



AGRIBIOINFORMATICS



<http://www.iari.res.in/divisions/usi/agribio.php>

Vol. 2, No 3 & 4

A Quarterly Newsletter for Disseminating ICT in Agriculture and Rural Development

July-December 2009



डॉ. अरविन्द कुमार
उप महानिदेशक (शिक्षा)

Dr. Arvind Kumar
Deputy Director General (Edn.)

भारतीय कृषि अनुसंधान परिषद
कृषि अनुसंधान भवन-II, पूसा, नई दिल्ली 110 012

INDIAN COUNCIL OF AGRICULTURAL RESEARCH

Krishi Anusandhan Bhavan-II, New Delhi -110 012
(Education Division)

Tel: 011-25841760; Fax: 011- 25843932;
Email: ddgedn@icar.org.in; Website: www.icar.org.in

MESSAGE



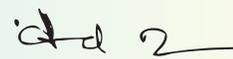
Bioinformatics is an interdisciplinary approach requiring sophisticated computer science, mathematics, and statistical methods, with a deep understanding of the biological and chemical context, problems, and methodology behind the data. It encompasses the storage, retrieval, and analysis of gene sequence, biological, pharmacological, and structural data, prediction of protein structure, function, and post-translational modification. It could have profound impact in fields as varied as from human health, agriculture, environment, energy and biotechnology to advance biomedical research and development. It has environmental benefits in identifying waste cleanup bacteria and in agriculture it can be used for producing high yield, low maintenance crops. The major impact of bioinformatics research has been to automate the genome sequencing, automated development of integrated genomics and proteomics databases, genome comparisons to identify the genome function, derivation of metabolic pathways, gene expression analysis to derive regulatory pathways, identification of candidate genes for vaccines and anti-microbial agents, and the whole genome comparison to understand the microbial evolution.

Education and training in bioinformatics must go beyond acquiring the skills to use existing tools. Bioinformaticists need to have a very strong understanding of the methods they must use and understand clearly the strengths and weaknesses of the tools.

I am happy to note that Unit of Simulation and Informatics, Indian Agricultural Research Institute is bringing out this newsletter from a grant received from Department of Biotechnology, Government of India. The information contained in this series must be of immense benefit to the scientists, teachers and students.

My congratulations to the Director, the Editor and the editorial board for this service.

Date: 23rd December, 2009
New Delhi


(Arvind Kumar)

CONTENTS

In Focus.....	2
Bioinformatics.....	3
Crop Bioinformatics.....	5
Pest Bioinformatics.....	6
Chemoinformatics.....	6
News Roundup.....	9
Online Bioinformatics Education.....	10
Obituary.....	12

EDITORIAL BOARD

Patrons

Dr H S Gupta, Director, IARI
Dr T Madan Mohan, Advisor, DBT, Govt. of India

Editor

Dr C Devakumar, ADG, ICAR

Members

Dr P Ananda Kumar, Director, NRCPB
Dr H Chandrasekharan, IARI
Dr P S Pandey, IARI
Dr Sridhar Gutam, IARI
Dr A Sarangi, IARI

In Focus

Open Access: Relevance to National Agricultural Research System

(Sridhar Gutam, H. Chandrasekharan and C. Devakumar)

According to the Agricultural Science and Technology Indicators (ASTI) Policy brief published by National Center for Agricultural Economics and Policy (NCAP) in 2008, India ranks fourth in terms of total investments in public agricultural R&D in the world following United States, Japan, and China. The agricultural R&D generates a lot of research data, which is distributed by the scholars to the world by publishing in peer-reviewed journals, in order to establish their own claim and for peer recognition. This data is assimilated and interpreted propelling new research to generate new knowledge and new data. Thus this is a cyclic process, which is facilitated by peer-reviewed journals published by various scholarly societies in the NARS system. There are more than 100 journals, which are being published by various scientific societies formed to promote and advance the interest of their academic disciplines. These societies are non-profit organizations and registered under the Societies Registration Act XXI of 1860. They hold regular conferences for the presentation and discussion of new research results and publish or sponsor peer-reviewed academic journals in their discipline. Some of these societies are very old such as Indian Society of Agricultural Statistics and Horticultural Society of India which were formed way back in 1947. However, these societies publish their journals as traditional print only mode, so these are apparently less visible to the world. These journals lag time between the submission and publication and are only available upon subscription or as complimentary copy to the society's members.

Now with the latest web 2.0 technologies, there is a possibility for the world-wide electronic distribution of the peer-reviewed journal literature via internet completely free

and with unrestricted access so that all the scientists, scholars, teachers, students, and other curious minds could be benefited. Removing the access barriers to this scholarly literature can accelerate research, which would throw light to enrich the education. However, for various reasons, there are limitations for online, free and unrestricted availability. Therefore, professional societies should react quickly, embrace the open source technologies and internet as much as possible and remove all the restrictions and barriers for the wide distribution of the journals online thereby making possible the peer-reviewed literature immediately accessible, searchable, and reusable to anyone in the world which is called 'Open Access' (OA).

Open access can be provided in two main ways: A researcher can place a copy of each article in an 'open access repository' (OAR) or can publish articles in 'open access journals'. In addition, a researcher may place a copy of each article on a personal or departmental website. All these routes to OA ensure that more users can access such articles than the subscription-based journals. The OARs are digital collections of research articles deposited by their authors, which are known as 'self-archiving'. In case of journal articles this may be done either before (preprints) or after publication (postprints). These repositories expose the metadata of each article (the title, authors, and other bibliographic details) in a format compliant with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Google, Google Scholar and others, such as web search engines can access and systematically harvest the contents of the archives worldwide and would form a database of current global research (Swan, 2009). OARs may be multidisciplinary and located in universities or other research-based institutions, or they may be centralized and subject-based, like for example, Crop Sciences, Horticulture, Animal Sciences, Fisheries and other related disciplines. Eprints@IARI (<http://eprints.iari.res.in>), an OAR has been established by IARI to secure agricultural research benefits for all and also to showcase the institute's research output to the world.

Obstacles for Open Access: In India and abroad, different initiatives have shown that open access is economically feasible, which gives readers extraordinary power to find and make use of relevant literature, and also it gives authors and their research works 'vast and measurable new visibility', readership, and impact. Many of the medical journals and professional bodies' journals in India are all open access. However, this kind of initiatives is not seen in Indian agricultural and allied science disciplines. The National Academy of Agricultural Sciences (NAAS) could produce a policy brief on the improvement in the societies' functioning and publication of journals and rate the journals



(Source: blogs.openaccesscentral.com)

which Thomson Reuters Inc. had left out in its ISI Master Journal list but has not declared open access policy as other professional bodies like Indian National Science Academy (INSA) and Indian Academy of Sciences (IAS). Apprehension of revenue loss from print subscriptions if journal becomes OA (even as it increases readership and circulation of knowledge), lack of IT infrastructure or the funds to develop it and uncertainty around copyright issues may also be regarded as the obstacles in making research literature open access.

