



Bioinformation up to Date

(Bioinformatics Infrastructure Facility, Biotechnology Division)

North-East Institute of Science & Technology

CONTENTS

Cover Story 1

Genomics 2

Bioserver 2

Bioinfo Quiz 3

Special Edition 3

Proteomics 3

Software Mania 4

Molecule of the

Month 4

Bioinfo Cartoon 4

Contact us 4

COVER STORY: [Human Microbiome Project](#)

Within the body of a healthy adult, microbial cells are estimated to outnumber human cells by a factor of ten to one. These communities, however, remain largely unstudied, leaving almost entirely unknown their influence upon human development, physiology, immunity, and nutrition. To take advantage of recent technological advances and to develop new ones, the NIH Roadmap has initiated the Human Microbiome Project (HMP) with the mission of generating resources enabling comprehensive characterization of the human microbiota and analysis of its role in human health and disease.



By leveraging both the metagenomic and traditional approach to genomic DNA sequencing, the Human Microbiome Project will lay the foundation for further studies of human-associated microbial communities. Broadly, the project has set the following goals:

- Determining whether individuals share a core human microbiome
- Understanding whether changes in the human microbiome can be correlated with changes in human health
- Developing the new technological and bioinformatic tools needed to support these goals

Addressing the ethical, legal and social implications raised by human microbiome research

Advisor:

Dr. P. G. Rao

Editors:

Dr. R. L. Bezbaruah

Alakesh Das

Sanchaita Rajkhowa

UPCOMING EVENTS

Instituted By
Department of Biotechnology
Government of India
New Delhi



Bioinformatics National Certification

Examination

Monday, December 27, 2010 | 3 : 53 : 55 PM

[Home](#)

[Eligibility & Syllabus](#)

[Application Instructions](#)

[Apply](#)

[Important Dates](#)

[Centres](#)

[FAQs](#)

Registration starts: 13 December 2010 | Registration closes: 03 February 2011 | Examination: 26-27 Febru

BIOSERVER:

ExPASy Proteomics Server

ExPASy is a home to a number of databases specializing in different aspects of protein molecular biology and proteomics

The screenshot shows the ExPASy Proteomics Server interface. At the top, there is a search bar and a navigation menu. The 'Translate tool' is selected, and the 'Translate' button is highlighted in red. The main content area shows a text input field with a DNA sequence, an 'Output format' dropdown set to 'Verbose', and a list of protein sequences with their start and stop codons.

Select only one of the following frames:

- 53' Frame 1
LRPWECCSLCLKLPQVSDPSFASIMet SLLPCHPLNRAIQPSYKPPTPDLALLSSFAWSMet FLVLFVTFNCStop HTGSLIFHGLLYIVLFPWCWVNSFLRAVIFYLHPQDPVQCPAHSKCLISA
ELCTPSSQLWAGTERVFLSEGStop MetNELMetNCSCWE
- 53' Frame 2
SDLGSAVPSAStop NSPKSHMet TPPLPPLCLCSHVLTSTGSSLLTNLPPQIWSHStop APLPGLCFSLStop HFLTASTPDHStop SFMetVYCISSYSPAGMet SASStop GQStop SFICIPRIQYSA
QHIVSAStop YLLNYVHPAVSSGQAQRGYFStop VKAEStop MetNStop StopIVLAGK
- 53' Frame 3
QTLGVLFPLPETPSSLIStop PLLCLHYVSAPMet SPSQQGHFALQTSHPRSCTPELLCLVYVSLCPNIFStop LLAHRIIDLWSFIVYRLIPLLECCQLLKGSDLLSAPGSSVTPSTStop StopVLD
ICStopIMetYTQOQALGRHREGISEStop RLNEStopINELFLLG
- 35' Frame 1
FPSKNNSLIHFSLSHSEIPSLCPRADCWYIIQISSYTYVLTGVLDPGDADKRSLPLRSStop HSSRGIRRYTINHERSMetIRCASSStop KMetLGQRETStop TRQRSSGVPDLGWEVCKKA
GWPCStopEGDStop MetGAETStop WRQRRGHMetRLGGVSGRGNSTPKVStop
- 35' Frame 2
FPARTIHStop FIHSFTQKYLPCACPETAAGCTStopFSRYQALTMetCWALYWLGMetQIKDHPStopEADIPAGEStopDDIQStopTMetKDOStopSGVLAVKCKStopDKEKHRPGKGAOECOIW
GGRFYRRLDGPVERVTWEQRHNGKGGVIStopDLGEFOAEGTALPRSE
- 35' Frame 3
SQOQDFINSFIQPSLRNLTSLVPAOSStopLLGVHNSADIKHLLCAGHCTGSWGCRStopKITALKKLTFOQGNKIYNKPStopKINDPVCStopOLKNVRTKRNI DOAKELRSARSGVGLStopEG
WMetALLRGStopHGSROIMetEAKESGSETWGSFRQREQHSQGL

