

Global Conference on

“Technological Challenges and Human Resources for Climate Smart Horticulture”

NAVSARI AGRICULTURAL UNIVERSITY, NAVSARI .

28<sup>th</sup> -31<sup>st</sup> May, 2014

## Application of Bioinformatics in Climate Smart Horticulture

**Dr. Hetalkumar Panchal** (Associate Professor)

**College of Agricultural Biotechnology**

Navsari Agricultural University,  
Ghod Dod Road, Surat-395007.

Email: [swamihetal@nau.in](mailto:swamihetal@nau.in)

Ph: 9427457245

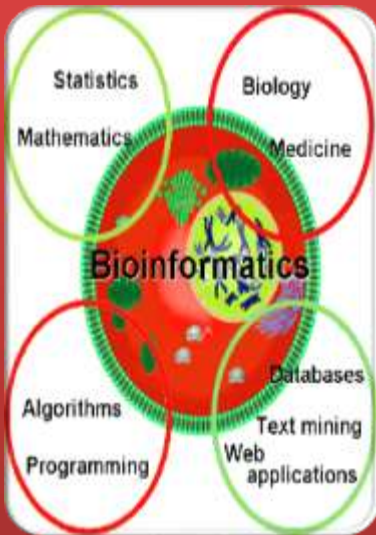
Introduction

Bioinformatics recourses for horticulture crops

Case studies

Horticulture is one of the important sectors of agriculture which consists of fruits, flowers, vegetables, spices & condiments, plantation crops, medicinal and aromatic plants. The importance of horticultural crops is widely acknowledged in many aspects of innovation, production, quality maintenance, for uplifting economic condition of farmers, entrepreneurs and in providing nutritional security to the people.

With the increasing fast growth of population, improving economic condition and awareness of importance of fruits, vegetables and flowers, demand for horticultural products is gradually increasing. To meet the demand of horticulture product genomic technologies can be used effectively in horticultural crop improvement programmes. Improvement in horticulture crop is also required to sustain in the climate change condition and make the Climate Smart Horticulture crop.



Bioinformatics is a combination of computer, mathematics algorithm and statistics with concept in life science to solve biological problem. The main task of bioinformatics is to manage and analyse the biological data. Bioinformatics has a number of applications in plant biology. Computational tools are very much helpful in identification of ergonomically important gene by comparative analysis between crop plant and model species.



Bioinformatics is a tool that is used to analyse the Biological data. We can compare Bioinformatics with Statistics. As statistics is a tool to analyse data by different formula and derive certain conclusions. (we can use either our own data or data generated by others)

Data Mining (someone's data is now mine)

# Role of Bioinformatics in Horticultural Crops

Crop  
Improvement

```
graph TD; A[Crop Improvement] --- B[Plant Breeding]; A --- C[Improve Nutritional Quality]; A --- D[Development of Stress Tolerant Varieties]
```

Plant Breeding

Improve  
Nutritional  
Quality

Development of  
Stress Tolerant  
Varieties

# CROP IMPROVEMENT

Omics-based research and application development provide crucial resources to promote research in model and applied plant species. A combinatorial approach using multiple omics platforms and integration of their outcomes is now an effective strategy for clarifying molecular systems integral to improving plant productivity.

Promotion of comparative genomics among model and applied plants allows us to grasp the biological properties of each species and to accelerate gene discovery and functional analyses of genes.

Bioinformatics platforms and their associated databases are also essential for the effective design of approaches making the best use of genomic resources, including resource integration.

Species-specific nucleotide sequence collections provide opportunities to identify the genomic aspects of phenotypic characters based on genome-wide comparative analyses and knowledge of model organisms.

# “Ten Important Bioinformatics Databases”

GenBank	<a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a>	nucleotide sequences
Ensembl	<a href="http://www.ensembl.org">www.ensembl.org</a>	human/mouse genome (and others)
PubMed	<a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a>	literature references
NR	<a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a>	protein sequences
SWISS-PROT	<a href="http://www.expasy.ch">www.expasy.ch</a>	protein sequences
InterPro	<a href="http://www.ebi.ac.uk">www.ebi.ac.uk</a>	protein domains
OMIM	<a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a>	genetic diseases
Enzymes	<a href="http://www.chem.qmul.ac.uk">www.chem.qmul.ac.uk</a>	enzymes
PDB	<a href="http://www.rcsb.org/pdb/">www.rcsb.org/pdb/</a>	protein structures
KEGG	<a href="http://www.genome.ad.jp">www.genome.ad.jp</a>	metabolic pathways



Source: *Bioinformatics for Dummies*

# Status of genome sequencing projects in horticultural crops.

Crop	Size (Mb)	Chr.	Status	Centre
<i>Allium cepa</i>	15,000	8	Initiated	USDA-ARS and University of Wisconsin
<i>Aquilegia coerulea</i>	350	-	In progress	Joint Genome Institute(JGI)
<i>Aquilegia formosa</i>	350	7	In progress	DOE Joint Genome Institute
<i>Beta vulgaris</i>	760	9	Initiated	Hokkaido University, Japan, Max Planck Institute, Michigan State University,
<i>Brassica juncea</i>	-	18	Initiated	Multinational Brassica Genome Project
<i>Brassica oleracea</i>	600	9	In progress	TIGR
<i>Brassica rapa</i>	500	10	In progress	<i>Brassica</i> IGF Project, The Multinational <i>Brassica rapa</i> Sequencing Project
<i>Capsicum annuum</i>	3000	12	Initiated	Korea Research Institute of Bioscience and Biotechnology, Seoul National University, Cornell University
<i>Carica papaya</i>	370	9	Draft assembly	The Papaya Genome Sequencing Consortium
<i>Carya illinoensis</i>	-	16	Initiated	University of Georgia, USA
<i>Citrus sinensis</i>	380	9	In progress	U.S. Citrus Genomics Initiative, University of California
<i>Coffea arabica</i>	-	11	Initiated	University of Central Florida
<i>Corylus avellana</i>	-	11	Initiated	Department of Horticulture
<i>Cucumis sativus</i>	370	7	Draft assembly	The Cucumber Genome Initiative
<i>Daucus carota</i>	470	9	Initiated	University of Central Florida, Vegetable Crops Research Unit
<i>Kadua centranthoides</i>	-	-	Initiated	Lita Annenberg Hazen Genome Sequencing Center
<i>Lactuca sativa</i>	-	9	Initiated	Research Institute of Innovative Technology for the Earth, Japan, University of California, Davis (UCD)
<i>Malus × domestica</i>	750	17	In progress	IASMA research center, The Horticulture and Food Research Institute of New Zealand
<i>Musa acuminata</i>	600	11	In progress	Global Musa Genomics Consortium (GMGC)
<i>Prunus armeniaca</i>	-	8	Initiated	Instituto Valenciano de Investigaciones Agrarias, INRA Avignon
<i>Prunus dulcis</i>	-	8	Initiated	Beijing Agro-biotechnology Research Center at CUGI
<i>Prunus persica</i>	290	8	Initiated	Clemson University, CUGI
<i>Rosa multiflora</i>	-	7	Initiated	Rosaceae International Genomics Initiative (RosIGI)
<i>Solanum bulbocastanum</i>	-	12	In progress	University of Central Florida, Molecular Biology and Microbiology, USA, Orlando, TIGR
<i>Solanum demissum</i>	-	12	In progress	TIGR
<i>Solanum lycopersicum</i>	950	12	Complete	International Tomato Genome Sequencing Project; solgenomics.net
<i>Solanum melongena</i>	1100	12	Initiated	Cornell University
<i>Solanum tuberosum</i>	840	12	Draft assembly	Potato Genome Sequencing Consortium, TIGR, The Canadian Potato Genome Project
<i>Vitis vinifera</i>	500	19	Draft assembly	International Grape Genome Program, IASMA Research Center



# PLANT BREEDING

Plant genomics helps in understanding the genetic and molecular basis of all biological process which helps in developing new cultivars with improved quality and reduced economic and environmental cost.

