Indian J. Applied & Pure Bio. Vol. 40(2), 973-982 (2025).
ISSN: 0970-2091
A Web of Science Journal

Bridging Genomes and Greenery: The vital role of Bioinformatics in Plant Sciences

***Dipali Gupta**

Faculty of Basic and Applied Sciences, RNB Global University,
Bikaner - 334601 (India)
dipali.gupta@rnbglobal.edu.in

Abstract

Bioinformatics has emerged as an indispensable tool for the advancement of plant sciences by connecting the intricate genetic data found in plant genomes. This article highlights the contribution of Bioinformatics by value enhanc ement and progression in genome sequencing, comparative genomics, functional annotation, transcriptomics, and metabolomics of plants. Through computational analysis and data integration, bioinformatics helps researchers decode plant genetic blueprints, understand their evolutionary history, identify critical genes and pathways, and illuminate diverse biological processes. Additionally, bioinformatics accelerates breeding programs, increases crop yields, and promotes sustainable agriculture. By harnessing bioinformatics, scientists can unlock plant genomics' potential, leading to significant discoveries impacting food security, environmental conservation, and human health. Plant-based biological databases are digital resources containing extensive information on plant genetics, physiology, and ecology. These databases support educational and research purposes, offering details on plant DNA, anatomy, physiology, and biochemistry, as well as plant-animal and plant-environment interactions. This information aids in understanding species interactions, developing new treatments and technologies, and enhancing agricultural productivity.

Key words : Bioinformatics, Biological databases, Plant Sciences, Plant Information Resources. Plant Databases.

Bioinformatics is a swiftly expanding discipline that has transformed the study of plant sciences[35]. It employs genomics and proteomics data to comprehend how plants function and cope with their surroundings. Bioinformatics tools generate hypotheses, evaluate change in the gene expression due to environmental duress, and classify and predict gene functions. In addition, bioinformatics is used to develop new technologies and methods

---

*Professor,

for identifying, analyzing, and interpreting plant data. Examples include the creation of algorithms to identify and predict gene networks, developing novel approaches to detect and analyze metabolite profiles, and designing genetic identifiers. These bioinformatics techniques can improve the quality of traditional medicinal plants, crop nutrition, single gene analysis, sequence similarity, protein modeling, crop breeding, insect resistance, and drought-resistant varieties[3,9].

Plant Genome Database :

A Plant Genome Database is a biological database containing information regarding plant species' genetic composition. This includes information regarding the proteins, genes, and other genetic components of plants. The database's data is exploited for diversified works, such as research, reproduction, and biotechnology. Gene sequences, gene expression patterns, and genetic maps are typical examples of the categories into which the database's information is typically organized. The gene sequences contain information about the nucleotide sequences of the genes, whereas the gene expression patterns describe how the genes are expressed in various tissues and stages of development. The genetic maps contain information regarding the locations of genes and their relationships: Plant Genome Database, Genome Database for Rosaceae, Gramene, Plant Direct, etc[29,36].

The Plant Genome Database (PGDB) is a public database for studying plant biology and genomics. It includes data on gene sequences, gene expression, metabolic pathways, genome annotations, genetic and physical maps, genetic markers, and plant-associated microorganisms. The Plant Genome Database (PGDB offers tools and resources for gene expression, sequence analysis, and genetic mapping. It stores expression data from numerous plant species, such as Arabidopsis, rice, and maize[7,20,30].

The data in the Plant Genome Database can be utilized for research purposes, including investigating evolutionary relationships between different species, tagging disease-specific genetic markers, genes for susceptibility, and creating novel plant varieties. Additionally, the database can facilitate the creation of new agricultural products and increase crop yields. Moreover, the information can be used to identify novel sources of genetic material for use in biotechnology[12,16,49].

Plant species Database :

The Plant Species Database is a comprehensive source of plant information, including information on species distribution, habitat, and anatomy. It also includes an extensive catalog of essential plant species, images, and descriptions[4]. A Plant Species Database contains detailed information about various plant species. It includes scientific and common names, physical characteristics, distribution, ecology, and other data. The database is intended for scientists, botanists, conservationists, and plant enthusiasts. The database is divided into four categories; taxonomy, ecology, geography, and application[23]. These categories include extensive descriptions of several plant species, like their scientific and common names, physical characteristics, distribution, ecology, and other data. The database also contains photographs, drawings, and illustrations depicting various plant species[13]. This allows users to quickly identify and observe the physical characteristics of a plant species. The database also includes multiple search tools, such as searching by scientific name, common name, or geographic location. This enables users to locate the desired information rapidly[40].

