
Proceedings of the
NIPGR Sub-DIC
National Workshop
2014

Our Team

Introduction

14 November 2014 NIPGR

The BTIS-NET Sub-DIC at NIPGR organized the annual Bioinformatics Workshop titled 'Computational Biology in Plant Science' on 13-14th November 2014.

The Sub-Distributed Information Centre (Sub-DIC) at NIPGR was established in early 2007 and aims to serve as a support structure for all IT related issues at the institute, in addition to providing computational facilities and services related to bioinformatics for researchers in various laboratories at the Institute. The major objectives of the centre are to develop software, database and other tools required for creating infrastructure in the field of plant genomics, proteomics and plant molecular biology. The centrespecialises in plant comparative genomics and plant stress biology. A major effort of the centre is to conduct annual training programs to make users aware of available facilities in plant computational biology and its applications.

Thesetraining activities are conducted as short and long term programs, at both individual and group levels, and this includes an annual bioinformatics workshop conducted for Ph.D. scholars, postdoctoral researchers and teachers from various colleges and universities in India, in order to spread awareness about the use and applications of bioinformatics in plant biology.

The endeavour of the workshop organising committee is to provide an opportunity to all participants to benefit from the rich experience and expertise available in the area of bioinformatics in India. 30 participants were selected for the current workshop based upon abstracts received from all candidates.

The Inaugural session was held at 10 am on the 13th of November, 2014. After the welcome address by the coordinator, lamp lighting was conducted followed by an ode to the Goddess Saraswati by Ms. Sangita Kumari and Ms. Piyush Priya. Dr. Gitanjali Yadav delivered a lecture on the theme of the workshop and stressed upon the diverse applications of computational biology in plant sciences giving cases studies from the the ongoing research activities at NIPGR. The remaining sessions of the workshop included four hands-on tutorial sessions, as described in this booklet.Specially designed demonstration based practicals were imparted in each of the the tutorial sessions. Participants were encouraged to interact with each other as well as with the organising team. Summary slides from all tutorial sessions have been provided in this booklet.

The valedictory session included distribution of certificates, proceedings and prizes by Coordinator, Sub-DIC, followed by aninformal feedback sessionon the workshop. Finally, Dr. J.K. Thakur, Dy Coordinator of the Sub-DIC delivered the vote of thanks and the workshop was brought to a close.

On behalf of the Sub-DIC, I thank all participants as well as the organising team and NIPGR administration for making this event a success. We hope to continue and improve upon this effort in the coming years.

Gitanjali Yadav
(Scientist & Workshop Coordinator)

A BIOLOGIST'S APPROACH TO ANALYZE MICROARRAY DATA

(By Kunal Chatrath)

Global transcriptome analysis is of growing importance in understanding the true complexity and dynamics of genome. Currently, the most widely used method to analyze global patterns of gene expression is the DNA microarray. Due to large data set , it is a big struggle to find useful information. This exercise is aimed to help in extracting desired information from microarray experiments present in database. Two different Expression databases are used : NCBI GEO DataSet, ArrayExpress.

A search of a transcriptome database can give researchers a list of all the tissues in which a gene is expressed, providing clues to its possible function. The transcriptome data gives researchers a good place to start in the search for a new gene's function.

The current tutorial includes the following parts :

Part 1: Selecting the experiment of interest.

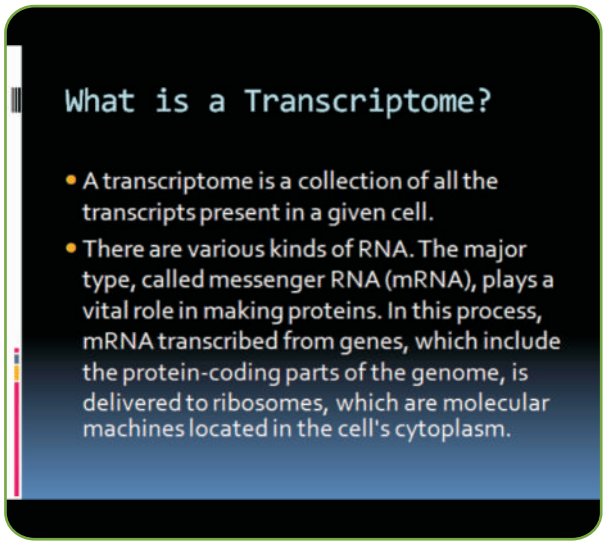
Part 2: Analysis of experiment using NCBI GEO dataset.

Part 3: Extracting the information about the differential expression of your gene of interest using ArrayExpress.

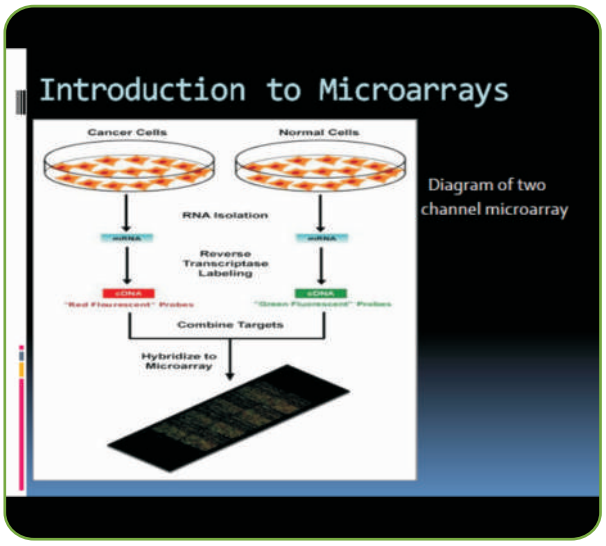
Part 4: Compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions.

Part 5: Finding known interactions for gene of interest and co-expressed genes using STRING database.