Now-a-days the Genome program is an important tool for the plant improvement. This genome programme helps in identifying the key genes and their function. This genome project generates data, which includes sequence information, markers etc. These data are then distributed to the multinational research community.

The bioinformatics tools helps in the submission of all data through ENTREZ Global Query Cross-Database Search System to the public domain. This helps in retrieving sequence from the NCBI. The bioinformatics tools helps in providing rational annotation of genes, proteins and phenotypes.

Interspecific and intergeneric hybridization is an important research approach creating usable variability for breeding. **Desirable traits like resistance against pathogens or abiotic stress** within varying environments, composition of metabolites or morphological traits and their responsible genes are often found only within wild species, related species and genera of the cultivars. It is possible to use the existing variability of the primary and secondary gene pool by interspecific and intergeneric hybridization.

Interspecific and intergeneric hybridization can be carried out both-way, generatively by crossing or somatically by protoplast fusion. The Institute for Breeding Research on Horticultural Crops has acquired wide experience in the development and adaption of the generative as well as somatic hybridization of vegetables, ornamentals and medical and aromatic plants for many years. Both methods are very labour-intensive, long-lasting and fraught with risk. Therefore, the breeding lines are chosen very carefully and are proved for defined valuable traits in advance.

# Examples of interspecific or intergenetic transferability of genic molecular markers in horticultural crops

Crops	Species used for validation	Markers used	References
Tomato ( <i>Solanum lycopersicum</i> )	<i>Solanaceous</i> members	EST-SSR, SNP, Gene specific	Frary et al. (2005)
Grape ( <i>Vitis vinifera</i> )	<i>Vitaceae</i> and <i>Rosaceae</i> species	EST-SSR, SNP	Scott et al., 2000; Arnold et al., 2002; Decroocq et al., 2003; Vezzulli et al., 2008b; Rossetto et al., 2002
Apricot ( <i>Prunus ameniaca</i> )	<i>Vitaceae</i> and <i>Rosaceae</i> species	EST-SSR	Decroocq et al. (2003)
Strawberry ( <i>Fragaria vesca</i> )	<i>F. gracilis</i> , <i>F. nilgerrensis</i> , <i>F. nipponica</i> , <i>F. iinumae</i>	EST-SSR	Bassil et al. (2006)
Citrus ( <i>Citrus sinensis</i> )	<i>Poncirus trifoliata</i>	EST-SSR	Chen et al. (2006)
Coffee ( <i>Coffea</i> spp.)	16 species of coffee and <i>Psilanthus</i>	EST-SSR	Bhat et al., 2005; Poncet et al., 2006; Aggarwal et al., 2007
Sweet cherry ( <i>Prunus avium</i> L.)	<i>Prunus</i> sps	SSR, CAPs	Olmstead et al. (2008)
White Campion ( <i>Silene latifolia</i> )	<i>Silene</i> species	EST-SSR	Moccia et al. (2009)
Litchi ( <i>Litchi chinensis</i> )	16 Ackee trees, Pulasan ( <i>Nephelium ramboutan-ake</i> L.)	SSR	Sim et al., 2005; Ekué et al., 2009
Iris	26 Louisiana <i>Iris</i> species ( <i>I. brevicaulis</i> , <i>I. fulva</i> , <i>I. hexagona</i> , and <i>I. nelsonii</i> ) and 13 ( <i>I. germanica</i> ), yellow-flag ( <i>I. pseudacorus</i> ), and Siberian ( <i>I. sibirica</i> ) <i>Iris</i>	EST-SSR	Tang et al. (2009)
Brassica	<i>Brassica</i> sps	SSR	Yadava et al., 2009; Bhati et al., in press
<i>Lactuca sativa</i> L.	96 accessions representing all major horticultural types and 3 wild species ( <i>L. serriola</i> , <i>Lactuca saligna</i> , and <i>Lactuca virosa</i> )	EST-SSR	Simko (2009)
Potato ( <i>Solanum tuberosum</i> )	65 <i>Solanum tuberosum</i> lines and 14 other species	SSR	Grover et al. (2009)
Sunflower	Safflower, <i>Compositae</i> sps	SSR, InDel, gene based markers	Heesacker et al., 2008; García-Moreno et al., 2010
<i>Cucumis melo</i> L.	<i>C. melo</i> , <i>Citrullus lanatus</i> , <i>Cucurbita maxima</i> , <i>C. moschata</i> , <i>C. pepo</i> and <i>Cucumis sativus</i> , Sps	SSR	Ritschel et al. (2004)
<i>Vicia faba</i> (Fabaceae)	Faba beans, <i>Pisum sativum</i>	EST-SSR	Gong et al. (2010)

# IMPROVE NUTRITIONAL QUALITY

There are various ways of improving nutrition quality. It can be improved by the redirection of the cellular activity, by the modification of the enzymatic transport and by regulating the function of the cell.

Various tools are available to identify the genes. With the advances in the proteomics and glycomics, there are various tools for the analysis of primary and secondary metabolic pathways.

The current advancement in plant biology research encompassing: generation of huge amount of molecular-genetic data, development of impressive methodological skills in molecular biology experimentation, and systems analyses, has set the stage to search for ways/means to utilize the available resources to strengthen interdisciplinary efforts to find solutions to the challenging goals of plant breeding efforts (such as abiotic stress tolerance) ultimately leading to gainful applications in crop improvement.

A positive fall out of such a realization and efforts has been the identification/development of a new class of very useful DNA markers called genic molecular markers (GMMs) utilizing the ever-increasing archives of gene sequence information being accumulated under the EST sequencing projects on a large number of plant species in the recent years. These markers being part of the cDNA/EST-sequences, are expected to represent the functional component of the genome i.e., gene(s), in contrast to all other random DNAbased markers (RDMs) that are developed/generated from the anonymous genomic DNA sequences/domains irrespective of their genic content/information.

