

**Proceedings of the  
National  
Workshop on  
‘use of Bioinformatics  
in Agriculture and  
Plant Sciences’,  
2010**



NIPGR

Phylogenetic  
Making Sense of Biology in t

R. Geet  
Professor  
Department of  
University of  
1100

# The BTIS-NET Sub-DIC at NIPGR organized the second Bioinformatics Workshop titled 'Use of Bioinformatics in Agriculture/ Plant Sciences' on 12-13th March 2010.

The endeavor of the organizing committee is to make this workshop an annual event, as per the BTIS-NET scheme, for Ph.D. scholars, postdoctoral researchers and teachers from various colleges and universities, in order to provide an opportunity to all participants to benefit from the rich experience and expertise available in bioinformatics in India.

Twenty two participants were selected for the current workshop based upon abstracts received from all candidates (page X). The Inaugural session of the event was held at 10 am on the 12th of March, 2010. The welcome address was delivered by the Sub-DIC coordinator, while Prof. R. Geeta, renowned plant biologist and phylogeneticist from Univ of Delhi, Dept of Botany, delivered the Inaugural lecture titled "Phylogenetic Biology: Making Sense of Biology in the Light of Evolution". She stressed upon the need for young scientific minds to hypothesize and to remain aware of past research when designing experiments, giving examples from her wide experience and her work in the evolution of the monocot leaf and the Angiosperm endosperm.

The remaining sessions of the workshop included, alternatively, three lecture sessions and five tutorial sessions (Page X). Young scientists working in the area of plant computational biology were invited for the lecture sessions, whereas previously designed hands-on practicals were imparted in the tutorial sessions. These included tutorials on Sequence Analysis, Data Interpretation, EST data analysis and structure prediction as well as modeling exercises.

During the First lecture session, Dr. G. Yadav, NIPGR, gave a lecture titled "Bioinformatics in Agriculture", describing the varied applications of computational biology in emerging areas like precision agriculture, crop improvement and evolutionary studies. Dr. S. M. Leighton, Univ of Delhi, in her lecture entitled "Bioinformatics and the pedagogy of Plant Sciences" described with examples, various steps in the design and implementation of a bioinformatics problem and its solution using freely available online programs and databases.

The first tutorial session gave the participants insights into sequence data retrieval and analysis, using web based tools. The next practical session included an exercise to enable the participants to carry out this analysis, using the example of wheat protein agglutinin.

On the second day of the workshop, participants were given a tutorial on EST data analysis and the methods involved therein. Later they were introduced to the concept of protein structure. Exercises were designed using the example of wheat protein. The last hands-on session involved protein family analysis and phylogenetics. Summary slides from all lectures have been provided on Page X.

The second lecture session held on this day included a lecture by Dr. D. Chattopadhyay, NIPGR. His talk was titled 'Next generation sequencing and its applications in genomics and epigenomics'. Dr. Chattopadhyay elaborated upon the history of genome sequencing, the methods involved and described in details the applications of the new and futuristic methods with specific examples from plant genomes.

The valedictory session included an informal interaction with all participants who appreciated all the sessions and gave positive feedback on the workshop. On behalf of the Sub-DIC, I thank all participants as well as the NIPGR administration for making this event a success.

Gitanjali Yadav  
(Staff Scientist & Workshop Coordinator)

# Protein structure prediction and visualization

This exercise is concerned with various online tools available to study a protein structure. It is shown where you can find these tools, how to use them and most importantly how to interpret the results.

**Part I: Predict the secondary structure of a protein sequence**

**Part 2: From the Primary Structure to the 3-D Structure**

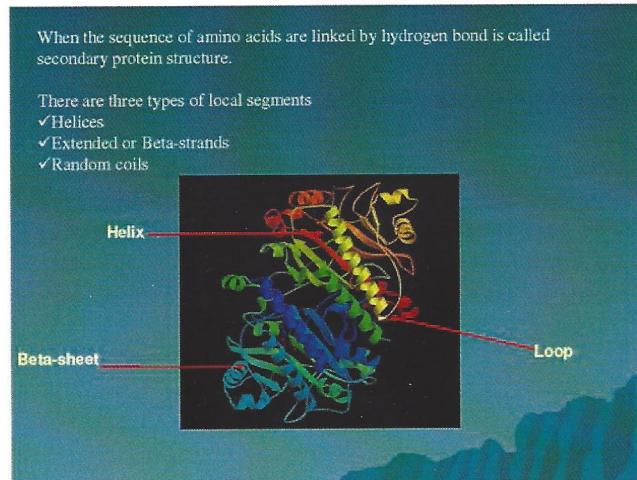
**Part III: Visualization of protein 3D structure**

**Part IV(a): Gene and Protein sequences**

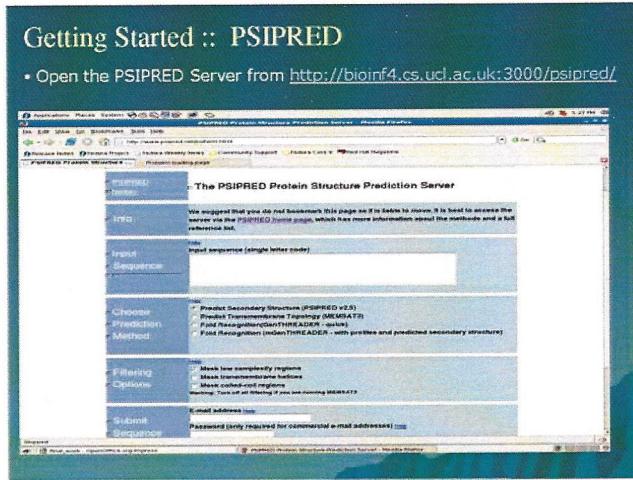
**Part IV(b): Translation of the nucleotide sequence**

**Part V: View the overall information of PDB**

1.)



2.)



3.)

Retrieve protein from PDB

- Open Protein Data Bank (PDB) [www.rcsb.org](http://www.rcsb.org).
- Enter a "Keyword" search by typing a keyword in the text box on the search bar at the top of the search box: keyword like "AGII\_WHEAT"
- Click the adjacent SEARCH button
- The result page will contain a list of PDB's entry name i.e. of four letters associated with this protein name called AGII\_WHEAT" eg 2UVO, 2UWG.

