

## Worked example - IGS Cassava Website

Go to the IGS cassava website: <http://cassava.igs.umaryland.edu>

**Introduction**

Cassava is one of the most important crops in unfavorable environments in developing countries, where poverty is common and severe. Because of its high productivity, even in extreme conditions, cassava constitutes a source of food and income for poor farmers in Africa, Asia and Latin America. Although cassava is fairly resistant to water stress, the molecular basis for this tolerance is poorly understood. Several traits have been associated with its drought tolerance, such as regulation of stomata activity, changing leaf expansion rates due to decrease in cell proliferation, and modifications of photosynthetic pathways to maintain high photosynthetic activity. Improving cassava's tolerance to drought is important to help increasing yields in the semi-arid Sub Saharan African regions where cassava is an essential crop. Cassava's natural stress tolerance can be substantially improved by breeding, especially by marker-assisted selection of key physiological traits associated with drought tolerance. Under this project we have constructed a BAC-based fingerprint map of an inbred cultivar of cassava. A BAC library of 70,000 BAC clones has been fingerprinted and a minimum tiling path (MTP) of clones has been selected. The ends of the BACs in the MTP are being sequenced and selected low-copy sequences spread throughout the genome will be re-sequenced in a panel of 10 cassava genotypes to identify single nucleotide polymorphisms (SNPs) and a linkage map will be constructed using two mapping populations generated at the Center for Tropical Agriculture (CIAT). This project will deliver SNP markers for uniformly distributed around the genome that will be useful for identifying quantitative trait loci (QTL) associated with drought tolerance.

**Funding**

A project funded by the [Generation Challenge Programme](#) of the [Consultative Group on International Agricultural Research](#)  
GCP Project G3007.03: "Development of Genomics Resources for Molecular Breeding of Drought Tolerance in Cassava"

**Investigators**

**Principal Investigator**

*Pablo Rabinowicz*  
Institute for Genome Sciences, Department of Biochemistry and Molecular Biology  
University of Maryland School of Medicine  
Baltimore MD, USA

**Co-Investigators**

*Alexander Myburg*  
African Center for Gene Technologies  
Department of Genetics  
Forestry and Agricultural Biotechnology Institute  
University of Pretoria  
Pretoria, South Africa

*Jane Morris*  
African Center for Gene Technologies  
University of Pretoria  
Pretoria, South Africa


*Chris Rey*  
African Center for Gene Technologies  
School of Molecular and Cell Biology  
University of the Witwatersrand  
Johannesburg, South Africa

*Mingcheng Luo*  
University of California  
Davis CA, USA

Click on the Blast tab at the top right

Open a new browser window and go to Entrez EST and retrieve the sequence for accession DB938508. Paste the sequence in FASTA format in the search box of the BLAST page. Select the "cassava ESTs" database and the program "blastn". Uncheck the filter "Low complexity" and click "Search".

Choose program to use and database to search:

Program:  Database:  

Enter sequence below in [FASTA](#) format for:  
 Assembled Cassava ESTs  
 Whole Genome Shotgun Sequences  
 BAC End Sequences

Or load from disk

Set subsequence: From  To

\*results will open in new window

The query sequence is filtered for low complexity regions by default.  
 Filter:  Low complexity  Mask for lookup table only

Expect:  Matrix:   Perform ungapped alignment

Query Genetic Codes (blastn only):

Database Genetic Codes (tblastn only):

Frame shift penalty for blastn:

Other advanced options:

Graphical Overview Alignment view:

Descriptions:  Alignments:  Color schema:

\*results will open in new window

Comments and suggestions to: [blast-help@ncbi.nlm.nih.gov](mailto:blast-help@ncbi.nlm.nih.gov)

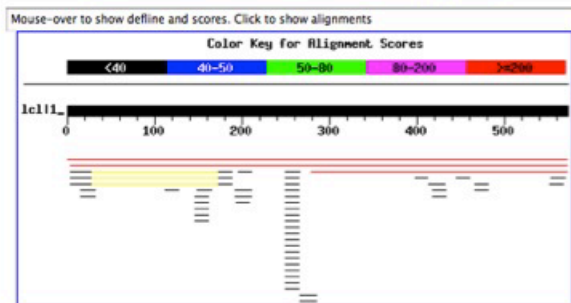
There will be two perfect matches.