# QTL cloned in important horticultural crops

Crops	QTL/gene function	Molecular identification	Candidate gene	Method used	References
Tomato	Fruit shape (Ovate)	Unknown	Positional cloning	Transformation	Liu et al. (2002)
	Fruit sugar content (Brix9-2-5/Lin5)	Invertase	Positional cloning	Complementation	Fridman et al. (2002, 2004)
	Fruit weight (fw2.2)	Unknown	Positional cloning	Transformation	Frary et al., 2000; Cong et al., 2002
	Sw4.1 QTL	ABC transporter gene	Positional Cloning	RNAi Transformation	Orsi and Tanksley (2009)
Potato	Resistance to Ro1 <i>Globodera ostochiensis</i> ( <i>Gro1-4</i> )	Unknown	Positional cloning	-	Paal et al. (2004)
	Quality trait cold-sweetening (invGE/GF)	Invertase	Positional cloning	-	Li et al. (2005)
	Flavonoid 3,5-hydroxylase ( <i>f3_5_h</i> )	-	Positional cloning	Transformation	Jung et al. (2005)
Apple	Scab resistance gene <i>VF</i>	Leucine rich repeat domain Transmembrane domain	Positional cloning	-	Xu and Korban (2003)
Sugarbeet	Nematode resistance <i>HS<sup>1pro-1</sup></i>	Leucine rich repeat domain	Positional cloning	-	Cai et al. (1997)
Pepper	Virus resistance <i>e1F4E</i> gene	Leucine rich repeat	Candidate gene	Complementation	Ruffel et al. (2002)
Cauliflower	Orange gene ( <i>Or</i> )	Dna J Cysteine Rich domain	Positional Cloning	Complementation	Lu et al. (2006)
Papaya	<i>CpCYC-b</i>	Lycopene $\beta$ cyclase	Candidate gene	Complementation	Blas et al. (2010)

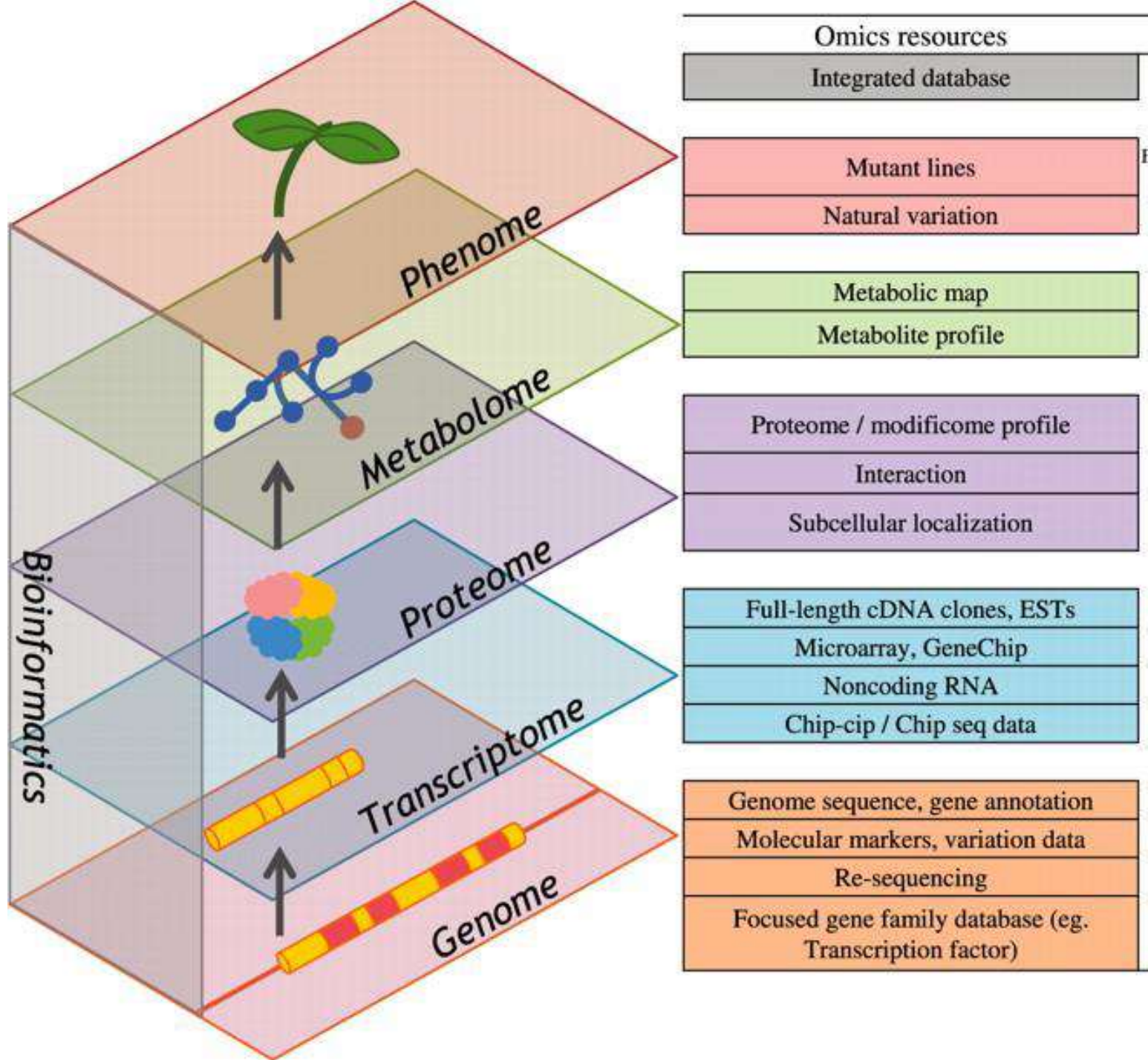
# DEVELOPMENT OF STRESS TOLERANT VARIETIES

Stress tolerant varieties can be developed by identifying the Stress tolerance genes and alleles. Various tools have been developed to study the physiology, expression profiling, comparative genomics.

