

A cassava physical map for genome-wide SNP discovery

Pablo Rabinowicz

Institute for Genome Sciences (IGS)
University of Maryland School of Medicine
Baltimore, MD



***GCP ARM Meeting, Bamako
September 2009***



UNIVERSITY OF MARYLAND
SCHOOL OF MEDICINE

INSTITUTE FOR GENOME SCIENCES

Focus on improving drought tolerance in cassava

- Cassava is rather tolerant to water stress
- Little known of molecular basis of tolerance
- Genome-wide SNP marker development for QTL mapping and breeding

Objectives

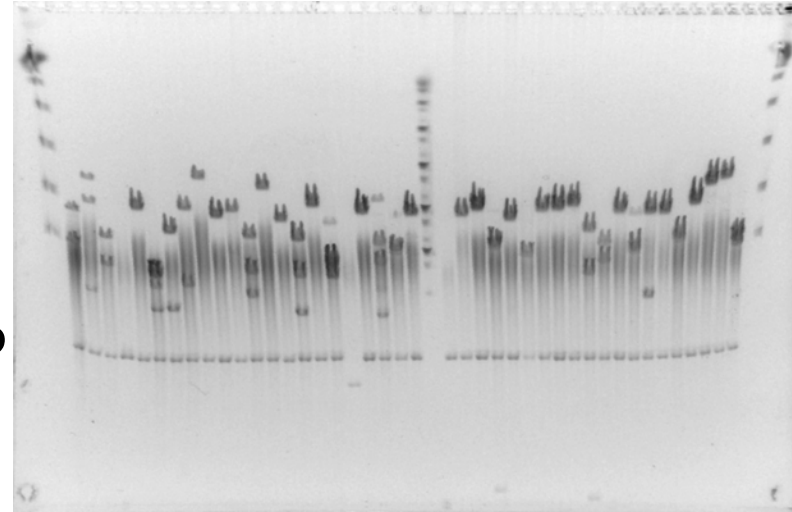
- Construct and fingerprint a cassava AM560-2 BAC library (~10X)
- BAC-end sequencing and re-sequence 10 genotypes to identify SNPs (TMS30572, CM2177-1, TME3, MTAI1, MNG19, MCOL22, MGUA76, MBRA12, MVEN309, and MMEX17)
- Genotype the MCol1734 x MVen77 mapping population (235 individuals will be phenotyped in Brazil and Colombia for drought tolerance)
- Linkage map and QTL mapping, and anchoring linkage and physical maps
- Making all data public through a project website
- Training National Programs in SNP genotyping

Synergism with the Cassava Genome Sequencing Consortium

- U. of Arizona, Roche, DOE-JGI sequenced the AM560-2 genome with 454
- Assembly & annotation expected later this year
- Coming project to add transcriptome, genome, and reduced-representation sequencing of additional genotypes
- Sources for additional SNPs

