulata</i> var. Ponkan	Peel of fruits 5 cm in diameter, from plant grown in the field
CL06-04-500	<i>C. limonia</i>	Roots from seedlings grown in nutritive solution
CL06-04-501	<i>C. limonia</i>	Roots from seedlings after drought stress
LT33-C1-003	<i>C. latifolia</i>	Leaves from plants grown in greenhouse
CG32-C1-003	<i>C. aurantifolia</i>	Leaves from plants grown in greenhouse
CM30-C1-401	<i>C. limettioides</i>	Leaves collected 48h after CiLV inoculation, grown in greenhouse
CA26-C1-002	<i>C. aurantium</i>	Leaves from plants grown in the field
PT11-C1-900	<i>Poncirus trifoliata</i>	Leaves from plants grown in greenhouse
PT11-C1-901	<i>P. trifoliata</i>	Leaves from plants inoculated with CTV, grown in greenhouse
PT11-C2-300	<i>P. trifoliata</i>	Bark from plants grown in greenhouse
PT11-C2-301	<i>P. trifoliata</i>	Bark from plants inoculated with <i>P. parasitica</i> , grown in greenhouse
PT11-C9-005	<i>P. trifoliata</i>	Seeds from fruits at different developmental stages
TS27-C2-300	<i>C. sunki</i>	Bark from plants grown in greenhouse
PP14-C7-801	<i>Phytophthora parasitica</i>	Mycelium from non-pathogenic isolate
PP14-C7-802	<i>Phytophthora parasitica</i>	Mycelium from pathogenic isolate

from field trees, trees growing under greenhouse and phytotron.

P. parasitica isolates IAC-01/95 G-40 and IAC-01/95 II, avirulent and virulent forms, respectively, were obtained from the collection of the Phytopathological Laboratory of the Centro APTA Citros Sylvio Moreira - IAC.

RNA isolation and cDNA library construction

Total RNA was extracted from 1 g of tissue using the TRIzol Reagent, according to the instructions of the manufacturer (Invitrogen). Due to the high content of carbohydrate, total RNA was extracted from seeds according to modifications of the procedure described by Naito *et al.* (1988). *P. parasitica* was obtained by filtration of liquid culture and total RNA was extracted from 1 g of the filtrate.

Poly A+ RNA was isolated from 0.5 mg of total RNA using the mRNA Isolation System (Promega Corporation, Madison, WI). cDNA libraries were constructed with the SuperScript Plasmid System with Gateway Technology for cDNA Synthesis and Cloning kit (Invitrogen). Complementary DNA was synthesized from mRNA using a primer consisting of an oligo(dT) sequence with a *NotI* restriction site. *SalI* adapters were ligated to the blunt-ended cDNA fragments followed by *NotI* digestion. The cDNA fragments were size fractionated in Sephacryl S cDNA Size Fractionation Columns (Invitrogen) and cloned into the *NotI-SalI* restriction site of the pSPORT 1 vector. The ligated cDNA fragments were transformed into *E. coli* DH5 α cells by the ice-cold RbCl/CaCl₂ solution method (Hanahan, 1983). White colonies were inoculated in 200 μ L of liquid Circle Grow medium (Molecular Biology Certified Bacterial Growth Media, QBiogene - Bio 101 Systems - USA) containing 8% (v/v) glycerol and 100 μ g/mL of ampicillin, in 96-well-microtiter plates, incubated overnight at 37 °C and stored at -80 °C.

The citrus EST collections were catalogued by two characters indicating the genera and species, followed by two numbers indicating the variety, a character and a number indicating the tissue source, and three numbers indicating the conditions. As an example: TS27-C2-300, bark tissue of *Citrus sunki* under greenhouse.

Shotgun library preparation

For the shotgun library, ten grams of young leaves were collected from Pera Olimpia sweet orange grown in greenhouse and used for total DNA preparation according to Dellaporta *et al.* (1983). The DNA was purified in a cesium chloride gradient. Twenty micrograms of DNA were partially digested with 0.1 U of *Sau3AI* (Invitrogen) for 10 min at room temperature. The DNA was separated in a 0.8% agarose gel, the fragments of 1.5 to 3 kb were isolated with the GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and cloned in pGEM3Z open at the *BamHI* restriction site.