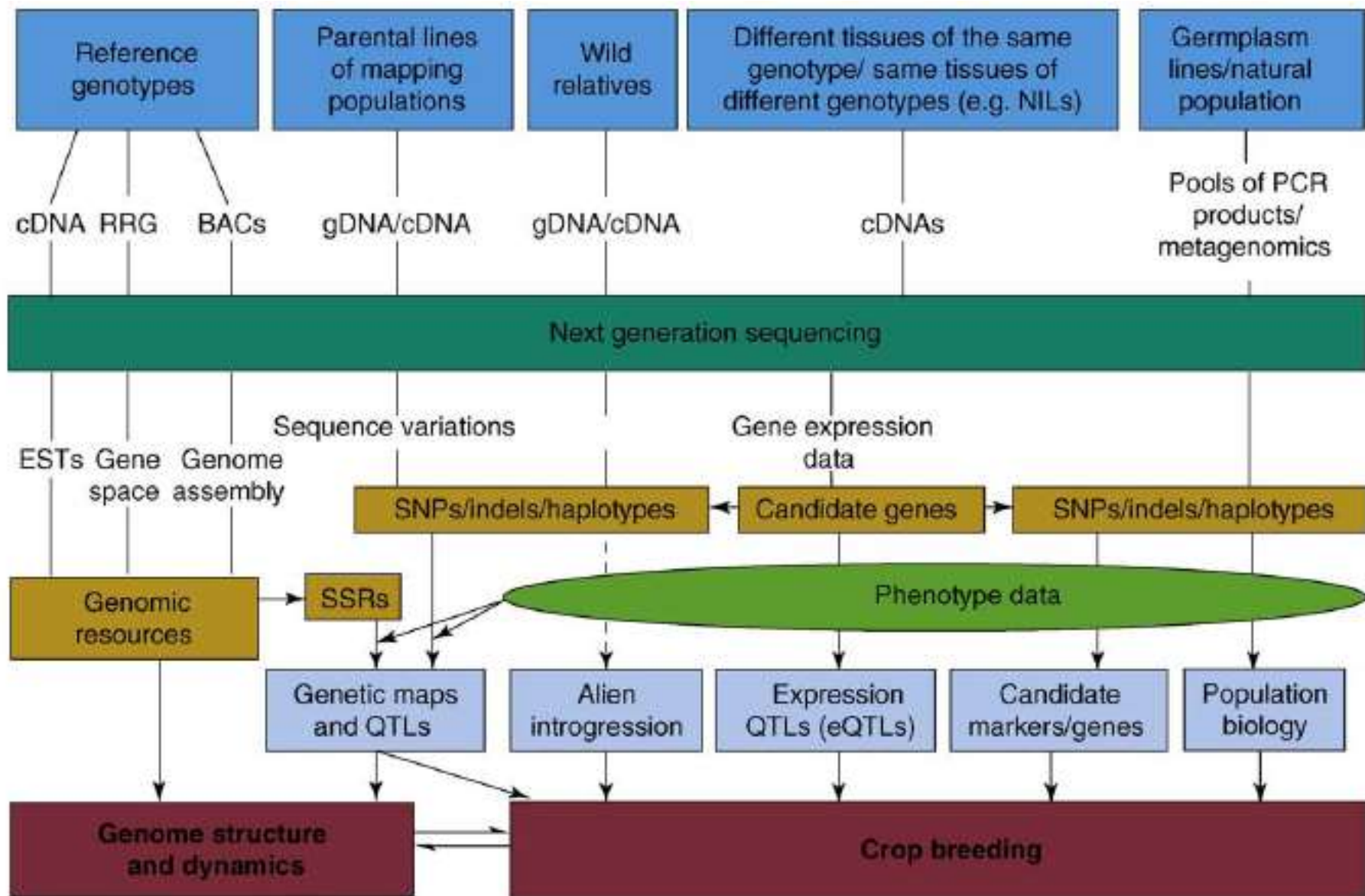
The KEGG database contains all the metabolic pathways like the pathway for the carbohydrate production. Genes in the ABA production pathway are important for the development of drought resistant varieties.

KEGG databases can be important in identifying the pathway for carbohydrate production and ABA production. After the identification of the pathway the genes involved in the same pathway are studied for the development.











## SGN data overview

This page presents a summary of all currently available data on SGN sorted by species. Click a species name for details about available its available maps, loci, libraries, ESTs, metabolic annotation and phenotyped accessions.

### Solanaceae

[click on an organism name to see more details](#)

[show in tree browser](#)

### Rubiaceae

[click on an organism name to see more details](#)

[show in tree browser](#)

### Plantaginaceae

[click on an organism name to see more details](#)

[show in tree browser](#)

<http://solgenomics.net/organism/all/view>



[search](#)

[maps](#)

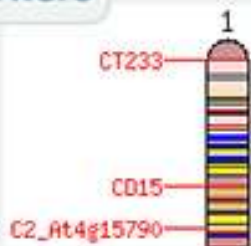
[genomes](#)

[tools](#)

[sol search](#)

[log in](#) | [new user](#)

## Maps & Markers



## Genes



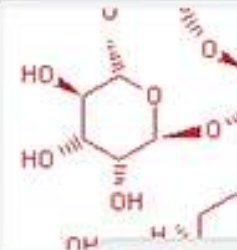
## Phenotypes



## Breeders Toolbox



## Genomes & Sequences



## Pathways

<http://www.sgn.cornell.edu/>

## PlantsDB databases:



## Projects:



[The Triticeae genome project](#)



[transPLANT](#)



[CrowsNest - Comparative Map Viewer](#)

<http://mips.helmholtz-muenchen.de/plant/genomes.jsp>



## PlantGDB

Home

## Sequence

Download  
Search  
EST Assemblies  
FTP Server  
More...

## Genomes

Genome Browsers  
Annotation

## Tools

BioExtract   
BLAST  
GeneSeqer  
More...

## Datasets

AcDs Tagging  
ASIP  
SRGD  
More...

## Outreach

PGROP   
Native American

## Help

Help & Tutorials

## Welcome to plantgdb.org! Tools and resources for plant genomics

▶ [Quick tips for using PlantGDB...](#)

### Public Plant Sequence NEW Release 187 is current.



Species-parsed [Viridiplantae sequences](#) from GenBank and UniProt.

▶ [More about Release 187...](#)

### Sequence Assemblies



PlantGDB-assembled unique transcripts ([PUT](#)) and Genome Survey Sequence assemblies ([GSS](#)).

▶ [More about Assemblies...](#)

### Tools



[Online tools](#) for pattern matching, sequence query, and spliced alignment.

▶ [More about Tools](#)

### Genome Browsers



[Genome browsers](#) for emerging & completed plant genomes.

[Video Tutorial \(6 min\)](#) [View Quicktime](#)

[View Flash](#)

▶ [More about Genome Browsers...](#)

### Community Annotation



Tools and databases for [gene structure annotation](#).

[Video Tutorial \(5 min\)](#) [View Quicktime](#)

[View Flash](#)

▶ [More about Community Annotation...](#)

### Special Datasets



[Special datasets](#) for transposon insertion sites, splicing-related genes, and more.

## New & Noteworthy

Click below or [view all news](#) |

[Twitter](#)

▶ [New Location for PlantGDB](#)  
(July 23, 2012)

▶ [BrGDB - Brassica rapa chromosome-based genome browser](#)  
(Mar. 16)

▶ [StGDB - Solanum tuberosum new genome browser](#)  
(Mar. 16)

▶ [BrGDB - Brassica rapa chromosome-based genome browser](#)  
(Mar. 16)

▶ [VcGDB - Volvox carteri new genome browser](#)  
(Mar. 14)

▶ [Medicago genome updated](#)  
(Feb 27)

▶ [Rice genome updated](#)  
(Feb 27)

▶ [Cassava genome updated](#)  
(Feb 27)

▶ [Populus annotation updated](#)  
(Feb 27)

▶ [Mimulus annotation updated](#)  
(Feb 27)

▶ [Chlamydomonas annotation updated](#)  
(Feb 27)

<http://www.plantgdb.org/>

- Tools
  - Alignment Tools
  - BRENDA
  - BioMart
  - CpGAT
  - Edit Tools
  - Emboss
  - Information Tools
  - iPlant
  - Kegg Search
  - Nucleic Tools
  - Phylogeny Tools
  - Protein Tools
  - Sabio - Rk
  - Similarity Search Tools



<https://www.bioextract.org/clean/index.jsp>

# The Floral Genome Project



[Home](#) [Methods](#) [Taxa](#) [Resources](#) [Publications](#) [Contacts](#) [Links](#) [Tools!](#)

**PlantTribes - ESTstat: Web-based software for statistical analysis of EST library samples**

[Home](#)  
[Software Download](#)  
[Documentation](#)  
[Run](#)

## CAP3 Files

Singleton  No file chosen

[What is this?](#)

Ace  No file chosen

[What is this?](#)

Direction  5'  3'

ISO Error

[What is this?](#)

## Gene Cluster Profile

n file  No file chosen

[What is this?](#)

