



EXPLORING PLANT RESEARCH WITH THE HELP OF VITAL BIOINFORMATICS DATABASES AND ONLINE RESOURCES

Noreen Fatima¹, Haris Abdul Rehman², Mah Noor Hassan³, Iqra Naeem⁴, Babur Ali Akbar⁵, Muhammad Nabeel Aslam⁶, Zanib Fatima⁵, Aqsa Qurban⁷, Muhammad Umar⁸, Tayyaba Arif^{9*}

¹Department of Plant Pathology, University of Sargodha, Pakistan

²Department of Microbiology, University of Central Punjab, Lahore, Pakistan

³Department of Biochemistry, University of Central Punjab, Lahore, Pakistan

⁴Department of Environmental Sciences, COMSATS University Vehari Campus, Pakistan

⁵Centre of Agricultural Biochemistry and Biotechnology, University of Agriculture Faisalabad, Pakistan

⁶Department of Plant Breeding and Genetics, University of Sargodha, Pakistan

⁷Department of Life Sciences, University of Management and Technology, Lahore, Pakistan

⁸Government College Women University Faisalabad, Pakistan

⁹Kinnaird College for Women, Lahore, Pakistan

***Corresponding author:-** Tayyaba Arif
Email:-tayyabaarif47@gmail.com

Abstract

Bioinformatics plays a role, in the field of plant science today. With an increase in data volume, there is a growing demand for tools and methods for managing, visualizing, implementing, evaluating, modeling, and predicting this data. However many biology researchers may lack familiarity with the bioinformatics resources, which can lead to missed opportunities and misinterpretation of the data. In this review article, we highlighted the web resources that offer analysis capabilities for plant research data including genomics, transcriptomics, comparative genomics, bio-ontologies, sequence and structural comparisons plant disease related databases well as proteomics databases. Additionally we provide insights into integrated modules found within these resources that are specifically tailored for analyzing plant associated data. Overall this review aims to assist plant researchers in accessing bioinformatics resources for their data analysis needs while promoting the use of bioinformatics tools to effectively address experimental challenges, within the field of plant sciences.

Keywords: Databases, Genomics, Bioinformatics, Plant, Crop, Web Servers.

1. Introduction

The Plant Kingdom is very beneficial not only to humans but to other living organisms as well[1]. A key role for plants is to provide an enormous amount of food. Plants are often used in the manufacturing of several human drugs, and were used as model organisms to research transposable elements in heterochromatin and epigenetic regulation[2]. Consequently, the research of plant biology has been carried out extensively since the early stages of human existence owing to its essential position. Technological innovations propelled plant biology studies to a higher level than before. High-throughput sequencing methods innovation gives researchers the opportunity to manipulate the

genetic material structure at the molecular level known as "genomics." Study on plant genomics has recently exploded and is becoming the central feature in plant research due to the rapid growth of sequenced genomes of several plant species[3]. The immense increase of plant genome research directly improve commercially significant plants[4]. Greater availability to this plant genomic knowledge and continuous changes build a favorable atmosphere for the development of plant science. This needs good liaison and cooperation between the global biological community[5]. This sea of data is too much for the human brain to comprehend and therefore, computational methods are increasingly needed to store and analyze these data. Thus to organize, store, retrieve and analyze such data bioinformatics has come up with a solution to develop biological databases, which could help researchers to easily store such big data.

Bioinformatics refers to the study of biological information using concepts and methods in computer science, statistics, and engineering. For example about 1685[6] online databases are currently available to the public at NAR online, Molecular Biology Database Collection. Such databases can be classified according to the data type, data curation processes, data distribution area, and database availability. Several such publicly funded resource host data (raw, annotated, analyzed) for different species includes crops, model and non-model plants. In contrast, others are related to a class of taxonomic species and may contain some form of data. Furthermore, there is an array of resources and web applications that enable the processing, analysis, and visualization of various genomic, proteomics, sequence and structural data[6, 7]. Bioinformatics has a fundamental place in many aspects of plant biology, including stress biology [8]. Every organism used to face stress weather it is biotic stress or abiotic stress in life, which leads to have significant impacts on its growth and development. It has been reported that plant productivity has been affected in a sufficient amount that's why stress biology has become an important and famous field and massive computational analysis tools generated yet to deal with the researches related to stress biology. For instance a number of local and international projects are dedicated to maintain the available public databases and web servers. For example, Genbank[9], NCBI[10], INSDC[11], EMBL[12], DDBJ[13], Ensemble for sequences and annotation, PDB[14] and DDBJ for protein structures while UniProt[15] for protein information. The databases are sustainable archives for valuable research data related to plants and other organisms. Community-specific databases such as those for plants, animals, and humans answer essential questions. In the case of plants, many important databases for a specific purpose, including PlantGDB[16], ENSEMBL Plants[17], Gramene[18], PLEXdb[19], Gene Expression Atlas and Planteome are developed to store data and answer related questions[16]. Besides, each database may be allocated to one or more categories based on its content, e.g., databases for gene expression, small-RNA databases, molecular interaction databases, protein or nucleotide databases, phenome, genome annotation, pathway and genomic variation databases[20].

Available bioinformatics resources, including software, web access, and databases have brought a significant revolution in the area of plant research. For better analysis and interpretation of experiments, these tools considered to be the best approach of the era [14, 15]. Thus the purpose of this review article is;

1. To collect the available webresources including both generic and specialized genomic, immunity, proteomics, disease related and other databases and analytics tools to support plant researchers.
2. The review also characterized each database and grouped into a specific based on the specified features such as the purpose of the database, the dataset (samples) and its used either plant specific or general.
3. This review also provides information regarding all the available databases and their applications in different studies.

4. Plant Genomics Databases (General)

Genomics is a biological discipline that focuses on the structure, functionality, evolution, visualization, and genome editing. Recent advancements in science have significantly increased our capacity to study and recognize plant genomes and reduce the difference between genotype and

phenotype. The rapidly growing genomics area helps scientists to simultaneously study thousands of genes, to grasp the genetic design of plant genomes, and even to distinguish the gene variants. Plant genomics is essential to recognize traits related to nutritional quality and weed, disease, and resistance to drought; help breeding activities. Advances of genomics techniques and bioinformatics play a significant role in the exploration of valuable genetic variation. Innovative technologies are providing bioinformatics to process, manage, and analyze sequencing data more efficiently as the commonly used Shotgun sequencing method for genome sequencing. For arranging data, like putting together the overlapping random segments and parallel sequencing, software has been developed [21]. Software like GAP4 (<http://staden.sourceforge.net/overview.html>) Arachne (<http://www.broad.mit.edu/wga/>) and Phred/Phrap/Consed (<http://www.phrap.org>), are also available for sequence assembly. A new open-source modular AMOS in TIGR (<http://www.tigr.org/software/AMOS/>), is developed for comparative genome assembly. Lately, with emerging technologies, sequencing such as polymorphism ratio sequencing along with oligonucleotide probes differential hybridization is very cost-effective [22-24]. However, Bioinformatics still have to encounter important analytical challenges with these methods in data management like its concurrence with other data as well as its interpretation and designing. Though numerous tools for different purposes like tools for finding new genes, and identification of genes that code for a specific protein have been developed still, plant science faces challenges in computational gene modeling specifically in large genomes (eukaryotic). More significantly, tools and software for identification and structural annotation of similar or nearly similar sequences of DNA (or repetitive DNA) are also available. Still, there is a need for more advanced technologies or prediction of that repetitive DNA sequences present in an organism. Thus to help the plant researchers to get a clear view of all these genomics problems and design an effective pipeline for their experimental studies, integration of bioinformatics is requisite. Hence, many of the tools which help to provide important information regarding the plant genomes are developed and published by the bioinformatics community. These tools are primarily available online for free or commercial use. Here, we have collected such tools and defined their data points with key features implemented in it. The shortlisted tools which could help the researcher to analyze their genomics data are given in **Table 1**.

i.GCP (The Generation Challenge Programme)

The GCP is a database (<http://www.generationcp.org>) that is world-widely used for investigating crops and aims for crop improvement through characterizing genetic assets for plant breeding[25].

ii.Graphical genotype (GGT)

GGT accessible at (www.plantbreeding.wur.nl) concentrates on simple association studies along with genetic analysis and diversity analysis on plant's genetic data and shape marker patterns. The similarity of markers also can be estimated based on similarities of genotypes. Visualization and analyses via GGT 2.0 involve molecular marker scores[26].

iii.GabiPD

GabiPD (<http://www.gabipd.org/>) database uses GABI projects and different 'Omics' such as genomics, transcriptomic, proteomics, and metabolomics to collect, process and visualize data originating from 14 different model crops species[27]. For several plants, GabiPD appoints a repository analysis platform for a broad set of heterogeneous arrays data via high throughput analysis and allows reciprocal genetic maps for potato barley and provides 2DE gels for proteins from *A. thaliana* and *B. napus*. MapManWeb can be used to visualize metabolic-profiling data and Gene expression.

iv.GrainGenes

GrainGenes is a global database (<http://www.graingenes.org>) for cereal crops like wheat, oat, rye, and barley genomes with 81 full-genome maps and serves as storage for primary information about QTLs genetic mapping, primers and probes [28].

v.GDR (Genome Database for Rosacea)