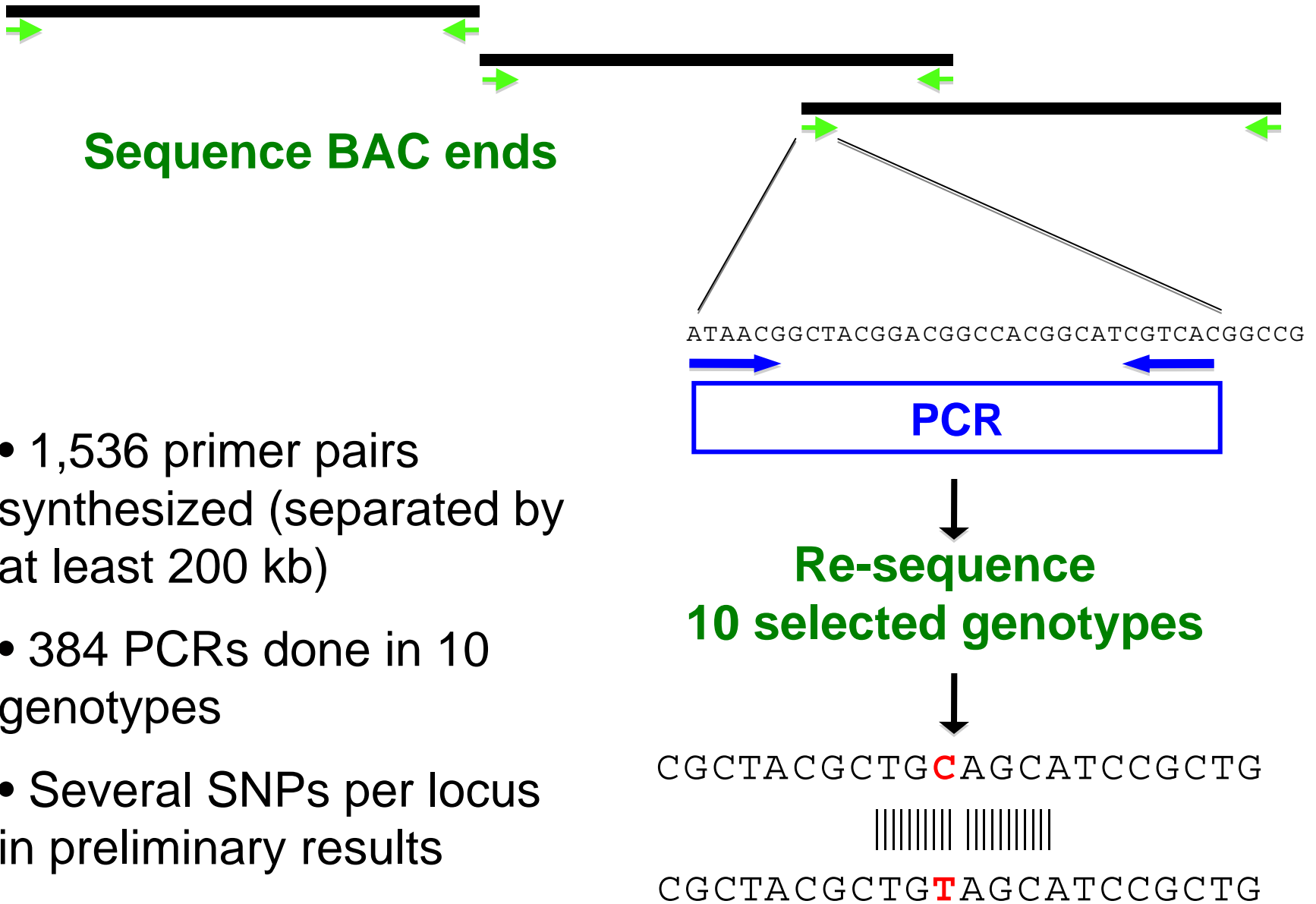
Fingerprinted AM560-2 BAC library

- 72,000 BAC clones, ~10X coverage
- 115Kb average BAC insert size, 7% empty clones
- 58,244 BACs assembled
- 2,104 contigs and 5,054 singletons
- Longest contig ~4.8 Mb with 357 clones
- Average contig: 25 BAC clones and 0.41 Mbp
- MTP: 6,868 clones, 710 Mbp
- Sequenced all BAC-ends in MTP and ~2,000 singletons