**Reference:**  
 Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

**Database:** cassava\_ests.fasta  
 79,444 sequences; 40,318,471 total letters

**Query=** gi|164392540|gb|DB927724.1|DB927724 DB927724 full-length enriched cassava cDNA library Manihot esculenta cDNA clone CAS01\_022\_e17\_5', mRNA sequence (572 letters)

**Distribution of 53 Blast Hits on the Query Sequence**



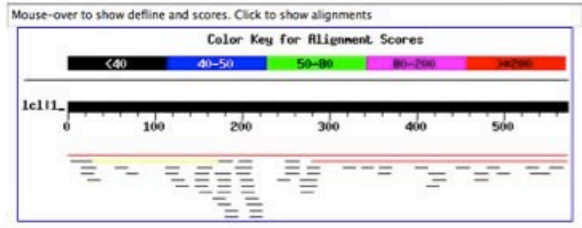
Sequences producing significant alignments:

	Score	E
	(bits)	Value
gb DB927724.1 DB927724 DB927724 full-length enriched cassav...	1134	0.0
gb DB938508.1 DB938508 DB938508 full-length enriched cassav...	1126	0.0
gb DB945690.1 DB945690 DB945690 full-length enriched cassav...	509	e-144
gb CF641326.1 CF641326 mb nosah.01.B18.u MBra685 cassava no...	34	1.1
gb DV441202.1 DV441202 CV01003B1G01.f1 CV01-normalized libr...	34	1.1
gb DV458957.1 DV458957 CV02028B2D06.f1 CV02-normalized libr...	34	1.1
ah na026553.1 na026553 na026553 full-length enriched cassav...	34	1.1

Repeat the search of this sequence but this time against the "Assembled cassava ESTs" database.

Query= gi|164392540|gb|DB927724.1|DB927724 DB927724 full-length enriched cassava cDNA library Manihot esculenta cDNA clone CAS01\_022\_E17 5', mRNA sequence (572 letters)

**Distribution of 69 Blast Hits on the Query Sequence**



Sequences producing significant alignments:

Sequence	Score (bits)	E Value
CL5886Contig1	1134	0.0
gb DB945690.1 DB945690	509	e-144
gb DV441202.1 DV441202	34	0.57
CL6124Contig1 nitroreductase family protein GO:0008152	34	0.57
gb CK641526.1 CK641526	34	0.57
gb DB923262.1 DB923262 mitochondrial substrate carrier fami...	34	0.57
CL5442Contig1	34	0.57
gb DV443115.1 DV443115 mitochondrial substrate carrier fami...	32	2.2
gb CK642441.1 CK642441	32	2.2
CL9002Contig1 transducin family protein wd-40 repeat family...	32	2.2
CL8862Contig1 ---NA--- GO:0003794, GO:0016020	32	2.2
CL10096Contig1 phosphofructokinase family protein GO:000609...	32	2.2

See that the first match corresponds to an assembly of the two ESTs hit in the previous search. Click in the component ESTs and open their GB records. See for example, the tissue in which these sequences are expressed.