GENOMICS: PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses

Whole-genome association studies (WGAS) bring new computational, as well as analytic, challenges to researchers. Many existing genetic-analysis tools are not designed to handle such large data sets in a convenient manner and do not necessarily exploit the new opportunities that whole-genome data bring. To address these issues, we developed PLINK, an open-source C/C++WGAS tool set. With PLINK, large data sets comprising hundreds of thousands of markers genotyped for thousands of individuals can be rapidly manipulated and analyzed in their entirety. As well as providing tools to make the basic analytic steps computationally efficient, PLINK also supports some novel approaches to whole-genome data that take advantage of whole-genome coverage. We introduce PLINK and describe the five main domains of function: **data management, summary statistics, population stratification, association analysis, and identity-by-descent estimation.** In particular, we focus on the estimation and use of identity-by-state and identity-by-descent information in the context of population-based whole-genome studies. This information can be used to detect and correct for population stratification and to identify extended chromosomal segments that are shared identical by descent between very distantly related individuals. Analysis of the patterns of segmental sharing has the potential to map disease loci that contain multiple rare variants in a population-based linkage analysis

PROTEOMICS: Trans-Proteomic Pipeline

The Trans-Proteomic Pipeline (TPP) is a suite of software tools for the analysis of MS/MS data sets. The tools encompass most of the steps in a proteomic data analysis workflow in a single, integrated software system. Specifically, the TPP supports all steps from spectrometer output file conversion to protein-level statistical validation, including quantification by stable isotope ratios. We describe here the full workflow of the TPP and the tools therein, along with an example on a sample data set, demonstrating that the setup and use of the tools are straightforward and well supported and do not require specialized informatic resources or knowledge

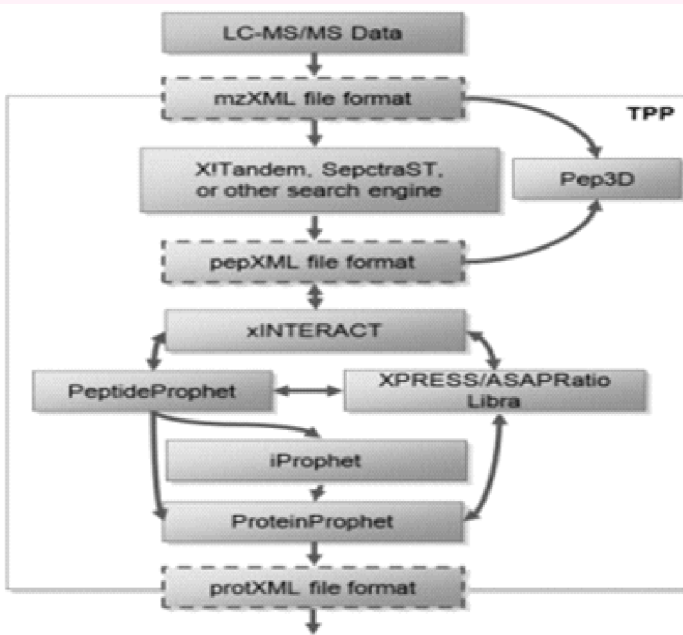


Figure 1. Schematic overview of the TPP workflow. Raw MS/MS data files are first converted to an open XML format such as mzXML or mzML, then analyzed with a search engine, either embedded in the TPP or used externally. Pep3D can allow visualization of the data. The search results, in pepXML format, are processed with tools PeptideProphet for initial spectrum-level validation, iProphet for peptide-level validation, and finally ProteinProphet for protein-level validation and final protein inference. Quantification tools like XPRESS, ASAPRatio, or Libra can be used on labeled data. The final output is protXML, which can be imported into a variety of analysis tools.