Resources for Open Access: There are many resources available to make research results freely available online to the whole research community. The author pay model has been implemented by various societies abroad whose journals were traditionally print and subscriptions based and transformed their journals into (OA) journals. Adaptation of such a model would bring back the eroded revenue if journals go open access. The societies can also generate funds from other activities and initiatives at their annual meetings to compensate revenue loss. By making use of recent innovations in online journal production and dissemination, the publishing costs could be reduced dramatically. The *Public Knowledge Project* (PKP), Canada, developed *Free & Open Source Software* (FOSS), *Open Journal Systems* (OJS) which would increase the efficiency of electronic journal publication and management. The OJS software is free and has online support. For the institutional repositories or open access repositories, e-prints and Dspace are the popular FOSS products developed by University of Southampton and Duraspace, respectively. In India, *Open Knowledge Society* formed for assisting the societies in transformation of their traditional print journal into open access journal. With the OJS journal management and publishing system and Open Archives Harvester (OAH) metadata indexing system and the scholarly communications could be used for expanding and improving the access to research.

In order to make easy for sharing the copyright material legally, the Creative Commons (CC) provides free tools that let authors and scientists to easily mark their work with the freedom they want it to carry and can use CC license to change their copyright terms from 'All Rights Reserved' to 'Some Rights Reserved'. The Science Commons (SC) apply the philosophies and activities of CC in the realm of science and aims to clear the legal and technical pathway for accelerating discovery worldwide with its open licenses for copyrighted works, building open source platforms for knowledge management and data visualization. Current publisher policies on self-archiving and copyright are detailed on the SHERPA project, OAKList and SCPJ. A similar kind of database for the societies copyright policies can be developed for the agricultural science publishers in India with the help of the upcoming

'ICAR Data Centre' under National Agricultural Innovation Project (NAIP).

Open Access in India & World: All the journals of the Indian Academy of Sciences (IAS) and Indian National Science Academy (INSA) are open-access journals. Recently, the Council of Scientific and Industrial Research (CSIR), the International Crops Research Institute for Semi-Arid Tropics (ICRISAT) and Consultative Group on International Agricultural Research (CGIAR) have adapted open access policy for their publications. The United States of America (USA) 'Bayh-Dole (BD) Act of 1980 is the forerunner for the introduction of the 'Protection and Utilization of Public Funded Intellectual Property Bill 2008' in the Rajya Sabha. There is a hope that the recently introduced 'Federal Research Public Access Act 2009' in USA may also inspire for introduction of a similar kind of bill which would make mandatory open access to all the peer-reviewed publications arising from the projects funded by public grants within six months.

Agricultural research scholars must come forward to deposit their peer-reviewed articles in open access repositories and support their scientific societies to make their journals open access. Societies, which embrace developments in scholarly publications, would realize that their membership and publications thrive more than the societies that cling to the potentially unstable status quo (Doyle, 2004).

1. Doyle H, Gass A, and Kennison R (2004). Open access and scientific societies. *PLoS Biol* 2 (5): e156.
2. Lawrence S (2001). Free online availability substantially increases a paper's impact. *Nature* 411:521.
3. Berry R S (1997). A very brief introduction to open access. accessed at *Scientific Information in the Electronic Era. The Professional Ethics Report X*: 3.
4. Suber P (2004). A very brief introduction to open access. Accessed at <http://www.earlham.edu/~peters/fos/brief.htm>.
5. Swan A (2009). What is open access? briefing papers accessed at http://openoasis.org/images/stories/briefing_papers/Open_Access.pdf.

BIOINFORMATICS

(C. Devakumar and H. Chandrasekharan)

Lossless Compression Algorithm for Genomic Data

We are familiar with compression software that can remove "redundant" code from a music file — to produce an mp3 — or an image — to make a jpg. However, these compression methods lose information. Less familiar to many users are lossless compression methods such as FLAC for sound files, TIFF for images, and the "zip" format for documents and other files. Lossless compression exploits the repeats, palindromes and patterns present in

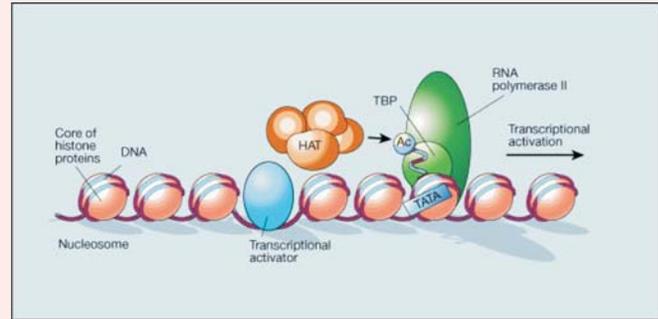


the digital data to reduce the overall size of the file in question. Researchers in Egypt have developed a technique to compress DNA sequences without loss of information. Molecular sequence databases, such as those at EMBL, GenBank, and Entrez contain millions of DNA sequences filling many thousands of gigabytes of computer storage capacity of sequences. With almost every new scientific publication in genetics and related sciences, a new sequence is added and the rate at which the data is accumulating is on the rise. These sequences play a vital role in medical research, disease diagnosis, and the design and development of new drugs. However, DNA sequences are comprised of just four different bases labelled A, C, G, and T. Each base can be represented in computer code by a two character binary digit, two bits in other words, A (00), C (01), G (10), and T (11). At first glance, one might imagine that this is the most efficient way to store DNA whose sequences are not random but contain repeating sections, palindromes, and other features that could be represented by fewer bits than is required to spell out the complete sequence in binary. A repeat pattern could be abbreviated to say the binary equivalent of “six times G” for instance, which would be a few bits shorter than explicitly writing “GGGGGG” in binary. Similarly, palindromes could be abbreviated in code relative to their complementary pattern in the DNA sequence. The authors have developed a Lossless Compression Algorithm that works with digitized DNA sequences to reduce the amount of computer storage needed for each sequence. LCA achieves a better compression ratio than existing compression algorithms for DNA, such as GenCompress, DNACompress, and DNAPack. The same approach could also be used for protein sequences.

Soliman T H A, Gharib T F, Alian A A, Sharkawy M A E (2009). A lossless compression a algorithm for DNA sequences. *International Journal of Bioinformatics Research and Application* 6: 593 – 602.