It's a database (<http://www.rosaceae.org>) with genetic information about economically significant plant's family like Rosacea includes databases of easily accessible Rosacea ESTs annotation and markers with genetic maps of Rosaceae[29]. The ESTs are collected for the production of sets for unigene of each of Rosacea family. Different annotations like ORFs, SNPs, GO, etc. are also applicable. The Comparative map (CMap) viewer can be used to view and compare the available Rosacea genetic maps. In contrast, a physical map for peach can be viewed by WebFPC/Web Chrome, or through another integrated GDR map viewer that serves as a portal for information about the combined genetic and can search through different categories. The query result sites are coupled with visualization tools. Additionally, for micro-satellites and primer detection, an online sequence assembly server GDR database is also provided that uses a BLAST/FASTA sever.

vi.MaizeGDB

MaizeGDB database (<https://www.maizegdb.org>) in present days house 12 completely sequence from maize along the inbred lines (B73, PH207, F7, W22, Mo17, EP1) and is expecting to be assembling 40 genomes by 2020[30].

i.PREP (predictive RNA editors for plants)

The PREP (<http://prep.unl.edu/>)[31] suite that predicts RNA sites editing computationally, established on the familiar phenomenon that conservation of proteins across species increases by editing plant organelles. Predictive RNA editors are provided by PREP suite for genes of different plant organelles such as chloroplast (PREP-Cp), mitochondria (PREP-Mt), and (PREP-AIn) for users submitted alignments. These servers come with key advantages such as fastness, minimal input requirement, and the most important one, high accuracy. Usage of PREP-Mt has been found very desirable in many research findings, and the recently established servers, PREP-AIn and PREP-Cp should be of more help for studies that necessitate information of the locality of sites of RNA editing.

ii.PMRD

PMRD (<http://bioinformatics.cau.edu.cn/PMRD>) is a curated database for genes and mutants reported to be involved in the reproduction of male plants [32]. It contains information on 4203 (male-sterile genes) MS genes, 3697 (male reproduction-related genes) MR genes, and 484 mutants of 33 plant species.

Table 1: List of plant genomics databases with URLs and key features highlighted.

iii. Plant MPSS databases

S. No	Database / Server Names	Link	Features		
			Molecular and Genetic Markers/Maps/ RNA Analysis	Sequence Comparison	Mutations and EST Analysis
i.	GCP	https://www.generationcp.org/	✓	✓	✓
ii.	GGT	http://www.plantbreeding.wur.nl	✓	✓	x
iii.	GabiPD	http://www.gabipd.org/	✓	✓	✓
iv.	GrainGenes	https://wheat.pw.usda.gov/GG3/	✓	✓	✓
v.	GDR	http://www.rosaceae.org/	✓	✓	✓
vi.	GeneCAT	http://genecat.mpg.de/	✓	✓	✓
vii.	MaizeGDB	http://www.maizegdb.org/	✓	✓	✓

viii.	PREP Suite	http://prep.unl.edu/	x	✓	✓
ix.	PMRD	http://bioinformatics.cau.edu.cn/PMRD	✓	X	x
x.	Plant MPSS databases	http://mpss.udel.edu/	✓	X	x
xi.	Plant snoRNA database	http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snoRNA/home	✓	X	x
xii.	PlantCARE	http://bioinformatics.psb.ugent.be/webtools/plantcare/html/	✓	X	x
xiii.	PlantProm DB	http://www.softberry.com/berry.phtml?topic=plantprom&group=data&subgroup=plantprom	✓	X	✓
xiv.	PmiRKB	http://bis.zju.edu.cn/pmirkb/	✓	X	x
xv.	Panzea	http://www.panzea.org/	✓	✓	x
xvi.	PoMaMo	https://www.gabipd.org/projects/Pomamo/	✓	✓	✓
xvii.	PlncRNADB	http://bis.zju.edu.cn/PlncRNADB/	✓	X	x
xviii.	The Plant DNA C-values Database	http://data.kew.org/cvalues/homepage.html	x	X	✓
xix.	UK CropNet	http://ukcrop.net/	✓	✓	✓
xx.	pssRNAMiner	http://bioinfo3.noble.org/pssRNAMiner/	✓	X	✓
xxi.	psRNATarget	http://plantgrn.noble.org/psRNATarget/	✓	X	x
xxii.	TPIA	http://tpia.teaplant.org/	✓	✓	✓
xxiii.	AtCircDB	http://genome.sdau.edu.cn/circRNA	✓	X	x
xxiv.	PVsiRNADB	http://www.nipgr.res.in/PVsiRNADB	✓	X	x
xxv.	PlantNATsDB	http://bis.zju.edu.cn/pnatdb/	✓	X	x
xxvi.	TAIR	http://www.arabidopsis.org/	✓	✓	✓
xxvii.	Plant Homolog Database	http://phd.big.ac.cn	x	✓	✓

MPSS databases calculate most gene expression levels and give results about new transcripts such as intergenic regulatory transcripts, antisense transcripts, and alternative splice isoforms. For deep profiling of Arabidopsis' small RNAs, an altered version of MPSS has been utilized. Small RNA MPSS data interpretation is facilitated by the presence of widespread repeat data in the genome viewer. All tools and data presented in this article are available at (<http://mpss.udel.edu>)[33].

iv. Plant snoRNA database

This database of plants (http://www.scri.sari.ac.uk/plant_snoRNA/) gives data regarding small nucleolar RNAs of Arabidopsis along with the other 18 species of plant. Pseudouridylation target modification sites, expression data and methylation, sequences, and gene organization (single

gene, intronic and polycistronic,) and variant genes frequency related information are provided by this valuable database[34].

v. PlantCARE

For enhancers, cis-regulatory elements, and repressors, the PlantCARE database has been developed. Essential regulatory elements, positional matrices with specific places on the promotor region are symbolized. Data related to transcription start sites is supplemented with an enlarge amount of in silico predicted data obtained from literature mining. This database also gives immensely information related to levels of confidence for the specific experimental evidence as well, along with a general description for transcription factor sites and position on promotor. The plant care database can be access at (<http://sphinx.rug.ac.be:8080/PlantCARE/>)[35].

vi. PlantProm DB

PlantProm DB is a database and can be accessed at (<http://mendel.cs.rhul.ac.uk/>)[36]. PlantProm DB, a database of promoter sequences for RNA polymerase II, comprised of well-characterized not-redundant group of proximal promoter sequences. This database contain experimentally reported transcription start site(s)(TSS) obtained from various plant sources. It can be used to find the promoter sequence and Nucleotide Frequency Matrices (NFM) for promoter elements such as CCAAT-box, TATA-box, and TSS-motif (Inr). However, analysis of TSS-motifs shown that their conformation is diverse in dicots and monocots, as well as for TATA-box and TATA-less promoters. This database also provides a way for the prediction of novel promoters that are not previously characterized.

vii. PmiRKB

Plant miRNA knowledge base (PmiRKB) (<http://bis.zju.edu.cn/pmirkb/>)[37] based on four main functional modules. A diverse data from different sources such as SNP data seven Arabidopsis species and 21 Oryza sativa subspecies are available under the 'SNP module'. This information can be used then to scrutinize the SNPs within precursor microRNAs and miRNA-target (RNA duplexes). To identify the tissue-specific precursor and primary microRNAs, a second module named 'Pri-miR' can be accessed. To authenticate the miRNA-target pairs, another integrated module named MiR-Tar has been added, which used equivalent analysis of RNA end (PARE) data miRNA-target pairs validation. Alternatively, the Self-reg module which also uses the same PARE data and examines the metabolism of miRNA precursors.

viii. Panzea

A useful bioinformatics arm for the Molecular and functional analysis of the Maize genome is Panzea (<http://www.panzea.org>). Panzea found the user-friendly web-resource to access genotype to phenotype information with comprehensive services related to maize assortment[38].

ix. PoMaMo

This database consists of approximately 1000 mapped elements with sequence data and gene function that results from BLAST analysis of all twelve chromosomes in the potato genome. This potato genome analysis based database is helpful SNP and InDel information from diverse kinds of potato genotypes with publication references and links to other databases. PoMaMo database can be accessed at (<https://gabi.rzpd.de/PoMaMo.html>).

x. PlncRNADB

The PlncRNADB is a portal or a database for recording the lncRNA sequencing and annotation of plants of species (Arabidopsis lyrata, Arabidopsis thaliana, Zea mays, and Populus trichocarpa). This database presents the link between lncRNAs and different RNA-binding proteins (RBPs) through a user-friendly interface. lncRNAs interaction with RNA-binding proteins in plants can also be studied. PlncRna is available at <http://bis.zju.edu.cn/PlncRNADB/>[39].

xi. The Plant DNA C-values database

The Plant DNA C-values database is a web-based database, available at (<https://cvalues.science.kew.org/>) that has genome size data of 12,273 species. Ten thousand seven hundred seventy species of angiosperms, C-values of major land plants gymnosperms, 334 bryophytes (mosses-209, liverworts-102, and hornworts-23) and 303 pteridophytes (ferns (monilophytes)-246 and lycophytes-57), 445 algae, Chlorophyta, Rhodophyta, Phaeophyta, the streptophyte green algae within Kingdom Plantae, and Heterokonta within the Stramenopiles are recorded and displayed in a user-friendly interface that facilitates the user to browse and acquire output[40].

xii. UK CropNet

UK CropNet (<http://synteny.nott.ac.uk/>) consists of a group of different tools for the identification and handling of agriculturally important genes by developing a basis for comparative analysis of crop plants with model species. It also makes the illustration and analysis of the data easy[41].

xiii. pssRNAMiner

pssRNAMiner [42] is a web server for the identification of the group of phased small RNAs and the potential phase-initiator. The pssRNAMiner can be accessed at (<http://bioinfo3.noble.org/pssRNAMiner/>).