BIO-INFY Quiz:

1. Crossing -over can occur between homologues during :
 - a. zygotene
 - b. pachytene
 - c. leptotene

2. Part of mitosis associated with separation of chromatids:
 - a. Interphase
 - b. prophase
 - c. anaphase

3. Lac operon induced by
 - a. lactose
 - b. promoter gene
 - c. lac I

4. which amino acid have got maximum no of codons
 - a. Arginine
 - b. Methionine
 - c. tryptophan

SPECIAL INTEREST: Ants lead way to speedier computer network

An analysis of how ants quickly find new routes in a changing maze reveals techniques that could be useful to systems engineers.

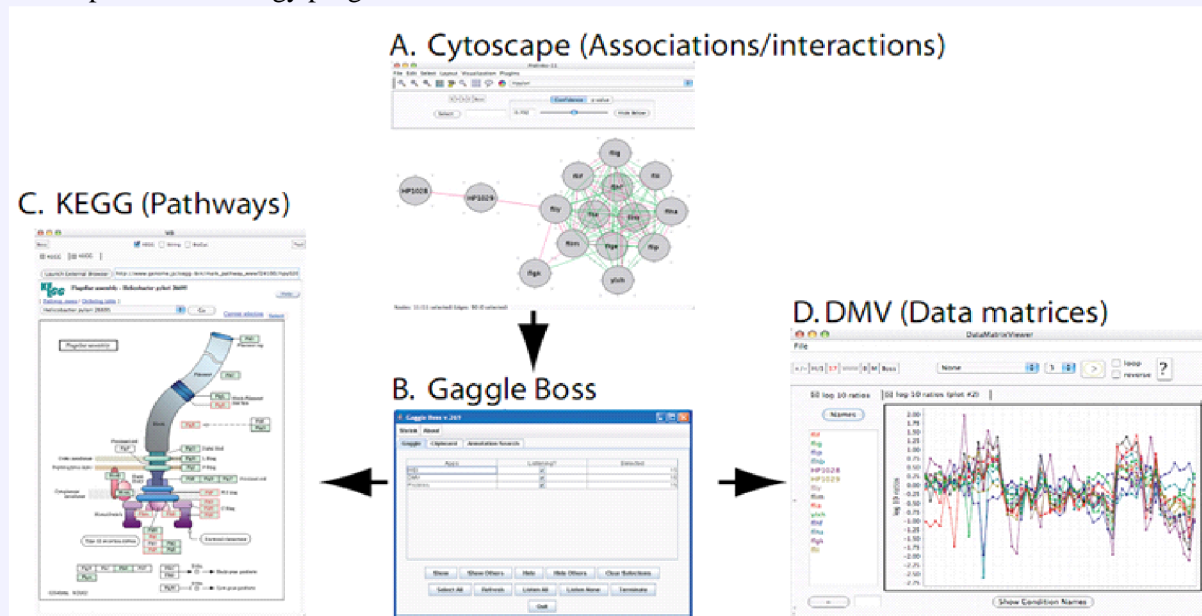
The research, reported in the *Journal of Experimental Biology*, shows that Argentine ants (*Linepithema humile*), do not just retrace their steps when presented with a barrier as might be expected. Instead, the ants begin a localized search that seems to take into account the direction in which they were planning to go. Because there are many network-management programs that mimic the search behaviour of this ant species, systems engineers are taking notice and wondering what they can learn.