In addition, the Plant Species Database contains connections to additional resources, such as scientific journals and websites publishing latest research about various plant species. This enables users to locate valuable information for their studies or research quickly. The Plant List, NCBI Taxonomy, Arabidopsis Information Resource (TAIR), TropGen, GRIN-Global, and TROPICOS are examples of such resources.

Plant Metabolomics Database :

The Plant Metabolomics Database is a publicly accessible resource for plant metabolomics research. It includes information on compounds, metabolites, metabolite pathways, other plant metabolites, metabolic networks, and maps. PMDB is an exhaustive database of plant metabolomics data. It is a free and open-access database that provides a platform for storing, analyzing, and sharing plant metabolomics data.[10,17] The database contains over 500 plant species and cultivars, including model and non-model species. It includes a variety of metabolite data sets, such as plant primary and secondary metabolites, as well as a variety of microarray and high-throughput sequencing data sets. The PMDB enables researchers to search for information on particular metabolites, species, and cultivars and compare metabolite levels across species[1,45].

Moreover, the database provides comprehensive information on the metabolic pathways of various species and cultivars and tools for predicting and analyzing metabolic pathways. In addition, it provides instruments for identifying metabolic biomarkers and developing metabolic engineering strategies. Researchers in the field of plant metabolomics18 utilize PMDB as an essential resource. It provides a platform for collaboration and data exchange, allowing scientists to store and analyze data comprehensively and in an organized manner. Constantly updated with new data sets, the database is an indispensable tool for studying plant metabolism[32,33,38].

Plant expression Database :

The Plant Expression Database is a resource for gene expression research on plants.[11,15] The PEDb is a comprehensive database of data on the expression of plant genes. The Institute of Plant Sciences at the University of Bonn, Germany, and the Institute of Molecular Biology in Vienna, Austria, developed it. PEDb contains expression data from diverse plant taxa, such as *Arabidopsis thaliana*, maize, poplar, rice, and wheat. It also includes information from experiments involving microarrays, proteomics, and metabolomics[14].

The database contains experimental and predicted expression data, enabling users to differentiate gene expression in various species[44]. The database also contains gene regulation data, along with transcription factor's binding sites and promoter regions. The Plant expression database is an essential resource for plant gene expression researchers[27]. It provides a comprehensive view of genetic expression in several species, enabling more

Table-1. List of Important Plant Databases

Sno.	Name of Database	Features
General Plant Databases		
1.	TAIR (The Arabidopsis Information Resource)	Arabidopsis thaliana
2.	Gramene	Comparative genomics for grasses
3.	Phytozome	Comparative genomics for green plants
4.	PlantGDB	Plant genome databases
5.	Ensembl Plants	Genomic information for plants
Functional Genomics Databases		
6.	PlantTFDB (Plant Transcription Factor Database)	Transcription factors in plants
7.	PlantCARE (Plant Cis-Acting Regulatory DNA Elements)	Regulatory elements in plant genomes
8.	PPDB (Plant Proteomics Database)	Proteomics data for plants
Metabolic Pathway Databases		
9.	MetaCyc	Metabolic pathways and enzymes in plants and other organisms
10.	PlantCyc	Plant-specific metabolic pathways
11.	Arabidopsis Reactome	Curated knowledge base of plant biological pathways
Expression Databases		
12.	ePlant (Electronic Plant Information Center)	Plant gene expression and other functional data
13.	PLEXdb (Plant Expression Database)	Gene expression data in plants
14.	miRBase	MicroRNA sequences, targets, and expression
15.	PMRD (Plant MicroRNA Database)	MicroRNA data in plants
Pathogen and Disease Databases		
16.	PHI-base (Pathogen-Host Interaction Database)	Pathogen-host interactions
17.	Plant Resistance Gene Database	Plant resistance genes
Specialized Databases		
18.	LEGUMEinfo	Genomic data for legume species
19.	Sol Genomics Network (SGN)	Genomic and phenotypic data for Solanaceae family