Home Gbrowse Blast Downloads

**EST Assembly Info**

**CL5886Contig1**

There was no annotation for this molecule

**Sequence:**

```
GAATGATGGCTATGGCGATGATGGAGAAGTGCAGAGCTAGTAATCCATCTTTAGCCACT
CCTTCAAATCGCTTCTTAATFATCGATCAATCTCCGATAATCCACCTCCGCTCCTCTT
CAGGCAAGGATTTTTTGGCCGATTTCTTATTTTTTCATTTTAAGTGATTTTGATCATCAT
CGCCATCGCAATCTGTGTCAAGCTCTCTTTGGCTTTGGATTCAGGCTTTGGCGATC
ACTTCACTGGAAATGGCCAACGCTGCAATACCAATTTTCCCATTTTCATAGCCGGACTGTG
TGACTTGTGGCACTTCTATAAGAAATATAAACAGACTCTCTGGCTTCTTTGGAAAGAG
GCCCTGGAGGAAATGATGACTCTCTTCAGTGTGGAAGTTTGGCTCGGCTGGACTAAAA
TTCAGAGAAATCGTGTGTGGACCTTGAATTCGATGATATAACTTATTTAAGGGTG
TGTGGGTGGAGAAATGTGCTTCAGGATGAAGATAGGTGGAGTTTGTAAATGCAGATT
GGATGCATAAATACAAAGGTTCAAGCAGCTGCTTCTTACTACGACCCTGACGAGG
TCTTCAAATCTTAAATATTGTAATTT
```

**Component Reads:**

- [gi|164392540|gb|DB927724.1|DB927724](#)
- [gi|164404535|gb|DB938508.1|DB938508](#)



## DB927724 full-length enriched cassava cDNA library *Manihot esculenta* cDNA clone CAS01\_022\_E17 5-, mRNA sequence

GenBank: DB927724.1

[GenBank](#) [FASTA](#)

### IDENTIFIERS

dbEST Id: 52600167  
EST name: DB927724  
GenBank Acc: DB927724  
GenBank gi: 164392540

### CLONE INFO

Clone Id: CAS01\_022\_E17 (5')  
DNA type: cDNA

### PRIMERS

PolyA Tail: Unknown

### SEQUENCE


```
GAATGATGGCTATGGCGATGATGGAGAAGTGCAGAGCTAGTAATCGCATCTTTAGCCACT
CCTTCAAATCGCTTCTTAATTATCGATCAATCFCGGATAATCCACTCTCCGCTCCTCTT
CAGGCAAGGATTTTTTGGCCGATTTCTATTTTCATTTTAAAGTGATTTGATCATCAT
CGCCATCGCAATTTCTGTGTCAAGCTCTTCTTTGGCTTTGGATTGAGGCTTTGCGCATC
ACTTCACTGGAAATGGCCAACGCTGCATACCAATTTTCCCATTTTCATAGGCCGGACTTGT
TGACTTGTGCGAACTTCTATAGGAATATAAACAGACTCTCTGGCCTTCTTTGGAAAGAG
GCCTTGGAGGAATGATGTACTCTCTTCAGTGTGGAAGTTTGGCTCGGCTGGGACTAAAA
TTCAGAGAAATCCGTCGTTTGGGACCTTGAATTCTGATGATATAACTTATTTAAGGGTG
TGTGGGTGAGAAAAATCTCGTTCAGGATGAAGATAGCTTGGAGTTTGTAAATGCAGATT
GGATGCATAAATACAAGGTTCAAGCAGGCTG
```

Entry Created: Dec 31 2007  
Last Updated: Nov 24 2008

### COMMENTS

Tetsuya Sakurai and Kazuo Shinozaki are at RIKEN Plant Science Center, Yokohama, Kanagawa 230-0045, Japan. This library was constructed from several development stages and tissues. The development stage and tissue of each entry are unknown.

### LIBRARY

Lib Name: full-length enriched cassava cDNA library  
Organism: [Manihot esculenta](#)  
Cultivar: MTA116  
Tissue type: mixture of leaf, root   
Description: This library was constructed from several development stages and tissues. The development stage and tissue of each entry are unknown.

Click in the "Gbrowse" tab.