Plasmid DNA minipreparation and sequencing

Plasmid DNA was extracted by the boiling method (Marra *et al.*, 1999). The sizes of the cloned fragments from the EST libraries were evaluated by digesting the plasmid DNA with *PvuII*. Visualization of the products was done by electrophoresis in 1.2% agarose gels. The sequencing reactions were prepared according to the instructions of the manufacturer (Applied Biosystems) for the DNA sequencing kit Big Dye Terminator cycle sequencing ready reaction v3 and v3.1. Sequencing was done in ABI 3700 and 3730 DNA Analyzers (Applied Biosystems).

Sequence data analysis

The methods used to submit, process and analyze the ESTs are described elsewhere in this issue (Reis *et al.*, in this issue).

Results and Discussion

A total of 33 cDNA libraries were constructed from different tissues of citrus species and genera at different developmental stages (fruits), or under biotic or abiotic stresses (pathogens and drought), as listed in Table 1. A total of 16, 9, 1, 1, 2 and 4 libraries were prepared from leaves, fruits, flowers (mixture of four developmental stages), seeds (from fruits at different developmental stages), roots and bark, respectively. Two cDNA libraries from mycelia of *P. parasitica* and one shotgun library from leaves of *C. sinensis* var. Pera Olimpia were also prepared. The fragment size of the cDNA clones was evaluated for each library and ranged from 1 to 2 kb, depending on the library (average fragment size was 1.5 kb). Vettore *et al.* (2001) reported an average fragment size of 1,250 bp for clones of sugarcane cDNA libraries.

A total of 13 *C. sinensis* var. Pera cDNA libraries were constructed, mainly because this is the most important cultivar in Brazil. Leaves from *X. fastidiosa*, CTV (*Citrus tristeza virus*), CiLV-C (*Citrus leprosis virus* cytoplasmic type) and *P. parasitica*-infected plants were used as RNA sources to evaluate the genes induced/repressed by these pathogens. Two cDNA libraries from Rangpur lime (*C. limonia*) were constructed, because this *Citrus* species was the most commonly used as rootstock in Brazil. To compare gene expression in different tissues, libraries were constructed using tissues of the same plant.

cDNA libraries from peel of *C. sinensis* and *C. reticulata* fruits with diameters of 1, 2.5, 5, 7, 8, 9 cm and 1, 2.5, 5 cm, respectively, were constructed with the purpose of identifying differentially expressed genes in each stage. Factors associated with productivity and fruit quality are strongly dependent on fruit development. The understanding of the molecular mechanisms by which citrus plants regulate this complex process represents a unique opportunity to improve our understanding of all physiological mechanisms determining fruit setting, fruit size, organic

acid accumulation, carbon flow, peel color and morphology. A database containing nonredundant sequences from all cDNA libraries was constructed and used to perform comparisons with all libraries and evaluate individual gene expression levels.

The processing of ESTs is a fundamental step to obtain high quality sequence reads from raw sequencer trace

Table 2 - General data obtained from each CitEST library.

Specie*	Library	# Submitted reads	# Valid reads	Efficiency (%)
CS	CS00-C1-100	10,268	9,536	92.87
CS	CS00-C1-101	6,777	6,408	94.56
CS	CS00-C1-102	10,464	9,243	88.33
CS	CS00-C1-401	5,664	4,852	85.66
CS	CS00-C1-650	5,280	4,918	93.14
CS	CS00-C2-003	10,560	7,952	75.30
CS	CS00-C3-700	10,558	9,624	91.15
CS	CS00-C3-701	9,694	8,760	90.37
CS	CS00-C3-702	10,452	9,167	87.71
CS	CS00-C3-703	10,560	8,195	77.60
CS	CS00-C3-704	10,079	7,330	72.73
CS	CS00-C3-705	10,560	8,653	81.94
CS	CS00-C5-003	10,080	9,082	90.10
CS	CS12-C1-001	3,072	2,603	84.73
CS	CS13-C1-001	2,592	2,288	88.27
CR	CR05-C1-100	9,600	8,461	88.14
CR	CR05-C1-102	9,564	8,452	88.37
CR	CR05-C1-103	9,600	8,516	88.71
CR	CR05-C3-700	11,328	9,367	82.69
CR	CR05-C3-701	9,792	8,924	91.14
CR	CR05-C3-702	9,600	7,959	82.91
PT	PT11-C1-900	9,600	8,941	93.14
PT	PT11-C1-901	9,600	8,926	92.98
PT	PT11-C2-300	8,160	5,988	73.38
PT	PT11-C2-301	6,816	4,685	68.74
PT	PT11-C9-005	4,800	4,097	85.35
CA	CA26-C1-002	9,600	8,439	87.91
CG	CG32-C1-003	9,600	8,219	85.61
LT	LT33-C1-003	9,600	8,756	91.21
CL	CL06-C4-500	7,871	5,438	69.09
CL	CL06-C4-501	7,584	5,607	73.93
CM	CM30-C1-401	9,600	8,188	85.29
TS	TS27-C2-300	7,584	5,216	68.78
PP	PP14-C7-801	9,600	7,601	79.20
PP	PP14-C7-802	9,600	8,799	91.66