xiv. psRNATarget

The psRNATarget is a web-based server, which can be accessed at (<http://plantgrn.noble.org/psRNATarget/>)[43]. It integrates identification of plant miRNA target recognition, i.e., it differentiates translational inhibition and post-transcriptional repression, and it describes small RNA and target site pairs, which can influence small RNA binding to a target. The psRNA Target is responsible for efficient next-generation data analysis with a systematic back-end computing pipeline operated on a Linux cluster. The server combines three rationalized and convenient interfaces to obtain preloaded or user-submitted sequences of small RNAs and transcripts. It results in a detailed short RNA/target pairs list alongside the web-based tools for keyword searching, batch downloading, and results sorting.

xv. Tea Plant Information Archive (TPIA)

TPIA (<http://tpia.teaplant.org/>) is generally used to clarify tea quality by using the tea plant genome as a model system and merge efficient orthologues, transcriptomes, gene expression, and abundant metabolites across different tissues, species, and stresses. It has a huge collection of polymorphic simple sequence repeats, correlations, transcription factors, globally collected germplasm, and manually assisted functional genes information[44].

xvi. AtCircDB

AtCircDB is freely accessible at (<http://genome.sdau.edu.cn/circRNA>), which is a database of circular RNA of specific Arabidopsis tissue[45]. This has 622 Arabidopsis RNA sequencing data sets from 87 independent NCBI SRA studies, and AtCircDB has been developed to classify, store and access circular RNAs systemically. This comprises of 30,648 tsc-RNAs (tissue-specific circular RNAs), 84 685 cRNAs (circular RNAs) and 3486 miRNA-cRNA (microRNA-circular RNA) interactions.

xvii. PVsiRNadb

PVsiRNadb available at (<http://www.nipgr.res.in/PVsiRNadb>), is a manually design plant-exclusive database containing detailed associated with vsiRNAs, based on probing of currently available scientific data of virus-infected plants[46]. This database comprises of 282 549 different sequences of vsiRNAs and 322 214 of total entries. In this database, comprehensive and detailed information present for all sequences of vsiRNA. Besides the primary information, which consists of the plant, tissue, name of the virus, and vsiRNA sequence, additional details of each vsiRNA

(map location, length, coordinates, strand details and the structure predicted) can be very useful to the user. PVsiRNAdb offers various types of search and navigation modules with three different tools: BLAST, Smith-Waterman Align, and Mapping.

xviii. PlantNATsDB

Approximately 2 million natural antisense transcripts (NAT) pairs are stored in PlantNATsDB (<http://bis.zju.edu.cn/pnatdb/>) database obtained from 69 plant species[47]. Moreover, for functional analysis of statistically significant overrepresented GO categories, a gene set analysis module antisense transcripts (NAT) specific is an integrated part of this database.

xix. TAIR

The genetic, as well as molecular data of *Arabidopsis thaliana*, has been maintained in Arabidopsis Information Resource (TAIR)[48]. The genetic and molecular data covers genome sequences, gene structure, gene expression, gene product, genetic markers, DNA stocks, seed stocks, publications, and information about Arabidopsis research. The data is usually updated weekly bases.

5. Functional Genomics databases

Functional genomics is a notion used to characterize data collection to help explain the cell's type and function. Microarrays are being used in the first instance since the late twentieth century to understand the transcriptional process of a biological sample. Analytical studies on transcriptome are done significantly for the insight of the development of an organism and its interaction with its environment as well as transcript abundance growth control. In this regard, DNA microarrays are the most efficient tool for genome-wide transcriptional studies [49, 50]. These studies are also done through combining microarrays with other technologies, for example, pathway information, gene-ology, and analysis of regulatory sequences to perform co-regulated processes. At the same time, arrays linked with the whole-genome are tailored to observe transcription with accuracy in alternative spliced variants and in the known or detected structure for a gene.

Additionally, techniques like DNA and Chromatin immunoprecipitation chip (DNA-chip and ChIP-chip) are also used to predict DNA binding and protein-DNA interaction sites in a genome [51, 52]. Contrary to the gene-by-gene approach of the classical molecular biology technique, functional genomics deals with the whole structure, function, and regulation of a gene. It seeks to connect the gene expression-level phenotype and genotype, including pathways such as transcription, translation, PPI, and epigenetic regulation.

Because of the massive amounts of data generated by these methods and the attempt to know biologically useful information, bioinformatics is particularly crucial for the interpretation of functional gene expression data. Examples of strategies in this area include data clustering or PCA for unsupervised machine learning (area detection), and also artificial neural networks or support vector machines for controlled machine learning (event prediction, classification). Analysis of functional enrichment is used to assess the degree of functional groups over- or under-expression (positive or negative- regulators in case of RNAi screens) compared to a context collection[53]. Study of gene ontology dependent enrichment is given by DAVID and an overview of gene set enrichment (GSEA), pathway-based analysis by Ingenuity and Pathway studio, and protein complex based analysis by COMPLEAT[54]. Thus to highlight the key developments of bioinformatics in this domain, we have shortlisted the key functional genomics tools given in **Table 2** and explained their key features with model plant species. These shortlisted databases are primarily useful for functional genomics data retrieval and analysis, which could help researchers get a clear picture of the experimental data.

Table 2: List of plant functional genomics specific databases with URLs and key features highlighted.

S. No	Database / Server Names	Link	Features		
			Visualization of gene expression data	Functional/comparative genomics analysis	Gene Ontology
i.	BarleyBase	http://www.barleybase.org/	✓	✓	x
ii.	FLAGdb++	http://genoplante-info.infobiogen.fr/FLAGdb/	✓	✓	x
iii.	Narcisse	http://narcisse.toulouse.inra.fr/	x	✓	x
iv.	CoP	http://webs2.kazusa.or.jp/kagiana/cop0911/	✓	✓	✓
v.	SGMD	http://bioinformatics.towson.edu/SGMD/	✓	✓	✓
vi.	The Adaptive Evolution Database (TAED)	http://www.bioinfo.no/tools/TAED	✓	✓	x
vii.	Phytome	http://www.phytome.org/	x	✓	x
viii.	openSputnik x97	http://sputnik.btk.fi/	x	✓	x
ix.	PLAZA 4.0	https://bioinformatics.psb.ugent.be/plaza	✓	✓	x
x.	SALAD	https://salad.dna.affrc.go.jp/salad/en/	✓	✓	X
xi.	EXpath	http://expath.itps.ncku.edu.tw	✓	✓	✓
xii.	PmiRExAt	http://pmirexat.nabi.res.in	✓	✓	✓

i. BarleyBase

A database (www.barleybase.org) for data visualization and statistical analysis of plant microarrays with unified tools and extensive sets of query offering options at a different levels of data sets, from experiments to individual hybridization and from probe sets to single probes[55]. BB comprises data from easily available Affymetrix genomic arrays like Barley1 plus Arabidopsis ATH1 and plans for assembling data from forthcoming Affymetrix 61K array for Wheat, Rice, Maize, and rice as soon they are provided. Moreover, observed expression profiles based cross-experiment can be performing on probes set, which are visualized with analysis tools like a range of plot types, Data filters, R toolbox. Also, it enhances Cross-species comparison by allowing to devour identification of gene's function and EST alignments via Barely1 exemplar sequence with related vocabularies for plant genetics in PlantGDB, Gramene, and GrainGenes.

ii. FLAGdb++

FLAGdb ++ is a database (<http://genoplante-info.infobiogen.fr/FLAGdb/>) which usefully sequence genomes (like Arabidopsis) for visualization and high-throughput functional investigation[56]. FLAGdb++ helps in processing raw information from various HTP structural approaches as

Transcriptome studies; T-DNA mutants generated massive sequencing and characterization for gene families at the genomic level.

iii. Narcisse

This database (<http://narcisse.toulouse.inra.fr>) performs the comparative analysis based on similarities and rearrangements of sequenced genomes of animals, plants, bacterial kingdoms, and fungi[57].

iv. CoP

CoP (<http://webs2.kazusa.or.jp/kagiana/cop0911/>) is A database for the characterization of co-expressed gene modules utilizing a massive dataset(10,022 assays) of biological knowledge in plants[58]. The database includes the gene modules for seven crops *Vitis vinifera* (grape), *Oryza sativa* (rice), *Triticum aestivum* (wheat), *Arabidopsis thaliana* (thale cress), *Glycine max* (soybean), *Hordeum vulgare* (barley) *Zea mays* (maize) and *Populus trichocarpa* (poplar),.

v. The Adaptive Evolution Database (TAED)

The Adaptive Evolution Database (TAED) is designed to record and collect the information that drives the molecular basis of evolution for the chordate and embryophytes tree of life[59]. It has the cluster of protein families where at a minimum, one branch in the reconstructed molecular record has a Ka/Ks value greater than 1 or 0.6. It can be accessed at <http://www.sbc.su.se/~liberles/TAED.html>.

vi. Phytome

Phytome (<http://www.phytome.org/>)[60], an online resource for genomic comparison and can be used for multiple type studies such as molecular breeding, functional plant genomics, and evolutionary analysis. It comprises of protein family assignments, sequences of predicted proteins, phylogenies, MSA, and functional annotations for proteins related to a large, phylogenetically varied set of plant taxa's. Phytome works as a link or thread between different databases of plant genes, both by recognizing the evolutionary interactions among paralogous and orthologous protein sequences of different species and by allowing cross-references between multiple kinds of the same gene curated independently by various database groups.

vii. PLAZA

For evolutionary, comparative, and functional genomics PLAZA (<https://bioinformatics.psb.ugent.be/plaza>) database could be used based on plant-oriented resources. The PLAZA stage depends upon different autonomous occurrences considering various plant clades, while additionally giving access to a continuous arrangement of varying reference species. The latest version, PLAZA 4.0 has 71 species of both monocot and dicot[61].

viii. EXPath

EXPath, accessible at (<http://expath.itps.ncku.edu.tw>), this database is a tool that under diverse conditions, gathers and uses microarray gene expression profiles to infer metabolic pathways for plants[62]. EXPath has been developed to not only compile information from more than 1000 samples for public microarray expressions in abiotic stress, biotic stress, and hormone secretion but also to make use of this precious resource for analyzing the coexpression and identifying DEGs (differential expression of genes), Finally inferring three models of plants into enriched KEGG pathways and GO (gene ontology): *Zea mays*, *Oryza sativa*, and *Arabidopsis thaliana*.

ix. PmiRExAt

PmiRExAt, a free interactive repository that provides information regarding the plant miRNA expression atlas. The web-based server consists of the 1859 wheat, 2330 barley, and 283 maize miRNA expression profiles and analysis resources for miRNA. The repository interface provides open and easy access to miRNA expression information and enables us to determine preferential,

differentiated, and constitutive expressing miRNA in different tissues. It also implements a feature allowing the analysis of conserved miRNA expression across multiple species. The custom expression analysis option allows a total of 117 datasets for the expression analysis of novel miRNA. The latest sRNA dataset for analysis of miRNA expression profiles for 73 plant species may also be uploaded. A simple object access protocol web service, PmiRExAt application program interface allows other programmers to remotely invoke the methods written on the PmiRExAt database to do programmatic search operations.

