

Overviews

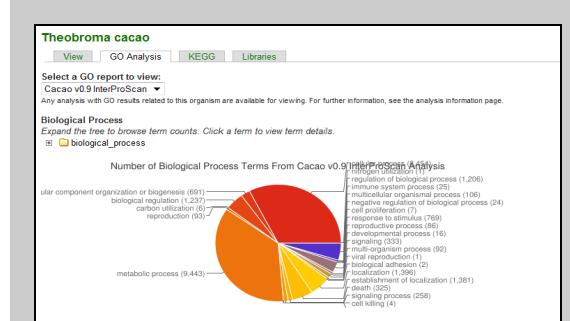
The Cacao Genome Database (CGD) provides a comprehensive data mining resource for sequence and annotations from the *Theobroma cacao* cv. Matina 1-6 cacao genome sequencing project. Intuitive interfaces have been developed to allow the users to access data and tools. This brochure introduces CGD and the tools and data available there.

Current Data

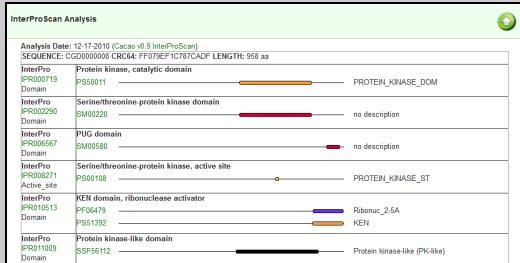
- Annotated *T. cacao* cv. Matina1-6 genome sequence (GBrowse)
- Nine genetic maps and one physical map (CMap)
- Annotated transcriptome
- SHRS database
- Community mailing lists
- Secure private portal for project communication

Resources

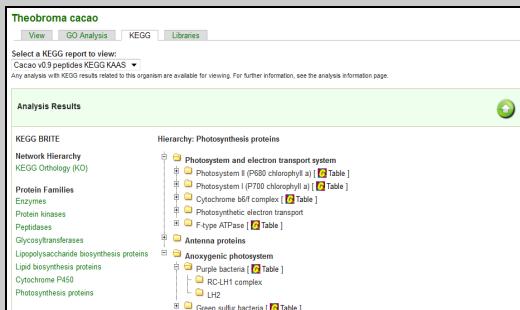
- BLAST sequences against the cacao genome v0.9 datasets
- View/search the cacao genome v0.9 in GBrowse
- View/search the cacao predicted gene KEGG and GO reports.
- Synteny between *T. cacao* cv. Criollo vs *T. cacao* cv. Matina1-6



Search and view GO terms



Search and view protein domains



Search and view KEGG pathways

Contact Us

Main Bioinformatics Laboratory
Washington State University
Pullman, WA99164, USA
dorrie@wsu.edu
ping_zheng@wsu.edu



The Cacao Genome Database

The Genomics, Genetics and Breeding Resource for Cacao Improvement

A collaboration among MARS, USDA-ARS, IBM, NCGR, Clemson University, HudsonAlpha Institute for Biotechnology, Indiana University and Washington State University



www.cacaogenomedb.org

Funded by

USDA-ARS

Acknowledgements

Thanks to our project collaborators who provide data and input for development of this database. We also wish to thank Washington State University for support.

Tools

Search Tools

CGD's search tools allow users to retrieve and analyze bulk data.

Cacao Genome Database

Logged in as: p_public | Log Out

Home About Groups Tools Genome Sequence Databases Resources Contact Mailing List Search

Search Analysis

Gene Homology Analysis InterProScan Domain Sequence Gene Ontology

Search database

- search Genes
- search Homology Analysis
- search InterProScan Domains
- search Sequences
- search GO Terms

MARS Incorporated USDA Agricultural Research Service TBM CUGI HudsonAlpha Institute for Biotechnology NCGR PIPRA Indiana University Washington State University

download the results

ID	Genome Location	Match	Sequence	Database	Description
CGD0000007	super_1:149017-156371	MLO1_ARATH	view	Swiss-Prot	MLO-like protein 10 OS=Arabidop..
CGD0000007	super_1:149017-156371	MLO8_ARATH	view	Swiss-Prot	MLO-like protein 8 OS=Arabidop..
CGD0000007	super_1:149017-156371	MLO7_ARATH	view	Swiss-Prot	MLO-like protein 7 OS=Arabidop..
CGD0000007	super_1:149017-156371	MLO5_ARATH	view	Swiss-Prot	MLO-like protein 5 OS=Arabidop..
CGD0000007	super_1:149017-156371	MLO9_ARATH	view	Swiss-Prot	MLO-like protein 9 OS=Arabidop..

SSR Server

Identify microsatellites and primers in uploaded sequence data sets. Results returned in excel files.