The screenshot shows the PDB search results for the keyword "AGII\_WHEAT". It lists two entries: 2UVO and 2UWG. The 2UVO entry is selected, showing details such as Authors (Schneeloch, D., Williamson, W., Haderlein, K., Weier, W.), Title (HIGH-RESOLUTION CRYSTAL STRUCTURE OF WHEAT GERM AGGLUTININ IN COMPLEX WITH N-Acetyl-D-Glucosamine), and Authors (Schneeloch, D., Williamson, W., Haderlein, K., Weier, W.).

4.)

• Click on the 2UVO for further analysis of its structure  
• Download PDB file save it on your computer

The screenshot shows the detailed view of the PDB entry 2UVO. It includes sections for Summary, Authors, Depositors, References, and Tables. The Summary section highlights the "HIGH-RESOLUTION CRYSTAL STRUCTURE OF WHEAT GERM AGGLUTININ IN COMPLEX WITH N-Acetyl-D-Glucosamine". The Authors section lists Schneeloch, D., Williamson, W., Haderlein, K., Weier, W. The Depositors section shows Williamson, W. The References section lists several publications. The Tables section provides structural details, including a 3D ribbon diagram of the protein structure.

5.)

- Point your web browser to the following URL: <http://www.ncbi.nlm.nih.gov/>
- In the top search box select Nucleotide and enter the text "Agglutinin Isolectin wheat"

Click on the Go button

The screenshot shows the NCBI Nucleotide search results page. The search term 'Agglutinin Isolectin wheat' is entered in the search bar. Below the search bar, there are several search filters and a 'Recent activity' section. The main results table lists several entries, with the first entry being 'Wheat (T. aestivum) genes agglutinin isolectin A, complete cds'. The details for this entry include the organism (T. aestivum), gene name (LOC\_Ots13252), and a link to the full sequence.

6.)

Obtain the FASTA sequence of Agglutinin isolectin wheat

The screenshot shows the NCBI Nucleotide search results page for the same search term. The results table shows the same entry for 'Wheat (T. aestivum) genes agglutinin isolectin A, complete cds'. The details section now displays the FASTA sequence of the gene, starting with '>Wheat (T. aestivum) genes agglutinin isolectin A, complete cds LOC\_Ots13252'. The sequence is presented in a standard FASTA format.

7.)

Translate gene to protein sequence by expasy tool.

The screenshot shows the ExPASy Translate tool interface. The user has pasted the FASTA sequence from step 6 into the input field. Below the input field, there is a dropdown menu for selecting the translation frame, with 'Frame 6' selected. At the bottom of the page, there is a 'TRANSLATE SEQUENCE' button.

8.)

Translated protein sequence in six Open Reading Frame

The screenshot shows the results of the protein translation. It displays six different open reading frames (ORFs) for the gene. Each frame is represented by a long string of amino acid residues. The frames are labeled 'Frame 1', 'Frame 2', 'Frame 3', 'Frame 4', 'Frame 5', and 'Frame 6' at the top of their respective sequences.

9.)

- Click on any frame six of them.
- The result page highlights the Methionine residues, or the starting point of the protein sequence.
- Click the first 'M' in the sequence.

The screenshot shows the ExPASy Virtual Sequencer tool interface. The user has selected 'Frame 6' and pasted the protein sequence from step 8 into the input field. The sequence is displayed with methionine (M) residues highlighted in red. The user has clicked on the first 'M' in the sequence, which is also highlighted. At the bottom of the page, there is a 'Virtual Sequencer' button.

10.)

The screenshot shows the ExPASy ProtParam tool interface. The user has pasted the protein sequence from step 8 into the input field. The results table shows various properties of the protein, including its molecular weight (62.0 kDa), isoelectric point (PI 5.2), and hydrophilicity (I 45.0). On the right side of the page, there is a detailed analysis of the protein's physicochemical properties, including a 'Surface' plot showing hydrophobicity across the sequence.

## Biological Data Retrieval & Analysis

1.)

2.)

3.)

**Family: Calreticulin (PF00262)**

Summary	
Domain organization	Calreticulin family
Alignments	<a href="#">InterPro entry PF00262</a>
MMW logo	
Trees	<a href="#">Phylogenetic tree</a>
Curators & models	<a href="#">Calreticulin</a>
Species	Primate: Catarrhini, Cetacea, Primates
Annotations	Calreticulin is a 200 kDa protein which binds calcium ions in a 1:1 ratio and is located at the periphery of the endoplasmic reticulum (ER) and the sarcoplasmic reticulum (SR). It contains a signal sequence and a membrane spanning domain. It has been implicated in the regulation of ER and SR Ca <sup>2+</sup> levels and may play a role in muscle contraction.
Structures	Structure of calreticulin in complex with a peptide substrate
Jump to...	<a href="#">Jump to top</a>   <a href="#">Jump to bottom</a>
<b>Jump to:</b> <a href="#">Jump to top</a>   <a href="#">Jump to bottom</a>	
<b>Gene Ontology</b>	
<a href="#">Predicted function</a>   <a href="#">GO term</a>   <a href="#">GO class</a>	
<b>External database links</b>	
<a href="#">PANTHER</a>   <a href="#">EPRINT</a>   <a href="#">PROSITE</a>   <a href="#">SCOP</a>   <a href="#">SMART</a>	

4.

5.)

6.

7.)

8.)

EMBL-EBI Home Search All Databases Enter Text Here Go Recent Advanced Search Give us Feedback

Databases Tools EBI Group Training Industry About Us Help Site Index Help

Quick Search Library Page Query Form Tools Results Projects Views Databases Help

**Reset** search UniProtKB/Swiss-Prot

---

### Search Options

Combine search terms with:  & (AND)  OR

Use wildcards

Get results of type:

Fields you can search		Your search terms
In a single field, you can separate multiple values by: ; ,   ?		
<input type="checkbox"/>	Description	<input type="text" value="Calnexin"/>
<input type="checkbox"/>	AllText	<input type="text"/>
<input type="checkbox"/>	AllText	<input type="text"/>
<input type="checkbox"/>	AllText	<input type="text"/>

---

### Result Display Options

View results using:

UniProtView  UniProtKB

Create a view

Show

results per page

### Create a view

Select the fields you want displayed in your view and choose the format

Choose 1 or more fields:

ID  EntryName  Data Class  Accession Number  Primary Accession Number  Sequence Version  Creation Date

Display As:  Table  List

Sequence Formats:

9.)

10.)