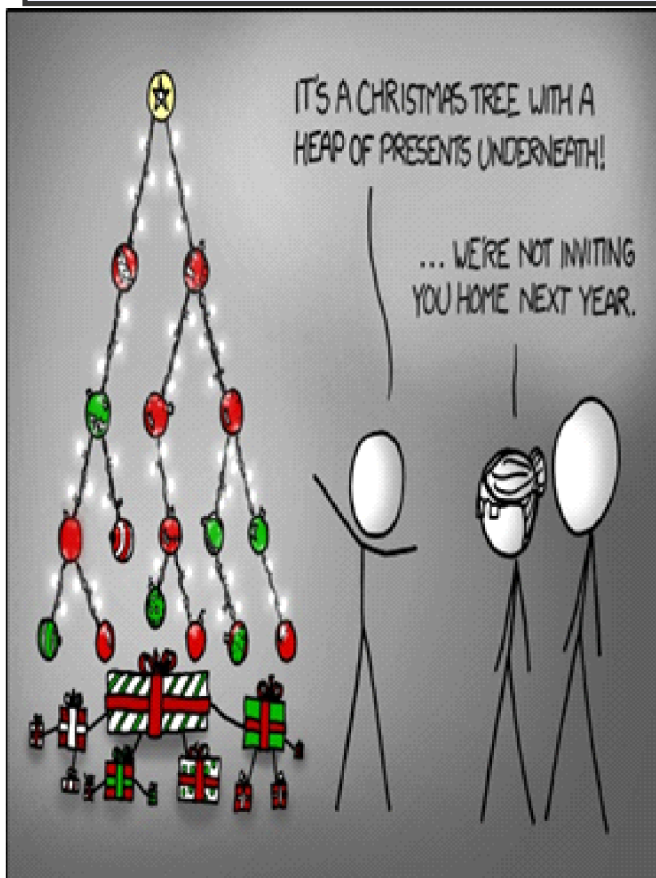
In the wild, ant scouts deposit pheromones along the trails between food and the nest. Nest mates to follow the trail, laying their own pheromones, amplifying the markings along the path. Because the pheromones gradually evaporate, longer trails on which ants are spread more thinly carry lower pheromone concentrations than short trails. Since pheromone strength is what draws an ant to follow a specific path, longer trails that have weaker pheromones are abandoned in favour of shorter ones

SOFTWARE MANIA: **Gaggle**

The Gaggle uses a minimalist approach to integrated data and software. It is written in Java and uses standard Java libraries. A small server program (the `-Gaggle Boss`) provides communication among analysis and display programs (the `-geese`) which are modest adaptations of existing (or novel) bioinformatics and computational biology programs and web resources

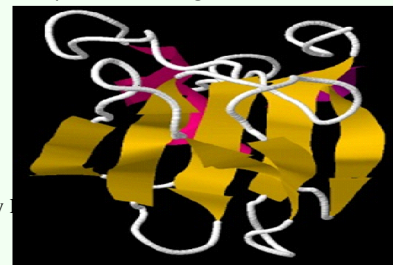


A simple introductory example for use of Gaggle. A set of genes (circular nodes with edges represents associations/interactions) selected in Cytoscape (A) are broadcasted to the Gaggle Boss (B). The Gaggle Boss re-routes the broadcast to a Java web browser connected to KEGG (C), further exploration wherein localizes *H. pylori* proteins to relevant subunits in the flagellar apparatus map. A second goose that receives the broadcast is the DMV (D). A plot function therein provides mRNA levels of the 15 *H. pylori* genes in 57 experimental conditions.



MOLECULE OF MONTH

Hepatocyte growth factor (HGF) is a multipotent growth factor that transduces a wide range of biological signals, including mitogenesis, motogenesis, and morphogenesis. The N terminal (N) domain of HGF, containing a hairpin-loop region, is important for receptor binding and the potent biological activities of HGF. The N domain is also the primary binding site for heparin or heparan sulfate, which enhances, receptor /ligand oligomerization and modulates receptor-dependent mitogenesis



PDB ID : 2HGF
Amino acids : 97
Exp. Method : X-Ray
Chains : 1 A
Type: Polypeptide L

For suggestions & contributions contact:

Sanchaita Rajkhowa (sanch.raj2008@gmail.com)

Alakesh Das (adas1757@gmail.com)

BIF, NEIST, JORHAT

ANSWERS: 1-b,2-c,3-a,4-a