**CASSAVA Genome Database**

Home | **Gbrowse** | Blast | Downloads

### Introduction

Cassava is one of the most important crops in unfavorable environments in developing countries, where poverty is common and severe. Because of its high productivity, even in extreme conditions, cassava constitutes a source of food and income for poor farmers in Africa, Asia and Latin America. Although cassava is fairly resistant to water stress, the molecular basis for this tolerance is poorly understood. Several traits have been associated with its drought tolerance, such as regulation of stomata activity, changing leaf expansion rates due to decrease in cell proliferation, and modifications of photosynthetic pathways to maintain high photosynthetic activity. Improving cassava's tolerance to drought is important to help increasing yields in the semi-arid Sub Saharan African regions where cassava as an essential crop. Cassava's natural stress tolerance can be substantially improved by breeding, especially by marker-assisted selection of key physiological traits associated with drought tolerance. Under this project we have constructed a BAC-based fingerprint map of an inbred cultivar of cassava. A BAC library of 70,000 BAC clones has been fingerprinted and a minimum tiling path (MTP) of clones has been selected. The ends of the BACs in the MTP are being sequenced and selected low-copy sequences spread throughout the genome will be re-sequenced in a panel of 10 cassava genotypes to identify single nucleotide polymorphisms (SNPs) and a linkage map will be constructed using two mapping populations generated at the Center for Tropical Agriculture (CIAT). This project will deliver SNP markers for uniformly distributed around the genome that will be useful for identifying quantitative trait loci (QTL) associated with drought tolerance.

### Funding

A project funded by the [Generation Challenge Programme](#) of the [Consultative Group on International Agricultural Research](#)  
 GCP Project G3007.03: "Development of Genomics Resources for Molecular Breeding of Drought Tolerance in Cassava"

### Investigators

**Principal Investigator**  
 Pablo Rabinowicz  
 Institute for Genome Sciences, Department of Biochemistry and Molecular Biology  
 University of Maryland School of Medicine  
 Baltimore MD, USA

**Co-investigators**

Alexander Myburg  
 African Center for Gene Technologies  
 Department of Genetics  
 Forestry and Agricultural Biotechnology Institute  
 University of Pretoria  
 Pretoria, South Africa

Jane Morris  
 African Center for Gene Technologies  
 University of Pretoria  
 Pretoria, South Africa

Chris Rey  
 African Center for Gene Technologies  
 School of Molecular and Cell Biology  
 University of the Witwatersrand  
 Johannesburg, South Africa

Mingcheng Luo  
 University of California  
 Davis CA, USA

Type ctg631 in the "Landmark or Region" box and click "Search" to browse contig number 631.

Showing 100 kbp from ctg1, positions 800,000 to 900,000

**Instructions**  
 Searching: Search using a sequence name, gene name, locus, or other landmark. The wildcard character \* is allowed.  
 Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position.

Examples: ctg1.

[Bookmark this] [Upload your data] [Hide banner] [Share these tracks] [Link to image] [High-res image] [Help] [Reset]

**Search**

Select a Contig to View

**Landmark or Region:** ctg631 Search

**Data Source:** Cassava Database (Chromosome)

Scroll/Zoom: <<< << >> >>> Show 100 kbp Flip

**Legend:**  
 Default: dark blue, Minimum Tiling Path: yellow, BAC with SNP: red

You can also look for a contig using the pull down menu under "Search" (there are around 2000 contigs).



Scroll down to the "Tracks" section toward the bottom of the page. Uncheck "BACs" Under "General" and click "Update image".

The screenshot shows a genomic browser interface. At the top, there is a search bar with 'ctg631-8033198.1966801' entered. Below the search bar, there are navigation controls for 'Scroll/Zoom' and 'Show 1.967 Mbp'. A legend indicates that dark blue represents the default, yellow represents the minimum tiling path, and red represents BACs with SNPs. The main view shows a horizontal axis with various BAC clones represented as colored bars. A yellow arrow points to a specific BAC clone labeled 'MH109K10' in the 'BACs' track. Below the tracks, there are controls for 'Tracks' (Overview, Contigs, General, BACs) and 'Disclaimer Settings'.

Click on a yellow BAC see the sequence. Click on a red BAC to see the sequence and the SNP.