Search Results							
Basic		Advanced Search				Actions	
		<input style="width: 100%; height: 25px; border: 1px solid #ccc; margin-bottom: 5px;" type="text" value="Cylinder forming"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Search"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Reset"/>					
Apply Options Bar		View/Print/Export Page		Account	Customer	Description	Sort By
<input type="checkbox"/> Selected items only <input checked="" type="checkbox"/> Unselected results only		<input type="radio"/> All pages <input type="radio"/> Current page				Cylinder forming	<input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Edit"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Delete"/>
<b>Result Options</b>		<input type="radio"/> All items in list <input checked="" type="radio"/> Details				Product sharing (No user cost)	<input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Edit"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Delete"/>
<input type="checkbox"/> Status: Active <input type="checkbox"/> Status: Inactive <input type="checkbox"/> Status: Deleted		<input type="radio"/> All pages <input checked="" type="radio"/> Current page				Product sharing (No user cost)	<input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Edit"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Delete"/>
<input type="checkbox"/> Sort by modified date <input checked="" type="checkbox"/> Sort by name		<input type="radio"/> All pages <input checked="" type="radio"/> Current page				Product sharing (No user cost)	<input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Edit"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Delete"/>
<input type="checkbox"/> Display Options		<input type="radio"/> All pages <input checked="" type="radio"/> Current page				Product sharing (No user cost)	<input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Edit"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Delete"/>
<input type="checkbox"/> Show results <input checked="" type="checkbox"/> Show details		<input type="radio"/> All pages <input checked="" type="radio"/> Current page				Product sharing (No user cost)	<input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Edit"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Delete"/>
<input type="checkbox"/> Set results per page <input checked="" type="checkbox"/> Set results per page		<input type="radio"/> All pages <input checked="" type="radio"/> Current page				Product sharing (No user cost)	<input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Edit"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Delete"/>
<input checked="" type="checkbox"/> Sort by name <input type="checkbox"/> Sort by modified date <input type="checkbox"/> Sort by status		<input type="radio"/> All pages <input checked="" type="radio"/> Current page				Product sharing (No user cost)	<input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Edit"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Delete"/>
<input type="checkbox"/> Print friendly view						Product sharing (No user cost)	<input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Edit"/> <input style="border: none; background-color: #f0f0f0; padding: 0 10px;" type="button" value="Delete"/>

**CLUSTALW-temp.job1**

Query found 1 entries

**Apply Options to:**

- selected results only
- unselected results only

**Result Options**

Launch analysis tool:  
ConsP

Show tools relevant to these results:

Link to related information:

Save results:

**Display Options**

View results using:  
+ Complete entries

Show 30  results per page

Printer friendly view

```

CALX_ANARCH ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_AZADIC ---MQRRIITTPVPLVLLVALLVPPVGSVCD---  

CALX_ASPPU ---MGRVRAITLGLAVSSVATMQGRANKEKTEKKADATL---  

CALX_BAERI ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_CANPA ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_NEUTU ---MGRGKDFCVCYQFCQCVLJICGTSICGA---  

CALX_HUMAN ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_MURIN ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_PRA ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_PUNAS ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_SCOPO ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_SOYIN ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_YEGST ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_ZEBRA ---MGRGKDFVPPFALLVPAVPPVQDYG---  

CALX_ANARCH ---DGTFLWLTSEPTDFTG---  

CALX_AZADIC ---SPTTLELTTTSDPPTG---  

CALX_ASPPU ---VWPHPTTPTTPEEFLPDTG---  

CALX_BAERI ---VWPHPTTPTTPEEFLPDTG---  

CALX_CANPA ---PPTTAAVLPCTTPEEFLPDTG---  

CALX_NEUTU ---S3DIAJTYEFPSKESFG---  

CALX_HUMAN ---PPTTAAVLPCTTPEEFLPDTG---  

CALX_MURIN ---PPTTAAVLPCTTPEEFLPDTG---  

CALX_PRA ---EVDTAATTTTSDPPTG---  

CALX_PUNAS ---EVDTAATTTTSDPPTG---  

CALX_SCOPO ---EVDTAATTTTSDPPTG---  

CALX_SOYIN ---EVDTAATTTTSDPPTG---  

CALX_YEGST ---EVDTAATTTTSDPPTG---  

CALX_ZEBRA ---EVDTAATTTTSDPPTG---

```

11.)

12.)

The screenshot shows the 'Phylogenetic tree printer' application. At the top, there are five icons representing different tree types: rooted, unrooted, circular, star-like, and regular. Below these are two tabs: 'Tree styles' and 'Tree printer'. The 'Tree printer' tab is active, displaying a tree structure with nodes labeled A through F. Each node has a 'Print' button next to it. To the right of the tree, there is a 'Newick' text input field containing the tree's phylogenetic information. Below this is a 'Print' button. Further down, there is a 'Create tree' section with a 'Search sample (in .nwk file)' input field, a 'Print' button, and 'Save' and 'Close' buttons. At the bottom, there is an 'Extra options' section with an 'Output' dropdown set to 'NEXUS', a note about URL handling, and a 'Print' button.

The phylogenetic tree illustrates the evolutionary relationships between calnexin protein sequences. The tree is rooted on the left and branches to the right. The sequences are labeled as follows:

- Top branch: CALX\_SOYBN
- Second branch: CALX\_PEA
- Third branch: CALX2\_BBTR
- Fourth branch: CALX1\_BBTR
- Large internal branch: CALX\_HELTU
- Sub-branch of CALX\_HELTU: CALX\_SCHPO
- Sub-branch of CALX\_HELTU: CALX\_ASPPU
- Sub-branch of CALX\_HELTU: CALX\_YEAST
- Large internal branch: CALX\_HUMAN
- Sub-branch of CALX\_HUMAN: CALX\_PONAB
- Sub-branch of CALX\_HUMAN: CALX\_CNRFA
- Sub-branch of CALX\_HUMAN: CALX\_RAT
- Sub-branch of CALX\_HUMAN: CALX\_MOUSE
- Bottom branch: CLGN\_MOUSE

# Methods for Analysis of EST Data

The Practical exercises discussed below will help us analyze EST data available publicly. We will analyze the High-throughput data for an agricultural crop, *Triticum aestivum* for being specific to one organism related to agriculture. Various tools available online will be used to retrieve, analyze and annotate the EST data.

ESTs are short (200-800 nucleotide bases in length), unedited, randomly selected single-pass sequence reads derived from cDNA libraries. High-throughput ESTs can be generated at a reasonably low cost from either the 50 or 30 end of a cDNA clone to get an insight into transcriptionally active regions in any organism.