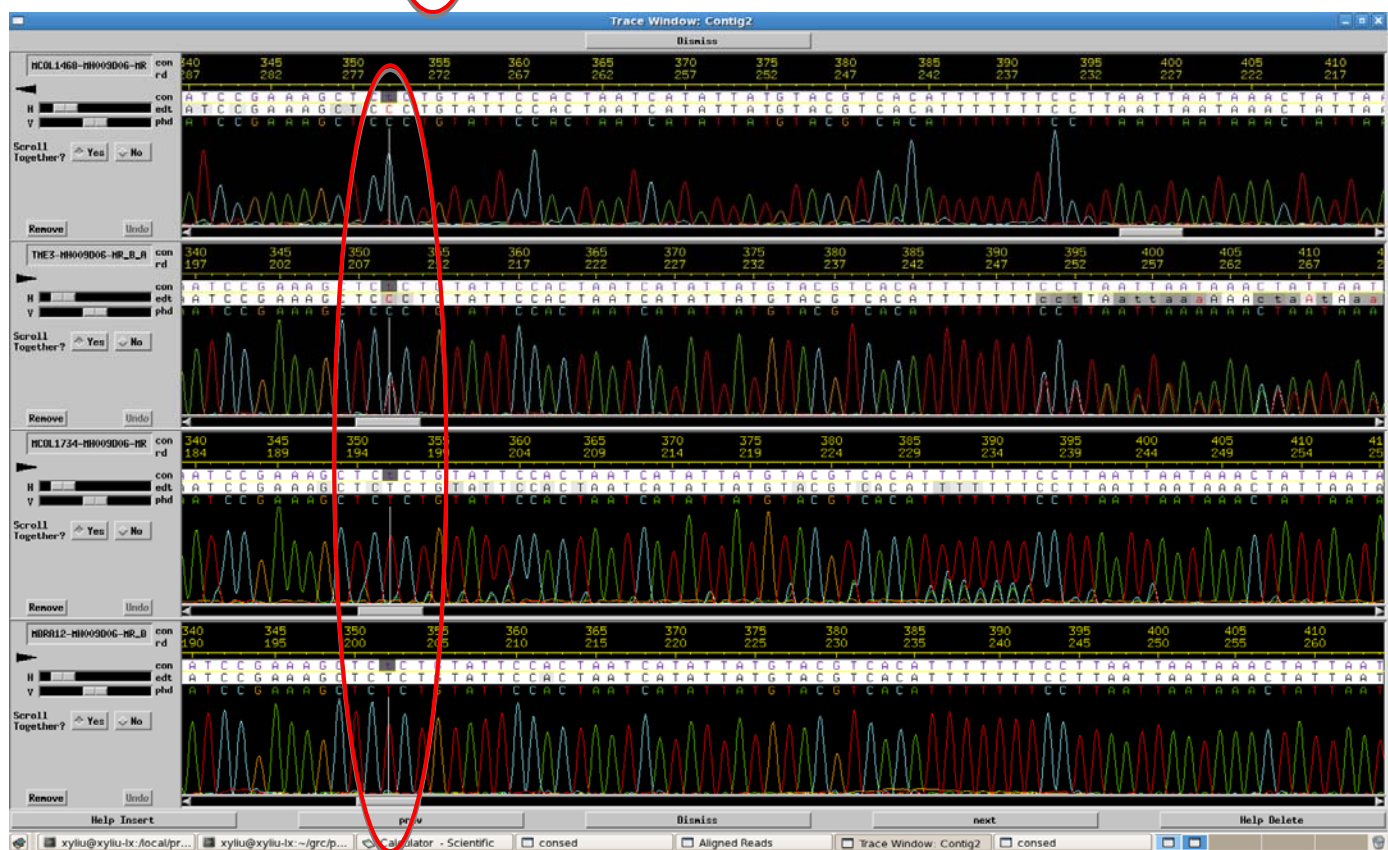
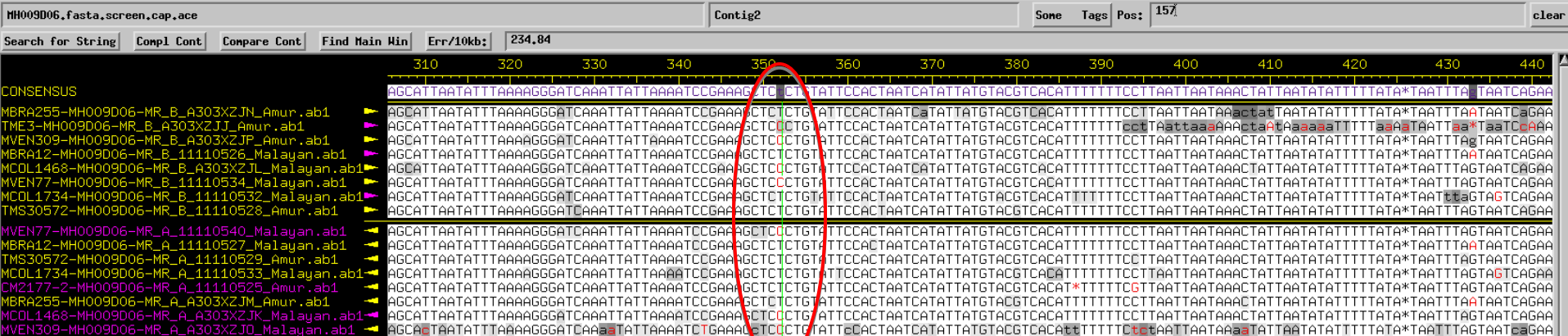
Library available from CUGI