20.	IPK Crop Genebank Information System	Crop genebank data
21.	Plant Reactome	Plant pathway knowledgebase
Ontology Databases		
22.	Plant Ontology (PO)	Controlled vocabulary for plant structures and growth stages
23.	Gene Ontology (GO)	Ontology for gene functions across species, including plants
Region-based Flora Databases		
24.	Go Botany	England
25.	New York Flora Atlas	New York
26.	Plant Finder Missouri Botanical Garden	Kemper Center, St. Louis
27.	UConn Plant Database	University of Connecticut
28.	USDA Plants Database	USA

precise and thorough analysis. Moreover, the database contains information on gene regulation, allowing researchers to understand better the regulation mechanism of specific genes.[46,47,50]

Agricultural Genebank Management System:

Agricultural Genebank Management System: The Global Crop Diversity Trust developed and maintained GRIN-Global as the premier agricultural genebank management system. It offers an extensive information management system for accumulating, storing, and utilizing crop genetic diversity[14,18]. GRIN-Global enables gene banks, research institutions, and other organizations to catalog, administer, and operate their genetic resource and information collections. More than 500 gene banks, research institutions, and organizations in over 100 countries use GRIN-Global. It is a free, open-source software application that anyone can use and modify. GRIN-Global facilitates user access to genetic information, collection management, and collaboration with other organizations[21,28]. In addition, it provides an integrated system for managing and sharing data about crop genetic resources, such as information about accession numbers, species, varieties, and associated characteristics. GRIN-Global is continuously evolving, and new features are introduced frequently. Recent additions include a data portal for sharing data with external partners, a mobile app for field data collection, and enhanced reporting and analysis tools. GRIN-Global is also entirely integrated with the Crop Wild Relative Database of the Global Crop Diversity Trust, which provides access to data on crop species' wild relatives[5,6].

Plant-based Biological Database Applications:

- Genome Sequencing and Assembly:

Bioinformatics facilitates the sequencing and assembly of plant genomes, yielding valuable insights into plant evolution and genetics. It employs next-generation sequencing (NGS) and de novo assembly algorithms to produce high-quality genome sequences. GenBank, Ensembl, and RefSeq, among others, provide access to assembled genome sequences and associated data for numerous organisms[43].

- Functional Annotation: Bioinformatics tools facilitate the annotation of plant genomes by identifying genes, determining their functions, and predicting their regulatory elements. This data aids in comprehending gene expression patterns, gene interactions, and metabolic pathways, thereby contributing to the overall understanding of plant biology. TAIR, Phytozome, PlnTFDB, PlantCyc, etc., allow researchers to understand plant genes' and proteins' functions and roles. They contribute to comprehending biological processes, pathways, and regulatory networks, thereby supporting plant research and crop improvement[29].

- Comparative Genomics: Comparative genomics involves comparing the genomes of various plant species to identify conserved regions, evolutionary relationships, and genetic variations. Comparative analyses are aided by bioinformatics tools, which emphasize similarities and differences between plant genomes and provide insights into the evolution of plant species. PLAZA, Gramene, CoGe, GreenPhylDB, Plant Genome Duplication Database (PGDD), etc., are widely used comparative genomics databases specializing in plants. These databases on comparative genomics allow researchers to investigate and analyze plant species' genomic characteristics, gene families, and evolutionary relationships. They provide invaluable insights into the evolution of the genome, the diversification of gene functions, and the impact of genomic alterations on plant biology and adaptation[22].

- Transcriptomics: Bioinformatics enables the analysis of transcriptomic data, which entails investigating a plant's complete set of RNA transcripts. It aids in identifying differentially expressed genes, alternative splicing events, and non-coding RNAs, thereby facilitating the comprehension of gene regulation, development, and environmental response. Gene Expression Omnibus, Array Express, Plant Expression Database (PLEXdb):, Electronic Fluorescent Pictograph (eFP), PlantRegMap, ePlant, etc., are valuable resources for plant transcriptomics researchers, allowing them to investigate gene expression patterns, identify differentially expressed genes, analyze co-expression networks, and gain insights into the regulatory mechanisms underlying plant biology[35].