1.)

This screenshot shows the NCBI Nucleotide database search interface. The search term 'EST' has been entered. The results page displays information about the dbEST database, which is a division of GenBank containing expressed sequence tags. It also provides a brief history of human ESTs in GenBank and links to the 'Genome Directory' issue of Nature. The main search results table includes columns for ID, Sequence, and Description, showing several entries for T. aestivum.

2.)

This screenshot shows a detailed view of an EST entry from the NCBI Nucleotide database. The entry is for 'M1126922 SSH cDNA library of T. aestivum cv. Chuanmai 28 Three pistils mutation Trichium aestivum cDNA mRNA sequence'. The details page includes sections for Identifiers, Clone Info, Features, and Sequence. The sequence is shown in FASTA format. A right-click context menu is open over the sequence, with the 'Send to' option highlighted.

3.)

This screenshot shows a detailed view of an EST entry from the NCBI Nucleotide database. The entry is for 'M1126922 SSH cDNA library of T. aestivum cv. Chuanmai 28 Three pistils mutation Trichium aestivum cDNA mRNA sequence'. The details page includes sections for Identifiers, Clone Info, Features, and Sequence. The sequence is shown in FASTA format. A right-click context menu is open over the sequence, with the 'Send to' option highlighted.

4.)

This screenshot shows a detailed view of an EST entry from the NCBI Nucleotide database. The entry is for 'M1126922 SSH cDNA library of T. aestivum cv. Chuanmai 28 Three pistils mutation Trichium aestivum cDNA mRNA sequence'. The details page includes sections for Identifiers, Clone Info, Features, and Sequence. The sequence is shown in FASTA format. A right-click context menu is open over the sequence, with the 'Send to' option highlighted.

# Working with a Single Protein Sequence

This exercise is concerned with various online tools available to study a protein before designing a new experiment in the lab. It is shown where you can find these tools, how to use them and most importantly how to interpret the results.

**Part I:** Predicting the main physico-chemical properties of a protein

**Part II:** Digesting a protein in a computer

Protease digestions — where you use an enzyme to cut your protein in specific ways — can be useful if you're only interested in carrying out experiments on a portion of your protein.

**Part III:** Primary Structure Analysis

A primary structure analysis is conducted to find segments in a protein that display a special composition. These segments can reveal some interesting properties of your protein.

**Note :** For the following analysis, obtain the protein sequence Q39817 (CALX\_SO4BN) from SwissProt.

**Part IV:** Predicting post-translational modifications in your protein.

**Part V:** Finding Known Domains in your Protein

**Note:** In many interesting domains, a particular type of amino acid is overrepresented. For instance, there are more leucines than expected by chance in the leucine zippers, or more glycines than expected by chance in the glycine-rich domains — and so on — for many domains. Because repeated residues make a sequence simpler, sequences that contain them are described as low-complexity. If you keep the Apply Low Complexity Filter box checked, you may lose these domains.

1.)

Enter the SwissProt ID for WGA protein at <http://www.expasy.ch/sprot/>

Predicting the main physico-chemical properties of a protein

The page shows the EXPASY Proteomics Server logo and links to Swiss-Prot, TrEMBL, and UniProt databases.

2.)

Retrieve sequence for WGA in Fasta format

The page shows the UniProtKB interface with the search term "WGA" entered. The results show a single entry for "Reviewed, UniProtKB/Swiss-Prot P10968 (A9H1\_WHEAT)".

Details for P10968 (A9H1\_WHEAT):  
Last modified March 2, 2010, Version 96, [Details...](#)  
Names and origin: Protein attributes - General annotation (Comments), Ontologies, Sequence annotation (Features), Sequences, References - Cross-references  
Protein names: Recommended name: Agamidin-like 1; Alternative name(s): WGA1; Isoform A;  
Organism: Triticeum aestivum (Wheat); Taxonomic identifier: 4565 (NCB)

3.)

The screenshot shows the ProtParam tool interface. At the top, it displays the Swiss Institute of Bioinformatics logo and the ExPASy Proteomics Server header. Below this, the URL is shown as "You are here: ExPASy > Tools > Primary structure analysis > ProtParam". The main content area is titled "ProtParam" and "Selection of endpoints on the sequence". It shows the protein identifier "Agt\_WHEAT (P10968)". A note states: "Please select one of the following features by clicking on a pair of endpoints, and the computation will be carried out for the corresponding sequence fragment." A note below says: "Note: Only the features corresponding to subsequences of at least 5 residues are highlighted." A table lists various protein features with their start and end positions and descriptions:

FT	SEIGRAD	1-26	Agglutinin isolectin 1.
FT	CHAHN	27-197	
FT	PROSP	198-212	Chitin-binding type-1 1.
FT	DOMAIN	27-68	Chitin-binding type-1 2.
FT	DOMAIN	69-111	Chitin-binding type-1 3.
FT	DOMAIN	112-154	Chitin-binding type-1 4.
FT	REGION	155-197	Substrate binding.
FT	REGION	36-38	

Below the table, a note says: "The sequence motif AGTAVATT consists of 212 amino acids." A "RESET" button is present.

4.)

The screenshot shows the PeptideCutter tool interface. At the top, it displays the Swiss Institute of Bioinformatics logo and the ExPASy Proteomics Server header. Below this, the URL is shown as "You are here: ExPASy > Tools > Other proteomic tools > PeptideCutter". The main content area is titled "PeptideCutter" and "PeptideCutter references / documented peptide cleavage sites cleaved by proteases or chemicals in a given protein sequence. PeptideCutter returns the query sequence with the possible cleavage sites mapped on it and/or a table of cleavage site positions." A text input field contains the identifier "P10968". A "Perform" button is present at the bottom.

5.)

The screenshot shows the ProtScale tool interface. At the top, it displays the Swiss Institute of Bioinformatics logo and the ExPASy Proteomics Server header. Below this, the URL is shown as "You are here: ExPASy > Tools > Primary structure analysis > ProtScale". The main content area is titled "Please, select" and "Please indicate the way you would like the cleavage sites to be displayed". It includes sections for selecting enzymes and chemicals, and for specifying display preferences. A note says: "Please indicate which enzymes to include in the display." A "Perform" button is present at the bottom.