<http://fgp.bio.psu.edu/cgi-bin/eststat/eststat.cgi?action=form>



## RESEARCH

## RESOURCES

## INFORMATION

HOME > information > about MBPG

### about the MBGP

#### Objectives

The Multinational Brassica Genome Project was established in 2002 by a research community who were at that time involved in developing a [genome](#) available [here](#), together with the subsequent [Concept note](#) [for Brassica](#) which is periodically updated

#### Achievements

The MBGP has been successful in ensuring that a range of valuable public domain. These now include:

- Illumina Infinium 50k SNP array
- 2011: *Brassica rapa* genome published in *Nature Genetics*
- 2010: Affymetrix GeneChip *Brassica Exon 1.0 ST Array*
- 2008: Standardisation of nomenclature (e.g. [linkage groups](#))
- Genetic maps and populations
- Collation of public domain SSR markers
- ESTs in public repositories
- BAC libraries and physical (BAC) contig maps
- Analysis of *B. oleracea* shotgun sequences
- MBGP "White Paper"
- [www.brassica.info](http://www.brassica.info) available since 2002 (see the [original page](#) at the Wayback Machine)

#### Future challenges

- Consistent genome annotation; information registries and data exchange standards
- Integration of genetic and genomic maps: from QTL to gene
- Standardised screening of [genetic diversity sets](#)

#### Organisation

The MBGP arose through the recognition of the benefits of international collaboration and co-ordination, especially with respect to development and dissemination of experimental resources and information. The [MBGP Steering Committee](#) [has](#) an agreed defined remit, with clear set of responsibilities. Several sub-committees exist to manage specific tasks. These include the *Brassica rapa* Genome Sequencing Project (BrGSP) [Management Committee](#) [has](#), the MBGP TILLING consortium, and others focused on development of a common bioinformatics framework. In Jan 2010, it was agreed that the MBGP SC and BrGSP would merge. The current Chair is [Rod Snowdon](#).

#### Meetings

The MBGP Steering Committee aims to meet at least once a year. This usually includes a main meeting at the annual [Plant Animal Genome](#) conference in San Diego, held in January. Minutes of meetings are available [here](#).

[clone libraries](#)

[DNA sequences](#)

[databases](#)

[genetic maps](#)

[genetic markers](#)

[genome](#)

[genomics](#)

[metabolomics](#)

[plants](#)

[proteomics](#)

[sequencing](#)

[SNPs](#)

[TILLING](#)

[transcriptomics](#)

[traits & QTL](#)

[AtEnsembl with Brassica tracks](#)

[BRAD - Brassica Database](#)

[BrassEnsembl](#)

[BrassicaDB](#)

[CropStore](#)

[physical mapping \(UK\)](#)

[TAGdb](#)

### information menu

#### brassica crop & end-uses

- [Canola production](#)
- [crop production stats](#)
- [fodder brassica - USA](#)
- [nutritional benefits](#)

#### education & outreach

- [Carleton College lab](#)

#### events / meetings

- [conferences](#)
- [MBGP meetings](#)

#### mailing list

#### news

#### organisations, networks

- [OREGIN](#)
- [UK-BRC](#)

#### posters

#### presentations from

#### conferences

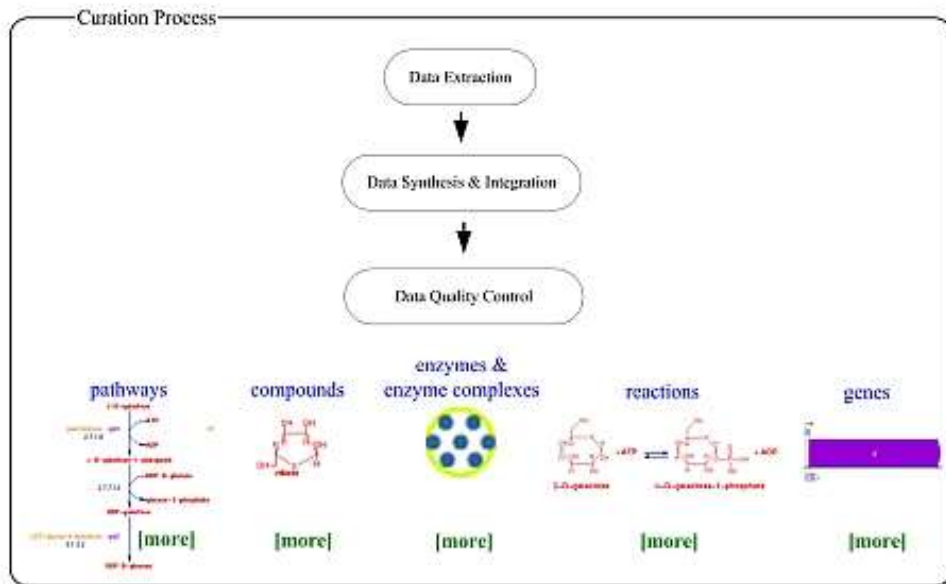
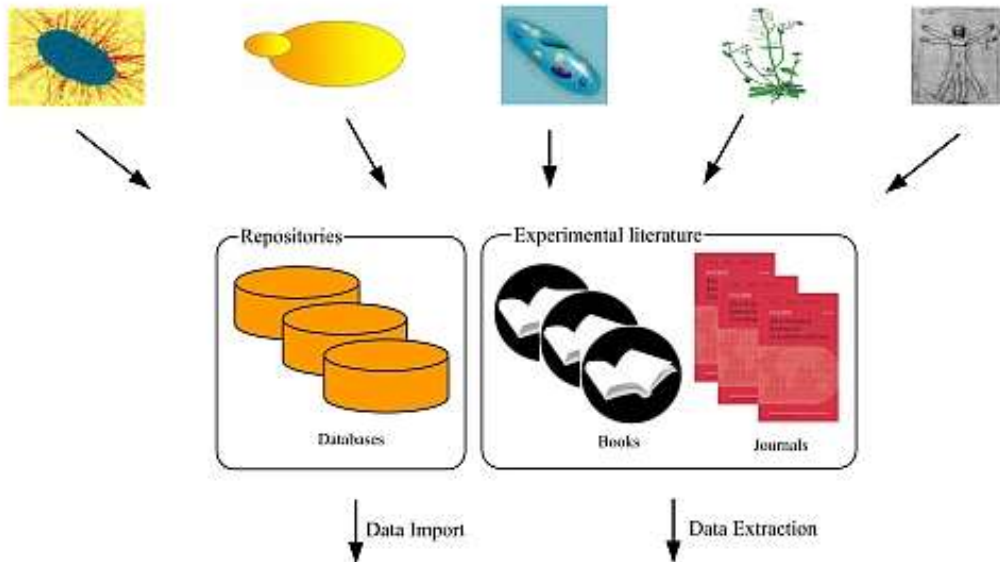
#### publications (to download)

- [Cruciferae Newsletter](#)
- [Cruciferae Trait Genetics](#)
- [Guide to Wild Germplasm](#)
- [White Paper](#)

#### reference information

- [gene nomenclature](#)
- [genome sizes](#)
- [image collection](#)
- [LG assignments](#)
- [MBGP nomenclature](#)
- [SOP](#)
- [trait descriptors](#)





Capture indications of *contradictory results* from published reports; present both viewpoints presented, with proper attribution.