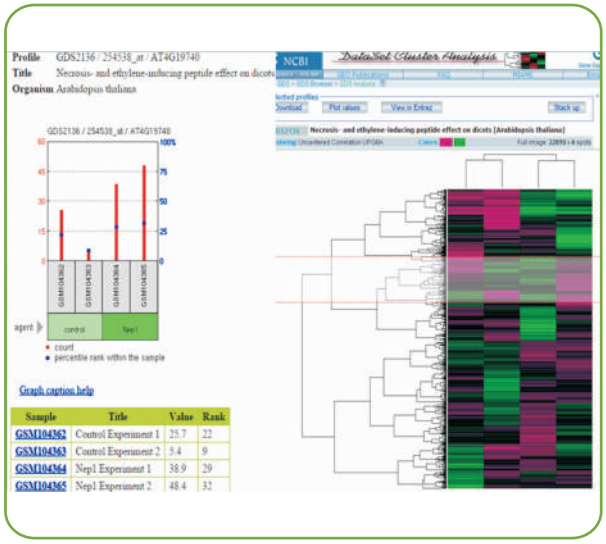
(1)



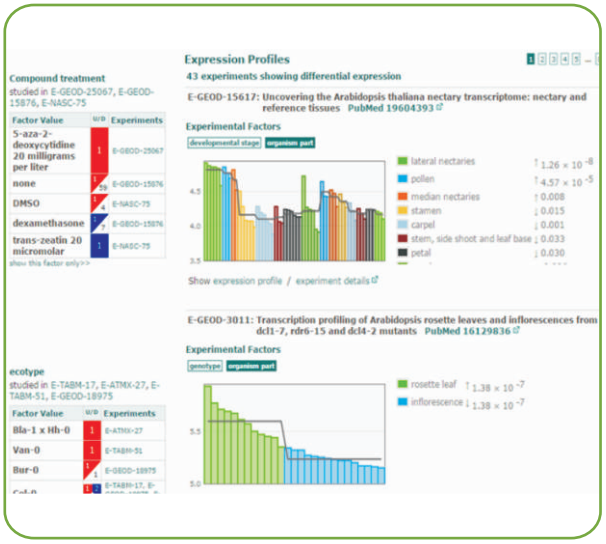
(2)



(3)


Profile GDS2136 / 254538_at / AT4G19740
Title Necrosis- and ethylene-inducing peptide effect on dicots
Organism Arabidopsis thaliana
NCBI
DataSet Cluster Analysis
Graph caption help
Sample Title Value Rank
GSM104362 Control Experiment 1 25.7 22
GSM104363 Control Experiment 2 5.4 9
GSM104364 Nep1 Experiment 1 38.9 29
GSM104365 Nep1 Experiment 2 48.4 32

(4)


Expression Profiles
Compound treatment
43 experiments showing differential expression
E-GEOD-15617: Uncovering the Arabidopsis thaliana nectary transcriptome: nectary and reference tissues PubMed 19604393
Experimental Factors
Show expression profile / experiment details
E-GEOD-3011: Transcription profiling of Arabidopsis rosette leaves and inflorescences from dcl1-7, rdr6-15 and dcl4-2 mutants PubMed 16129836
ecotype

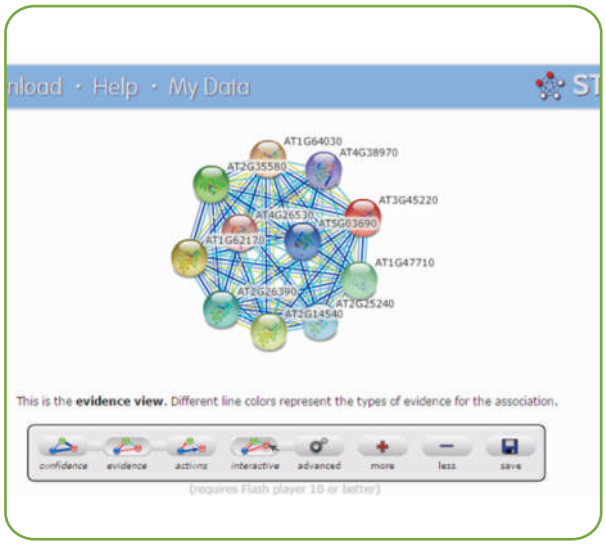
(5)


Genes/Design Elements
Conditions
Search Results:
Gene Design Element Experimental Factor
AT4G19740 254538_at organism part
AT4G19740 254538_at developmental stage
AT4G19740 254538_at organism part
AT4G19740 254538_at developmental stage
AT4G19740 254538_at organism part

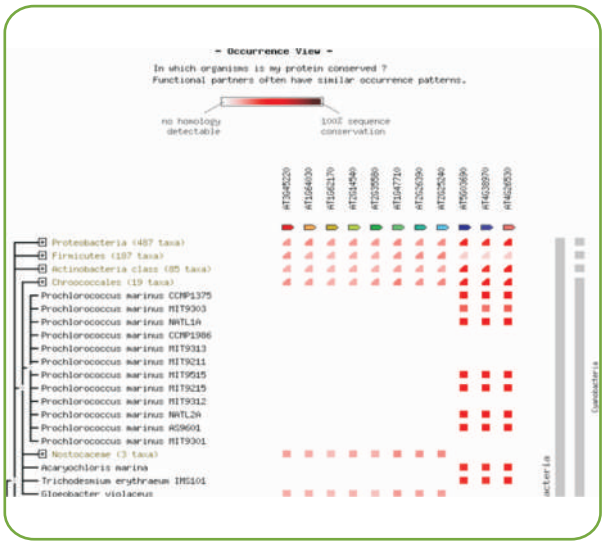
(6)