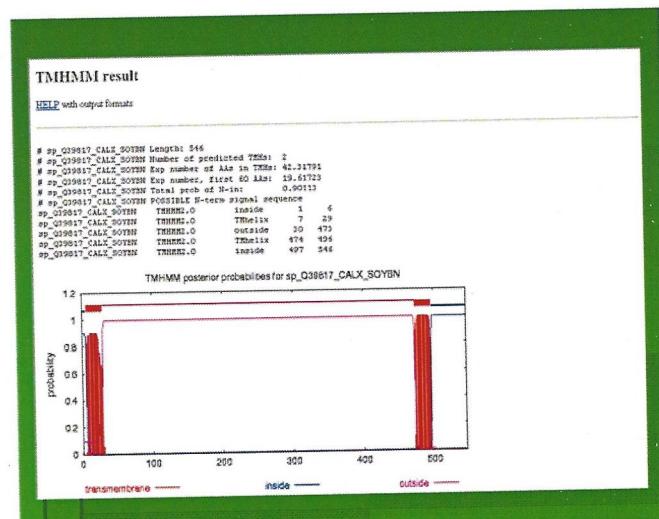
6.)

The screenshot shows the TMHMM result interface. At the top, it displays the Swiss Institute of Bioinformatics logo and the ExPASy Proteomics Server header. Below this, the URL is shown as "You are here: ExPASy > Tools > Primary structure analysis > ProtScale". The main content area is titled "Looking for transmembrane segments (TMS)" and "An Example Protein for TMS detection UniProtKB/Swiss-Prot ID Q39817 (CALX\_SOYBN)". A note says: "Calnexin is a calcium-binding protein that interacts with newly synthesized glycoproteins in the endoplasmic reticulum. It may act in assisting protein assembly and/or in the retention within the ER of unassembled protein subunits. It seems to play a major role in the quality control apparatus of the ER by the retention of incorrectly folded proteins." A text input field contains the identifier "Q39817". A note says: "Or you can paste your own sequence in the box below." A "Perform" button is present at the bottom.

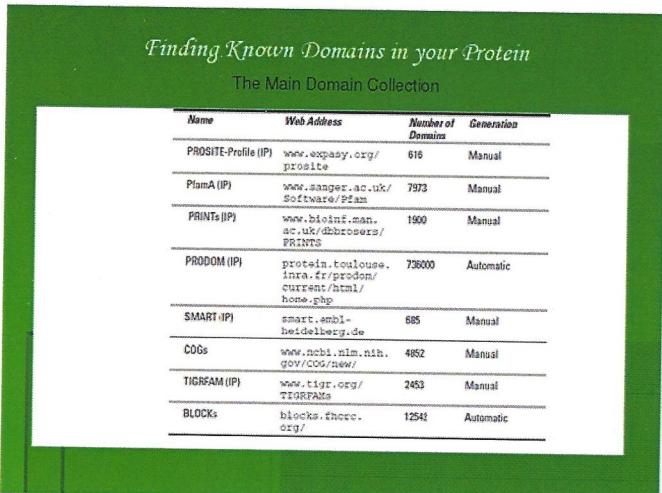
7.)

The screenshot shows the TMHMM Server v. 2.0 interface. At the top, it displays the Center for Biotechnology and Bioinformatics logo and the TMHMM Server v. 2.0 header. Below this, the URL is shown as "You are here: TMHMM > TMHMM Server v. 2.0". The main content area is titled "Prediction of transmembrane helices in proteins". A note says: "NOTE: You can submit many proteins at once in one fasta file. Please limit each submission to at most 4500 proteins. Please tick the 'One line per protein' option. Please leave space between each large submission." A "Submit" button is present at the bottom.

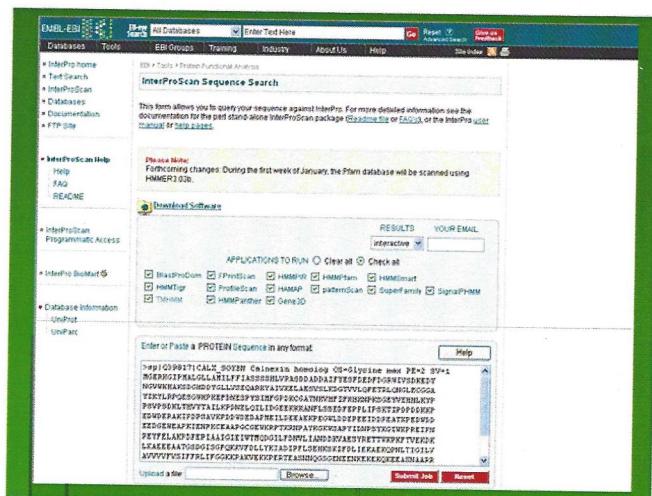
8.)



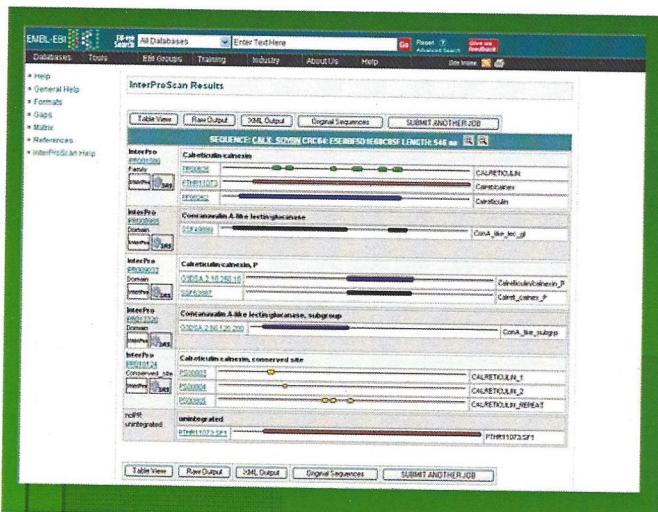
9.)



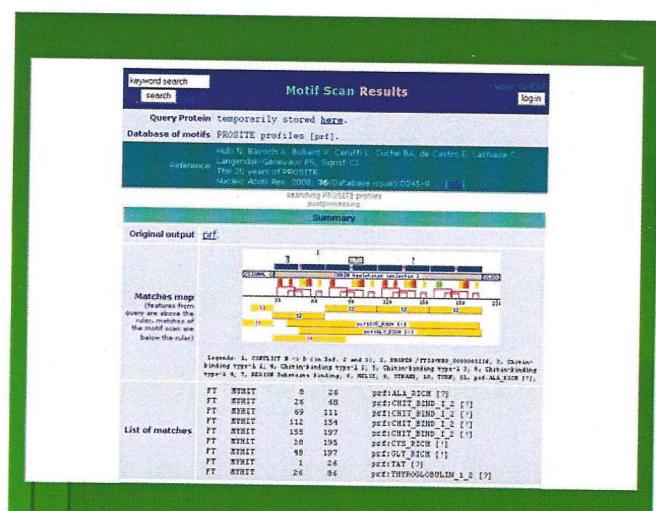
10.)



