

Bioinformatics Tools in Agriculture: An Update

Madiha Zaynab¹, Sonia Kanwal⁵, Safdar Abbas⁶, Fiza Fida⁷, Waqar Islam², Muhammad Qasim², Nazia Rehman⁸, Faiza Fida⁷, Muhammad Furqan³, Muhammad Rizwan¹, Muhammad Anwar⁴, Ansar Hussain³, Muhammad Tayab²

¹College of Life Science, ²College of Plant Protection, ³College of Crop Science, ⁴College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou 350002, PR. China.

⁵Institute of Horticulture Sciences, University of Agriculture Faisalabad, Pakistan.

⁶Department of Biochemistry, Quid-e-Azam University, Islamabad, Pakistan.

⁷Department of computer science, Government College University, Faisalabad, Pakistan.

⁸National Institute of Genomics and Advance Biotechnology, Islamabad, Pakistan.

Received: 28.Mar.2017; Accepted: 15.June.2017; Published Online: 06.Aug.2017

*Corresponding author: Madiha Zaynab; Email: madiha.zaynab14@gmail.com



Abstract

Bioinformatics plays an important role in agriculture science. As the data amount grows exponentially, there is a parallel growth in tools and methods demand in visualization, integration, analysis, prediction and management of data. At the same time, many researchers in the field of plant sciences are unfamiliar with available methods, databases, and tools of bioinformatics which could lead to missed information opportunities or misinterpretation. Some key concepts of software packages, methods, and databases used in bioinformatics are described in this review. In this review, we have discussed some problems related to biological databases and biological sequence analyses. Gene findings, genome annotation, type of biological database, how to data represent and store was deliberated. Future perspective of bioinformatics tools was also discussed in this review.

Keywords: Biological databases, Gene findings, Homology, Integration, Sequence.

Cite this article: Zaynab, M., Kanwal, S., Abbas, S., Fida, F., Islam, W., Qasim, M., Rehman, N., Fida, F., Furqan, M., Rizwan, M., Anwar, M., Hussain, A., Tayab, M., 2017. Bioinformatics Tools in Agriculture: An Update. PSM Biol. Res., 2(3): 111-116.

INTRODUCTION

Recent technologies developments and instrumentation allow large-scale as well as nano-scale biological samples probing for generation of unprecedented data in life sciences (Noman *et al.*, 2016a,b). For human brain, it is too much difficult to process this sea of data. There is an increasing need for computational methods to process and contextualize these data. Model plants genomic studies promoted the discovery of gene and gene function (Islam *et al.*, 2017; Noman *et al.*, 2017a). Knowledge of gene and gene function has provided in the 21st-century first decade. These technological advances have accelerated genome-scale development in model species of plants (Mochida and Shinozaki, 2010). Feasible applications provide by next-generation sequencing (NGS) technology such as, for the variation analysis re-sequencing of the whole genome, RNA sequencing (RNA-seq) for transcriptome analysis, non-coding RNAome analysis epigenomic dynamics quantitative detection and Chip-seq analysis for DNA– protein interactions (Lister *et al.*, 2009; Noman and Aqeel, 2017). Many other approaches have been developed including

networks analysis formed by protein–protein interaction (Arabidopsis Interactome Mapping Consortium., 2011), phytohormone-mediated cellular signaling approach for hormone analysis (Kojima *et al.*, 2009; Ali *et al.*, 2017) and metabolome analysis approach for metabolic systems (Saito and Matsuda, 2010). To extract valuable knowledge and manage effectively various types of genome-scale data sets, bioinformatics has been crucial in every aspect of omics-based research. Accumulation of omics integration outcomes will update understanding and facilitate the exchange of knowledge with other model organisms (Shinozaki and Sakakibara, 2009; Noman *et al.*, 2017b).

The plant genetic information is translated with the help of direct transcription (mRNA) into a protein. Therefore, amongst various technologies for estimation of individual genes' expression as well as their functional mechanisms, is to explore the proteins decoded from those genes, which is known as proteomics. The information of significant proteins that assume a fundamental role in the appropriate plant propagation is necessary to lead towards the upgrading of biotechnology related to plants (Noman *et al.*, 2017b).