SSR Job Information: Tool Description

Job Title

Email Address Browse...

Upload Sequence File

Specify Motif Frequency
Motifs that occur less frequently than indicated below will be ignored

Mononucleotides 0

Dinucleotides 5

Trinucleotides 4

Tetranucleotides 3

Pentanucleotides 3

Hexanucleotides 0

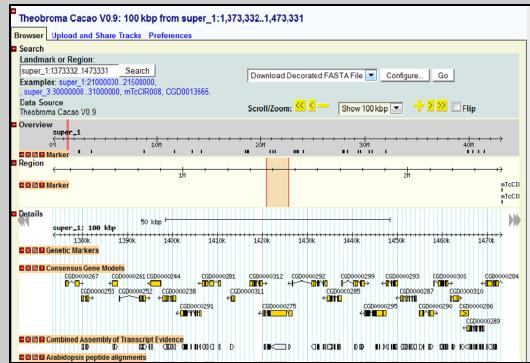
Generate Primers

Submit

Tools (continued)

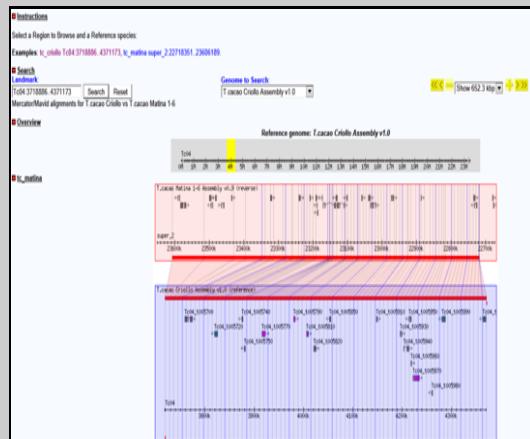
Cacao GBrowse

Use GBrowse to view gene models, transcripts, the results of genome-wide SNP association studies, and linkage studies.



Cacao Synteny Browser

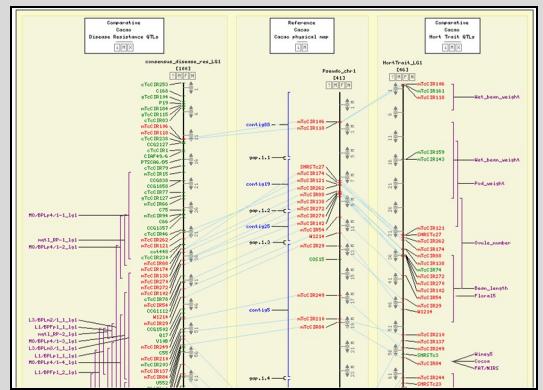
Use GBrowse_syn to display multiple genomes, with a central reference species compared to two or more additional species.



Tools (continued)

Comparative Map Viewer

View and compare genetic maps



BATCH BLAST

CGD BLAST tool allows users to conduct homology searches between their sequences and available datasets.

BATCH BLAST SERVER

Title of Search

BLAST type blastn: DNA query vs DNA db

Email Address

Database cacao_dbest_all_15932.fasta

FASTA Sequence File Browse...

Submit

Cacao Blast Tool BLAST Results

1 Top-level alignment of my-cp-brefrgtblast

2 Database size: about 15,932,000

3 Query File: demo_blast_10.fasta

4 Tue Jan 5 15:46:11 2010

5

6

7 Top 10 Matches Per Query

Query	Query Length	Match	Description	Orga	Exp.	Score	PID	Align Length	Query Start	Query Stop	Match Start	Match Stop
9 E5441179 TS	237	E5441179	The C5441179	The	1e-121	454	100%	237	1	237	1	237
10 E5441179 TS	237	E5427284	The C5427284	The	0.032	38.8	96%	25	213	237	6	29
11 E5441178 TS	454	E5441178	The E5441178	The	0.0	831	100%	454	1	454	1	454
12 E5441178 TS	454	E5457827	The C5457827	The	0.0	826	99%	454	1	454	52	505
13 E5441178 TS	454	C546360	The C546360	The	0.0	813	99%	454	1	454	44	497
14 E5441178 TS	454	C5474567	The C5474567	The	0.0	804	98%	454	1	454	76	78
15 E5441178 TS	454	C5452470	The C5452470	The	0.0	799	98%	454	1	454	171	624
16 E5441178 TS	454	C5452470	The C5452470	The	0.0	781	98%	454	1	454	171	624
17 E5441178 TS	454	C5452470	The C5452470	The	0.0	781	98%	454	1	454	171	624
18 E5441178 TS	454	C5452470	The C5452470	The	0.0	781	98%	454	1	454	171	624