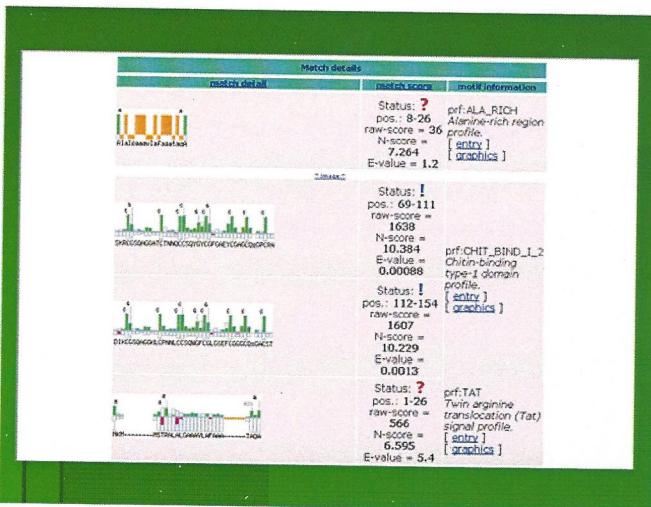
11.)



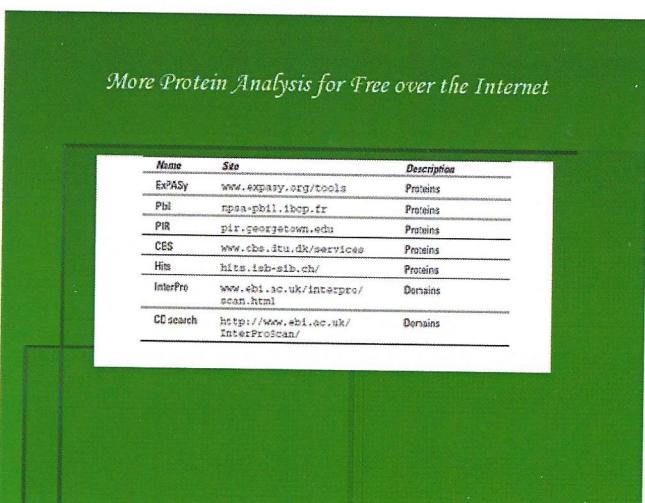
12.



13.)



4.)

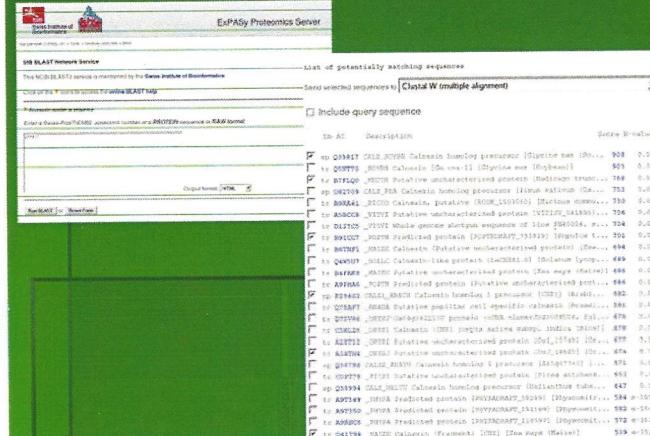


# Protein Family Analysis & Phylogenetics

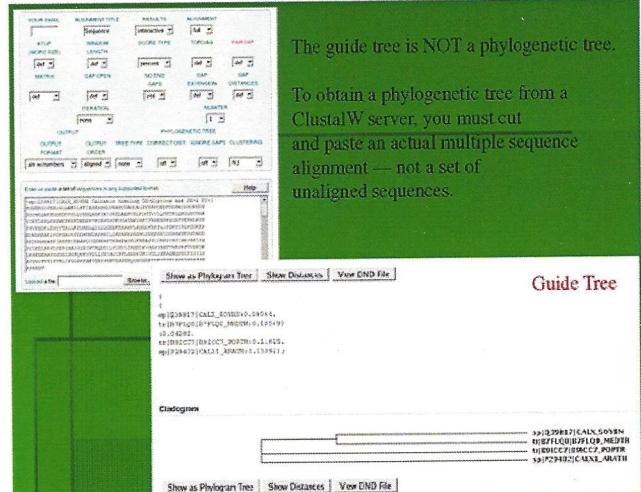
1.)

Gathering sequences with online BLAST servers  
<http://www.expasy.ch/tools/blast/>

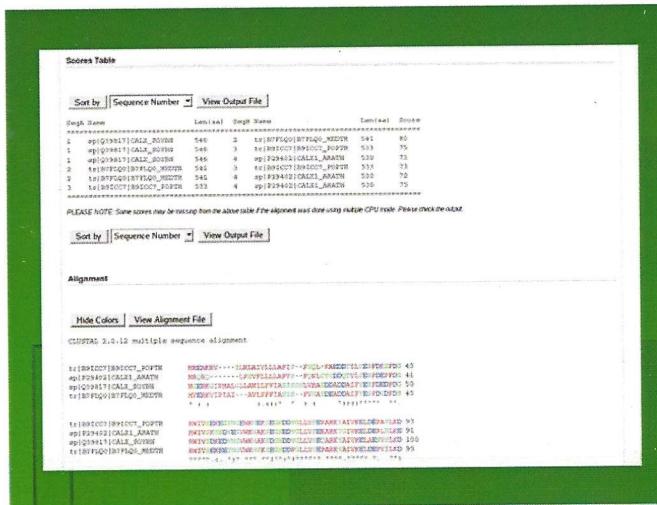
<http://www.expasy.ch/tools/blast/>



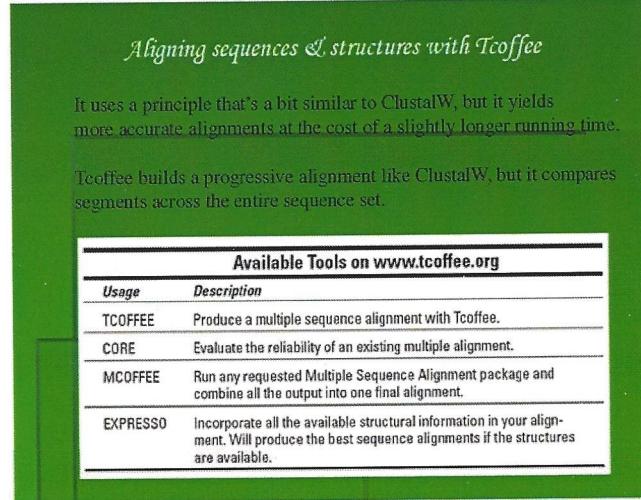
2.)