However, these proteins work under various physiological and biochemical pathways accordingly, to sustain the plant growth (Eldakak *et al.*, 2013). Furthermore, research innovations disclosed that genomics and proteomics are the two important shifts, those are discovering novel genes, which can eventually help to update the agriculture via improvement of biotechnological programs (Noman *et al.*, 2017a,b). Similarly, proteomics is supported by two-dimensional electrophoresis (2-DE) and mass spectroscopy (MS), which are being used to categorize proteins, and advancement in 2-DE is useful to accommodate the proteomics into latest biotechnological programs. Recently, main quantitative proteomics techniques linked with nano-LC-MS/MS have specified the proteomic analysis in the model plant, *Arabidopsis thaliana* (Niehl *et al.*, 2013) as well as other non-model plants, like *Zingiber zerumbet* (Mahadevan *et al.*, 2014) and *Nicotiana attenuata* (Weinhold *et al.*, 2015). Several advantages of top broadband technique comprise the gel-free handling of proteins, digestion of trypsin and use of an internal peptide, for finest quantification of protein, and thus it can be used for documentation of distinctive proteins from non-model organisms, in spite of limited genome information.

Bioinformatics is the study of biological information by utilizing ideas and strategies in software engineering and statistics. It can be categorized into two classes: (1) management of biological data and (2) computational biology, and has many applications in agriculture (Figure 1). In this review, we discuss sequence-based analyses and comparative genomics approaches, biological databases and representation of data and storage of data access and exchange (Table 1).

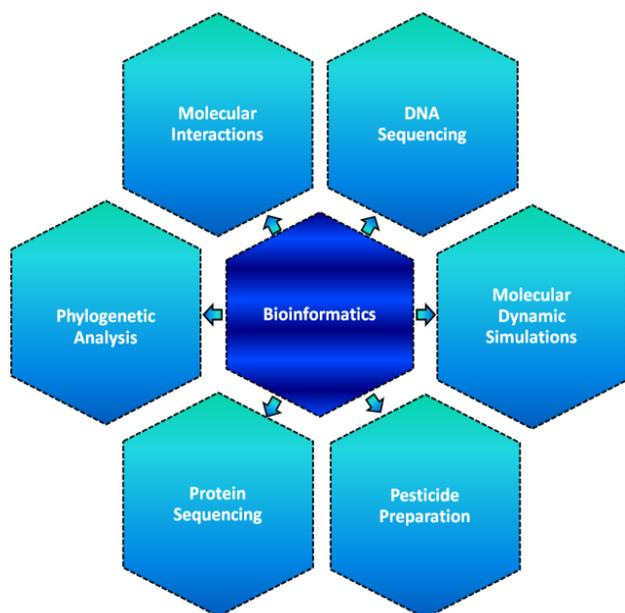


Fig.1. Basic bioinformatics tools in modern agriculture

Sequence Analysis

A biological sequence is principal biological system object at the molecular level like DNA, RNA, and protein. Genomes of *A. thaliana* (The, 2000) and rice (Goff *et al.*, 2002) plants have been sequenced. Lotus (<http://www.kazusa.or.jp/lotus/>) and poplar (<http://genome.jgi-psf.org/Poptr1/>), a draft of genome sequences are available. Many other plants like maize, tomato *Medicago truncatula* and sorghum genome sequencing efforts are in progress (Bedell *et al.*, 2005). The researcher has created expressed sequence tags (ESTs) from plants such as sorghum, wheat, cotton, beet, soybean and wheat. (<http://www.ncbi.nlm.nih.gov/dbEST/>).

Genome Sequencing

Sequencing technologies advances provide opportunities for processing, managing, and analyzing sequence in bioinformatics. For the sequencing of genome most common method is a shotgun. DNA pieces are randomly sheared, cloned and sequenced in parallel. There is software which can place together with the overlapping sequences which are sequenced separately (Myers, 1995). Many packages of software exist for sequence assembly (Gibbs *et al.*, 2003) such as Phred/Phrap/Consed (<http://www.phrap.org/>), Arachne (<http://www.broad.mit.edu/wga/>), and GAP4 (<http://staden.sourceforge.net/overview.html>). Package of a modular, open-source developed by TIGR has known AMOS (<http://www.tigr.org/software/AMOS/>), which can be used for assembly of the comparative genome (Pop *et al.*, 2004).