GSE4638/254519_at/AT4G19940
Sample values for ID = 254519_at
Sample Title Value
GSM104362 Control Experiment 1 34.6
GSM104363 Control Experiment 2 36.4
GSM104364 Nep1 Experiment 1 5.7
GSM104365 Nep1 Experiment 2 8.1
Sample values
control
sample
expression value

(7)


AT1G64030
AT4G38970
AT2G35580
AT3G45220
AT4G26530
AT5G03690
AT1G62170
AT1G47710
AT2G26390
AT2G25240
AT2G14540
This is the evidence view. Different line colors represent the types of evidence for the association.
confidence
evidence
actions
interactive
advanced
more
less
save

(8)


- Occurrence View -
In which organisms is my protein conserved ?
Functional partners often have similar occurrence patterns.
no homology detectable
100% sequence conservation
Proteobacteria (487 taxa)
Firmicutes (107 taxa)
Actinobacteria class (85 taxa)
Chroococcales (19 taxa)
Prochlorococcus marinus CCMP1375
Prochlorococcus marinus MIT9303
Prochlorococcus marinus NATL1A
Prochlorococcus marinus CCMP1986
Prochlorococcus marinus MIT9313
Prochlorococcus marinus MIT9211
Prochlorococcus marinus MIT9515
Prochlorococcus marinus MIT9215
Prochlorococcus marinus MIT9312
Prochlorococcus marinus NATL2A
Prochlorococcus marinus AS9601
Prochlorococcus marinus MIT9301
Nostocaceae (3 taxa)
Acaryochloris marina
Trichodesmium erythraeum IMS101
Gloeobacter violaceus
archaea
cyanobacteria

COMPARATIVE GENOMICS

(By Zohra Firdos Jafaree)

The comparison of sequence similarity, gene numbers, gene locations, the length and number of coding regions (called exons) within genes, the amount of noncoding DNA in each genome, and highly conserved regions maintained in organisms as simple as bacteria and as complex as humans is defined as Comparative Genomics.Because all modern genomes have arisen from common ancestral genomes, the relationships between genomes can be studied with this fact in mind. This commonality means that information gained in one organism can have application in other even distantly related organisms. Comparative genomics enables the application of information gained from facile model systems to agricultural and medical problems. In the present workshop, VISTA: an online tool for comparing genomes has been used.

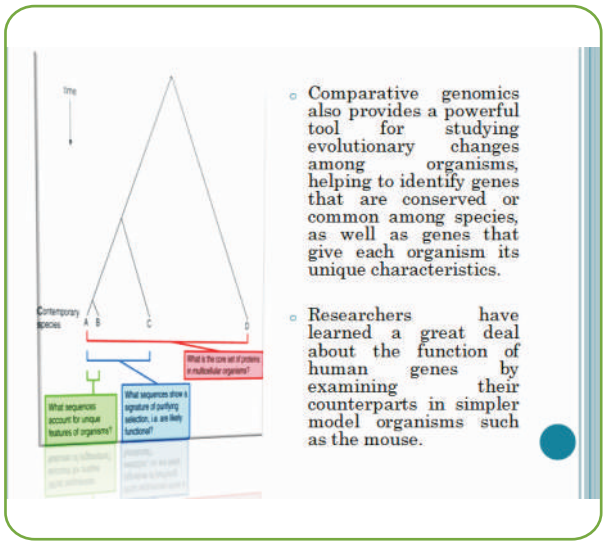
Step1: Download the sequences of interest or copy the already provided sequences to your directory.

Step2:Open and upload all the required files on specified comparative genomics portal (wgVISTA).

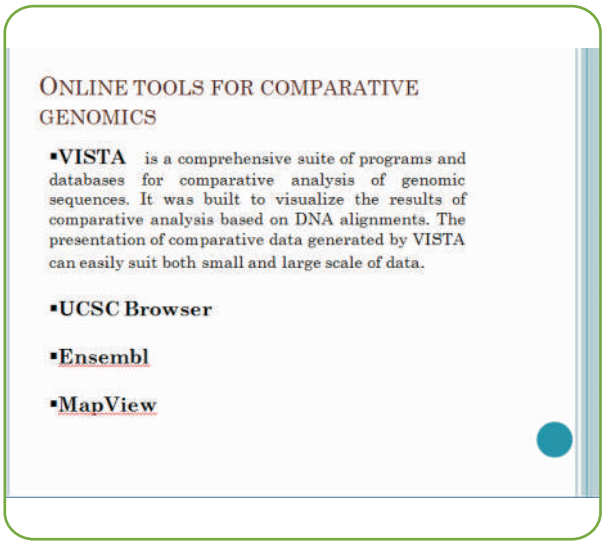
Step 3:Wait for sometime after successful submissionto the online comparative genomics portal.

Step4: Observe and perform comparative analysis of the results.

(1)



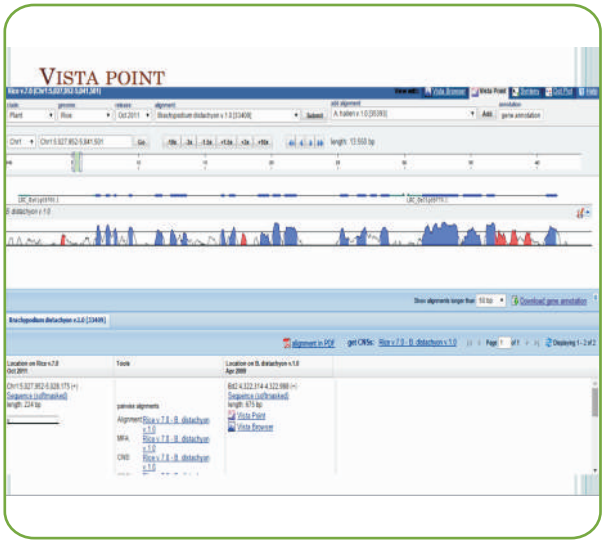
(2)