Technique Finds Gene Regulatory Sites without Knowledge of Regulators

Proper gene regulation depends upon certain combinations of molecules being attached to segments of DNA at the right place and at the right time. Such combinatorial regulation is arguably the most prominent theme in cell development. The clusters of binding sites to which the necessary molecules attach are called cis-regulatory modules. Finding the appropriate cis-regulatory modules is the first step toward unraveling the gene regulatory process. A new statistical technique is reported which would scan a genome for specific gene-regulatory regions without requiring prior knowledge of the relevant transcription factors. The technique has been experimentally validated in both the mouse and *Drosophila* genomes. Beginning with a short list of sample cis-regulatory module



A commonly held view of gene regulation (Source: Nature.com)

sequences (obtained through genetic assays) the computer algorithms search genome-wide for other sequences with similar-looking binding sites and similar functionality. Given a representative subset of cis-regulatory module sequences, it is possible to predict most of other modules. Such an exercise would help in gene sequencing. The authors carried out extensive cross-validations representing hundreds of modules in the mouse and the fruit fly genomes. Seven predictions were experimentally validated through genetic techniques from one species to very diverse species.

Kantorovitz M R, Kazemian M, Kinston S, Saavedra D M, Zhu Q, Robinson G E, Götting B, Halfon M S, Sinha S (2009). Motif-blind, genome-wide discovery of cis-regulatory modules in *drosophila* and mouse. *Developmental cell* 17:568 – 579.

Note: Cis-regulatory module: a stretch of DNA, 500-1000 basepairs in length, where a number of transcription factors can bind and regulate expression of nearby genes.

Error free Genome mapping

Efforts to identify the genomic location of all the gene products involved in the complex biological processes in a single organism have been less fruitful. Bioengineers at San Diego, USA claim a breakthrough development that will now allow scientists to perform full delineation of the location and use of genomic elements. They describe a four-step systems approach that integrates multiple genome-scale measurements on the basis of genetic information flow to identify the organizational elements and map them onto the genome sequence. They have applied this approach to the *Escherichia coli* genome to generate a detailed description of its transcription unit architecture giving us a fine resolution of the contents of a genome and location of its elements with higher resolution and better accuracy. Hopefully this ‘metastucture’ of a genome that they have reported would eliminate the trial and error and enable to reach new metabolic designs faster with lower failure rates.

Cho B K, Zengler K, Qiu Y, Park Y S, Knight E M, Barrett C L, Gao Y and Palsson B (2009). The transcription unit architecture of the *Escherichia coli* genome. *Nature Biotechnology* 27: 1043-1049.

CROP BIOINFORMATICS

(H. Chandrasekharan, C. Devakumar and Minakshi)

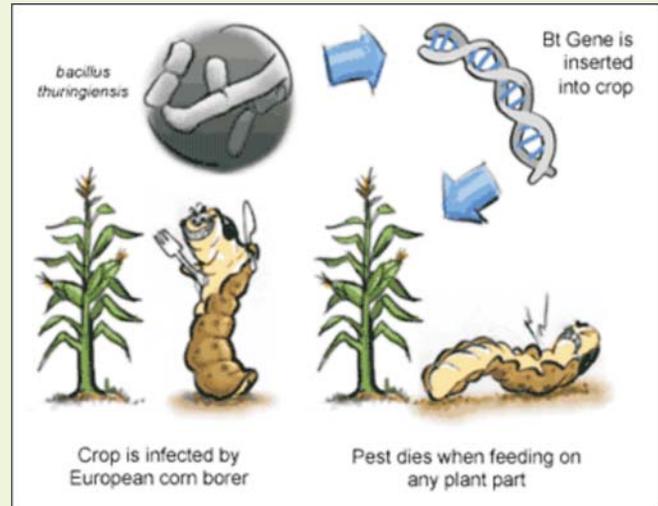
Carbon Nanotubes to Promote Plant Growth

A study from Arkansas found that tomato seeds exposed to Carbon Nanotubes (CNTs) germinated faster and grew into larger, heavier seedlings than other seeds. Mariya Khodakovskaya and colleagues note that considerable scientific research is underway to use nanoparticles in agriculture. The goals of “nano-agriculture” include improving the productivity of plants for food, fuel, and other uses. The scientists report the first evidence that CNTs penetrate the hard outer coating of seeds, and have beneficial effects. Nanotube-exposed seeds sprouted up to two times faster than control seeds and the seedlings weighed more than twice as much as the untreated plants. Those effects may occur because nanotubes penetrate the seed coat and boost water uptake, the researchers state.

Khodakovskaya M, Dervishi E, Mahmood M, Xu Y, Li Z, Watanabe F, Biris S A (2009). Carbon nanotubes are able to penetrate plant seed coat and dramatically affect seed germination and plant growth. *ACS Nano* 3: 3221–3227.

Management Options against Resistance onset by Bt Crops

A new study analyzes insect resistance data from five continents, as reported in 41 studies, and concludes that the existing theories and strategies can be used to predict, monitor, and manage insect resistance to *Bt* crops. Monitoring and surveillance studies have shown that resistance has emerged in a few pest populations, but not in most others. By systematically analyzing the extensive data, we can learn what accelerates resistance and what delays it. With this knowledge, we can more effectively predict and thwart pest resistance. Among the authors' conclusions are: The refugia strategy (growing non-*Bt* crops near the *Bt* crops) can slow the evolution of insect resistance by increasing the chances of resistant insects mating with non-resistant ones, resulting in non-resistant offspring. Crops that are “pyramided” to incorporate two or more *Bt* toxins are more effective at controlling insect resistance when they are used independently from crops that contain only one *Bt* toxin. Resistance monitoring can be especially effective when insects collected from the field include survivors from *Bt* crops. DNA screening can complement traditional methods for monitoring resistance, such as exposing insects to toxins in the lab. Despite a few documented cases of field-evolved resistance to the *Bt* toxins in transgenic crops, most insect pest populations are still susceptible. With *Bt* crop acreage increasing worldwide, incorporating enhanced understanding of observed patterns of field-evolved resistance into future resistance management strategies



Engineering resistant corn. Following the insertion of a gene from the bacteria *Bacillus thuringiensis*, corn becomes resistant to corn borer infection. This allows farmers to use fewer insecticides (Source: www.scq.ubc.ca).

can help to minimize the drawbacks and maximize the benefits of current and future generations of transgenic crops.

Tabashnik E B, Rensburg V and Carriere (2009). Field-evolved insect resistance to bt crops: definition, theory, and data. *J. Econ. Entomol.* 102(6): 2011-2025.