Gene Finding and Genome Annotation

Gene finding refers to introns and exons prediction in a DNA sequence segment. A Computer programs Dozens are available for identification of protein-coding genes (Zhang, 2002) such as (<http://genes.mit.edu/GENSCAN.html>), GeneMarkHMM (<http://opal.biology.gatech.edu/GeneMark/>), Genie (<http://www.fruitfly.org/seqtools/genie.html>), GRAIL (<http://compbio.ornl.gov/Grail-1.3/>), and Glimmer (<http://www.tigr.org/softlab/glimmer>). Many new tools of gene-finding are tailored for plant genomic sequences applications (Schlueter *et al.*, 2003). Prediction of Ab initio gene remains a challenging problem for large size eukaryotic genomes. A typical gene of *A. thaliana* with five exons, it is expected that at least one exon have at least one of its borders predicted incorrectly by the ab initio approach (Brendel and Zhu, 2002). Transcript evidence from full-length cDNA or EST sequences or similarity to homologs protein potential can reduce gene identification uncertainty significantly (Zhu *et al.*, 2003). In "structural annotation" of genomes, such techniques are widely used which refers to the features identification like genes and transposons in a sequence of the genome using ab initio algorithms and other

information. For structural annotation, many software packages have been developed (Allen *et al.*, 2003).

Genome comparison tools can be used to enhance gene identification accuracy like as SynBrowse (<http://www.synbrowser.org/>) and VISTA (<http://genome.lbl.gov/vista/index.shtml>). An important genome annotation aspect is the repetitive DNAs, the analysis which is identical copies or nearly identical to the sequences present in the genome (Lewin, 2003). Repetitive sequences are present in any genome and abundant in most of the plant genomes (Jiang *et al.*, 2004). The identification and characterization of repeats are crucial to shed light on the evolution of genomes, function, and organization to enable filtering for homology searches of many types. Plant-specific repeats small library can be found at [ftp://ftp.tigr.org/pub/data/TIGR Plant repeats/](ftp://ftp.tigr.org/pub/data/TIGR_Plant_repeats/); this is likely to grow as more genomes are sequenced substantially. To search genome repetitive sequences Repeat-Masker (<http://www.repeatmasker.org/>) can use. Working from a known repeats library, Repeat Masker is built upon BLAST and can screen sequences of DNA for interspersed repeats and regions of low complexity. Repeats with poorly conserved patterns or short sequences are hard to identify due to the limitations of BLAST using Repeat-Masker. Various algorithms were developed to identify different repeats some widely used tools such as RECON (<http://www.genetics.wustl.edu/eddy/recon/>) and RepeatFinder (<http://ser-loopp.tc.cornell.edu/cbsu/repeatfinder.htm>).

Sequence Comparison

Comparison of sequences can be vital to provide a foundation for many tools of bioinformatics and may allow genes and genomes, structure, and evolution function. For example, comparison of sequence provides a basis for a consensus gene model building like UniGene (Blueggel *et al.*, 2004). For identification of homology, many computational methods have been developed (Wan and Xu, 2005). Comparison of the sequence is highly useful; it is similarity based sequence between two text strings, which may not correspond to homology especially when the result confidence level of a comparison is small. Comparison of sequence methods can be mainly grouped into pair-wise, profile sequence and profile-profile comparison. Among researchers for pair-wise comparison of the sequence are BLAST (<http://www.ncbi.nlm.nih.gov/blast/>) and (<http://fasta.bioch.virginia.edu/>) are popular. To evaluate the level of confidence for an alignment to represent a homologous relationship, a statistical measure (Expectation Value) was integrated into pair-wise sequence alignments (Karlin and Altschul, 1990). Pair-wise sequence alignment missed remote homologous relationships due to its insensitivity. For detecting remote homologs, sequence-profile alignment is more sensitive. A profile of protein sequence is generated by a closely related proteins group of multiple sequence alignment. A multiple sequence alignment

builds correspondence among residues across all of the sequences simultaneously; sequences show the functional and structural relationship where it aligned in different positions. A profile of sequence is calculated using the occurrence probability for each amino acid at each of alignment. A famous example of a sequence-profile alignment tool is PSI-BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>). Proteomics is the main innovation for the qualitative and quantitative proteins characterization and their interactions on a genome scale. The proteomics targets large-scale identification and all protein types' quantification in a cell or tissue, post-translational modification analysis and association with other proteins, and protein activities characterization and structures.