### Clone Info

**MH109K10**

Located on: [ctg631](#)

Coordinates: 826801 - 939601

Length: 112801

#### BAC End Sequence: Forward

```
GGGGATCCCTCAGAGTGGACCTGCAGGCATGCAAGCAAAACGGGACGATGAAATACCCCA
GGGGAGCATTTGATGAGCCACCCAGGGTGGTTATGTTATTAGAGGGGACACATGTTTCAG
GTTAAGCTTGGAAAAGAAAGAAAAGTTTACATTCCTTATGTATGTTGATCATGTATGG
GATATTACAGGTTGACGGGGGGGGTGGCTTGGCTTCTACGGGTCCTCCGGCCCTTAAGCCG
ATCTGGATCCTAGCCGGCCGTAACGGGTGGGATTTCCGGGTCGTTACAGAGTGGTATCAGA
GCCTTTTTTTTATATGTTCCGACCTAGTATGTCGGGCTCATAGATGTTCTAGAAAGGTCAA
GCACACNGGAAAATCATGTCACACAGGATGGACCCATGTCCTGTCATATATATGAGA
TGTAGATGCCAAGATTATATGCAATG
```

#### BAC End Sequence: Reverse

```
AAATATGCCATATCAACGCTCCGCTCAACTAAAATAGGGGCACTGCTTGGCTCAGATAE
TTTTGAGTGAATTAGTACAGCCAGGCTTGTGACACATGCAAAATATTTTCAGCACAAT
GCACATGGTCTGTTAGCTTGACATTCAAATTTCAACCCCAACTACAATAATGAAATG
AATGTAGAAAGCTTTAAAATCAAGAGTACCAATTTAGAAATAGTGGTAGCCATTTTATA
TTATATGTCAGCATTTCAATCAAAAGCACGTTCAAATGATCTCTCAAGAAAATAGAAA
TACATGAGAAATGCCTTGAAATATTTGCTGATTTAGACAGGGGCCAACTTTAATGACAT
TAGCTGAAGGGTAAACGAATATGATATACGCCAAACATATTTGAAGACAGTCAAGACT
GGAGCTCATGCTCCGCTTGAATAATGGCTCAAGTTCAGCTTGACAAAATCATGAAAATCT
TGAACAGCTGAAAAGGGAATGTTGAGTTGTGACTGAATCAAATTAATTGCTCAAATC
CCATCCAGGCTTAGGTAAMAGACAGGCTTACTTAGTACCTACTGTTGTCATTTGACAT
TTGATCAGCAATAGTTGATGCTCAGTAGATGCAAAAATTTTATCCATTTATGCAATCAG
AGAAGTCTTTAAATGAGATATTTAAATTTCAAAATCTTGAAGAAAATCATAAAAAAGGA
ATAGAATAGTCTGTTGTTGGTAAAAAATATGAGAAACAGTTGCTTCAAAACTAACA
CAAGTTCATATCTCAGGAAATAACAGGCTAAAACCTTACCTCCGGGGTTC
```

#### Physical Mapping (raw FPC output file):

```
Approximate_match_to_oomid "MH109K10"
Creation_date 108 10 10 12 52
Modified_date 108 10 10 15 52
```

#### SNPs

Strand: reverse

TACGCAACATATTTGAAAGACAGTCAAGACTGGAGCTCATGCTCCGCT [\(r/a\)](#)GAAAAATGGCTCAAGTTCAGCTTGACAAAATCATGAAAATC



**Exercises:**

1. Search for EST number DB927724 in NCBI GenBank. Go to the Cassava IGS BLAST page and search the FASTA sequence of DB927724 versus the cassava BAC-end sequences.

Display the best hit in the IGS cassava Gbrowse (hint: copy and paste the first hit's name). Zoom out to see all the BACs in the contig. Click on a BAC that contains a SNP to see the SNP sequence.

2. Download the SNPs derived from the physical map and find MH070O17-MF. Display the contig that contains this BAC in Gbrowse.

3. Find where the "MH109K10: Reverse" sequence is located in the cassava genome sequence in Phytozome (hint: use blast in Phytozome). What gene does this sequence tag?