Adding One Single Gene to Yeast Dramatically Improves Bioethanol Production from Agricultural Waste

Bioethanol is made by the yeast *Saccharomyces cerevisiae* from plant sugars. With an annual world production of 65 billion liters, bioethanol is already the largest product of the fermentation industry. Bioethanol should of course preferably be produced from alternate resources preferably using agricultural residues such as wheat straw and corn stover. However, when the sugars from these raw materials are released, significant quantities of acetate are formed. Acetate is an auto inhibitor and therefore can slow down or even halt bioethanol production by yeast. Another limitation of the current bioethanol production process is that about 4% of the sugar is lost to formation of the byproduct glycerol. With the introduction of a single bacterial gene into yeast, researchers from Delft University of Technology in the Netherlands achieved three improvements in bioethanol production from agricultural waste material: By introducing a single gene from the bacterium *Escherichia coli*, conversion of acetate to ethanol was achieved. The invention has three in one advantages such as no glycerol formation, higher ethanol yields and utilization of toxic acetate'.

Guadalupe M V, Almering M J H, Van Maris A J A and Pronk J T (2009). Elimination of glycerol production in anaerobic cultures of *Saccharomyces cerevisiae* engineered for use of acetic acid as electron acceptor. *Applied and Environ. Microbiol.* 76: 190-195.



PEST BIOINFORMATICS

(Usha Dev, NBPGR)

Termites Create Sustainable Monoculture Fungus Farming

The termite fungus-farming symbiosis had a single African rain-forest origin and now comprises ca 330 species. It is of major ecological importance for decomposition and mineral cycling. The termite queen and king normally do not acquire their first garden until they have raised the first workers. These helpers collect *Termitomyces* spores while foraging, together with the plant material that they defecate in the nest to establish a garden substrate. These spores are amply available because the fungus gardens produce large mushrooms once a year on top of the termite mounds. Practice offers a paradox: According to evolutionary theory symbioses with multiple lineages per colony should be unstable, because these genotypes can be expected to compete for making mushrooms rather than collaborate to serve the termite farmers. A unique mechanism seems to prevent this from happening. All colonies were found to contain only a single fungal genotype in spite of gardens having been initiated from at least two and probably many genetically different spores. The authors show that genotypes that happen to be common in a garden become even more common at the expense of rarer genotypes. This happens because they have a higher chance of having an identical genotype as neighbour. Every time this happens, such genetically identical mycelia merge, which enhances the efficiency by which they produce asexual spores that the termites eat and deposit in new garden material of the colony. This process of positive reinforcement makes every colony end up with a single fungal symbiont in spite of the population at large having many fungal genotypes.

Aanen D K, Licht H H F, Debets M A J, Kerstes N A G, Hoekstra R F, and Boomsma J J (2009). High symbiont relatedness stabilizes mutualistic cooperation in fungus-growing termites. *Science* **326**: 1103-1106.

SPODOBASE: An EST Database for the Lepidopteran Crop Pest *Spodoptera*

The Lepidoptera *Spodoptera frugiperda* causes widespread economic damage on a variety of crop plants. It is also well known through its famous Sf9 cell line which is used for numerous heterologous protein productions. Species of the *Spodoptera* genus are used as model for pesticide resistance and to study virus host interactions. A genomic approach is now a critical step for further new developments in biology and pathology of these insects, and the results of Expressed sequence tags (ESTs) sequencing efforts need to be structured into databases providing an integrated set of tools and informations. The ESTs from five independent cDNA libraries, prepared from three

SPODOBASE

An integrated database for the genomics of the Lepidoptera *Spodoptera frugiperda*

[Sequences database](#)
[Blast](#)
[Publications](#)
[Contributors](#)
[Silkworm Database](#)
[Contact](#)

A fruitful collaboration :

Integrative biology and virology of insects (BIVI)
UMR INRA / UM II
Université Montpellier II - Case courrier 101 - Place Eugène Bataillon
34095 Montpellier cedex 5

Insect Genome Laboratory
National Institute of Agrobiological Sciences
1-2 Oiwashi, Tsukuba, Ibaraki
305-8634, Japon

UMR Biotic Interactions and Plant Health
UMR INRA CNRS / Université de Nice
400 route des Chappes - BP 167
06903 Sophia Antipolis Cedex

Homepage of Spodobase

different *S. frugiperda* tissues (hemocytes, midgut and fat body) and from the Sf9 cell line, are deposited in the database. So far, the SPODOBASE contains 29,325 ESTs, which are cleaned and clustered into non-redundant sets (2294 clusters and 6103 singletons). The SPODOBASE is constructed in such a way that other ESTs from *S. frugiperda* or other species may be added. User can retrieve information using text searches, pre-formatted queries, query assistant or blast searches. Annotation is provided against NCBI, UNIPROT or *Bombyx mori* ESTs databases, and with GO-Slim vocabulary. The SPODOBASE database provides integrated access to expressed sequence tags (ESTs) from the lepidopteran insect *S. frugiperda*. It is a publicly available structured database with insect pest sequences which will allow identification of a number of genes and comprehensive cloning of gene families of interest for scientific community. SPODOBASE is available from URL: <http://bioweb.ensam.inra.fr/>.

Nègre V, Hôtelier T, Volkoff A N, Gimenez S, Cousserans F, Mita K, Sabau X, Rocher J, López-Ferber M, d'Alençon E, Audant P, Sabourault C, Bidegainberry V, Hilliou F and Fournier (2006). SPODOBASE: An EST database for the lepidopteran crop pest *Spodoptera*. *BMC Bioinformatics* **7**:322.

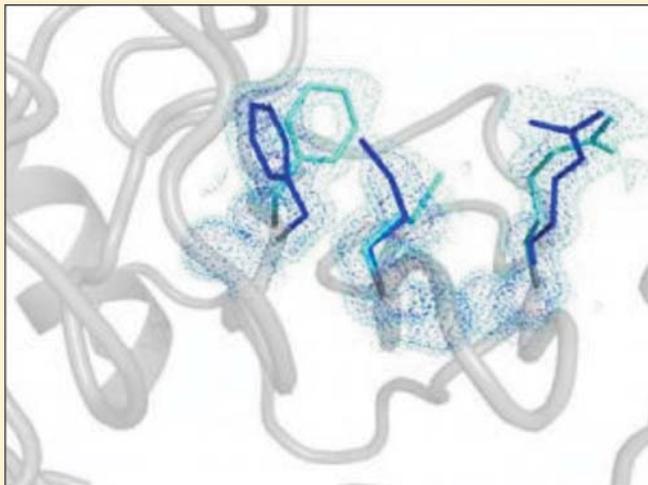
CHEMOINFORMATICS

(C. Devakumar and Minakshi)