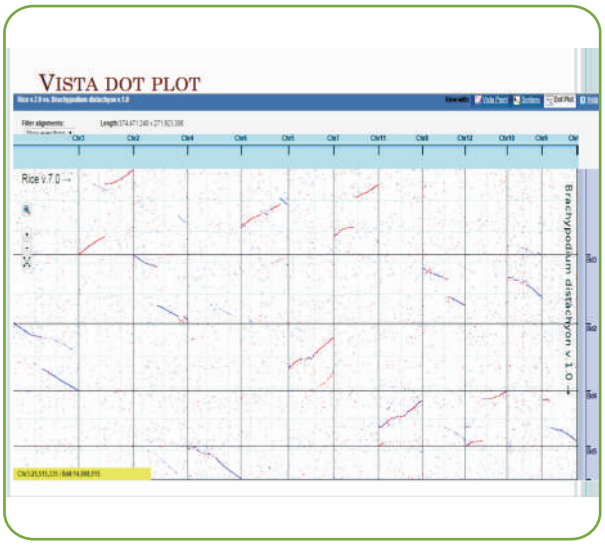
(3)


VISTA HOMEPAGE
VISTA
Tools for Comparative Genomics
About Us
Cite Us
Contact Us
VISTA Home
Custom Alignment
Browser
Enhancer DB
Downloads
Publications
Help
VISTA is a comprehensive suite of programs and databases for comparative analysis of genomic sequences. There are two ways of using VISTA - you can submit your own sequences and alignments for analysis (VISTA servers) or examine pre-computed whole-genome alignments of different species.
Submit Your Sequences
mVISTA
GACAC
GACAT
mVISTA
Align and compare your sequences from multiple species
rVISTA
Locate regulatory sequences in your data using comparative sequence analysis and transcription factor binding site search.
gVISTA
Compare your sequences against whole-genome assemblies.
wgVISTA
Align pair of sequences up to 10Mb long (finished or draft) including microbial whole-genome assemblies.
Precomputed Alignments
VISTA Browser
VISTA Point
Access complete data and visual presentation of pairwise and multiple alignments of whole genome assemblies.
VISTA Browser
Examine pre-computed pairwise and multiple alignments of whole genome assemblies.
Whole Genome rVISTA
Identify transcription factor binding sites that are conserved between species and over-represented in upstream regions of groups of genes.
Microbial Genomes
Access pre-computed full scaffold alignments for microbial genomes through the VISTA component of IMG
New tool from VISTA family!
VISTA Region Viewer (RViewer) is an interactive on-line tool for comparing and prioritizing genomic intervals.
Updates
April 2014
Updated the Sorghum, Monkey flower, Moss, Maize, Medicago, Switchgrass, and Soybean assemblies, and added 5 new plants: C. grandiflora, Drummond's rockcress, Turnip mustard, A. halleri, and Hall's petrocallis.
180 New whole-genome plant alignments are added to VISTA Browser.
August 2013
Updated the C. elegans and C. briggsae assemblies, and added 5 new worms: C. brenneri, C. remanei, C. japonica, C. sp. 11, and C. angaria

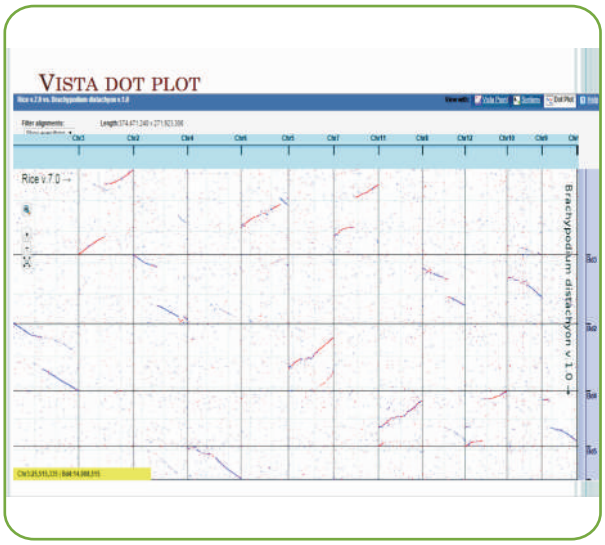
(4)


VISTA POINT

(5)


VISTA DOT PLOT
Rice v.7.0
Brachypodium distachyon v.1.0

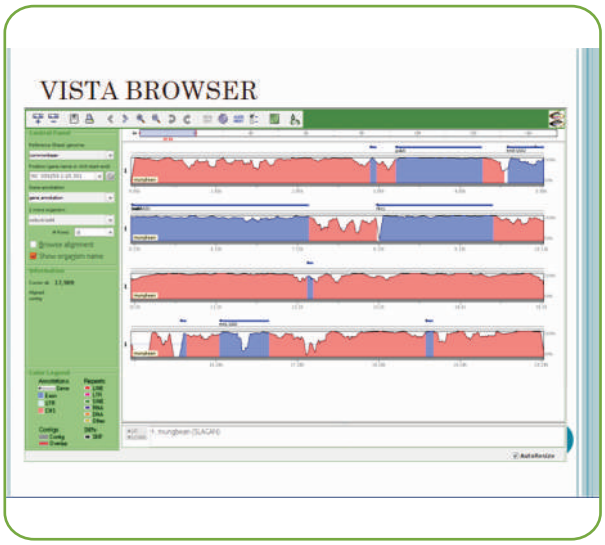
(6)


VISTA DOT PLOT
Rice v.7.0
Brachypodium distachyon v.1.0

(7)