**MetaCyc** <http://metacyc.org/MetaCycUserGuide.shtml>

## MetaCyc Overview

MetaCyc is a database of experimentally elucidated metabolic pathways from all domains of life. MetaCyc contains more than 2151 pathways from more than 2515 different organisms, and is curated from the scientific experimental literature.

MetaCyc contains pathways involved in both primary and secondary metabolism, as well as associated compounds, enzymes, and genes. The [goal](#) of MetaCyc is to catalog the universe of metabolism by storing a representative sample of each experimentally elucidated pathway.

MetaCyc [applications](#) include serving as an encyclopedia of metabolism, providing a reference data set for the computational prediction of metabolic pathways in sequenced organisms, supporting metabolic engineering, and helping to compare biochemical networks.

[Guide To MetaCyc](#)

### Accessing MetaCyc Data

MetaCyc data can be accessed in several ways:



### Run MetaCyc on Your Own Computer

Install SRI's Pathway Tools software locally to predict metabolic pathways from sequenced genomes, generate metabolic models, and analyze omics data.

[Learn More](#)

paused...

### How to Cite MetaCyc

Please cite MetaCyc as Caspi et al 2014.

<http://metacyc.org/>



# Spud DB

## Potato Genomics Resource

At Michigan State University

[Home](#)[Breeder's Assistant](#)[Genome Browser](#)[Search Tools](#)[Download](#)[Links](#)[Contact](#)

### Solanaceae Genomics Resource

March 28, 2014 - A [paper](#) describing the resources, tools and data available in Spud DB has been published in the journal *The Plant Genome*.

Spud DB is a collection of datasets and data-mining tools to view and analyze the potato genome, including tools to facilitate breeding improved cultivars.

### Potato Genomic Tools

Potato (*Solanum tuberosum* L.) is the fourth most important crop in the world, behind wheat, rice, and maize. In 2004, worldwide potato production exceeded 327 million metric tons. Potato also produces more food energy and value per unit of land area than any other crop. Compared to grain crops, potato is a superior source of nutrition; however, among major crop plants, potato is arguably the most intensively managed. In addition, tuber quality requirements are complex and must be maintained during harvest and storage. The genome sequence of potato was released and published in 2011 by the [International Potato Genome Sequencing Consortium](#) (PGSC). The assembled potato genome is that of the doubled monoploid *S. tuberosum* Group Phureja DM1-3 (DM; [AEWC00000000](#)); the genome assembly represents 726 MB, of which 86%, is anchored to the genetic map and contains 39,031 annotated genes. In addition, BAC and whole genome shotgun sequences were released from a second clone, *S. tuberosum* Group Tuberosum RH89-039-16 (RH; [ERP000627](#)).

### Quick Search

Gene ID:

Search terms:

### What's New

**April 4, 2014**

An update has been made to the Breeder's Assistant Tools. Searching for SNPs via gene annotation is no longer case sensitive, leading to more complete search results.

**March 28, 2014**

A [paper](#) describing the resources, tools and data available in Spud DB has been

<http://potato.plantbiology.msu.edu/index.shtml>

## Introduction

Brassica species represent important crops providing a major source of cooking oil, vegetables and condiments across many countries. The species relationship of cultivated Brassicas was described by the "triangle of U" with the three amphidiploid Brassica species *B. juncea* (AABB,  $2n=36$ ), *B. napus* (AACC,  $2n=38$ ) and *B. carinata* (BBCC,  $2n=34$ ) formed through interspecific hybridization between the diploid Brassica species, *B. rapa* (AA,  $2n=20$ ), *B. nigra* (BB,  $2n=16$ ) and *B. oleracea* (CC,  $2n=18$ ). Brassicas are closely related to the model species, *Arabidopsis thaliana*, for which the genome sequence was determined in 2000.

Brassica species include weedy types and cultivated varieties which demonstrate broad phenotypic plasticity. This is demonstrated by differences between root vegetable crops such as turnip, leafy forms such as Chinese cabbage, and oilseed varieties.

As well as the obvious economic reasons for studying Brassica genomes, with associated improvements in vegetable and oilseed crops, they provide also provide a model to study the effect of polyploidisation as well as extreme morphological diversity in plant species.

## Plant MetGenMAP

Plant MetGenMAP is a visualization and analysis package that allows to identify changed pathways and enriched GO terms from gene expression and/or metabolite profile data and to visualize the profile data in a biochemical pathway context. You need to [login](#) to upload and analysis your data. New users please [register](#) to obtain an account. More information of the system can be obtained from the [help](#) page.

Currently the system supports the following organisms and platforms (identifiers). [A list of example input files for each platform are provided in the help page.](#)

Organism	Pathways	Platform
Arabidopsis	AraCyc	Arabidopsis ATH1 array
		TAIR AGI Locus Number
Rice	RiceCyc	Affymetix genome array
		Locus Number
Maize	MaizeCyc	Genome locus
Tomato	LycoCyc	Genome locus
		TOM1 cDNA array
		TOM2 oligo array
		Affymetrix genome array
poplar	PoplarCyc	Genome locus
soybean		Genome locus
sorghum		Genome locus
Brachypodium		Genome locus
grape		Genome locus
papaya		Genome locus
apple		Genome locus
banana		Genome locus
watermelon		Genome locus
Chinese cabbage		Genome locus
melon		Genome locus
strawberry		Genome locus

## News

**NEW** *Gene annotation updated* -- Gene annotations were updated with the [AHRD](#) pipeline. [Sept. 2013]

*Melon and strawberry genomes added* -- Pathways predicted from melon and strawberry genomes were added to the system [April. 2013]

*Tomato and Chinese cabbage genomes added* -- Pathways predicted from tomato and Chinese cabbage genomes were added to the system [April. 2013]

*Apple, banana, and watermelon genomes added* -- Apple, banana, and watermelon pathways were added to the system [Nov. 2012]