Hidden Protein Structures are Essential for Catalysis

A study appearing in nature reports that for the first time, researchers used x-ray crystallography and nuclear





The hidden alternate structure of the human prolyl isomerase uncovered by ambient temperature high-resolution x-ray crystallography. (Source: Dorothee Kern).

magnetic resonance (NMR) techniques to directly visualize protein structures essential for catalysis at the rare high-energy state. The study demonstrated how the motions of these rare, or hidden, structures collectively and directly contribute to catalysis which could be new molecular sites for potential drug targets. The highlight of this work is the visualization of the high-resolution structure of the hidden, high-energy state for the first time. The high-resolution x-ray crystallography was further improved by analyzing electron density data previously discarded as “noise” and by collecting data at ambient temperature. The protein of interest is human cyclophilin A, an enzyme that is abused by the HIV virus to aid its own replication. Dynamic NMR spectroscopy provided direct experimental evidence that the hidden structures in the high-energy state are in fact essential for catalysis.

Fraser J S, Clarkson W M, Degnan S C, Erion R, Kern D, Alber T (2009). Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**: 669-673.

Automated Genome Mining for Natural Products

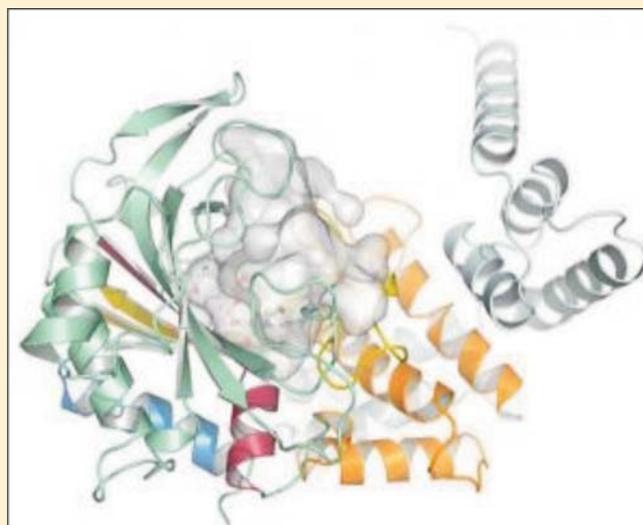
Discovery of new medicines from natural sources has largely been a tedious exercise based on screening of plant and microbial extracts combined with bioassay guided identification and structure elucidation. Increasingly rapid and more cost effective genome sequencing technologies coupled with advanced computational power have converged to transform this trend toward a more rational and predictive pursuit. A rapid method of scanning genome sequences for multiple polyketide, nonribosomal peptide, and mixed combination natural products with output in a text format that can be readily converted to 2D and 3D structures using conventional software. This open-source and web-based program can assemble various small molecules composed of twenty standard amino acids and twenty two other chain-elongation intermediates used

in nonribosomal peptide systems, and four acyl-CoA extender units incorporated into polyketides by reading a hidden Markov model of DNA. This process evaluates and selects the substrate specificities along the assembly line of nonribosomal synthetases and modular polyketide synthases. Using this approach the authors have predicted the structures of natural products from a diverse range of bacteria based on a limited number of signature sequences. In accelerating direct DNA to metabolomic analysis, this method bridges the interface between chemists and biologists and enables rapid scanning for compounds with potential therapeutic value.

Li M H, Ung P M, Zajkowski J, GarneauTsodikova S and Sherman D H (2009). Automated genome mining for natural products. *BMC Bioinformatics* **10(185)**: 1471-2105.

Spellchecker: Chemical Basis for Extra ‘Quality Control’ in Protein Production

Even small errors made by cells during protein production can have profound disease effects, and nature has developed ways to detect and correct them. The amino acid alanine nature has been extra careful, developing not one, but two checkpoints in its effort to make sure that this component is used correctly. Now, scientists at The Scripps Research Institute have discovered the chemical basis for why these extraordinary efforts are necessary. In the coming years, there will be an increasing awareness of the role of mistranslation in human diseases and of how nature has struggled to find solutions to attenuate mistranslation and its consequences. Normally transfer RNAs (tRNAs), transport specific amino acids to the ribosomes, the protein factories of cells, so that amino acids can be fixed to their correct place in a growing chain. At the beginning of this process, 20 tRNA enzymes, one for each of the 20 common amino acids, select the proper amino acid to be transported by a tRNA and stitch them



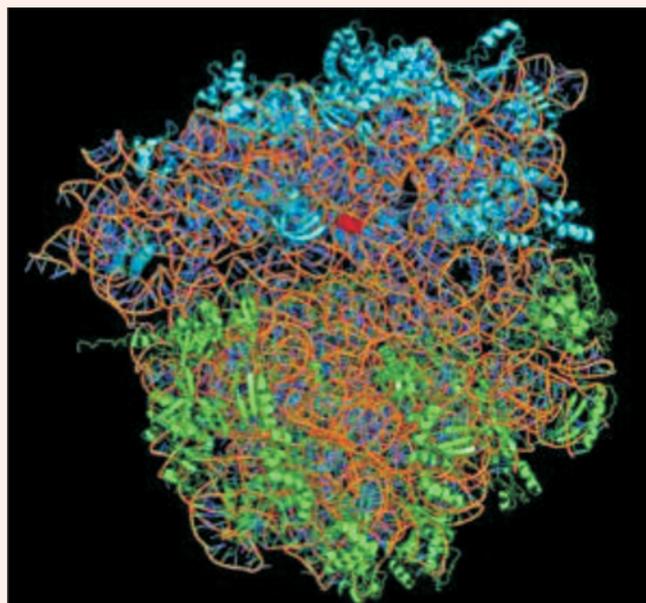
Structure of a protein (colored ribbons) and bound amino acids sticks in the cavity (Source: Scripps Research Institute)