- Proteomics and Metabolomics: Proteomic and metabolomic data, which examine the entire collection of proteins and metabolites in plants, are analyzed by bioinformatics tools. These analyses aid in comprehending protein functions, and metabolic pathways, and study the influence of surroundings on plant physiology. PRIDE (Proteomics Identifications Database), Plant Proteome Database (PPDB), Araport, Plant Metabolic Network (PMN), Metabolomics Workbench, MassIVE (Mass Spectrometry Interactive Virtual Environment), etc., enable the investigation of protein identifications, posttranslational modifications, metabolite profiles, metabolic pathways, and functional annotations, thereby facilitating the comprehension of plant molecular processes and metabolite dynamics[34].

- Functional Genomics: Bioinformatics enables the incorporation of various omics data (genomics, transcriptomics, proteomics, metabolomics) for a thorough comprehension of plant functions. It permits researchers to determine potential genes in charge of particular traits, predict gene functions, and devise experiments to validate these predictions. The Arabidopsis Information Resource (TAIR), Phytozome, Planteome, Plant TFDB, Co-expression Gene Networks in Plants (COB) and ATTED-II, Bio-analytic Resource for Plant Biology (BAR), etc. integrate diverse omics data and annotations, which facilitates the analysis and interpretation of plant molecular functions and regulatory mechanisms[37,42].

- Phylogenetics and Evolutionary Studies: Bioinformatics plays a crucial role in constructing phylogenetic trees, analyzing evolutionary relationships, and deducing the evolutionary history of plant species. Researchers can decipher plants' evolutionary dynamics and divergence patterns by comparing genetic sequences and employing computational algorithms. The Plant List, Tree Base, Phyto REF, One Thousand Plant Transcriptomes (1KP), Phylota, and Open Tree of Life, among others, provide valuable resources for plant phylogenetics and evolutionary biology researchers. They enable the exploration of phylogenetic trees, sequence alignments, taxonomic data, and evolutionary relationships, thereby facilitating comparative studies and advancing our knowledge of plant evolution[2,39].

- Data Management and Analysis: Bioinformatics provides tools and databases for managing, organizing, and analysis of plant-related data. These resources assist researchers in storing and gaining access to genomic data, gene expression data, and other plant-related datasets, fostering data sharing and collaboration within the scientific community[26,41].

- Crop Improvement and Plant Breeding: Bioinformatics contributes to crop enhancement and plant breeding through the exploration of molecular markers linked to desired characteristics. This knowledge can help in the conservation of Plant Genetic Resources, guide the selection of superior plant varieties, and aid in the creation of genetically modified commodities with enhanced yield and resistance against biotic and abiotic stress[19,24,48].

Summary :

Plant-based biological databases are online resources that offer organized and curated data on various aspects of plant biology. Scientists, researchers, and other interested fellows may access and analyze data on plant genetics, genomics, proteomics, metabolomics, and other aspects of plant biology through these databases, which gather and preserve the information. For instance, the Arabidopsis Information Resource (TAIR) provides information about the *Arabidopsis thaliana* plant species, including its genome, gene structure, and functions. The Phytozome database provides information of more than

70 plant species about the genomes, gene families, and metabolic pathways. The Plant Genome Database (PGD) contains information on plant genomes, gene expression, function, genome annotation, and analysis tools. The Plant Metabolic Network (PMN) provides information about the genes and enzymes involved in plant metabolic pathways. The Plant Transcription Factor Database (PlnTFDB) contains detailed data of transcription factors responsible for regulating plant gene expression. These databases are an invaluable resource for comprehending the complex biology of plant systems and play a vital part in advancing plant research.

References :