6. Plants SNPs and DiseaseDatabase

Using molecular markers has transformed the work rate and accuracy of plant genetic research, which results in improving molecular crop breeding. Discovering SNPs has expanded our understanding of genetic diversity and improved crop development[63]. For instance, a study by Hafiz., et al; used SNPs to identify *Colletotrichum gossypii* in Upland Cotton. On the other hand, to overcome the problem of waterlogging sensitivity in plants Jiangshuo Su et.,al, in a genome-wide studies, identified waterlogging tolerant SNPs in *Chrysanthemums*. Furthermore, Sajid., et al; defined the role of SNPs in cotton to attain favorable properties[64]. Hence, this common practice of identifying SNPs and their role in crop improvement and additional features to respond to both biotic and abiotic elements demands technology-based platform to perform these analyses with experimental data. With such an exponential increase in the experimental data, computational methods for SNPs discovery, processing and analysis to gain knowledge full insight into the plant species are also designed and deployed online. This portion shortlisted such important online biological repositories and platforms for discovering SNPs and their roles in different plants. Important online SNPs related databases with their integrated features are given in **Table 3**.

Table 3: List of Plants SNPs and Diseases related databases with URLs and key features highlighted.

S. No	Database / Server Names	Link	Features	
			Disease Gene / RNA /protein Therapeutic information	Mutation data
i.	Medicago	http://www.medicago.org/	X	
ii.	NIASGBdb	http://www.gene.affrc.go.jp/databases_en.php	✓	✓
iii.	DRASTICx97INSI GHTS	http://www.drastic.org.uk/	✓	x
iv.	PMRD	http://bioinformatics.cau.edu.cn/PMRD		
v.	PRGDB	http://prgdb.cbm.fvg.it/	✓	x
vi.	STIFDB2 (Stress Responsive Transcription Factor Database)	http://caps.ncbs.res.in/stifdb2	✓	✓
vii.	PVsiRNAdb	http://www.nipgr.res.in/PVsiRNAdb	✓	✓
viii.	RiceMetaSysB	http://14.139.229.201/ricemetasysb/	✓	✓
ix.	RiceSRTFDB	http://www.nipgr.res.in/RiceSRTFDB.html	✓	✓
x.	DroughtDB	http://pgsb.helmholtz-muenchen.de/droughtdb	✓	✓

xi.	CSTDB	http://pcsb.ahau.edu.cn:8080/CSTDB	✓	✓
xii.	HEATSTER	http://applbio.biologie.uni-frankfurt.de/hsf/heatster/	✓	✓
xiii.	PSPDB	http://www.bioclues.org/pspdb/	✓	✓
xiv.	SNP-Seek II	http://snp-seek.irri.org/	✓	✓
xv.	Cereon Corp	http://www.arabidopsis.org/cereon/	✓	✓
xvi.	CerealDB	https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/	✓	✓
xvii.	PlantMarkers	http://markers.btk.fi/	✓	✓
xix.	MOLEGEN		✓	✓
xx.	MaizeDB	http://www.agron.missouri.edu/	✓	✓

i. NIASGBdb

NIASGBdb (http://www.gene.affrc.go.jp/databases_en.php) is a database of genetic resource and plant diseases[65]. This database is searched for both plant and microorganism either by passport data or evaluation data. Another Japanese database (http://www.gene.affrc.go.jp/databases-micro_pl_diseases_en.php) that provides information on plant common disease names in Japan has also been developed. This database primarily contains 11,350 of diseases information that is related to a total of 1933 host plant and their specific disease-causing pathogens. It also includes information on hostnames, Japanese names of diseases (converted to Roman alphabet), English names, scientific names of pathogens, and other information from the book.

ii. DRASTIC

DRASTIC is an online database (<http://www.drastic.org.uk/>) use for constructing signal transduction pathways and contains more than 17,700 records extracted for 73 affecting plants from 306 different treatments peer-reviewed from 512 theoretical types of research mostly emphasizing on data from *Arabidopsis thaliana*[66].

iii. PMRD

PMRD (<http://bioinformatics.cau.edu.cn/PMRD>) is a curated database for genes and mutants reported to be involved in the reproduction of male plants [32]. It contains information on 4203 male sterile genes (MS genes), 3697 male reproduction-related genes (MR genes), and 484 mutants of 33 plant species.

iv. PRGdb

PRGdb database (<http://www.prgdb.org>) represents the first bioinformatics tool that provides a broad overview of plant resistance genes (R-genes)[67]. It holds more than 16,000 both known and putative R-genes in 192 various species of plant and 115 pathogens. There are a total of 73 curated reference R-genes, 6308 putative R-genes, and 10463 predicted putative R-genes.

v. STIFDB2

Arabidopsis thaliana, *Oryza sativa* subsp. *Japonica* and *Oryza sativa* subsp. *Indica* 31 transcription factors, 15 stress signals and 5,984 stress-responsive genes are added to STIFDB2 (<http://caps.ncbs.res.in/stifdb2>)[68]. Overall, STIFDB2 contains 38,798 stress-responsive genes, stress signals, and binding sites of transcription factor obtained by Stress-responsive Transcription Factor (STIF) algorithm, along with functional annotation. This database can be used for computational and high-throughput experimental studies to understand the principles of the stress regulome in dicots and Gramineae.

vi. PVsiRNAdb

PVsiRNAdb available at (<http://www.nipgr.res.in/PVsiRNAdb>), is a manual design plant-exclusive database containing detailed associated with vsRNAs, based on probing of currently available scientific data of virus-infected plants[46]. This database comprises of 282 549 different sequences of vsRNAs and 322 214 of total entries. In this database, comprehensive and detailed information present for all courses of siRNA. Besides the primary information, which consists of the plant, tissue, name of the virus, and vsRNA sequence, additional details of each vsRNA (map location, length, coordinates, strand details, and the structure predicted) can be very useful to the user. PVsiRNAdb offers various types of search and navigation modules with 3 different tools: BLAST, Smith-Waterman Align, and Mapping.

vii. RiceMetaSysB

RiceMetaSysB is an open-source database that provides comprehensive details about BB-RGs (bacterial blight- responsive genes) and rice blast[69]. The meta-analysis of microarray data in 241 and 186 samples, consisting of a various blast- and BB-related studies, reported 1,5135 single blast genes and 7,475 for BB. In Blast and BB RG for development of marker, a total of 9365 and 5375 single sequence repeats (SSRs) were detected.

viii. RiceSRTFDB

The RiceSRTFDB (<http://www.nipgr.res.in/RiceSRTFDB.html>)[70] offers detailed expression information for rice Transcription Factors during salinity and drought conditions and different stages of their production. This knowledge will be useful in determining the target TF(s) of stress response at a given stage of development. Curated information has also been provided for cis-regulatory elements present in their promoters, which are critical in investigating binding proteins. Furthermore, the available mutants and information about the phenotype of rice translational factors (TFs) are added. Each of these information has to be integrated into the database to enable selection of TFs for functional analysis of the targets of interest. This database aims to speed up the genomics investigations for rice TFs and to comprehend the regulatory mechanisms behind abiotic stress reactions.

ix. DroughtDB

DroughtDB available at (<http://pgsb.helmholtz-muenchen.de/droughtdb/>) is a manually organized catalog of appraisal genes that are involved in the response of drought stress[71]. DroughtDB provides the necessary information of the genes, also about its molecular and physiological function, mutant phenotype, and comprehensive detailed about computational orthologues gene of nine crops from model plants, including barley and maize species. All ornithologists reported are interlinked with the respective MIPS / PGSB PlantsDB reference entry, allowing extra information such as genome background and sequence data to be obtained. DroughtDB is a beneficial source and information tool for researchers working on drought stress. It will help identify, analyze, and characterize genes that are involved in drought stress tolerance in crop plants of agricultural importance.

x. CSTDB

Crops Stress-tolerance Database (CSTDB) available online at (<http://pcsb.ahau.edu.cn:8080/CSTDB>), which is an integrated database that comprises of stress-tolerance proteins and genes for various crops species[72]. The database was built based on neural network technology. This database has compressive detail for major crops genes and proteins that are involved in stress tolerance. The database currently reports four major crops that contain 1.371 genes or proteins for abiotic stress toleration and 207 genes or proteins linked to biotic stress. Each protein and gene have comprehensive sequence and functional information, such as Genbank ID, Pubmed ID, stress types, Protein ID, FASTA files, and 3D model picture. This database offers search, BALST, and file transfer functions as an easy-to-usage browsing resource. CSTDB can be a valuable tool to meet the specific needs of researchers involved in the experiments on crop stress tolerance.