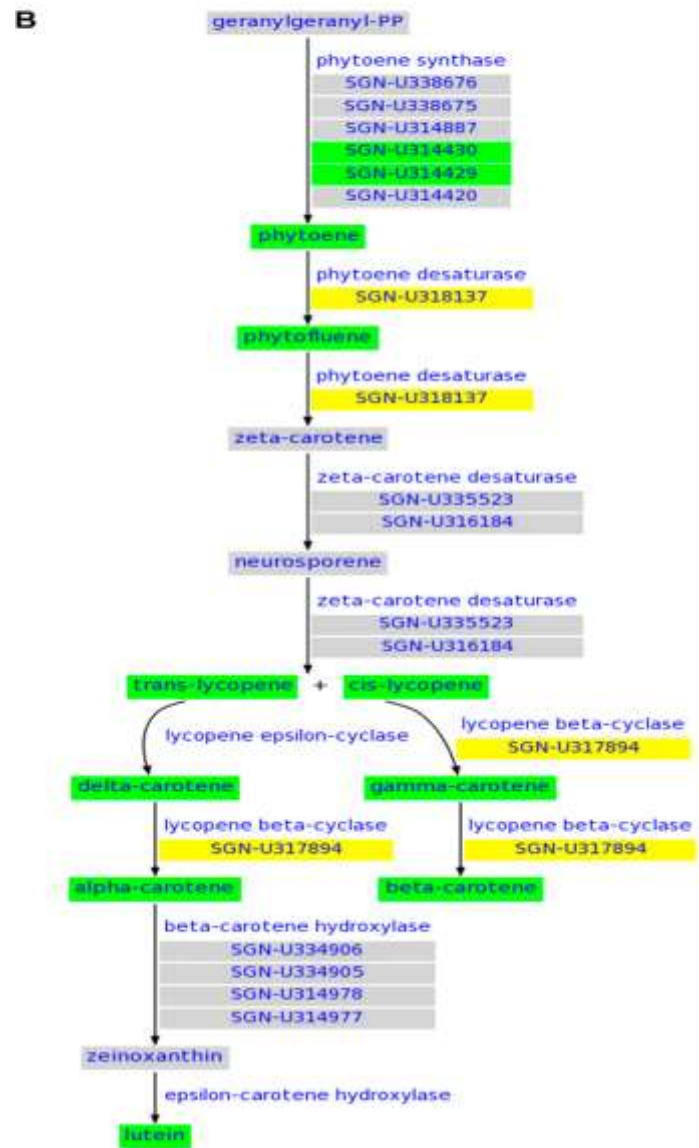
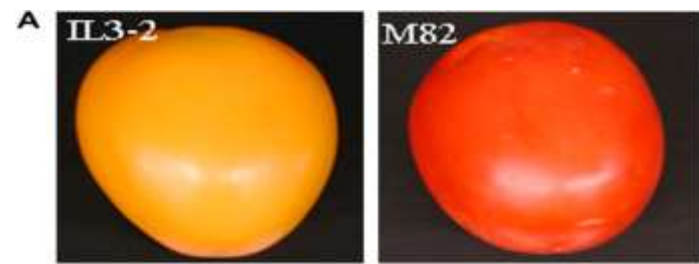
*Maize genome added* -- Maize pathways were added to the system [May 2012]

*Papaya genome added* -- Papaya pathways predicted from the [genome sequences](#) were added to the system [Jan. 2012]

*Sorghum pathways added* -- Sorghum pathways predicted from the [genome sequences](#) were added to the system [Sept. 2010]

*Arabidopsis pathways updated* -- Arabidopsis pathway were updated using the most recent version (7.0) of [AraCyc database](#). Arabidopsis datasets in the database need to be reprocessed. [Detail on how to reprocess the dataset...](#) [Aug. 2010]

*New organisms added to Plant MetGenMAP* -- Soybean, Grape and Brachypodium were added to Plant MetGenMAP based on their genome predicted genes [July 2010]



Changes of the carotenoid biosynthesis pathway in tomato introgression line IL3-2. (A) Ripened fruit of IL3-2 and its cultivated *S. lycopersicum* parent line M82. (B) Changes of gene expression levels and metabolite contents of the carotenoid biosynthesis pathway in IL3-2 compared with M82. Genes/metabolites with green background are down-regulated in IL3-2, those with yellow background are not changed, and those with gray background are not measured.

## Tomato Metabolite data

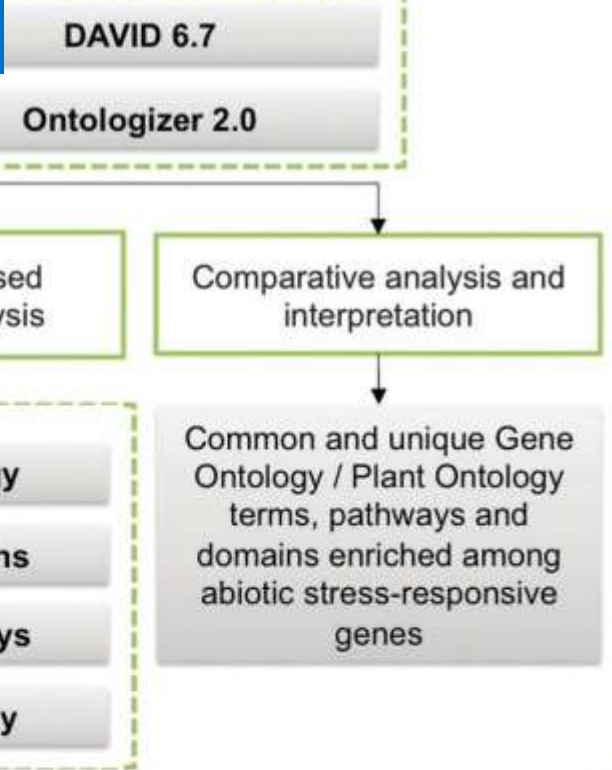
ID	IL3-2/M82	p value
trans-lycopene	-43.33	1.31E-73
phytoene	-3673.56	1.37E-41
Phytofluene	-4312.93	1.62E-34
cis-lycopene	-1600.89	2.07E-28
gamma-carotene	-26.65	1.89E-10
beta-carotene	-10.91	1.87E-06
total_phenolics	+1.38	3.32E-05
alpha-carotene	-17.09	0.025219871
delta-carotene	-21.36	0.028334675
lutein	-2.75	0.03217806
280nm_phenolics	+1.17	0.237474698
reduced_ascorbate	+1.17	0.293551165
total_ascorbate	-1.33	0.392909002
ORAC	+1.45	0.440811451
360nm_phenolics	+1.07	0.68962953
320nm_phenolics	+1.02	0.85775214

## Tomato Expression data

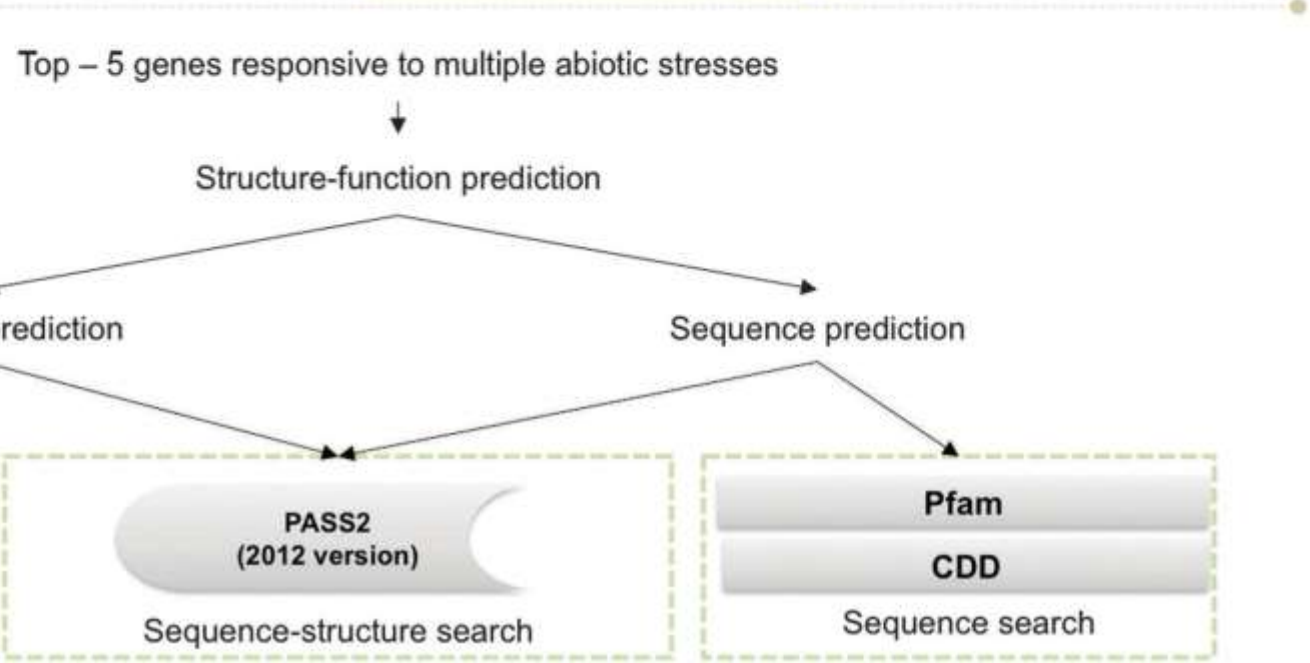
ID	IL3-2/M82	p value
1-1-5.1.12.20	0.05	2.43E-05
1-1-1.4.19.19	4.733	3.70E-05
1-1-7.4.19.9	0.1	7.03E-05
1-1-2.2.17.20	0.078	7.12E-05
1-1-8.2.17.20	0.086	7.18E-05
1-1-5.2.12.7	0.092	8.86E-05
1-1-5.1.19.11	0.084	9.02E-05
1-1-2.1.17.12	0.064	0.000102
1-1-4.2.1.6	0.067	0.000234
1-1-1.2.8.15	0.079	0.00458
1-1-5.2.8.4	0.092	0.000349
1-1-6.2.17.12	0.095	0.00137
1-1-7.1.12.20	0.111	0.000121
1-1-7.4.17.15	0.119	0.00129
1-1-7.1.3.4	0.121	0.00229
1-1-4.4.17.12	0.13	0.000739
1-1-8.4.17.16	0.159	0.000208
1-1-1.1.8.1	0.191	0.00884
1-1-3.3.3.12	0.194	0.00167
1-1-6.4.19.14	0.204	0.000214
1-1-2.4.19.14	0.205	0.00286