SNP development

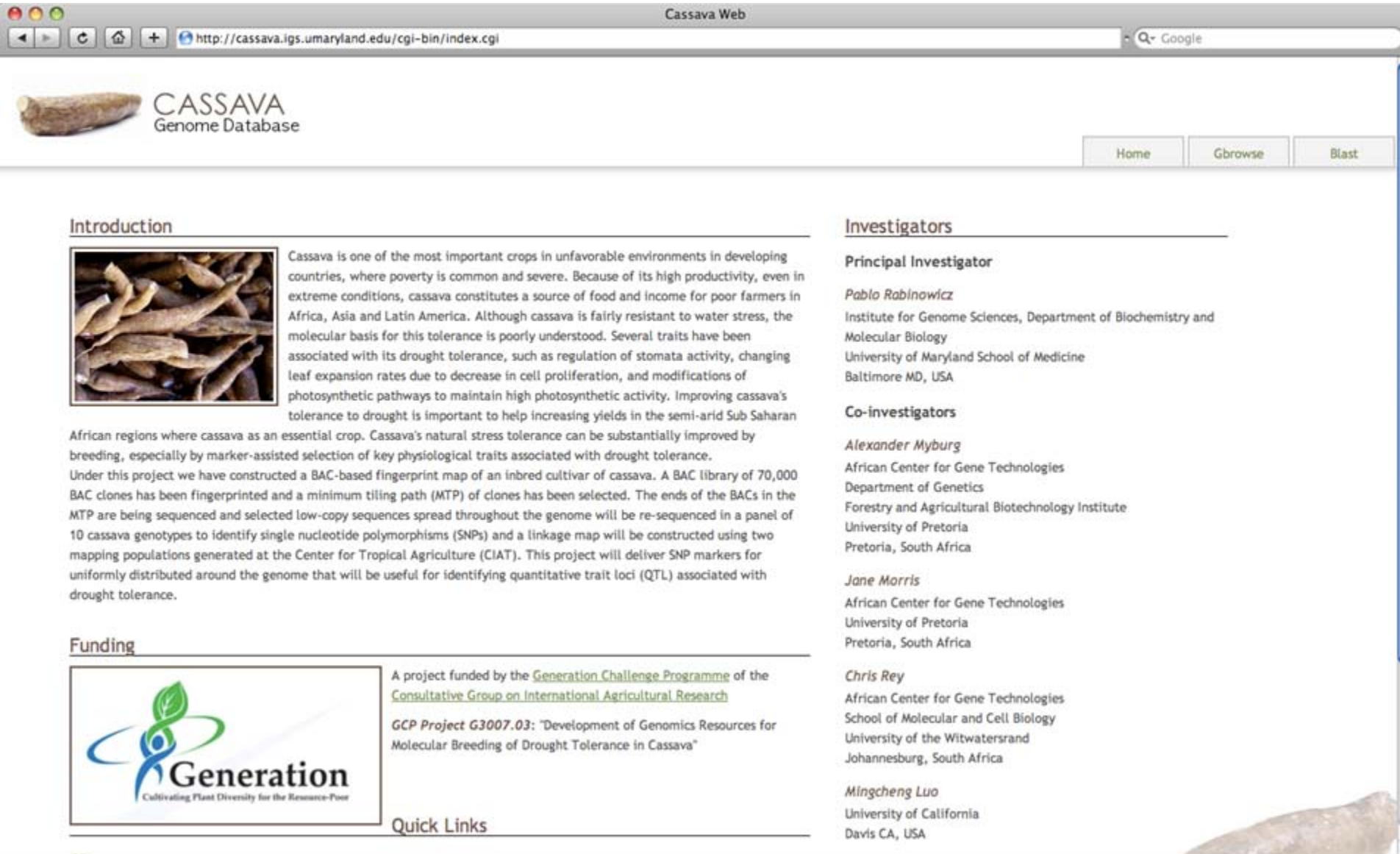


- 1,536 primer pairs synthesized (separated by at least 200 kb)
- 384 PCRs done in 10 genotypes
- Several SNPs per locus in preliminary results

SNP Discovery




Accessing the data: <http://cassava.igs.umaryland.edu>




The screenshot shows a web browser window titled "Cassava Web" with the URL <http://cassava.igs.umaryland.edu/cgi-bin/index.cgi>. The page features a header with the "CASSAVA Genome Database" logo and navigation buttons for "Home", "Browse", and "Blast". The main content is divided into three sections: "Introduction", "Funding", and "Investigators".

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"

Quick Links

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

Cassava BLAST server

BLAST Search

http://cassava.igs.umaryland.edu/blast/blast.html

Google



Home | Gbrowse | Blast

Choose program to use and database to search:

Program: blastn Database: Cassava ESTs

Enter sequence below in [FASTA](#) format

Or load it from disk no file selected

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: 10 Matrix: BLOSUM62 Perform ungapped alignment

Query Genetic Codes (blastx only): Standard (1)

Database Genetic Codes (tblast[nx] only): Standard (1)

Frame shift penalty for blastx: No OOF

Other advanced options:

Graphical Overview Alignment view: Pairwise

Descriptions: 100 Alignments: 50 Color schema: No color schema

*results will open in new window

Comments and suggestions to: < blast-help@ncbi.nlm.nih.gov >

Cassava ESTs
Assembled cassava ESTs
JGI's WGS sequences
BAC-ends

Search

Select a Contig to View

Landmark or Region:

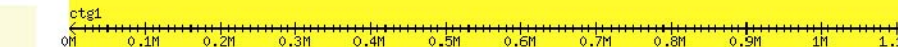
ctg1

Data Source

Cassava Database (Chromosome)