xi. HEATSTER

HEATSTER (<http://applbio.biologie.uni-frankfurt.de/hsf/heatster/>) is an internet-based resource for extracting, analyzing, and elucidating different HSFs in plants. A web-interface permits HSFs identification and class assignment, intuitive queries in the database, and conserved patterns visualization and novel HSFs classification domains[73].

xii. PSPDB

PSPDB (Plant Stress Protein Database)[74], is available at (<http://www.bioclues.org/pspdb/>), it includes 2,064 plant stress proteins, manually curated from a wide range of 134 plants with 30 different abiotic and biotic stress. The primary criteria for inclusion into the database are functional and experimental confirmation of proteins associated with biotic and abiotic stress.

xiii. SNP-Seek II

SNP-seek II is an *Oryza sativa* specific SNPs and EST database accessible at (<https://snp-seek.irri.org/>). SNP-seek II is an interactive tool which visualize million of SNPs in rice varieties[75].

xiv. Cereon Corp

Over 56,000 DNA polymorphisms were identified in *Arabidopsis* (specific) and added to Cereon Corp database (<http://www.arabidopsis.org/cereon/>). It was achieved by the Landsberg ecotype shotgun genomic sequencing and alignment of the genomes with the Columbia ecotype public DNA sample[76].

xv. CerealsDB Database

CerealsDB (<https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php>)[77] is a wheat specific (*Triticum aestivum*) genomics resources database for markers and markers assisted selection. CerealsDB has more than 100,000 SNPs, among which large proportion are experimentally validated. This database also contains EST sequences and DArT markers.

xvi. PlantMarkers

PlantMarkers[78] accessible at (<http://markers.btk.fi/>) is an online web repository which provides searching and processing of plants species specific markers. The database has SNPs but not limited to non-synonymous SNPs from 50 different plants. The identified markers are also mapped onto their respective protein sequence.

xvii. MELOGEN

MELOGEN database for Melon (*Cucumis melo* L.) has been developed that includes all the EST series, contiguous illustrations and other analytical and data mining methods. It has over 30K ESTs clustered into different categories [79].

7. Proteomics Databases

Proteomics is a widely used technology for protein interaction and characterization. Proteomics aims for identification, quantification, and characterization of protein's function and structures (in the cell or a tissue), association, and analysis of post-transcriptional modification[80]. Though promising progress has been made in the initiative in structural genomics of *Arabidopsis* (<http://www.uwstructuralgenomics.org/>), However, this field of plant science (proteomics) is yet juvenile especially in protein identification, its subcellular localization and structure protein-protein interaction and profiling activity of a protein are still not brought in to use. Proteomic problems such as PPI, function prediction, localization within the cell, and structure prediction specifically remained less practiced in plant sciences[81]. Bioinformatics resources for plant proteomics such as PlantTFDB, PlantP, PlantS, NBSPred, and many others considerably eased the knowledge related to plant proteomics by predicting the protein function, structure, and its comparison, cellular localization and its interaction with the other cellular partner. Many researchers developed online repositories for such results and predictions to help the researcher to access such information easily and map useful

knowledge for crop improvement. Hence, in this section of the review, we focused on the proteomics related web resources with their integrated data and its prominent analysis modules to ease for the researcher to exercise such tools in their experimental study. The available resources for the proteomics analysis with their links and integrated features are given in **Table 4**.

Table 4: The table shortlisted the available resources for the proteomics analysis with their links and integrated features.

S. No	Database / Server Names	Link	Features			
			TFs, stress and other proteins information	Sequence and information	Structural Information	Functional information
i.	PlantPIs	http://plantpis.ba.itb.cnr.it/	✓	✓	X	x
ii.	POGs/PlantRBP	http://cas-pogs.uoregon.edu/#/	✓	x	X	✓
iii.	ChromDB	http://www.chromdb.org/	✓	✓	✓	✓
iv.	MIPS	https://www.helmholtz-muenchen.de/ibis/	x	✓	X	✓
v.	P3DB	http://p3db.org/index.php	✓	✓	✓	✓
vi.	PHYTOPROT	http://genoplante-info.infobiogen.fr/phytoprot	✓	✓	✓	✓
vii.	SALAD	https://salad.dna.afrc.go.jp/salad/en/	✓	✓	X	✓
viii.	NBSPred	http://soilecology.biol.lu.se/nbs/	✓	✓	X	✓
ix.	Lectindb	http://nscdb.bic.physics.iisc.ernet.in/	✓	x	X	✓
x.	PlantsP	http://plantsp.sdsc.edu/	✓	x	X	x
xi.	PlantsT	http://plantst.sdsc.edu/	✓	✓	X	x
xii.	PlantMP	https://www.plantmp.com/	✓	x	X	✓
xiii.	MoonProt	http://moonlightinproteins.org/	✓	✓	✓	✓
xiv.	MultiTaskProtDB	http://wallace.uab.edu/multitaskII	✓	✓	✓	✓
xv.	CSTDB	http://pcsb.ahau.edu.cn:8080/CSTDB	✓	✓	X	✓
xvi.	PSPDB	http://www.biocloudes.org/pspdb/	✓	x	X	✓
xvii.	PlantTFDB	http://planttfdb.cbi.pku.edu.cn/	✓	✓	X	✓

xviii.	ERISdb	http://lemur.amu.edu.pl/share/ERISdb/home.html	✓	✓	X	✓
xix.	HEATSTER	http://applbio.biologie.uni-frankfurt.de/hsf/heatster/	✓	✓	X	✓
xx.	JASPAR 2018	http://jaspar.genereg.net/	✓	✓	X	✓
xxi.	PlnTFDB	http://plntfdb.bio.uni-potsdam.de/v3.0	✓	✓	X	✓
xxii.	FamNet	http://www.gene2function.de/famnet.html	✓	✓	✓	✓
xxiii.	Plant-PrAS	http://plant-pras.riken.jp	✓	✓	✓	✓

i. PLANT-PIs

PLANT-PIs (<http://bighost.area.ba.cnr.it/PLANT-PIs>) database is helpful to retrieve information of 351 plant protease inhibitors (PIs) and their related genes and other several iso-inhibitors[82].

ii. POGs/PlantRBP

POGs/PlantRBP (<http://plantrbp.uoregon.edu/>) is basically a relational database. In this database, the data obtained from rice, Arabidopsis, and maize are integrated into POGs (putative orthologous groups) by placing the complete proteomes of Arabidopsis and rice and maize sequences[83]. A web interface enables queries that incorporate search terms for different types of features such as location and domain content. The server displays the data in a form that helps orthology comparisons.

iii. ChromDB

This database at (<http://www.chromdb.org>) shows proteins associated with chromatin and RNAi for a vast amount of different organisms[84]. For making the database applicable for comparative studies and analysis of proteomes for chromatin researchers, model animals as well fungal proteins are included in the database. From several 100 plant proteins, ChromDB has grown to more than 7000 proteins comprising 7474 proteins in total in which 3328 are from plants, 1779 are from animals, 2143 from fungi, 167 from stramenopiles and 57 from protists.

iv. MIPS

MIPS is an online server (<http://mips.gsf.de>) contains information about sequence relating protein from the genomic analysis[85]. MIPS focuses on gathering and disciplined organization of knowledge through different algorithms and raw data from scientific literature. Moreover, from large-scaled published experiments, there are >10,600 records for inter protein interaction (in which ~9100 are physical and ~1500 genetic) are added. The annotated >1000 protein complexes can be divided into ~87 000 different putative paired interactions.

v. P3DB

P3DB (<http://www.p3db.org/>) is unique as it contains data on protein phosphorylation obtained from multiple plants with a lot of other querying capacities[86]. This database to date representing the largest collections and includes 14,670 phosphosites and 6382 phosphoproteins.

vi. PHYTOPROT

vii. PHYTOPROT (<http://genoplante-info.infobiogen.fr/phytoprot>) provides information on protein clustering[87]. In this dataset, the entire proteome of Arabidopsis thaliana and all the other plant sequences have been clustered into groups and compared.

viii. SALAD

Genome-based proteome data set information provided by the SALAD database (<http://salad.dna.affrc.go.jp/salad/>), which covers from 209,529 protein-sequence annotations of 10 different species including rice, Arabidopsis, sorghum, grape, thaliana, a moss, a lycophyte, yeast. SALAD database is considered to be user friendly provide best graphical representation and shows motif pattern shapes based on bootstrapped dendrogram of each protein category. It is considered to be the potential tool for comparative genome-based proteome sequence analysis[88].

ix. NBSPred

The NBSPred (<http://soilecology.biol.lu.se/nbs/>) pipeline has been used for differentiation between NBSLRR (also NBSLRR-like protein) and Non-NBSLRR protein from proteome, transcriptome, and genome[89]. Many plants such as Zea mays, Brachypodium distachyon, Boechera stricta, Solanum lycopersicum, and Arabidopsis thaliana sequences were tested with the help of this pipeline.

x. Lectindb

Lectindb, (<http://nscdb.bic.physics.iisc.ernet.in>), together with appropriate analytical tools, an online tool for plant lectins by collating and integrating diverse data[90]. Data for each lectin pertain to molecular sequence, biochemical, structural, domain architecture, and taxonomic details as well as carbohydrates are added here. Extensive links on specific bioinformatics tools and methodologies have also been given. It is anticipated that the development of diverse data incorporated into a shared framework would be of high importance not only for specific studies in lectin biology but also for general studies utilizing these molecules in pursuing many applications in biotechnology, immunology, and clinical practice.