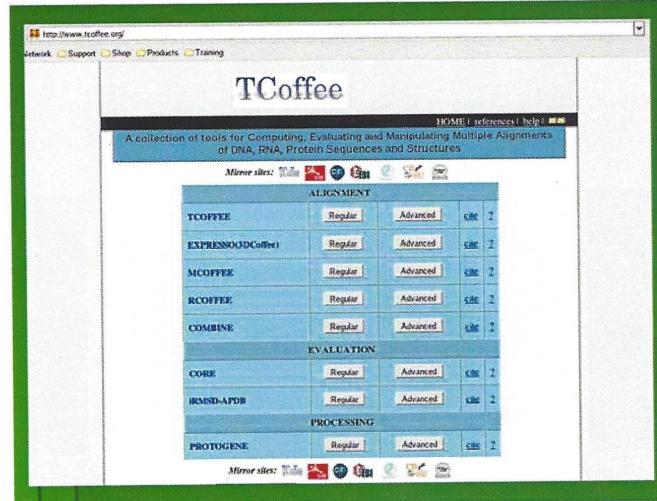
3.)



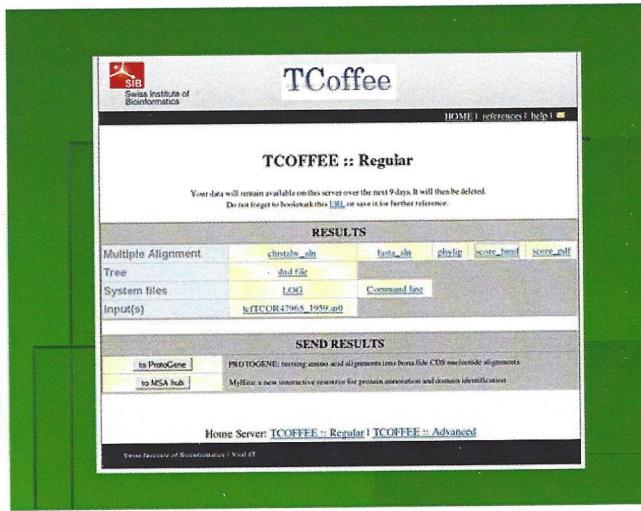
4.1



5.)



6.



7.)

**TCoffee**

---

**CORE::Regular**

**Results**: Your alignment and output. A colored version where unaligned positions are in **blue** and the reliable ones in **red**. The aligned motif contains about 400 sequences.

**Unaligned**: Maximum number of sequences is 16  
Minimum weight of sequences is 0.000

**Multiple Alignment**

**Output**: Multiple sequence alignment file (.aln), FASTA format (.fa), ClustalW format (.clw), XML format (.xml), and PDF (.pdf).

**Sequence Data**

**Sequence ID**:  **Sequence Name**:

**nr protein data**

**Sequence ID**:  **Sequence Name**:

**You may paste your e-mail address:**

**The Institute of Bioinformatics (IUB)**

**TCoffee**

---

**CORE :: Regular**

Your data will remain available on this server over the next 8 days. It will then be deleted.  
Do not forget to bookmark this [URL](#), or save it for further reference.

<b>RESULTS</b>	
Multiple Alignment	<a href="#">clustalw.aln</a>
System files	<a href="#">LOG</a>
Input(s)	<a href="#">tCOPR97493_13334.m0</a>
	<a href="#">FASTA_fn</a>
	<a href="#">phylog</a>
	<a href="#">score.html</a>
	<a href="#">score.pdf</a>

**Home Server: CORE - Regular | CORE :: Advanced**

The Institute of Bioinformatics - Page 12

8.

```
T-COFFEE Version.7.71 (Mon Feb 23 21:41:54 WEST 2009)
Cedric Notredame
CPU TIME: sec.
SCORES: 54

  100% IDENTICAL
sp|Q39817|CALX          : 55
tr|B7FLG01|B7FLG           : 53
tr|B59CC7|B59CC           : 53
sp|P29462|CALXI          : 54
tr|A3ATX4|A3ATX           : 54
tr|Q41798|Q4179           : 54
cons                         : 54

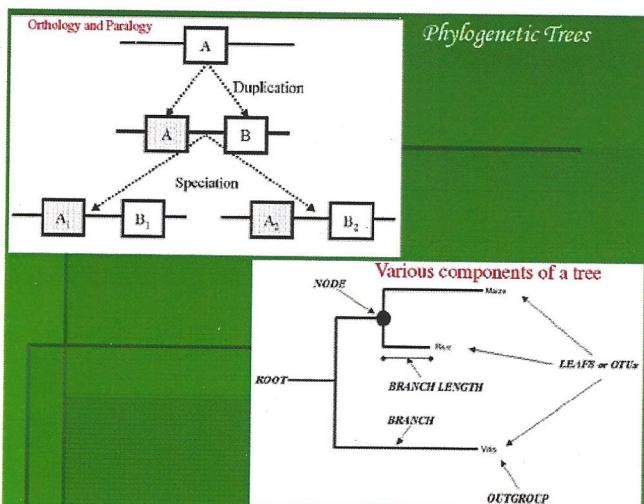
sp|Q39817|CALX          : MTCRIGGIDPMALVLLANLTTTAAAGGTTTGGG
tr|B7FLG01|B7FLG           : MTCRIGGIDPMALVLLANLTTTAAAGGTTTGGG
tr|B59CC7|B59CC           : MTCRIGGIDPMALVLLANLTTTAAAGGTTTGGG
sp|P29462|CALXI          : MTCRIGGIDPMALVLLANLTTTAAAGGTTTGGG
tr|A3ATX4|A3ATX           : MTCRIGGIDPMALVLLANLTTTAAAGGTTTGGG
tr|Q41798|Q4179           : MTCRIGGIDPMALVLLANLTTTAAAGGTTTGGG
cons                         : MTCRIGGIDPMALVLLANLTTTAAAGGTTTGGG

  100% IDENTICAL
sp|Q39817|CALX          : 55
tr|B7FLG01|B7FLG           : 53
tr|B59CC7|B59CC           : 53
sp|P29462|CALXI          : 54
tr|A3ATX4|A3ATX           : 54
tr|Q41798|Q4179           : 54
cons                         : 54

sp|Q39817|CALX          : 55
tr|B7FLG01|B7FLG           : 53
tr|B59CC7|B59CC           : 53
sp|P29462|CALXI          : 54
tr|A3ATX4|A3ATX           : 54
tr|Q41798|Q4179           : 54
cons                         : 54

sp|Q39817|CALX          : GRWIVRSOKEDZYDNWYTHAKSGDQEDSYGLLYLQVQARYTAIVKELAE
tr|B7FLG01|B7FLG           : GRWIVRSOKEDZYDNWYTHAKSGDQEDSYGLLYLQVQARYTAIVKELAE
tr|B59CC7|B59CC           : GRWIVRSOKEDZYDNWYTHAKSGDQEDSYGLLYLQVQARYTAIVKELAE
sp|P29462|CALXI          : GRWIVRSOKEDZYDNWYTHAKSGDQEDSYGLLYLQVQARYTAIVKELAE
tr|A3ATX4|A3ATX           : GRWIVRSOKEDZYDNWYTHAKSGDQEDSYGLLYLQVQARYTAIVKELAE
tr|Q41798|Q4179           : GRWIVRSOKEDZYDNWYTHAKSGDQEDSYGLLYLQVQARYTAIVKELAE
cons                         : GRWIVRSOKEDZYDNWYTHAKSGDQEDSYGLLYLQVQARYTAIVKELAE
```