1. Afendi, F. M., T. Okada, M. Yamazaki, A. Hirai-Morita, Y. Nakamura, K. Nakamura, and S. Kanaya, (2012). *Plant and Cell Physiology*, *53*(2): e1-e1. (2).
2. Arkhipova, I. R. (2017). *Mobile DNA, 8:* 1-14.
3. Aslam, Z., J. Z. K. Khattak, M. Ahmed, and M. Asif, (2017). *A role of bioinformatics in agriculture. Quantification of climate variability, adaptation and mitigation for agricultural sustainability,* 413-434.
4. Austin, M. P., L. Belbin, J. A. A. Meyers, M. D. Doherty, and M. Luoto, (2006). *Ecological modelling, 199*(2): 197-216.
5. Avant, Sandra. (2011) *Agricultural Research*, *59*(9): 13.
6. Bhadauria, V. (Ed.). (2017). Next-generation sequencing and bioinformatics for plant science (Vol. 10). Wymondham: Caister Academic Press.
7. Bielewicz, D., J. Dolata, A. Zielezinski, S. Alaba, B. Szarzynska, M.W. Szczesniak and W. M. Karlowski, (2012). *Nucleic acids research, 40*(D1), D191-D197.
8. Chandok, I. K., H. Afreen, R. Afreen, S. Haider, D. P. Moharana, T. Hussain, ... and M. Z. Malik, (2022). Functional genomics tools for studying microbe-mediated stress tolerance in plants. In Mitigation of Plant Abiotic Stress by Microorganisms (pp. 175-204). Academic Press.
9. Chen, S., Y. Lu, and S. Wang, (2011). *Trends in Genetics*, *27*(4): 173-177.
10. D. Vassilev, J. Leunissen, A. Atanassov, A. Nenov and G. Dimov (2014) *Biotechnology & Biotechnological Equipment*, *19:* 139-152.
11. Dash, S., J. Van Hemert, L. Hong, R. P. Wise and J. A. Dickerson, (2012). *Nucleic acids research, 40*(D1), D1194-D1201.
12. Duvick, J., A. Fu, U. Muppirala, M. Sabharwal, M.D. Wilkerson, C.J. Lawrence ... and V. Brendel, (2007). *Nucleic acids research, 36*(suppl_1), D959-D965.
13. Fuentes, N., A. Pauchard, P. Sánchez, J. Esquivel, and A. Marticorena, (2013). *Biological Invasions*, *15*(4): 847–858.
14. Gardner, C., P. Cyr, M. Millard, L. Gu, M. Reisinger, G. Emberland, and P. Bretting, (2012, October). Using the GRIN-Global System to Identify Useful Plant Genetic Resources and Information. In Agronomy Society of America, Crop Science Society of America, Soil Science Society of America Meeting (pp. 259-4).
15. Hawkins, C., D. Ginzburg, K. Zhao, W. Dwyer, B. Xue, A. Xu, and S. Y. Rhee, (2021). *Journal of integrative plant biology, 63*(11): 1888-1905.
16. Hidayatun, N., A. Risliawati, H. Kurniawan, N. Dewi, M. Sabda, Y. Wahyu, ... and S. Marwiyah, (2017). Conservation of

Indonesia plant genetic resources for food and agriculture in ICABIOGRAD gene bank. *Utilization of Genetic Resources for Productivity and Quality Improvement*, 84-93.