xi. PlantsP Database

This database is centered on the protein kinases and phosphates through the data collected from sequences with an experiment based genomic information. The database serves as a resource for T-DNA insertion mutants (knockouts) in each protein phosphatase and kinase of Arabidopsis thaliana. The curated view of protein providing the annotations of functionality related sequences, alignments, phylogenetic trees, sequence family definitions, and other descriptive information is available on the web at (<http://PlantsP.sdsc.edu>)[91].

xii. PlantsT Database

With the help of PlantsT, users quickly get a universal comprehension of plant membrane transporters and plant phosphoproteins from developmental relationships to biochemical function and a full understanding of molecular biology of genes and their products immediately. The purpose of PlantsT to make membrane transport proteins. PlantsT can be accessed at (<http://plantst.sdsc.edu>)[92].

xiii. PlantMP

PlantMP (<https://www.plantmp.com/>) is a Moonlighting Plant Protein database. PlantMP reportedly comprises 110 proteins for plant moonlighting, ten probable moonlighters, and 27 predicted moonlighters[93]. Arranging plant moonlighting proteins on one framework helps researchers to efficiently collect raw and refined plant-specific data such as molecular functions, biological roles, and systemic features that are important for concept formulations in scientific research and biotechnological advancement.

xiv. MoonProt

MoonProt Database (<http://www.moonlightingproteins.org>) contains experimentally reported ~200 moonlighting protein[94]. The availability of experimentally verified proteins provides a

comprehensive picture of moonlighting proteins. The database will provide significant insight to researchers working on genome sequencing projects for identification of genes functions, proteomics projects for data analysis, and protein annotation. Additionally, moonlighting proteins' structural and functional information would be helpful in understanding the evolution of novel protein functional sites.

xv. MultitaskProtDB

MultitaskProtDB, an online database (<http://wallace.uab.es/multitask>) is the repository of diverse multitasking proteins reported in the literature[95]. Multitask or moonlight proteins have the potential to perform more than on biological functions. In MultitaskProtDB, more than 288 proteins are stored containing protein accession numbers, biological functions, PDB codes, monomeric/oligomeric states, and bibliographic references. This user-friendly web would be helpful in providing insight into multitasking protein characteristics and phylogenetic conservation.

xvi. CSTDB

Crops Stress-tolerance Database (CSTDB) available online at (<http://pcsb.ahau.edu.cn:8080/CSTDB>), which is an integrated database that comprises of stress-tolerance proteins and genes for various crops species[72]. The database was built on the basis of neural network technology. This database has compressive detail for major crops genes and proteins that are involved in stress tolerance. The database currently reports four major crops that contain 1.371 genes or proteins for abiotic stress toleration and 207 genes or proteins linked to biotic stress. Each protein and gene have complete sequence and functional information, such as Genbank ID, Pubmed ID, stress types, Protein ID, FASTA files, and 3D model picture. This database offers search, BALST, and file transfer functions as an easy-to-usage browsing resource. CSTDB can be a valuable tool to meet the specific needs of researchers involved in the experiments on crop stress tolerance.

xvii. PSPDB

PSPDB (Plant Stress Protein Database)[74], is available at (<http://www.bioclues.org/pspdb/>), it includes 2,064 plant stress proteins, manually curated from a wide range of 134 plants with 30 different abiotic and biotic stress. The primary criteria for inclusion into the database are functional and experimental confirmation of proteins associated with biotic and abiotic stress.

xviii. PlantTFDB

PlantTFDB is a database, available at (<http://planttfdb.cbi.pku.edu.cn>)[96] is a TF database with 26,402 Transcription factors collected from 22 plant species. The data included in this database obtained from WGS and EST sequences five models of 17 plant species. Users can provide IDs to search a specific transcription factor. The database allows users to download data into the local machine and perform its analysis.

xix. ERISdb

ERISdb can be accessed at (<http://lemur.amu.edu.pl/share/ERISdb/>). It is a database of Plant Splicing Signals and Splice Sites. The database also includes polypyrimidine tracts, orthologous splice sites, and branch sites[97].

xx. JASPAR

JASPAR (<http://jaspar.genereg.net>)[98] is a web-based tool that provides knowledge about transcription binding profiles and serves as TF flexible models (TFFMs) and position frequency matrices (PFMs) for transcription factors among several species mainly involved in six taxonomic categories. In the 2018 JASPAR release, 322 new PFMs (262 for plants and 60 for vertebrates) and 33 PFMs were updated (8 for plants, 24 for vertebrates, and 1 for insects) have been added to the CORE collection. In comparison to the 2016 release, these new profiles show a 30% extension.

Furthermore, 316 TFFMs have also been introduced (218 for plants, 95 for vertebrates and 3 for insects).

8. Sequence and Structural Annotation Databases

Ontologies are biological expressions whose meanings are clearly described concerning other terms and are used for data annotation like sequences, strains, and gene expression clusters [99, 100]. These ontologies in databases are used in different ways, such as building a framework for result interpretation, connecting databases, and refining searching. A large number of linked ontologies as that of microarray experiments, metabolic pathways[101] and anatomies cell types and developmental stages of living organisms[102]. An example with community approval of bio-gene ontologies are GO (www.geneontology.org) an acyclic graph-based tree hierarchy, which is a set of above 16,000 terms for molecular functions and biological processes etc. and allows the concept of a single specific name with belonging to more general terms. All the available web resources for sequence and structure comparison are given in **Table 5**.

Table 5: The table shortlisted the available resources for sequence and structure comparison analysis with their links and integrated features.

S. No	Database / Server Names	Link	Features		
			Functiona l Annotatio n	Sequence informati on	Genomics/prot eomics and evolutionary annotation
i.	CR-EST	http://pgrc.ipk-gatersleben.de/cr-est/	✓	✓	x
ii.	CSRDB	http://sundarlab.ucdavis.edu/smrnas/	x	✓	✓
iii.	GeneSeqer	http://www.plantgdb.org/cgi-bin/GeneSeqer/index.cgi	x	✓	✓
iv.	MetaCrop	http://metacrop.ipk-gatersleben.de/	✓	x	x
v.	PLACE	https://www.dna.affrc.go.jp/PLACE/?action=newplace	x	✓	✓
vi.	PMRD	http://www.nipgr.res.in/plecdom.html	x	✓	✓
vii.	TIGR the Plant Repeat Database	http://plantrepeats.plantbiology.msu.edu/index.html	x	✓	✓
viii.	The TIGR Plant Transcript Assemblies database	http://plantta.tigr.org/	x	✓	✓
ix.	The Plant Ontology Database	http://www.plantontology.org/amigo/go.cgi	x	✓	✓
x.	The Planteome database	http://planteome.org/	✓	✓	✓
xi.	ShinyGO	http://ge-lab.org/go/	✓	✓	✓

xii.	CMAUP	http://bidd2.nus.edu.sg/CMAUP/	✓	✓	✓
xiii.	OrthoDB v10	https://orthodb.org/	✓	x	✓

i. Crop EST Database (CR-EST)

CR-EST is an online database (<http://pgrc.ipk-gatersleben.de/cr-est/>)[103] that currently contains more than 200,000 sequences of Barley, Wheat, Pea, and Potatoes gained from 41 cDNA libraries. CR-EST at the IPK provides accession to annotating, classifying and clustering sequences. It allows the sequence to be clustered by species-specific projects for generating a non-redundant set of ~22 600 consent and ~17,200 singletons, and shape the ground for given unigenes. This database carried out its queries from sequences similarities and annotations of data from GO and metabolic pathways via the BLAST repository.

ii. CSRDB

An online small RNA unified database (<http://sundarlab.ucdavis.edu/smrnas/>) of cereals contains smRNA sequences of a wide range of datasets from maize and rice derived through high-throughput pyrosequencing[104].

iii. GeneSeqer

GeneSeqer (<http://www.plantgdb.org/cgi-bin/GeneSeqer.cgi>)[105], offers a tool for prediction of gene structure with defined applications for genomic refinement of gene structure annotation also with de novo expansion of sketched genomic sequences. Predictions are made with source-native ESTs (using spliced alignment), probes from genes that are homologous, and cDNAs libraries.

iv. MetaCrop

MetaCrop is a database (<http://metacrop.ipk-gatersleben.de>) with information from 7 major and 2 model plants is a storage of factual data regarding plant metabolites and provides details about maps of metabolites function and enzymatic data[106]. MetaCrop aims to facilitate research for the betterment of crops for both nutrition and industrial use. It is easily approachable via web services as well as GUI software.

v. PLACE

PLACE (<https://www.dna.affrc.go.jp/PLACE/?action=newplace>), a motifs database of plant cis-acting regulatory DNA elements that were collected from previously published sources[107]. It also reports variations in other genes or other species of plants and includes some motifs of non-plant cis-elements that are homologs to plants.

vi. PMRD

PMRD (<http://bioinformatics.cau.edu.cn/PMRD>) is a curated database for genes and mutants reported to be involved in the reproduction of male plants [32]. It contains information on 4203 male sterile genes (MS genes), 3697 male reproduction-related genes (MR genes), and 484 mutants of 33 plant species.