WGVISTA: WHOLE GENOME VISTA INPUT PAGE
VISTA
Tools for Comparative Genomics
About Us
Cite Us
Contact Us
VISTA Home
Custom Alignment
Browser
Enhancer DB
Downloads
Publications
Training
Help
mVISTA
rVISTA
gVISTA
wgVISTA
Phylo-VISTA
wgVISTA
Submit
Instructions
About wgVISTA
Cite
wgVISTA Submit:
Inquiry
Your email address:
Genome #1:
Genome #2:
Choose File
No file chosen
OR The GENBANK identifier(s)
Reset
Submit
Additional options
Genome #1
Name: genome1
Annotation:
Soft-masked
Genome #2
Name: genome2

(8)


VISTA BROWSER

PROTEIN STRUCTURE VISUALIZATION AND ANALYSIS

(By, Archana Yadav, Ph.D. Student (SRF) NIPGR)

In recent years, there has been a great advancement in determination of three-dimensional structure of macromolecules by NMR and X-ray crystallography methods. This has created a huge gap in terms of available data and researcher's ability to analyze and use the conformational information and deduce functional insights from them. Extensive details about structure and function can be obtained through proper visualization of molecules in three dimensions.

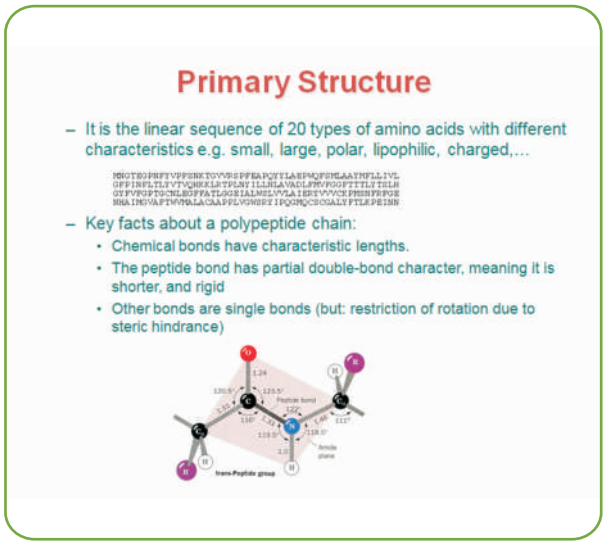
There are many powerful visualization tools available today which are able to provide minute stereotypical details about the molecule along with viewing the molecule in motion on the console of the computer.

This session deals with protein structure analysis starting from prediction of secondary structure from the primary sequence, exploring structure databases and visualization of the molecule in 3-D along with interpretation of the functional details.

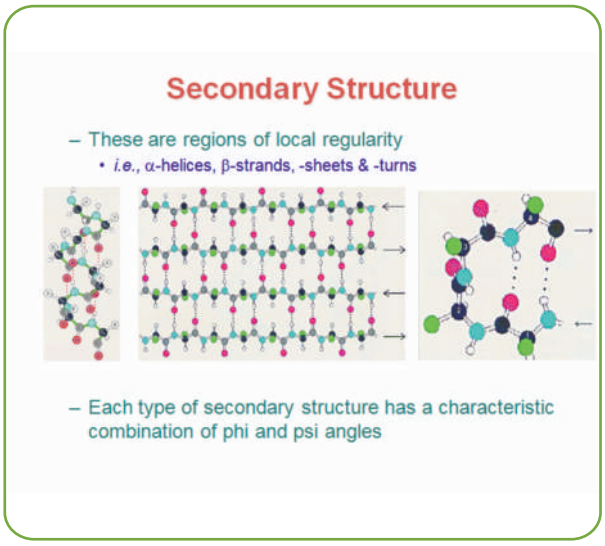
Following will be covered in this session:

Part1: Predict the secondary structure of a protein sequence
Part2: Primary Structure to the 3-D Structure
Part 3: View the overall information of PDB
Part 4: Visualization of protein 3D structure

(1)



(2)



(3)

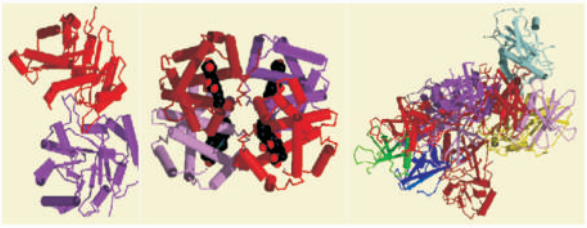
Tertiary structure

- the overall chain fold that results from packing of secondary structure elements
- Domains, repeats, zinc fingers...
- Domain: independently folded part of a protein. Average size, about 150 aa residues, lower limit ca 50 residues
- Repeats: several types: LRR, ANK, HEAT.... Composed of few secondary structure elements. Stabilized by interactions between repeats; can form large structures.
- Zinc fingers: several types; structure is stabilized by bound zinc ion

(4)

Quaternary structure

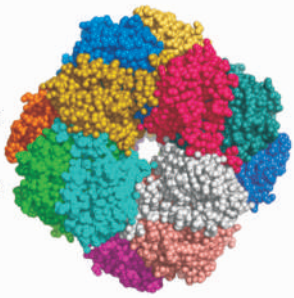
- the arrangement of separate chains within a protein that has more than one subunit
- the arrangement of separate molecules, such as in protein-protein or protein-nucleic acid interactions
  - *e.g.*, haemoglobin



(5)

RUBISCO

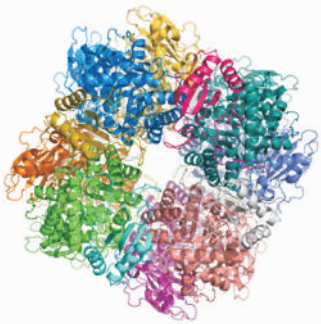
- Rubisco, ribulose bisphosphate carboxylase, is a protein molecule present in plant cells.
- It takes part in photosynthesis and converts inorganic CO_2 into organic forms containing C-C bonds and H atoms.
- These are used to sustain plants in the form of sucrose (table sugar) or stored as starch.
- Most abundant protein on earth



(6)

RUBISCO-structure (1rcx)

- Rubisco protein of spinach (pdb code 1rcx) is compose of total 16 chains, 8 large a 8 small.
- The arrangement of large a small chains is L8S8.
- The large chain has substra RUBP binding site.