Table.1. Applications of basic bio informatics tools in modern agriculture

DNA sequence analysis	<ul style="list-style-type: none"> ▪ Blast ▪ Clustal X ▪ Promoter analysis ▪ Gene prediction ▪ Regulatory elements ▪ Intron, exon findings ▪ Primer designing ▪ Codon usage optimization ▪ Virtual translation
Molecular dynamic simulations	<ul style="list-style-type: none"> ▪ Protein-DNA simulations ▪ Protein-ligand simulations ▪ Drug-DNA simulations
Pesticide preparations	<ul style="list-style-type: none"> ▪ Target identification ▪ Target validation ▪ Lead identification ▪ Lead optimization ▪ ADMET prediction
Protein sequence analysis	<ul style="list-style-type: none"> ▪ Molecular mass ▪ Amino acid composition ▪ Domain and motifs search ▪ Signal peptide identification ▪ Secondary structure analysis
Phylogenetic analysis	<ul style="list-style-type: none"> ▪ Reconstruction of evolution history ▪ Tracking gene flow ▪ Identification of conserved regions
Molecular modeling and interactions	<ul style="list-style-type: none"> ▪ 3D structure prediction ▪ Protein function prediction ▪ Protein-protein docking ▪ Finding inhibitors and activators ▪ Protein-DNA interactions ▪ Protein-ligand interactions ▪ Transcriptional factor identification

Ontologies Applications

Ontology is a set of vocabulary terms whose relations and meanings with other terms are stated explicitly and which are used to annotate data (Ashburner *et al.*, 2000). Ontologies are used for description of gene and protein function (Harris, 2004), types of cell (Bard *et al.*, 2005), organisms anatomies and stages of development (Garcia-Hernandez, 2002), metabolic pathways (Mao *et al.*, 2005) and microarray experiments (Stoeckert *et al.*, 2002). To annotate data ontologies are used such as sequences, experiments and strains cluster of gene expression.

Databases

Traditionally, biologists depend on research articles and textbooks published in scientific journals as the primary source of information. This has changed now dramatically, the Internet and Web browsers became commonplace. The Internet has become the first place for researchers, to find information these days.

Types of Biological Databases

There are three types of biological databases that have been established and developed: community-specific databases, large-scale public repositories and project-specific databases. Nucleic Acids Research (<http://nar.oxfordjournals.org/>) publishes an issue of the database in every year January, and Plant Physiology has started publishing databases describing articles (Rhee and Crosby, 2005). Government agencies or international consortia developed and maintained large-scale public repositories and places for long-term data storage. Examples include sequences GenBank (Wheeler *et al.*, 2005), UniProt (Schneider *et al.*, 2005) for information on protein, Protein Data Bank (PDB) (Deshpande *et al.*, 2005), for structure information of protein and Array Express (Parkinson *et al.*, 2005) and Gene Expression Omnibus (GEO) (Edgar *et al.*, 2002) microarray data. There are many community-specific databases, which typically contain high standards information and address the particular researchers' community needs. Prominent community-specific databases are an example of those that cater to researchers focused on model organisms study (Lawrence *et al.*, 2005) or clade-oriented comparative databases (Gonzales *et al.*, 2005). Databases focused on specific types of data such as metabolism (Zhang *et al.*, 2005) modification of protein (Tchieu *et al.*, 2003) are examples of community-specific databases. The community-specific databases concept is subject to change as researchers are widening their research scope. For example, databases focused on genome sequences comparing have recently emerged (e.g., <http://www.phytome.org>). Smaller-scale and short-lived are the third category of databases that are developed for management of data project during the funding period. These databases and web resources are reassured through the project funding period, and currently, there is no depositing or archiving standard way of these databases after the period of funding. Some issues are

observed in database management of the database. The major aim of the projects is to a generic organism database toolkit to allow researchers to a genome database "off the shelf" set up. There is a general infrastructure for supporting, managing and using digital data archived in databases and websites in the long term (Lord and Macdonald, 2003). Several projects are building systems of the digital repository that can be models for a repository such as DSpace (<http://dspace.org/>) and the CalTech Open Digital Archives (CODA; <http://library.caltech.edu/digital/>) Collection. Some additional challenges in long-term data archiving were articulated in a recent National Science Board report (<https://www.nsf.gov/pubs/2005/nsb0540/nsb0540>).