vii. TIGR

The TIGR (<http://www.tigr.org/tdb/e2k1/plant.repeats/index.shtml>) is a Plant Repeat Database of recurring series for 12 plant species: Arabidopsis, Glycine, Brassica, Lotus, Hordeum, Lycopersicon, Oryza, Medicago, Solanum, Triticum, Sorghum, and Zea. As a whole, the function of these databases is to provide a resource from the identification and classification to the analysis of recurrent series in plants[108].

viii. The Planteome database

The Planteome project (<http://www.planteome.org>) gives a cluster of reference and species-specific ontologies (vocabularies) for plants and elucidation to genes and phenotypes[109]. It provides

incorporated data on plant phenotypes, gene function, traits, and expression from 95 plant taxa, described with reference ontology terms. To facilitate community engagement, the Planteome project is establishing a plant gene elucidation platform, Planteome Noctua.

ix. ShinyGO

ShinyGO (<http://ge-lab.org/go/>) is a database that is capable of graphical advancement tools for plants and animals. It is based on a database obtained from STRING-db and Ensembl for 256 animals, 59 plants, 1678 bacterial, and 115 archaeal species. ShinyGO is an automatic, a graphical web resource that can assist researchers in getting reasonable comprehension from gene-sets[110].

x. CMAUP

CMAUP (Collective Molecular Activities of Useful Plants)[111] give a combine landscape of various targets (ChEMBL target classes). Activity levels (in 2D (two-dimensional) target-ingredient heatmap), in biological pathways (KEGG categories), regulated GO (gene ontology categories) and diseases for five thousand six hundred forty-five plants (5645) which include 170 food, 2567 medicinal, 1567 edible, 119 garden, and three agricultural plants, used traditionally or collected from one hundred fifty-three (153) countries and regions. The observed landscapes were extracted from 47 645 plant components. They were found active against six hundred forty-six (646) sites of two hundred thirty-four (234) KEGG pathways linked to 2473 genes and 656 anomalies. CMAUP is freely available on (<http://bidd2.nus.edu.sg/CMAUP/>) and also accessible through keywords, species families, plant usage classes, KEGG pathways, diseases (ICD code), targets, geographical locations, and gene ontologies.

xi. OrthoDB

OrthoDB (<https://www.orthodb.org>) is an orthologs database which present functional and evolutionary annotations[112]. OrthoDB catalogs groups of orthological genes, from more general to more fine-grain distinguishing groups among closely related species, with increasing radiation of the species phylogeny. A COG-like and inparanoid-like orthology mechanism for evaluating 58 eukaryotic genomes on the basis of all-against-all Smith-waterman sequence comparisons, which concentrate on vertebrates, insects, and fungi to help more comparative study.

9. Plants-Pathogen Interaction databases

Pathogen identification and signal transmission are essential for the activation of plant defense mechanisms during plant pathogenesis[113]. To facilitate the researcher working on plant immune responses to pathogens many online repositories are developed and deployed online to provide information regarding how the plant interacts with the pathogen, the expression pathways up- or down-regulated upon the interaction and the information on the virulence genes and proteins are provided. Thus, this section focuses on providing information regarding the available resources for plant-pathogen communications. Such repositories with their original features and available links are given in **Table 6**.

Table 6: The table shortlisted the available resources for plant-pathogen interaction analysis with their links and integrated features.

S. No	Database / Server Names	Link	Features		
			Plants-pathogens interaction information	Expression/pathways information	Virulence protein/genes information
i.	PHI-base	http://www.phi-base.org/	✓	✓	✓
ii.	PLEXdb	http://www.plexdb.org	✓	✓	✓
iii.	PathoPlant	http://pathoplant.de	✓	✓	✓

i. PHI-base

The PHI-base (pathogen-host interactions database) is an open-access database available at www.phi-base.org [114]. The database consists of curated biological and molecular information regarding different genes that have proven effect on pathogen-host interactions outcome, reported in peer-reviewed research articles. It also provides curated literature that describes specific gene alterations that do not have any effect on the disease interaction phenotype, to offer complete datasets for comparative purposes. Viruses have been excluded from the database because of being covered extensively in other databases. The PHI-base version 4.8, released in September 2019, consists of 3454 manually curated references, and make available information on 6780 genes from 268 pathogens, tested on 210 hosts in 13,801 interactions. The database represents Prokaryote and eukaryote pathogens in nearly the same numbers. Host species include of around 60% plants (split 50:50 between cereal and non-cereal plants) and 40% other species of environmental and medical importance.

ii. PathoPlant

PathoPlant (<http://www.pathoplant.de>) is a tool that provides interface for the analysis of microarray expression data of co-regulated genes responsible for plant defense[115]. PathoPlant, a plant-pathogen relations and signal passage reactions database, has currently been supplemented by data from *Arabidopsis thaliana*'s microarray gene expression imperiled to pathogen infection and elicitor action.

iii. PLEXdb

PLEXdb (<http://www.plexdb.org>), facilitates gene expression comparisons across the various plant and pathogen organisms, facilitating the transfer of genome-scale data sets to compete with historically collected data by individuals and/or federations. Methods at PLEXdb enable users to quickly perform complex analyses. The Model Genome Interrogator (MGI) method allows the mapping of gene lists from model plant species, such as rice, and arabidopsis, on related genes. MGI forecasts homologies depicts gene architectures and supports annotated gene details, and full-length cDNA. PLEXdb also offers features for users to use Gene OscilloScope to track how gene expression changes over many experimental tests. This platform can pinpoint interesting patterns of expression, such as up-regulation under different conditions, or evaluating the suitability of any gene as a control of the steady-state.

10. Other Databases

This section of the review focuses on the general purposes' databases which don't fit precisely in any of the above categories. This section provides multi-purpose resources for plant data analysis. The available web resources for multiple purposes are given in **Table 7**.

Table 7: The table shortlisted general purposes available resources for multiple analyses with their links and integrated features.

S. No	Database / Server Names	Link
i.	POBD	http://podb.nibb.ac.jp/Organellome/
ii.	PlantGBD	http://www.plantgdb.org/
iii.	SGN	http://sgn.cornell.edu/
iv.	TropGENE-DB	http://tropgenedb.cirad.fr/
v.	The TriForC database	http://bioinformatics.psb.ugent.be/triforc/
vi.	EXPath	http://expath.itps.ncku.edu.tw/
vii.	PED	http://bigd.big.ac.cn/ped
viii.	AllerBase	http://bioinfo.net.in/AllerBase/Home.html
ix.	MPIC	http://www.plantenergy.uwa.edu.au/applications/mpic

x.	The Bio-Analytic Resource	http://bar.utoronto.ca/
xi.	PlaMoM	http://www.systembioinfo.org/plamom/

i. PODB

The PODB (plant organelles database) accessible at <http://podb.nibb.ac.jp/Organellome> contained the visualized collections of organelles of the plant and their research protocols[116]. PODB helps understand the dynamics of organelles such as; their function, movement, differentiation, biogenesis, and interactions. The database has three functional domains; the organellome database, functional analysis database, and external links (for other databases and homepages).

ii. PlantGDB

PlantGDB (<http://www.plantgdb.org/>) is a genetic sequence data archive for all plant organisms that have substantial sequencing efforts. The database organizes sequences of EST into contiguous that reflect special genes of preliminary existence. Contigs are annotated and associated with their respective genomic DNA wherever possible. Likewise, fragments of the Genome sequence are assembled. The PlantGDB website intends to lay the groundwork for defining collections of genes common to all crops or species-specific by incorporating a range of bioinformatics resources that promote gene identification and cross-species correlations.

iii. Sol Genomics Network (SGN)

The SGN is an online site that is based upon the information regarding the Solanaceae family and its homologs. The increasing number of Solanaceae members and current members, inclusive of pepper, potato, tobacco, eggplant, and *Nicotiana benthamiana* is recorded in the form of their genomic map and the phenotypic characters. Tools like Jbrowse, BLAST, GBrowse for browsing, and analysis tools like QTL are available. The researchers use the web-based tools to record, edit and upload the databases and store information regarding loci and phenotypes[117].

iv. TropGENE-DB

TropGENE-DB, is a web-based database about crop information that stores phenotypic, genetic, and molecular data of the many tropical crop species[118]. The most commonly stored data in TropGENE-DB is the information on molecular markers, genetic resources (allelic diversity, agromorphological data, and parentages) and genetic maps. It results in genes, sequences, data from physical mapping, quantitative trait loci analyses, and other related references. TropGENE-DB is arranged on the basis of crops with three running modules (cocoa, sun, and banana) currently. It has a strategy to develop more modules for cotton, coconut, rice, oil palm, rubber tree, taro, sorghum, pineapple, and yam. The TropGENE-DB accessed through the internet at (<http://tropgenedb.cirad.fr>).

v. TriForC

TriForC database (<http://bioinformatics.psb.ugent.be/triforc/>) encompass an extensive inventory of triterpene biosynthesis proteins[119]. This massively interconnected database is a user-friendly passage to flexible data sets of enzyme and compound features that enables the filtering of a total archive of experimentally approved triterpene enzymes, their substrates, their pathways in different plants and products. The database can either be retrieved directly via URL or through suitable tools including, BLAST, substructure options, and plant species.

vi. PlaMoM

PlaMoM (Plant Mobile Macromolecules) repository, a repository that gives consumers with convenient and responsive search tools for retrieving, analyzing and also predicting mobile RNAs/proteins. Each PlaMoM entry provides extensive information, including sequences of nucleotides / amino acids, orthology partners, related experiments, gene features, and literature. Protein–protein interactions from mobile transcripts are presented as interactive molecular networks

for the model plant *Arabidopsis thaliana*. Also, PlaMoM offers an integrated tool for identifying potential RNA mobility signals, like structures similar to tRNA. The latest edition of PlaMoM utilizes existing research and literature to collect a minimum of 17 991 mobile macromolecules from 14 plant species / ecotypes. PlaMoM is available at <http://www.systembioinfo.org/plamom/>.