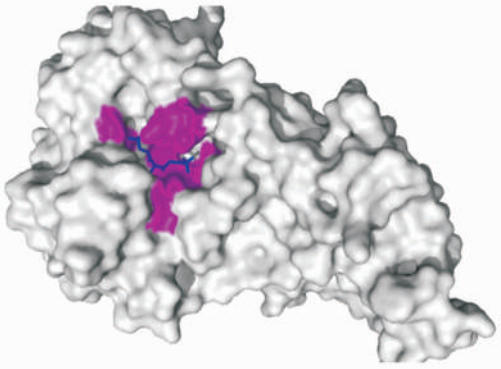
(7)

1RCX- active site

- The central structural motif in the catalytic subunit (L) is an eight-stranded parallel α/β-barrel, which forms a scaffold harboring the active site.
- Active site residues are situated in the loops that connect the β strands with the α helices.
- It is non-activated form

(8)

Binding pocket



Binding pocket of RUBP. Interacting residues are colored pink

GENOME ANNOTATION

(By Amish Kumar)

The past decade has seen the completion of numerous whole genome sequencing projects, however this is not the end of a genome project. In order to make use of these genome sequences, we need to understand all of its components.

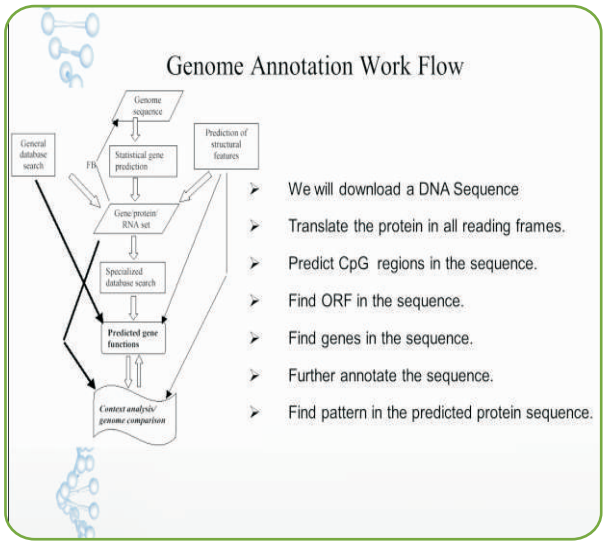
Assigning identities and functions to sequences within the genome is called genome annotation. It is the process of taking the raw DNA sequence produced by the genome-sequencing projects and adding the layers of analysis and interpretation necessary to extract its biological significance and place it into the context of our understanding of biological processes.

Genome annotation itself is a multi-step process this involves describing different regions of the code and identifying which regions can be called genes.

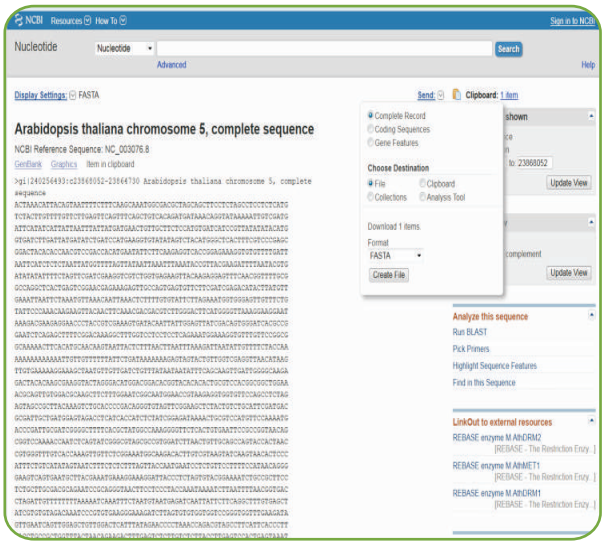
This tutorial aims to give an overview of various online tools which can be used for the annotation of plant genome as follows.

Part-1: We will download a DNA Sequence.
Part-2: Translate the protein in all reading frames.
Part-3: Predict CpG regions in the sequence.
Part-4: Find ORF in the sequence.
Part-5: Find genes in the sequence.
Part-6: Further annotate the sequence.
Part-7: Find pattern in the predicted protein sequence.

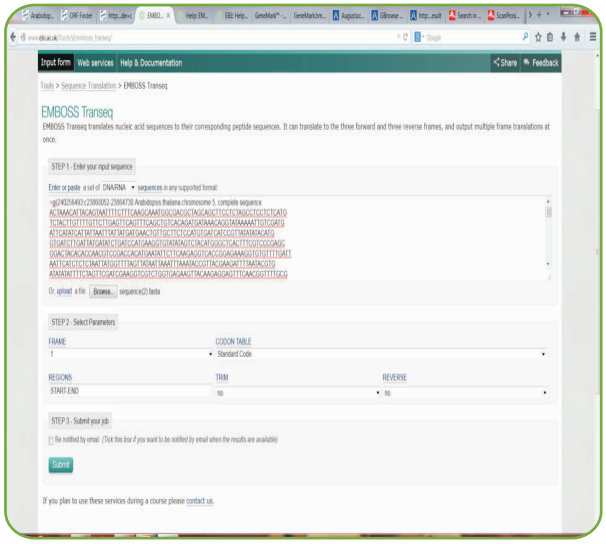
(1)