Representation of data and Storage

Different methods are used to developed databases such as software of object-oriented database, simple file directories and software of relational database. Because of the expanding information amount that should be deposited and made available utilizing the internet, the software of relational database management has become popular and has become the de facto standard in biology. Relational databases provide sufficient storing means and retrieving data of large quantities via indexes, normalization, triggers, referential integrity, and transactions. The notable software of relational database that is freely available and popular in bioinformatics is MySQL (<http://www.mysql.com/>) and PostgreSQL (<http://www.postgresql.org/>). Data are represented as attributes, entities and relationships between the entities in relational databases. This representation type is called Entity-Relationship (ER), and database schemas are described using ER diagrams (TAIR schema <http://arabidopsis.org/search/schemas.html>). Attributes and entities become columns and tables in the physical database implementation respectively. Data are the values that are stored in the tables' fields. For storing large data quantities, relational databases are powerful ways.

Data Access and Exchange

Structured query language (SQL) is easiest and powerful way of data accessing in a database (<http://databases.about.com/od/sql/>). SQL has an intuitive reasonably and simple syntax that requires programming knowledge and is suited to learn without a steep learning curve for biologists. Information accessing from a database is easy if one knows which database is to go, it is hard to find information if one does not know which database to search. There are many ways to solve this problem like database-driven pages content indexing developing software that will directly connect to individual databases or develop different types of a data warehouse or in one site database. Data formatting simple way is using a system of tag, and values are known markup language. For data exchanging and information via the web is Extensible Markup Language (XML) which is an emerging standard. It allows to information providers define new attribute names and tag at

will and to nest document structures to any complexity level, among other features. The document that defines tags meaning for an XML document is called Document Type Definition (DTD). The common DTD use allows different users and applications to exchange data in XML.

CONCLUSION AND FUTURE PERSPECTIVES

In the current review, we try to discuss some advances like gene expression, sequence and databases, and ontologies, in the key areas of bioinformatics field. Numerous unsolved issues exist in the field of bioinformatics today which includes data and integration of database, robust interpretation of phenotype from genotype, automated knowledge extraction, and established researchers. Bioinformatics is an approach that will be a major role in the research of plants. On the off chance that plant science can be summed into a single word, it would be "integration." The next 50 years we will see basic research integration with applied research in which plant biotechnology will assume a critical part in taking care of numerous issues such as reducing world hunger and poverty, developing renewable energy sources and preserving the environment. We will see disparate integration, specialized plant research area into more comparative connected, holistic views and approaches in the biology of plants. Bioinformatics will give the glue with which all of these types of integration will occur. More time of researchers will be spent on the internet and computer for description and generation of data to perform their experiments. They also analyze and find other people's data for comparison to find knowledge existing in the relevant field to publish results to the world.

CONFLICT OF INTEREST

The authors declare that they don't have any conflicts of interest and are also not interested in competing with anyone.