together. When the enzyme that adds the amino acid alanine to tRNAs called alanyl-tRNA synthetase (AlaRS) — mischarges its tRNA (tRNA^{Ala}), the error leads to the accumulation of misfolded proteins leading to severe neurological and other defects. Another study by the Schimmel-Yang lab showed that *E. coli* bacterium with a similar mutation becomes very sensitive to serine and glycine, dying when these amino acids (but not others) are added to the culture. Clearly, nature has a vested interest in avoiding such costly errors. It has been shown that alanine's tRNA synthetase, AlaRS, not only loads the tRNA with an amino acid, but also checks to make sure it attached the right one. In addition, many organisms have an extra freestanding "spellchecker" molecule in the form of a protein called AlaXp to ensure that alanine is not confused with other amino acids. The results showed that one reason for the difficulty AlaRS has in distinguishing alanine from serine and glycine is that the active site on the AlaRS molecule is a large, flexible pocket. Instead of acting as a rigid lock for a single key, the cavity flexes to hold not only its target alanine, but also similar-size molecules serine and glycine. But serine and glycine are not exactly the same size as alanine. None of the tRNA synthetases other than alanines have a problem misactivating both a smaller and a bigger amino acid," said Guo. Unexpectedly, the new study's results also revealed that within AlaRS's binding pocket the acidic group of Asp235 creates an extra hydrogen bond with the larger serine molecule. This additional bond turns out to be the major force that helps to secure the misplaced serine in the pocket, despite its larger size. However, x-ray analysis showed that Asp235 is also critical for holding the amino group of alanine. Attempts by the Schimmel-Yang lab to replace Asp235 with another residue failed. In fact, the scientists found that to make a change that would eliminate the interaction with serine would also impact negatively on the interaction with the correct amino acid, alanine. But there is another solution by creating AlaXp, which is specifically designed to provide a second check and eliminate any serine that is attached to tRNA (Ala). Together, AlaRS's large, flexible pocket and the additional hydrogen bond with serine explain the chemical basis for frequent confusion of glycine and serine for alanine and the need for additional checks to make sure that alanine, not one of its look-alikes, is incorporated into a protein when called for. Now we can understand why nature takes so much effort to invent two editing checkpoints. All organisms from bacteria to humans share the same biological solution that nature found to this critical chemical problem.

Guo M, Chong Y E, Shapir R, Beebe K, Yang X L and Schimmel P (2009). Paradox of mistranslation of serine for alanine caused by AlaRS recognition dilemma. *Nature* **46**: 808-812.

Nobel Prize in Chemistry: What Ribosomes Look Like and How They Function at Atomic Level

The ribosome translates the DNA code into life. The Nobel Prize in Chemistry for 2009 awards studies of one of life's core processes: the ribosome's translation of DNA information into life. Ribosomes produce proteins, which in turn control the chemistry in all living organisms. As ribosomes are crucial to life, they are also a major target for new antibiotics. This year's Nobel Prize in Chemistry awards Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath for having shown what the ribosome looks like and how it functions at the atomic level. All three have used X-ray crystallography to map the position for each and every one of the hundreds of thousands of atoms that make up the ribosome. Inside every cell in all organisms, there are DNA molecules which contain the blueprints for how an organism looks and functions. The blueprints become transformed into living matter through the work of ribosomes. Based upon the information in DNA, ribosomes synthesise proteins: which are tens of thousands of in the body and with different forms and functions. They build and control life at the chemical level. An understanding of the ribosome's innermost workings is important for a scientific understanding of life. This knowledge can be put to a practical and immediate use. Many of today's antibiotics cure various diseases by blocking the function of bacterial ribosomes. Without functional ribosomes, bacteria cannot survive. This is why ribosomes are such an important target for new antibiotics.



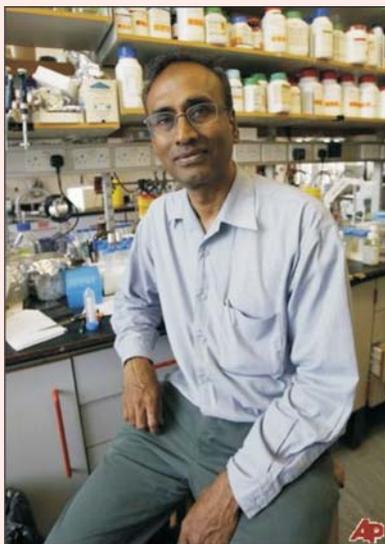
X-ray structure of a bacterium ribosome. rRNA (orange), small subunit (blue) and large subunit (green). An antibiotic molecule (red) bound to the small subunit. (Source: Image courtesy of Nobel Foundation).

NEWS ROUNDUP

(H. Chandrasekharan and C. Devakumar)

Professor Venkatraman Ramakrishnan wins Nobel Prize for Chemistry

Dr. Venkatraman Ramakrishnan is a first chemist of Indian origin to win Nobel Prize in chemistry for the year 2009 in addition to Dr C.V. Raman who won the Nobel Prize for his work carried in India. The other NRI scientists who won Nobel Prize; are Hargobind Khorana (1968), and Dr. S. Chandrasekhar (1983). Dr. Ramakrishnan



(Source: Alastair Grant/AP)

shared the Nobel Prize 2009 for chemistry with Thomas Steitz, and Ada Yonath. Their work has led to breakthrough in the development of antibiotics that disable infections by binding to specific pockets the ribosome structure of bacteria. Venkatraman Ramakrishnan was born in the temple town of Chidambaram famously known for dancing Nataraja in Tamil Nadu, in 1952. He earned his B.Sc. in Physics (1971) from Maharaja Sayaji Rao University of Baroda and Ph.D in Physics (1976) from Ohio University. As a postdoctoral fellow at Yale University, he worked on a neutron-scattering map of the small ribosomal subunit of *E. coli*. He has been studying ribosome structure ever since. At Brookhaven National Laboratory in USA he began his collaboration with Stephen White to clone the genes for several ribosomal proteins and determine their three-dimensional structures. He earned a Guggenheim fellowship to his work on X-ray crystallography of ribosome structure. Ramakrishnan, now a Senior Scientist at the MRC Laboratory of Molecular Biology in Cambridge published the structure of the small ribosomal subunit of *Thermus thermophilus*, a heat-stable bacterium. With this 5.5 Å resolution structure, Ramakrishnan's group identified key portions of the RNA, using previously determined structures, and positioned seven of the subunit's proteins. He presented the 3 Å structure of the 30S ribosomal subunit alone and the structures of the 30S subunit in complex with three antibiotics that target different regions of the subunit. Ramakrishnan elucidated the structural basis for the action of each of these drugs. More recently, he determined the atomic structure of the whole ribosome

in complex with its tRNA and mRNA ligands. He is thus known for his work on the determination of the 3D structure of the small ribosomal subunit and its complexes with substrates and antibiotics, which has shed light on the mechanism that ensures the fidelity of protein synthesis, and for his work on the structures of chromatin-related proteins.

1. Ramakrishnan V and White S W (1992). The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA. *Nature* **358** (6389):768-71.
2. Wimberly B T, Brodersen D E, Clemons W M J, Morgan-Warren R J, Carter A P, Vonnrhein C, Hartsch T and Ramakrishnan V (2000). Structure of the 30S ribosomal subunit. *Nature* **407**(6802): 327-339.
3. Carter A P, Clemons W M, Brodersen D E, Morgan-Warren R J, Wimberly B T and Ramakrishnan V (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407** (6802): 340-348.