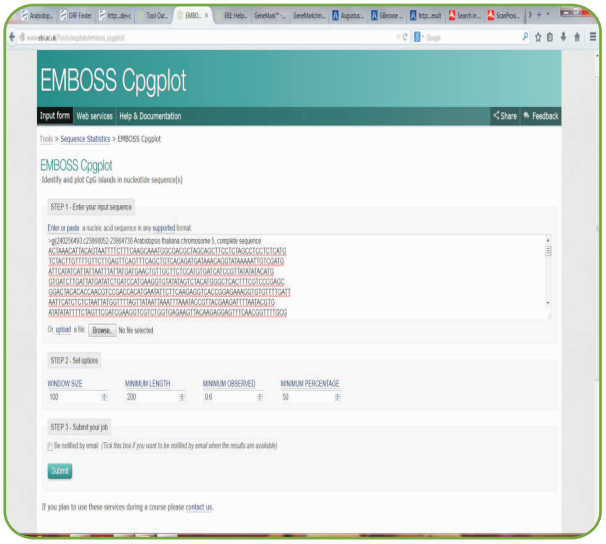
(2)



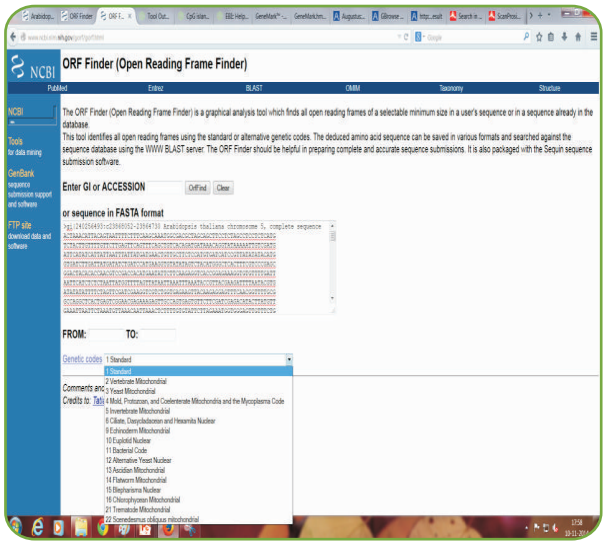
(3)


EMBOSS Transeq
Input form
Web services
Help & Documentation
Share
Feedback
EMBOSS Transeq translates nucleic acid sequences to their corresponding peptide sequences. It can translate to the three forward and three reverse frames, and output multiple frame translations at once.
STEP 1 - Enter your input sequence
STEP 2 - Select Parameters
STEP 3 - Submit your job
Submit

(4)


EMBOSS Cpgplot
Input form
Web services
Help & Documentation
Share
Feedback
EMBOSS Cpgplot
Identify and plot CpG islands in nucleotide sequence(s)
STEP 1 - Enter your input sequence
STEP 2 - Set options
STEP 3 - Submit your job
Submit

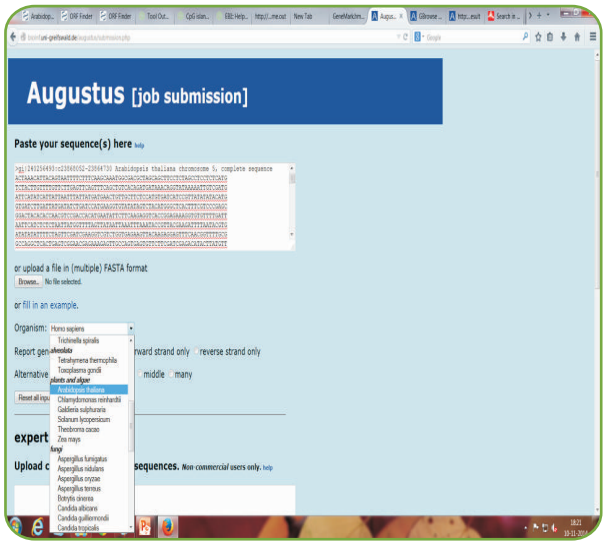
(5)


NCBI
ORF Finder (Open Reading Frame Finder)
The ORF Finder (Open Reading Frame Finder) is a graphical analysis tool which finds all open reading frames of a selectable minimum size in a user's sequence or in a sequence already in the database.
Enter GI or ACCESSION
or sequence in FASTA format
FROM:
TO:
Genetic codes

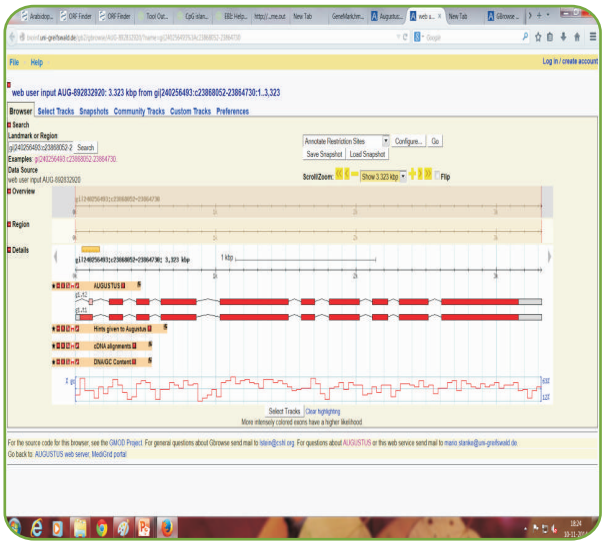
(6)


GeneMark
A family of gene prediction programs developed at Georgia Institute of Technology, Atlanta, Georgia, USA.
Gene Prediction in Bacteria, Archaea, Metagenomes and Metatranscriptomes
Gene Prediction in Eukaryotes
Gene Prediction in Transcripts
Gene Prediction in Viruses, Phages and Plasmids

(7)


Augustus [job submission]
Paste your sequence(s) here
or upload a file in (multiple) FASTA format
or fill in an example.
Organism:
Homo sapiens