REFERENCES

- Ali, Q., Javed M.T, Noman, A., Haider, M.Z., Waseem, M, Iqbal, N., Waseem, M., Shah, M.S., Shahzad, F., Perveen, R., 2017. Assessment of drought tolerance in mung bean cultivars/lines as depicted by the activities of germination enzymes, seedling's antioxidative potential and nutrient acquisition, Archives of Agronomy and Soil Science., DOI:10.1080/03650340.2017.1335393
- Arabidopsis Interactome Mapping Consortium., 2011. Evidence for network evolution in an Arabidopsis interactome map. *Sci.*, 333:601–607.
- Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, 25:25–29.
- ArrayExpress., a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.*, 33:553–55.
- Bard, J., Rhee, S.Y., Ashburner, M., 2005. An ontology for cell types. *Genome Biol.*, 6:21.
- Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., 2005. Sorghum genome sequencing by methylation filtration. *PLoS Biol.*, 3:13.
- Blueggel, M., Chamrad, D., Meyer, H.E., 2004. Bioinformatics in proteomics. *Curr. Pharm. Biotechnol.*, 5:79–88.
- Brendel, V., Zhu, W., 2002. Computational modeling of gene structure in *Arabidopsis thaliana*. *Plant Mol. Biol.*, 48:49–58.
- Deshpande, N., Address, K.J., Bluhm, W.F., Merino-Ott, J.C., Townsend-Merino, W., 2005. The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.*, 33:233–37.
- Edgar, R., Domrachev, M., Lash, A.E., 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, 30:207–10.
- Garcia-Hernandez, M., Berardini, T.Z., Chen, G., Crist, D., Doyle, A., 2002. TAIR: a resource for integrated *Arabidopsis* data. *Funct. Integr. Genomics.*, 2:239–53.
- Gibbs, R.A., Weinstock, G.M., 2003. Evolving methods for the assembly of large genomes. *Cold Spring Harb. Symp. Quant. Biol.*, 68:189–94.
- Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Sci.*, 296:92–100.
- Gonzales, M.D., Archuleta, E., Farmer, A., Gajendran, K., Grant, D., 2005. The Legume Information System (LIS): an integrated information resource for comparative legume biology. *Nucleic Acids Res.*, 33:660–65.
- Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, 32: D258–61.
- Islam, W., Zhang, J., Adnan, M., Noman, A., Zainab, M., Jian, W., 2017. Plant virus ecology: a glimpse of recent accomplishments. *App. Eco. Environ. Res.*, 15(1): 691–705.
- Jiang, N., Bao, Z., Zhang, X., Eddy, S.R., Wessler, S.R., 2004. Pack-MULE transposable elements mediate gene evolution in plants. *Nature.*, 431: 569–73.
- Karlin, S., Altschul, S.F., 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA.*, 87:2264–68.
- Kojima, M., Kamada-Nobusada, T., Komatsu, H., Takei, K., Kuroha, T., Mizutani, M. 2009 Highly sensitive and high-throughput analysis of plant hormones using MS-probe modification and liquid chromatography tandem mass spectrometry: an application for hormone profiling in *Oryza sativa*. *Plant Cell Physiol.*, 50: 1201–1214.