*CS - *Citrus sinensis*; CR - *Citrus reticulata*; PT - *Poncirus trifoliata*; CA - *Citrus aurantium*; CG - *Citrus aurantifolia*; LT - *Citrus latifolia*; CL - *Citrus limonia*; CM - *Citrus limettioides*; TS - *Citrus sunki*; PP - *Phytophthora parasitica*.

data (Kunne *et al.*, 2005). Table 2 presents the general data obtained from each CitEST library. A total of 286,559 cDNA clones from citrus were sequenced in their 5' end, generating 242,790 valid reads of citrus. A total of 19,200 cDNA clones from *P. parasitica* were sequenced, generating 16,400 valid reads. The success index is the percentage of reads with more than 150 bp with Phred quality above 20. Table 2 presents the percentage of efficiency per library (number of valid reads / number of submitted reads x 100). The minimum efficiency was observed in TS27-C2-300 and PT11-C2-301, from bark tissue of *C. sunki* and *P. trifoliata*, respectively. The maximum efficiency was observed in CS00-C1-100, the library from leaves collected from trees grown in greenhouse. In this case, the leaves used to construct the library were from new flushes of branches, and therefore more suitable for RNA extraction. The sequences are available at: biotecnologia.centrodecitricultura.br

Table 3 presents a summary of CitEST data. The average size of reads was 847.48 bp. Forment *et al.* (2005) reported an average sequence length of 500 nucleotides for citrus EST collection obtained from 25 cDNA libraries. The average EST length from 26 libraries constructed from different sugarcane tissue was 750 bp (Vettore *et al.*, 2001). The number of clusters/specie is also presented in Table 3 and is expressed by (# Contigs + # Singlets). The average efficiency of all libraries was 84.73%. Redundancy is calculated as number of clusters / number of Reads, expressed as a percentage. The high values in redundancy obtained for some species are related to the high number of reads sequenced for these particular species, since the libraries are being exhausted.

The efficiency of the genomic library (shotgun) of sweet orange was 78.2% based on the same quality analysis used for the ESTs. The objective of constructing this library was to identify putative promoter sequences as well as molecular markers that could be useful for breeding programs. We have produced a total of 9,504 sequences for this library and the valid reads were assembled using CAP3. In this procedure, we have obtained 1,131 contigs and 4,083 singletons. These clusters were analyzed for chloroplast or mitochondria contamination using the Blast tool to search for similarity against the chloroplast and mitochondria protein sequences from the Swiss-Prot knowledgebase. In this analysis, it was shown that 107 contigs and 203 singletons were, in fact, contaminant sequences from either chloroplast or mitochondria. This encompasses a total of 1,652 sequences, which corresponds to 17.38% of the total.

In reports from other sequencing projects in Brazil, a total of 237,954 sugarcane ESTs (Vettore *et al.*, 2003), and 123,889 Eucalyptus ESTs (Carrer, 2005) were generated. Forment *et al.* (2005) obtained 22,635 high quality citrus ESTs from 25 citrus libraries covering different tissues, developmental stages and stress conditions. In the CitEST project, a total of 242,790 citrus ESTs were generated. This

Table 3 - Summary of CitEST data.