CSIR completes first ever Human Genome Sequencing in India

The CSIR scientists at the Institute of Genomics and Integrative Biology (IGIB), Delhi have sequenced the Human Genome of an anonymous healthy Indian citizen. With the completion of the first Human genome sequence in India, the nation is now in the league of select few countries like United States, China, Canada, United Kingdom, and Korea who have demonstrated the capability to sequence and assemble complete human genomes. CSIR could achieve this by adopting new technologies and by effectively integrating complex computational tools with high throughput analytical capabilities. While the first Human Genome Sequence effort took more than a decade spending over a billion US dollars, CSIR scientists finished this task in much shorter time comparable with similar recent effort the world over. By using next generation sequencing technologies and skills, they successfully bridged the technological gap that existed a decade ago. The team at IGIB generated over 51 Gigabases of data using next generation sequencing technology, resulting in



(Source: abclive.in)



over 13x coverage of the Human Genome. This next-generation sequencing technology enables massively parallel sequencing of millions of genomic fragments of 76 base pairs, which are then mapped back to the reference genome. This feat in conjunction with Indian Genome variation programme opens newer vistas for low cost affordable healthcare and predictive medicine in future for the masses. This also opens up newer possibilities in disease diagnostics, treatment and sustaining low-cost drugs in the market.

New Guidelines for Organization-wide Password Management

Remembering a complex list of passwords is a sticky problem for an individual and organization alike. We need a password management policy that balances risk and complexity. The new Guide to Enterprise Password Management (NIST Special Publication 800-118) can be useful to industry as well as to aid in understanding common threats against character-based passwords and how to mitigate those threats within the organization. The guide covers defining and implementing password policy, educating users and measuring the effectiveness of password policies. Effective management reduces the risk of compromising password-based authentication mechanisms. Topics addressed in the guide include defining password policy requirements and selecting centralized and local password management solutions. Agencies need to consider using several mitigation strategies, including secure storage and transmission of passwords, user awareness activities, and secure password recovery and reset mechanisms. The guide also is designed to raise awareness of the changing threats against passwords. Using effective password management as described in the guide will reduce the likelihood and impact of password compromises. The guide recommends that users may be educated about threats against passwords and how they should respond. The publication also suggests that for some applications with high security needs, password-based authentication should be replaced with, or supplemented by, stronger forms of authentication such as biometrics or personal identity verification (PIV) cards.

National Institute of Standards and Technology (NIST), *New NIST guidelines for organization-wide password management*,

Note: (PIV) Personal Identity Verification card: It is required for all US Government employees and contractors to gain physical and logical access to government resources. The card will be used for access to secured buildings as well as to access computer resources.

ONLINE BIOINFORMATICS EDUCATION

KNOW YOUR BIOINFORMATICS TOOLS: PART II*

GRAMENE-A Database for Grass Genomics

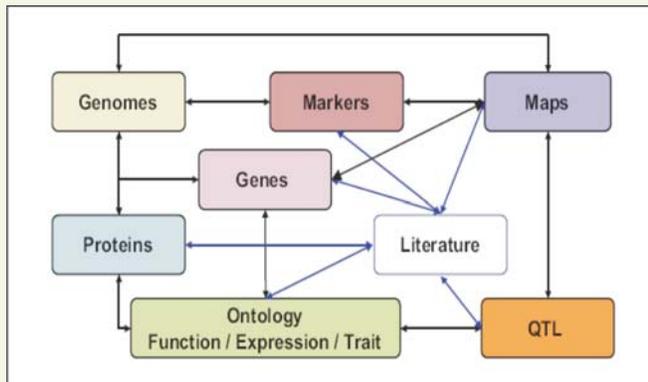
The screenshot shows the GRAMENE website interface. At the top, there is a navigation menu with links for Search, Genome, Species, Download, Resources, About, Help, and Feedback. Below this, the main content area is divided into two columns. The left column is titled 'A Resource for Comparative Grass Genomics' and includes a 'Release #30' section dated October 2009, and a 'News' section with several bullet points about job openings and conferences. The right column is titled 'Quick Start' and contains a list of links for various database features: GENOMES, PROTEINS, COMPARATIVE MAPS, MARKERS, TRAITS, GENETIC DIVERSITY, BIOCHEMICAL PATHWAYS, and LITERATURE. Each link is accompanied by a brief description of the feature.

Homepage of Gramene

Grasses are one of the largest agricultural crops, providing food, industrial materials and renewable energy sources. Due to their large genome size and the number of the species in the taxa, sequencing of many of the genomes are not yet completely sequenced. Only rice (*Oryza sativa*), with a compact genome of approximately 400 Mbp, has been sequenced and annotated. Rice, maize, sorghum, wheat, barley and the other major crop grasses from the family Poaceae (Gramineae) are mankind's most important source of calories and contribute billions of dollars annually to the world economy (FAO 1999, <http://www.fao.org>). Therefore, continued improvement of Poaceae crops is necessary in order to continue to feed an ever-growing world population.

Gramene (<http://www.gramene.org>) is a comparative genome mapping database for grasses and a community resource for rice. Gramene replaces the existing AceDB database 'RiceGenes' with a relational database based on Oracle. It provides curated and integrative information about maps, sequence, genes, genetic markers, mutants, QTLs, controlled vocabularies and publications. Its aims are to use the rice genetic, physical and sequence maps as fundamental organizing units, to provide a common denominator for moving from one crop grass to another and to serve as a portal for interconnecting with other web-based crop grass resources. Being an open-source project, all data and software are freely downloadable through the ftp site (<ftp://ftp.gramene.org/pub/gramene>) and available for use without restriction. Gramene's core data types include genome assembly and annotations, other DNA/mRNA sequences, genetic and physical maps/markers, genes, quantitative trait loci (QTLs), proteins, ontologies, literature and comparative mappings. Gramene is a collaborative project between Cold Spring Harbor





A flow diagram showing all the modules and datasets (genomes, maps, markers, QTL, genes, proteins, ontologies and literature) present in the Gramene database.