- Lawrence, C.J., Seigfried, T.E., Brendel, V., 2005. The maize genetics and genomics database. The community resource for access to diverse maize data. *Plant Physiol.*, 138: 55–58.
- Lewin, B., 2003. *Genes VIII*. Upper Saddle River, NJ: Prentice Hall.
- Lister, R., Gregory, B.D., Ecker, J.R. 2009. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr.Opin. Plant Biol.*, 12: 107–118.
- Lord, P., Macdonald, A., 2003. e-ScienceCuration Report–Data Curation for e-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision. Twickenham, UK: Digital Archiving Consultancy Ltd.
- Mahadevan,C.,Jaleel,A.,Deb,L.,Thomas,G.,Sakuntala,M., 2014.Development of an efficient virus induced gene silencing strategy in the non-model wild ginger - Zingiber zerumbet and investigation of associated Proteome changes. *PLoS ONE* 10:e0124518.doi:10.1371/journal.pone.0124518
- Mao, X., Cai, T., Olyarchuk, J.G,Wei, L., 2005. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics.*, 21: 3787–93.
- Mochida, K., Shinozaki, K., 2010. Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol.*, 51: 497–523.
- Moustafa, E., Sanaa, M., Milad, A., Nawar, Rohila., 2013. Proteomics: a biotechnology tool for crop improvement. *Front. Plant Sci.*, doi.org/10.3389/fpls.2013.00035.
- Myers, E.W., 1995. Toward simplifying and accurately formulating fragment assembly. *J. Comput. Biol.*, 2:275–90.
- Niehl,A.,Zhang,Z.,Kuiper,M.,Peck,S.,Heinlein,M.,2013.Label-free quantitative proteomic analysis of systemic responses to local wounding and Virus infection in Arabidopsis thaliana. *J. Proteome Res.*, 12:2491–2503.doi:10.1021/pr3010698
- Noman, A., M., Aqeel, J., Deng, N., Khalid, T., Sanaullah, H., Shuilin, H., 2017b. Biotechnological Advancements for Improving Floral Attributes in Ornamental Plants. *Front Plant Sci.*, 8:530.
- Noman, A., Fahad, S., Aqeel, M., Ali, U, Ullah, A., Anwer, S, Khan, S., Zainab, M, 2017a, miRNAs: Major modulators for crop growth and development under abiotic stresses. *Biotechnol Lett.* DOI 10.1007/s10529-017-2302-9
- Noman, A., Aqeel, M., 2017. miRNA-based heavy metal homeostasis and plant growth. *Environ Sci Poll Res.* 24:10068–10082 10.1007/s11356-017-8593-5
- Noman, A., Aqeel, M., He, S., 2016a, Crispr-cas9: tool for qualitative and quantitative plant genome editing. *Front Plant Sci* 7:1740. doi: 10.3389/fpls.2016.01740
- Noman, A., Bashir, R., Aqeel, M., Anwer, S., Iftikhar, W., Zainab, M., Zafar, S., Khan, S., Islam, W., and Adnan,M, 2016b, Success of transgenic cotton (*Gossypium hirsutum* L.): Fiction or reality?, *Cogent. Food Agri.* 2: 1207844
- Parkinson, H., Sarkans, U., Shojatalab, M., Abeygunawardena, N., Contrino, S., 2005.
- Pop, M., Phillippy, A., Delcher, A.L., Salzberg, S.L., 2004. Comparative genome assembly. *Brief Bioinform.*, 5:237–48.
- Rhee, S.Y., Crosby, B., 2005. Biological databases for plant research. *Plant Physiol.*, 138:1–3.
- Saito, K.,Matsuda, F., 2010.Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.*, 61: 463–489.
- Schneider, M., Bairoch, A., Wu, CH., Apweiler, R., 2005. Plant protein annotation in theUniProt Knowledgebase. *Plant Physiol.*, 138: 59–66.
- Schlueter, S.D., Dong, Q., Brendel, V., 2003.GeneSeqer@PlantGDB: gene structure prediction in plant genomes. *Nucleic Acids Res.*, 31:3597–600.
- Shinozaki, K., Sakakibara, H., 2009. Omics and bioinformatics: an essential toolbox for systems analyses of plant functions beyond 2010. *Plant Cell Physiol.*, 50: 1177–1180.
- Stoeckert ,C.J. J., Causton, H.C., Ball, C.A., 2002. Microarray databases: standards and ontologies.*Nat. Genet.*,32:469–73.
- Tchieu, J.H., Fana, F., Fink, J.L., Harper, J., Nair, T.M., 2003. The PlantsP and PlantsT Functional Genomics Databases. *Nucleic Acids Res.*, 31:342–44.
- The Arabidopsis Genome Initiative., 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature.*, 408:796–815.
- Wan, X., Xu, D., 2005. Computational methods for remote homolog identification.*Curr. Protein Peptide Sci.*, 6:527–46.
- Weinhold,A,Wielsch,N.,Svato,A.,Baldwin,I.,2015. Label-free nano UPLC-MSE based quantification of antimicrobial peptides from the leaf apoplast of *Nicotiana attenuata*. *BMC Plant Biol.*, 15:18.doi:10.1186/s12870-014-0398-9
- Wheeler, D.L., Smith-White, B., Chetvernin, V., Resenchuk, S., Dombrowski, S.M., 2005.Plant genome resources at the national center for biotechnology information. *Plant Physiol.*, 138:1280–88.
- Zhang, M.Q., 2002. Computational prediction of eukaryotic protein-coding genes.*Nat.Rev. Genet.*, 3: 698–709.
- Zhang, P., Foerster, H., Tissier, C.P., Mueller, L., Paley, S., 2005. MetaCyc and AraCyc.Metabolic pathway databases for plant research. *Plant Physiol.*, 138:27–37.
- Zhu, W., Schlueter, S.D., Brendel, V., 2003.Refined annotation of the Arabidopsis genome by complete expressed sequence tag mapping. *Plant Physiol.*, 132:469–84.