Species	# Submitted reads	# Valid reads	Average size of reads (bp)	# Contigs	# Singlets	# Clusters	Redundancy (%)	Efficiency (%)
<i>Citrus sinensis</i> (CS)	126,600	108,611	842,22	14,422	17,699	32,121	70.42	85.75
<i>Citrus reticulata</i> (CR)	59,484	51,676	842,63	8,307	10,566	18,873	63.48	86.88
<i>Poncirus trifoliata</i> (PT)	38,976	32,637	832,73	5,202	7,671	12,873	60.55	83.74
<i>Citrus aurantium</i> (CA)	9,600	8,439	849,58	1,359	4,248	5,607	33.55	87.91
<i>Citrus aurantifolia</i> (CG)	9,600	8,219	819,88	1,642	2,685	4,327	47.35	85.61
<i>Citrus latifolia</i> (LT)	9,600	8,756	863,55	1,672	3,211	4,883	44.23	91.21
<i>Citrus limonia</i> (CL)	15,455	11,045	803,16	2,385	3,560	5,945	46.17	71.47
<i>Citrus limettioides</i> (CM)	9,600	8,188	926,35	1,560	2,980	4,540	44.55	85.29
<i>Citrus sunki</i> (TS)	7,584	5,216	830,92	1,011	1,641	2,652	49.15	68.78
All species	286,559	242,790	847,48	-	-	-	-	84.73

number is, to our knowledge, the largest citrus sequence database in the world.

Acknowledgments

The authors would like to thank Marcelo Reis for assistance in bioinformatics issues. The CitEST project was sponsored by grants from CNPq/Institutos do Milênio/Citrus to M. A. M., process number 620054/01-8.

References

- Carrer H (2005) Seeing the FORESTs for the trees. *Genet Mol Biol* 283(Suppl.3):i-ii.
- da Silva AC, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Van Sluys MA, Almeida NF, Alves LM, *et al.* (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* 417:459-463.
- da Silva FG, Iandolino A, Al-Kayal F, Bohlmann MC, Cushman MA, Lim H, Ergul A, Figueroa R, Kabuloglu EK, Osborne C, *et al.* (2005) Characterizing the grape transcriptome. Analysis of expressed sequence Tags from multiple *Vitis* species and development of a compendium of gene expression during berry development. *Plant Physiol* 139:574-597.
- Dellaporta SL, Wood J and Hicks JB (1983) A plant miniprep: Version II. *Plant Mol Biol Rep.* 1:19-20.
- Deng Z, Tao Q, Chang YL, Huang Ling P, Yu C, Chen C, Gmitter Jr DFG and Zhang HB (2001) Construction of a bacterial artificial chromosome (BAC) library for citrus and identification of BAC contigs containing resistance gene candidates. *Theor Appl Genet* 102:1177-1184.
- FAO (2005) Food and Agriculture Organization of the United Nations. <http://www.fao.org>.
- Forment J, Gadea J, Huerta L, Abizanda L, Agusti J, Alamar S, Alos E, Andres F, Arribas R Beltran JP, *et al.* (2005) Development of a citrus genome-wide EST collection and DNA microarray as resources for genomic studies. *Plant Mol Biol* 57:375-391.
- Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166:557.
- Kunne C, Lange M, Funke T, Miede H, Thiel T, Grosse I and Scholz U (2005) CR-EST: A resource for crop ESTs. *Nucleic Acids Res* 33. Database issue D619-D621.
- Marra MA, Kucaba TA, Hillier LW and Waterston RH (1999) High-throughput plasmid DNA purification for 3 cents per sample. *Nucleic Acids Res* 27:e37.
- Naito S, Dubé PH and Beachy RN (1988) Differential expression of conglycinin α and β subunit genes in transgenic plants. *Plant Mol Biol* 11:109-123.
- Simpson AJG, Reinach FC, Arruda PA, Abreu FA, Acencio M, Alvarenga R, Alves LMC, Araya JE, Baia GS, Baptista CS, *et al.* (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature* 406:151-157.
- Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM, Charbonnel-Campaa L, Lindvall JJ, Tandre K, Strauss SH, *et al.* (2004) A *Populus* EST resource for plant functional genomics. *Proc Natl Acad Sci USA* 101:13951-13956.
- Vettore AL, da Silva FR, Kemper ED and Arruda P (2001) The libraries that made SUCEST. *Genet Mol Biol* 24:1-7.
- Vettore AL, da Silva FR, Kemper ED, Souza, GM, da Silva AM, Ferro MIT, Henrique-Silva F, Giglioti EA, Lemos MVF, Coutinho LL, *et al.* (2003) Analysis and functional annotation of an expressed sequence TAG collection for tropical crop sugarcane. *Genome Res* 13:2725-2735.

Associate Editor: Ivan de Godoy Maia