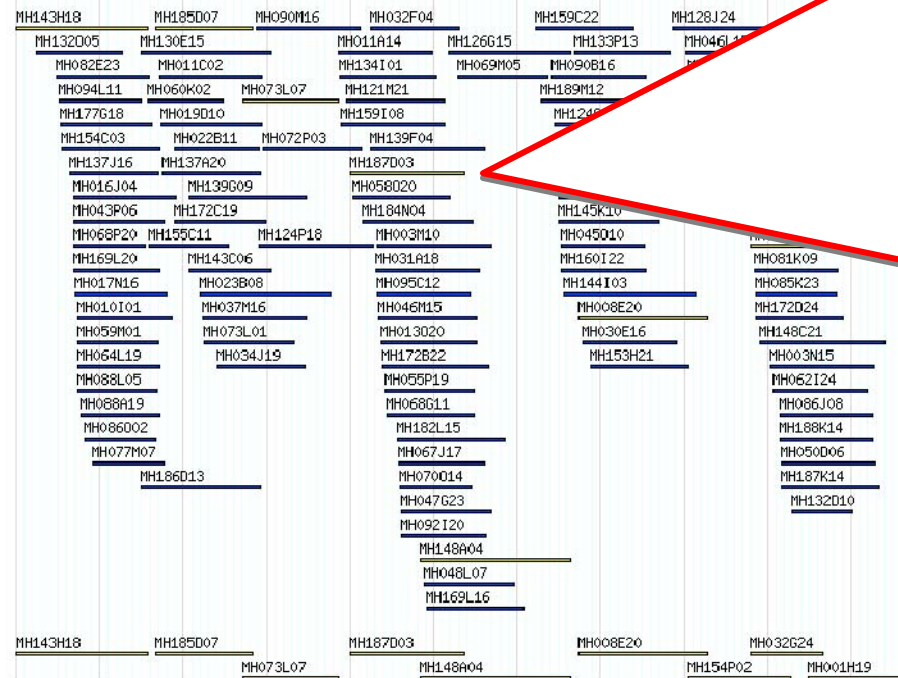
Scroll/Zoom: Show 1.286 Mbp

Overview

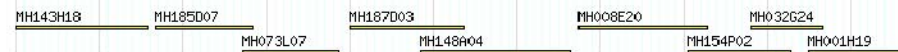


Details

BACs



Minimum Tiling Path



[Clear highlighting](#)



CASSAVA
Genome Database

Clone Info

MH187D03

Located on: [ctg1](#)
Coordinates: 400801 - 536401
Length: 135601

BAC End Sequence: Forward

```

TTCCAGCTCGGACCCCGGGATCCCTAGAGCTCGACCTGCAGCCATCCAACTACTCTTGC
AAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTACTGATATCAGGGATGTACTCTTGCAGAGGGTGTGTGTGTGTGTGTGTGTGTGTGT
ACTACTTTCAGAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
ATAGGGATGTACTCTTGCAGAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TGCAGAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GATGTAGGGATGTACTCTTGCAGAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TACTTGCAGAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

```

BAC End Sequence: Reverse

```

GCCGCTTTCGACTTGGAGCAGTGAACAGCCGCTTCTCGAGTGAATACAGCACTCA
CAATGTAGCAGCTGACGGACATCTG66TTTTGCGATGCGAGAGAACGCTACCTGC
CCAAATGCAATGTACTGACCTTCCAGAAAGTGTAGGCTTTTATGCACTAAGAGGGG
GACCACTCCATATGATGCCCATATTTTGGATGAGATTTCCAGAGAGGTGTGTGTGT
CATACCTTTCGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
CAATGACTTTCAGCAGCAGATGATGCTCCCTTCTGTAGCTTGTGTGTGTGTGTGTGT
CGGCTGAGCCCTTTCAGCTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
CTGAAATGCAAAATGCACTGATGAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
ATCTCCAGATGACTGATCCCTATGAGCAAACTTTCTGTGTGTGTGTGTGTGTGTGTGT
TAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TTATCTGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