# Stress-responsive transcription Factors DataBase (STIFDB)



# Database for Annotation, Visualization and Integrated Discovery (DAVID)



**STIFDB V2.0 [Stress Responsive Transcription Factor Database] :**

STIFDB V2.0 [Stress Responsive Transcription Factor Database] is a comprehensive collection of biotic and abiotic stress responsive genes in *Arabidopsis thaliana* and *Oryza sativa* L. with options to identify probable Transcription Factor Binding Sites in their promoters. In the response to biotic stress like Bacteria and abiotic stresses like ABA, drought, cold, salinity, dehydration, UV-B, high light, heat, heavy metals etc, ten specific families of transcription factors in *Arabidopsis thaliana* and six in *Oryza sativa* L. are known to be involved. HMM-based models are used to identify binding sites of transcription factors belonging to these families. We have also consulted literature reports to cross-validate the Transcription Factor Binding Sites predicted by the method.

Various  
Stress  
Signals  
Analyzed

ABA WOUNDING  
OSMOTIC STRESS HEAT  
LIGHT HEAVY METALS  
UV-B DEHYDRATION COLD  
DROUGHT SALINITY OXIDATIVE STRESS

Click on the Stress Signal Names to Explore

**STIFDB V2.0: Database Statistics as on October 2012**

Database contains various Stress Responsive Genes & Transcription Factor Binding Site Related Information : TFmap , TAIR Accession ID, TIGR Accession ID, RAPDB Accession ID, Gene Names [Including Aliases], Chromosome Position, References to Publication and Related Resources, Transcription Factor Family Name, Binding Sites, Orientation of Binding Sites, Stress-Signal, Z-Score, and Normalization Score.

Genome	No. of Stress Resp. Genes
<i>Arabidopsis thaliana</i>	3150
<i>Oryza sativa</i> subsp <i>japonica</i>	1118
<i>Oryza sativa</i> subsp <i>indica</i>	1716

**Quick Search**
 [ Example Input ]


**If you use STIFDB data, please cite as:**

- Mahantesha Naika, Khader Shameer, Oommen K. Mathew, Ramanjini Gowda and Ramanathan Sowdhamini.(2013) STIFDB2: An Updated Version of Plant Stress-Responsive Transcription Factor DataBase with Additional Stress Signals, Stress-Responsive Transcription Factor Binding Sites and Stress-Responsive Genes in Arabidopsis and Rice. Plant Cell Physiol. 54(2):e8(1-15). [ PMID: 23314754 ]
- K. Shameer, S. Ambika, Susan Mary Varghese, N. Karaba, M. Udayakumar and R. Sowdhamini1.(2009) STIFDB-Arabidopsis Stress Responsive Transcription Factor DataBase. Int J Plant Genomics. 2009; 2009: 583429 [ PMID: PMC2763139 ]

<http://caps.ncbs.res.in/stifdb2/>

**Shortcut to DAVID Tools**

**Functional Annotation**

Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and more

**Gene Functional Classification**

Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. More

**Gene ID Conversion**

Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. More

**Gene Name Batch Viewer**

Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. More

Recommending: A paper published in *Nature Protocols* describes step-by-step procedure to use DAVID!

**Welcome to DAVID 6.7**

2003 - 2014

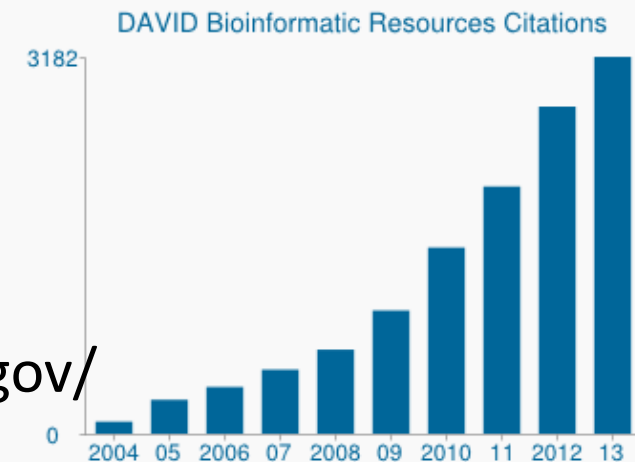
The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 is an update to the sixth version of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list
- List interacting proteins
- Explore gene names in batch
- Link gene-disease associations
- Highlight protein functional domains and motifs
- Redirect to related literatures
- Convert gene identifiers from one type to another.
- And more

**What's Important in DAVID?**

- Current (v 6.7) release note
- New requirement to cite DAVID
- IDs of Affy Exon and Gene arrays supported
- Novel Classification Algorithms
- Pre-built Affymetrix and Illumina backgrounds
- User's customized gene background
- Enhanced calculating speed

**Statistics of DAVID**



<http://david.abcc.ncifcrf.gov/>

## CONCLUSION

The ultimate goal of bioinformatics is to integrate large-scale data for understanding the molecular mechanism involved in various developmental processes. This understanding can help in producing the Climate Smart Horticulture Crops.

The explosion of genomic data along with other omics data has put a great responsibility on the bioinformatics community to aid in the management and analyses of such data. There is a need to not only analyze but also to integrate these data with functional information and a priori knowledge. The development of efficient and powerful in silico integrative analysis techniques can allow for higher levels of discovery and knowledge for producing the Climate Smart Horticulture Crops.



Q & A

QUESTIONS  
&  
ANSWERS

Thank you!  
Jimmy