(8)


web user input AUG-892832920: 3.323 kbp from gi|240256493:c23868052-23864730:1..3,323
Browser
Select Tracks
Snapshots
Community Tracks
Custom Tracks
Preferences
Overview
Region
Details

PROGRAM DETAILS FOR THE NATIONAL WORKSHOP ON
'COMPUTATIONAL BIOLOGY IN PLANT SCIENCE'
13 - 14th November 2014

DAY 01(13th November 2014)

INAUGURAL SESSION

9:00 am	Registration	Reception, NIPGR
9:30 am	Welcome Address	Co-ordinator, Sub-DIC
09:40 am	Chairman Address	Director, NIPGR
09:50 am	Deputy Co-ordinator Address	Dy Co-ordinator, Sub-DIC
10:00 am	Vote of Thanks	Co-ordinator, Sub-DIC
10:05 am	Tea	

MORNING SESSION

10:30 am	Computational Biology in Plant Science	Dr. Gitanjali Yadav, NIPGR
11:15 am	Transcriptome Analysis	Mr. Kunal Chatrath, NIPGR
14:00 pm	Lunch	

AFTERNOON SESSION

14:30 pm	Comparative Genomics	Ms. Zohra Firdos Jafaree, NIPGR
16:00 pm	Tea	
16:15 pm	Discussion on Tutorials	Mr. Kunal Chatrath and Ms. Zohra Firdos Jafaree, NIPGR

DAY 02 (14th NOV 2014)

09:15 am	Genome Annotation	Mr. Amish Kumar, NIPGR
11:00 am	Tea	
11:15 am	Genome Annotation (Contd.)	Mr. Amish Kumar, NIPGR
12:15 pm	Protein Structure Visualization and Analysis	Ms. Archana Yadav, NIPGR
14:00 pm	Lunch	

AFTERNOON SESSION

14:30 pm	Protein Structure Visualization and Analysis (Contd.)	Ms. Archana Yadav, NIPGR
16:00 pm	Tea	

VALEDICTORY FUNCTION

16:15 pm	Director's Address	
16:30 pm	Certificate & Prize Distribution	
17:00 pm	Vote of Thanks	Dy Co-ordinator, Sub-DIC
17:10 pm	Closing Address	Co-ordinator, Sub-DIC

PROGRAM DETAILS FOR THE NATIONAL WORKSHOP ON

'COMPUTATIONAL BIOLOGY IN PLANT SCIENCE'

13 - 14th November 2014

	Name	UNIVERSITY
1	Ms. Madhvi Mishra	UPTU
2	Mr. Vijaylaxmi Gupta	MGKVP Varanasi
3	Dr. Kaushal G. Modha	Navsari Agricultural University
4	Ms. Jananee Jaishankar	Maulana Azad National Institute of Technology
5	Ms. Tina Sharma	CCS University
6	Mr. Rohit Khandelwal	Centre for Converging Technologies, University of Rajasthan
7	Mr. B. Meganathan	Madurai Kamaraj University, Madurai
8	Mr. Ashirbad Guria	Madurai Kamraj University
9	Dr. Rana D.P. Singh	Sugarcane Research Station, Gorakhpur
10	Ms. Aditi Jain	University of Delhi
11	Ms. Vandana Jaggi	Maharshi Markandeshwar University
12	Mr. Kanak Rakshit	Maharshi Markandeshwar University
13	Ms. Sunakshi Rustagi	Maharshi Markandeshwar University
14	Dr. Pratibha Yadav	IIT Delhi
15	Ms. Mahima Tiwari	Jaipur National University
16	Mr. Rahul Michael	CSIR-National Botanical Research Institute
17	Dr. Muthukumar P	IARI, New Delhi
18	Ms. Seema Pradhan	NIPGR
19	Mr. Vimal Pandey	NIPGR
20	Mr. Abdulrazak Ado	Sharda University
21	Mr. Dauda Danlami	Sharda University
22	Mr. Idris Zubairu Sadiq	Sharda University
23	Mr. Rabiu Sani Shawai	Sharda University
24	Mr. Hassan Muhammad Ibrahim	Sharda University
25	Mr. Jamilu Yusuf Muhammad	Sharda University
26	Mr. Sale Ali Ibrahim	Sharda University
27	Dr Somdutta Sinha Roy,	University of Delhi
28	Ms. Shaista Parveen	NIPGR

PROGRAM DETAILS FOR THE NATIONAL WORKSHOP ON

'COMPUTATIONAL BIOLOGY IN PLANT SCIENCE'

13 - 14th November 2014

ORGANISERS

DR. GITANJALI YADAV	COORDINATOR, NIPGR
MR. SUBHASISH MONDAL	NIPGR
DR. RENU KUMARI	NIPGR
MRS. SANGITA	NIPGR
MRS. PIYUSH PRIYA	NIPGR
MS. ARCHANA YADAV	NIPGR
MR. AMISH KUMAR	NIPGR
MR. SANJEET KUMAR MAHTHA	NIPGR
MR. SATISH BARFA	NIPGR
MR. KUNAL CHATRATH	NIPGR
MS. ZOHRA FIRDOS JAFAREE	NIPGR

NIPGR SUB-DIC

CHAIRMAN	PROF. A. K. TYAGI
COORDINATOR	DR. G. YADAV
DY COORDINATOR	DR. J. K. THAKUR
STAFF	MR. SUBHASISH MONDAL
TECHNICAL ASSISTANT	MR. KUNAL CHATRATH

SPEAKERS

DR. GITANJALI YADAV	NIPGR
MR. KUNAL CHATRATH	NIPGR
MS. ZOHRA FIRDOS JAFAREE	NIPGR
MR. AMISH KUMAR	NIPGR
MS. ARCHANA YADAV	NIPGR