AATATTAACATGATGCTGAGCTGAAAGAGCTGATATTAACATGATGATGATGATGATG
CTTAGAGAGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
CTACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GAGATATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
CTTTTGAAGAAATTAATA

```

Physical Mapping (raw FPC output file):

Approximate_match_to_cosmid "MH187D03"
Creation_date 108 10 10 12 52
Modified_date 108 10 10 17 7

Illumina BeadExpress Genotyping

- Already installed at Zander Myburg's lab, University of Pretoria (UP), African Centre for Gene Technologies (ACGT)
- DNAs from progeny of mapping populations kindly provided by CIAT.
- Ready to test SNP platform with SNP data kindly provided by Morag Ferguson
- Cost-recovery SNP genotyping service at UP (48-, 96- and 384-plex assays)



Capacity building

- A SNP genotyping workshop scheduled for November 9-13, 2009 at the University of Pretoria, South Africa
- Participants will bring DNA from ~200 cassava genotypes
- The workshop will assess SNP diversity in cassava
- Participants from
 - EMBRPA, Brazil
 - CSIR, Ghana
 - National Root Crop Research Institute, Nigeria
 - NaCRRI, Uganda
 - ACCI, South Africa

Acknowledgements

IGS-UMB, MD:

Pablo Rabinowicz
Kimm O'Brien
Kevin Galens

ACGT, South Africa:

Zander Myburg, University of Pretoria
Jane Morris, University of Pretoria
Chris Rey, University of the Witwatersrand

UC Davis, CA:

Mingcheng Luo
Yaqin Ma

CIAT, Cali, Colombia:

Hernan Ceballos
Emmanuel Okogbenin

Danforth Center, St. Louis, MO:

Martin Fregene

Funding: GCP-CGIAR