17. Jin, J., P. Lu, Y. Xu, Z. Li, S. Yu, J. Liu, ... and P. Cao, (2021). *Nucleic Acids Research, 49*(D1): D1489-D1495.
18. Jin, J., H. Zhang, L. Kong, G. Gao, and J. Luo, (2014). *Nucleic acids research, 42*(D1), D1182-D1187.
19. Jon Duvick *et al.*, (2008) *Nucleic Acids Research*, *36*(1): D959–D965,
20. Kattge, J., S. Diaz, S. Lavorel, I.C. Prentice, P. Leadley, G. Bönisch, and C. Wirth, (2011). *Global change biology, 17*(9): 2905-2935.
21. Klepikova, A. V., and A. A. Penin, (2019). *Gene Expression Maps in Plants: Current State and Prospects. Plants. 8:* 309.
22. Kudo, T., S. Terashima, Y. Takaki, K. Tomita, M. Saito, M. Kanno, and K. Yano, (2017). *Plant and Cell Physiology, 58*(1): e1-e1.
23. Kushwaha, U. K. S., I. Deo, J. P. Jaiswal, and B. Prasad, (2017). *Glob J Sci Front Res D Agric Vet, 17*(1): 13-23.
24. Louafi, S., M. Thomas, E. T. Berthet, F. Pélissier, K. Vaing, F. Jankowski, and M. Leclercq, (2021). *Agronomy, 11*(9): 1893.
25. Ma, X., H. Yan, J. Yang, Y. Liu, Z. Li, M. Sheng and Z. Su, (2022). *Nucleic Acids Research, 50*(D1), D1456-D1467.
26. Mitra, D., D. Mitra, M. S. Bensaad, S. Sinha, K. Pant, M. Pant, and P. K. D. Mohapatra, (2022). *Computational Toxicology, 24:* 100248.
27. Oppermann, M., S. Weise, C. Dittmann, and H. Knüpffer, (2015). GBIS: the information system of the German Genebank. Database, 2015, bav021.
28. Osterrieder, A., *et al.* (2015). *Plant Physiology*, *169*(3): 1599-1609.
29. Postman, J., K. Hummer, T. Ayala-Silva, P. Bretting, T. Franko, G. Kinard, ... and B. Weaver, (2009, July). GRIN-Global: An international project to develop a global plant genebank information management system. In *International Symposium on Molecular Markers in Horticulture 859* (pp. 49-55).
30. Qunfeng Dong and others (2005). *Plant Physiology*, *139*(2): 610–618.
31. Qunfeng Dong and others (2004). Plant GDB, plant genome database and analysis tools, *Nucleic Acids Research*, *32*(1): D354–D359,
32. Qunfeng Dong, and Lori Kroiss, *et al.*, (2005). *Methods in Enzymology, Academic Press*, *395:* 400-418.
33. Rensink, W. A., and C. R. Buell, (2005). *Trends in plant science, 10*(12): 603-609.
34. Ruzankina, Y. V., and C. Rausch, (2018). *Metabolites*, *8*(6): 75.
35. Schauer, N., and C. Rausch, (2017). *Frontiers in Plant Science*, *8:* 821.
36. Schlueter, S. D., Q. Dong, and V. Brendel, (2003). *Nucleic acids research, 31*(13): 3597-3600.
37. Seung Yon Rhee, Dickerson Julie, and Xu Dong (2006) *Annual Review of Plant Biology*, *57:* 335-360
38. Shafi, A., I. Zahoor, E. Haq, and K. M. Fazili, (2019). Impact of bioinformatics on plant science research and crop improvement. Essentials of Bioinformatics, Volume III: In *Silico Life Sciences: Agriculture,* 29-46.
39. Shaw, P. D., S. Weise, M. Obreza, S. Raubach, S. McCouch, B. Kilian, and P. Werner, (2023). Database solutions for genebanks and germplasm collections. In

Plant genetic resources for the 21st century (pp. 285-309). Apple Academic Press.

40. Shi, H., X. Wu, Y. Zhu, T. Jiang, Z. Wang, X. Li and X. Li, (2024). *Nucleic Acids Research, 52*(D1), D1614-D1628.
41. Singh V.K., A.K. Singh, R. Chand and C Kushwaha (2011) *International Journal of Bioinformatics*, *3*(2): 221-226.
42. Subba, P., C. Narayana Kotimoole and T.S.K. Prasad, (2019). *Omics: a journal of integrative biology, 23*(4): 190-206.
43. Sun, R., T. Sun, and Q. Wang, (2015). *Plant Methods*, *11*(1): 21.
44. Thuiller, W., G. F. Midgley, M. Rougeti, and R. M. Cowling, (2006). *Ecography, 29*(5): 733-744.
45. Trinder, S., N. Fenby, P. Brown, and C. J. Allender, (2023, October). A window into a genebank: GRIN-Global as a tool to support use of the UK Vegetable Genebank collections. In III International Symposium on Carrot and Other Apiaceae 1393 (pp. 227-230).
46. Tseng, K. C., G. Z. Li, Y. C. Hung, C. N. Chow, N. Y. Wu, Y.Y. Chien, ... and W.C. Chang (2020). *Plant and Cell Physiology, 61*(10): 1818-1827.
47. Tyagi, R. K., and A. Agrawal, (2015). *The Indian Journal of Agricultural Sciences, 85*(2): 157-165.
48. Warnow, T. (Ed.). (2019). Bioinformatics and phylogenetics: seminal contributions of Bernard Moret (Vol. 29). Springer.
49. Weise, S., U. Lohwasser and M. Oppermann (2020). *Plants, 9*(8): 1050.
50. Yim, W. C., Y. Yu, K. Song, C. S. Jang, and B. M. Lee, (2013). *BMC plant biology, 13:* 1-9.