Laboratory, Cornell University and the rice community. The information provided via the database is curated using both manual and computational methods. The technological core of Gramene is the MySQL database management system, an open source relational database system that is stable and well supported. The Gramene database takes advantage of the known genetic colinearity (synteny) between rice and the major crop plant genomes to provide maize, sorghum, millet, wheat, oat and barley researchers with the benefits of an annotated genome years before their own species are sequenced. Software tools are available for comparative analysis among the grasses and contribute to our understanding of the anatomy, development, environmental response and the factors influencing agronomic performance of cereal crops. The database has web interface in order to provide a more user-friendly browsing experience, including a dropdown navigation menu system, unified web page for markers, genes, QTLs and proteins, and enhanced quick search functions. Other features include rice pathways for functional annotation of rice genes; genetic diversity data from rice, maize and wheat to show genetic variations among different germplasm; large-scale genome comparisons among *O. sativa* and its wild relatives for evolutionary studies; and the creation of orthologous gene sets and phylogenetic trees among rice, *Arabidopsis thaliana*, maize, poplar and several animal species (for reference purpose).

Gramene's 'maps' section provides CMap tool which presents maps as a linear array of interconnected features that correspond to a single linkage group in the case of a genetic map, to a single contig for a physical map, or to a contig or scaffold in the case of an annotated sequence. To set up a comparison between different map sets from either the same or different species and/or map types, the

researcher first selects a reference map set, and then selects a reference map (chromosome, linkage group or contig) from within the set. This reference map serves as the basis for any comparison that one chooses to make.

The 'Genome' section provides a graphical display of annotations on the rice genome, and includes various tracks describing genes, transcripts, peptides, SNPs, repeats, ESTs, genetic markers (RFLP, SSR), flanking sequence tags (FSTs) from the mutant insertion lines and other features of interest. The annotated rice genome and its pre-computed comparisons with the maize and *Arabidopsis* genomes help users familiar with the function(s) or phenotype(s) of known gene(s) to traverse between these genomes and find the expressed, known and/or predicted gene sequence(s) based on either orthology or on gene function(s).

The 'Proteins' section provides curated information on about 55,000 Swissprot-Trembl protein entries from rice. Protein entries are annotated using the Gene Ontology (GO) for biochemical characterization and the Plant Ontology (PO) for gene expression and phenotype associations.

The addition of several new tools to Gramene has greatly facilitated the potential for comparative analysis among the grasses and contributes to our understanding of the anatomy, development, environmental responses and the factors influencing agronomic performance of cereal crops.

1. Liang C, Jaiswal P, Hebbard C, Avraham S, Buckler E S, Casstevens T, Hurwitz B, McCouch S, Ni J, Pujar A, Ravenscroft D, Ren L, Spooner W, Teclé I, Thomason J, Tung C, Wei X, Yap I, Youens-Clark K, Ware D, and Stein L (2008). Gramene: a growing plant comparative genomics resource. *Nucleic Acids Research* **36**:947–953.
2. Jaiswal P, Ni J, Yap, Ware D, Youens-Clark W S, Ren L, Chengzhi I, Liang, Zhao1 W, Ratnapu1 K, Canaran B F P, Fogleman M, Hebbard C, Schmidt S A S, Casstevens T M, Edward S, Buckler, Stein L and McCouch S (2006). Gramene: a bird's eye view of cereal genomes. *Nucleic Acids Research* **34**:717–723.
3. Jaiswal P, Ware D, Ni J, Chang K, Zhao W, Schmidt S, Pan X, Clark K, Teytelman L, Cartinhour S, Stein L and McCouch S (2002). Gramene: development and integration of trait and gene ontologies for rice. *Comparative and Functional Genomics* **3**: 132–136.

Minakshi

Senior Research Fellow,
Unit of Simulation and Informatics, IARI,
New Delhi-110 012

*Part I: SWISS MODEL-A Fully Automated Protein Structure Homology-Modelling Server, *Agribioinformatics*, Vol. **2** (1 & 2) 2009.



OBITUARY

Norman Borlaug, the Plant Scientist

President Pratibha Patil and Prime Minister Manmohan Singh on Monday led “a grateful nation” in mourning the death of Noble Laureate and “Father of India’s Green Revolution,” Norman Borlaug, who died in Texas, U.S., after a battle with cancer. “Dr. Borlaug had a special place in the hearts of all Indians as he, along with M.S. Swaminathan, brought the Green Revolution to India marking a watershed in India’s agrarian economy and rural development. In his passing away the world has lost a famous agricultural scientist and a friend of India,” the President said in a message.



“More than any other single person of this age, he has helped provide bread for a hungry world,” “We have made this choice in the hope that providing bread will also give the world peace.” the Nobel committee.

In his condolence note the Prime Minister said Dr. Borlaug’s life and achievements were testimony to the contribution that one man’s towering intellect, persistence and scientific vision could make to human peace and progress. “With the passing away of Dr. Borlaug, an era has ended, in which he spearheaded a scientific revolution in agriculture. At a time in the sixties when the country was facing the spectre of severe food shortages, his introduction of high yielding varieties of [wheat] seeds set in motion a technological revolution in Indian agriculture that led eventually to the country achieving self-sufficiency in food grains. The Green Revolution lifted the spirits of the Indian people and gave them new hope and confidence in their ability to tackle the country’s daunting economic

challenges. “Dr. Borlaug’s impact on India’s science and economy went much beyond the Green Revolution. A science-based approach to the problems of agriculture was a fundamental tenet of his thinking and the success of the Green Revolution spawned other successful interventions in areas such as animal husbandry, dairying and agriculture. One of Dr. Borlaug’s favourite quotations was to ‘reach for the stars.’ In doing so, he helped millions of people escape from a life of hunger and deprivation. On behalf of a grateful nation, I convey my deepest condolences to the family and friends of Dr. Norman Borlaug,” Dr. Singh said.



Borlaug visiting IARI Farm

Expressing “personal grief” over the death of the 95-year-old agriculture scientist, Union Agriculture and Food Minister Sharad Pawar said that the world had lost not only an eminent agricultural scientist but also a man dedicated to the cause of humanity. “The Father of the Green Revolution, Norman Borlaug is credited with, what he himself described as, ‘a temporary success in man’s war against hunger and deprivation.’ Having known him since 1974, it is with a profound sense of personal grief that I mourn his passing away.” Mr. Pawar noted that Dr. Borlaug’s contribution would always be remembered for the world peace he heralded through increasing food supply and saving over 245 million lives worldwide. “India, amongst many other nations of the world, owes a debt of gratitude to this outstanding personality. As India moves towards the second Green Revolution, his enduring vision will be a source of inspiration and sustenance for everyone,” he added.

We welcome contributions from readers

Kindly send us your feedback to head_usi@iari.res.in

Published by : Dr H Chandrasekharan, Head and Principal Scientist, Unit of Simulation and Informatics, A0 Block, LBS Building, on behalf of Director, Indian Agricultural Research Institute, New Delhi and funded by Department of Biotechnology, Govt. of India

Printed by : Ashok Computers, Phone: 9868842167, 9582097885