9.)



10.)

Initial Alignment									
Column	1	2	3	4	5	6	7	8	9
seq1	A	B	C	D	E	F	G	H	I
seq2	A	A	B	C	B	A	C	A	A
seq3	C	C	A	C	B	A	C	A	B

Bootstrap Alignment 1					Bootstrap Alignment 2				
1	1	8	1	2	5	1	8	2	1
AAH	A	B	E	A	H	B	A	D	E
AA	C	A	B	C	A	C	A	G	A
CC	A	C	C	B	C	A	C	A	C

11.)

121

# PROGRAM DETAILS FOR THE NATIONAL WORKSHOP ON 'Use of Bioinformatics in Agriculture & Plant Sciences'

12-13th March 2010

DAY 01 (12th March 2010)

## INAUGURAL SESSION

9:30 am	REGISTRATION	Reception, NIPGR
10:00 am	WELCOME ADDRESS	Dr. G. Yadav, Coordinator, Sub-DIC
10:05 am	KEYNOTE ADDRESS	Dr. T. Madhan Mohan, DBT
10:10 am	INAUGURAL ADDRESS	Prof. R Geeta, University of Delhi
11:00 am	VOTE OF THANKS	Coordinator, Sub-DIC
11:05 am	HIGH TEA	

## WORKSHOP SESSION - I

11:30 am	BIOINFORMATICS IN AGRICULTURE	Dr. G. Yadav, NIPGR
12:15 pm	BIOLOGICAL DATA RETRIEVAL & ANALYSIS: DNA	Sachin Pandit
14:00 pm	LUNCH	

## WORKSHOP SESSION - II

14:30 pm	PROTEIN SEQUENCE ANALYSIS: WHEAT PROTEIN	Smriti Shridhar
15:45 pm	TEA	
16:00 pm	BIOINFORMATICS & THE PEDAGOGY OF PLANT SCIENCES	Dr. S.M. Leighton, University of Delhi

DAY 02 (13th March 2010)

## WORKSHOP SESSION - III

9:30 am	METHODS FOR ANALYSIS OF EST DATA	Sachin Pandit
11:00 am	TEA	
11:15 am	NEXT GEN SEQUENCING & APPLICATIONS	Dr. D. Chattopadhyay, NIPGR
13:00 pm	LUNCH	

## WORKSHOP SESSION - IV

14:00 pm	3-D STRUCTURAL ANALYSIS OF PROTEINS	Daljit Singh
15:45 pm	TEA	
16:00 pm	PROTEIN FAMILY ANALYSIS & PHYLOGENETICS	Smriti Shridhar

## VALEDICTORY FUNCTION

17:30 pm	CERTIFICATE DISTRIBUTION	Prof. A.K Tyagi, DIRECTOR, NIPGR
17:45 pm	CLOSING CEREMONY	

**PARTICIPANTS OF THE NATIONAL WORKSHOP ON**  
**'Use of Bioinformatics in Agriculture & Plant Sciences'**

12-13th March 2010

Name of Participant	Affiliation
1 Ravi Kant	Haryana Agricultural University
2 Poonam Sharma	CCS Haryana Agricultural University
3 Harmeet Kaur	NIPGR
4 Dr. Neelam R. Yadav	CCS Haryana Agricultural University
5 Rehna Augustine	NIPGR
6 Pooja Verma	NIPGR
7 Rahul Gautam	National Institute of Technology
8 Renu Kamari	NIPGR
9 Swati Chaudhary	NIPGR
10 Ashutosh Pandey	Shobhit University
11 Anurag Srivastava	Shobhit University
12 Abhay Pratap	Shobhit University
13 Pritika Singh	GNDU
14 Dr Parveen Chhuneja	Punjab Agricultural University
15 Khagendra Kumar	National Institute of Technology
16 Praveena K.	NIPGR
17 Meenu	NIPGR
18 Dr. Sridhar Gutam	Indian Agricultural Research Institute
19 Bhana Prakash Petla	NIPGR
20 Tanima Datta	GNDU
21 Harsimran Kaur	GNDU

**THE NATIONAL WORKSHOP ON  
Use of Bioinformatics in Agriculture & Plant Sciences'**

12-13th March 2010

**Organisers**

Dr. Gitanjali Yadav

Coordinator, NIPGR

Sachin Pandhir

NIPGR

Smriti Shridhar

NIPGR

Subhasish Mondal

NIPGR

**Speakers**

Prof. R. Geeta

University of Delhi

Dr. Sudeshna M. Leighton

University of Delhi

Dr. Gitanjali Yadav

Coordinator, NIPGR

Dr. Debasis Chattopadhyay

NIPGR