Conclusion

In this review, we tried to showcase some of the latest events in bioinformatics in the fundamental areas of sequence, proteins, gene expression, diseases, databases, and ontologies, current limitations in these areas, and some emerging areas. We describe some of the key concepts, methods, software applications, and bioinformatics repositories, with a focus on all those pertinent to plant sciences. It also provides knowledge regarding the integrated modules in different web resources for plants related to data analysis. Thus, this review article will help the plant researchers to access the related bioinformatics resources for their data analysis easily and will increase the practice of bioinformatics tools to solve experimental problems pertinent to plant sciences.

References

1. Millstone, E. and T. Lang, The atlas of food: who eats what, where, and why. 2008: University of California Press.
2. Mann, J., Natural products in cancer chemotherapy: past, present and future. *Nature Reviews Cancer*, 2002. **2**(2): p. 143-148.
3. Lippman, Z., et al., Role of transposable elements in heterochromatin and epigenetic control. *Nature*, 2004. **430**(6998): p. 471-476.
4. Schuster, S.C., Next-generation sequencing transforms today's biology. *Nature methods*, 2008. **5**(1): p. 16-18.
5. Govindaraj, M., M. Vetriventhan, and M. Srinivasan, Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genetics research international*, 2015. **2015**.
6. Rigden, D.J., X.M. Fernández-Suárez, and M.Y. Galperin, The 2016 database issue of *Nucleic Acids Research* and an updated molecular biology database collection. *Nucleic acids research*, 2016. **44**(D1): p. D1-D6.
7. Kind, T., M. Scholz, and O. Fiehn, How large is the metabolome? A critical analysis of data exchange practices in chemistry. *PloS one*, 2009. **4**(5): p. e5440.
8. Edwards, D. and J. Batley, Plant bioinformatics: from genome to phenome. *Trends in biotechnology*, 2004. **22**(5): p. 232-237.
9. Benson, D.A., et al., GenBank. *Nucleic acids research*, 2012. **41**(D1): p. D36-D42.
10. Coordinators, N.R., Database resources of the national center for biotechnology information. *Nucleic acids research*, 2017. **45**(Database issue): p. D12.
11. Cochrane, G., et al., The international nucleotide sequence database collaboration. *Nucleic acids research*, 2016. **44**(D1): p. D48-D50.
12. Kanz, C., et al., The EMBL nucleotide sequence database. *Nucleic acids research*, 2005. **33**(suppl_1): p. D29-D33.
13. Tateno, Y., et al., DNA Data Bank of Japan (DDBJ) for genome scale research in life science. *Nucleic acids research*, 2002. **30**(1): p. 27-30.
14. Rose, P.W., et al., The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic acids research*, 2016: p. gkw1000.
15. UniProt: the universal protein knowledgebase. *Nucleic acids research*, 2017. **45**(D1): p. D158-D169.
16. Garg, P. and P. Jaiswal, Databases and bioinformatics tools for rice research. *Current Plant Biology*, 2016. **7**: p. 39-52.
17. Bolser, D., et al., Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data, in *Plant bioinformatics*. 2016, Springer. p. 115-140.

18. Youens-Clark, K., et al., Gramene database in 2010: updates and extensions. *Nucleic acids research*, 2010. **39**(suppl_1): p. D1085-D1094.
19. Dash, S., et al., PLEXdb: gene expression resources for plants and plant pathogens. *Nucleic acids research*, 2012. **40**(D1): p. D1194-D1201.
20. Ong, Q., et al., Bioinformatics Approach in Plant Genomic Research. *Current genomics*, 2016. **17**(4): p. 368-378.
21. Myers, E.W., Toward simplifying and accurately formulating fragment assembly. *Journal of Computational Biology*, 1995. **2**(2): p. 275-290.
22. Patil, N., et al., Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*, 2001. **294**(5547): p. 1719-1723.
23. Hinds, D.A., et al., Whole-genome patterns of common DNA variation in three human populations. *Science*, 2005. **307**(5712): p. 1072-1079.
24. Frazer, K.A., et al., Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome research*, 2003. **13**(3): p. 341-346.
25. Bruskiwich, R., et al., Generation Challenge Programme (GCP): standards for crop data. *Omic: a journal of integrative biology*, 2006. **10**(2): p. 215-219.
26. van Berloo, R., GGT 2.0: versatile software for visualization and analysis of genetic data. *Journal of Heredity*, 2008. **99**(2): p. 232-236.
27. Riaño-Pachón, D.M., et al., GabiPD: the GABI primary database—a plant integrative ‘omics’ database. *Nucleic acids research*, 2009. **37**(suppl_1): p. D954-D959.
28. Carollo, V., et al., GrainGenes 2.0. An improved resource for the small-grains community. *Plant physiology*, 2005. **139**(2): p. 643-651.
29. Jung, S., et al., GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. *Nucleic acids research*, 2007. **36**(suppl_1): p. D1034-D1040.
30. Lawrence, C.J., et al., MaizeGDB, the community database for maize genetics and genomics. *Nucleic acids research*, 2004. **32**(suppl_1): p. D393-D397.
31. Mower, J.P., The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic acids research*, 2009. **37**(suppl_2): p. W253-W259.
32. Zhang, Z., et al., PMRD: plant microRNA database. *Nucleic acids research*, 2010. **38**(suppl_1): p. D806-D813.
33. Nakano, M., et al., Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic acids research*, 2006. **34**(suppl_1): p. D731-D735.
34. Brown, J.W., et al., Plant snoRNA database. *Nucleic acids research*, 2003. **31**(1): p. 432-435.
35. Lescot, M., et al., PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic acids research*, 2002. **30**(1): p. 325-327.
36. Shahmuradov, I.A., et al., PlantProm: a database of plant promoter sequences. *Nucleic acids research*, 2003. **31**(1): p. 114-117.
37. Meng, Y., et al., PmiRKB: a plant microRNA knowledge base. *Nucleic acids research*, 2010. **39**(suppl_1): p. D181-D187.
38. Canaran, P., et al., Panzea: an update on new content and features. *Nucleic acids research*, 2007. **36**(suppl_1): p. D1041-D1043.
39. Bai, Y., et al., PlncRNADB: A Repository of Plant lncRNAs and lncRNA-RBP Protein Interactions. *Current Bioinformatics*, 2019. **14**(7): p. 621-627.
40. Zonneveld, B., I. Leitch, and M. Bennett, First nuclear DNA amounts in more than 300 angiosperms. *Annals of botany*, 2005. **96**(2): p. 229-244.
41. Dicks, J., et al., UK CropNet: a collection of databases and bioinformatics resources for crop plant genomics. *Nucleic acids research*, 2000. **28**(1): p. 104-107.
42. Dai, X. and P.X. Zhao, pssRNAMiner: a plant short small RNA regulatory cascade analysis server. *Nucleic acids research*, 2008. **36**(suppl_2): p. W114-W118.
43. Dai, X., Z. Zhuang, and P.X. Zhao, psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic acids research*, 2018. **46**(W1): p. W49-W54.

44. Xia, E.H., et al., Tea Plant Information Archive: a comprehensive genomics and bioinformatics platform for tea plant. *Plant biotechnology journal*, 2019. **17**(10): p. 1938-1953.
45. Ye, J., et al., AtCircDB: a tissue-specific database for Arabidopsis circular RNAs. *Briefings in bioinformatics*, 2019. **20**(1): p. 58-65.
46. Gupta, N., et al., PVsiRNADB: a database for plant exclusive virus-derived small interfering RNAs. *Database*, 2018. **2018**.
47. Chen, D., et al., PlantNATsDB: a comprehensive database of plant natural antisense transcripts. *Nucleic acids research*, 2012. **40**(D1): p. D1187-D1193.
48. Lamesch, P., et al., The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic acids research*, 2012. **40**(D1): p. D1202-D1210.
49. Bedell, J.A., et al., Sorghum genome sequencing by methylation filtration. *PLoS Biol*, 2005. **3**(1): p. e13.
50. Coughlan, S.J., V. Agrawal, and B. Meyers, A comparison of global gene expression measurement technologies in Arabidopsis thaliana. *Comparative and Functional Genomics*, 2004. **5**(3): p. 245-252.
51. Buck, M.J. and J.D. Lieb, ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics*, 2004. **83**(3): p. 349-360.
52. Liu, X., et al., DIP-chip: rapid and accurate determination of DNA-binding specificity. *Genome research*, 2005. **15**(3): p. 421-427.
53. Drăghici, S., *Data analysis tools for DNA microarrays*. 2019: CRC Press.
54. Eilbeck, K., et al., The Sequence Ontology: a tool for the unification of genome annotations. *Genome biology*, 2005. **6**(5): p. R44.
55. Shen, L., et al., BarleyBase—an expression profiling database for plant genomics. *Nucleic Acids Research*, 2005. **33**(suppl_1): p. D614-D618.
56. Samson, F., et al., FLAGdb++: a database for the functional analysis of the Arabidopsis genome. *Nucleic acids research*, 2004. **32**(suppl_1): p. D347-D350.
57. Courcelle, E., et al., Narcisse: a mirror view of conserved synteny. *Nucleic acids research*, 2008. **36**(suppl_1): p. D485-D490.
58. Ogata, Y., et al., CoP: a database for characterizing co-expressed gene modules with biological information in plants. *Bioinformatics*, 2010. **26**(9): p. 1267-1268.
59. Roth, C., et al., The Adaptive Evolution Database (TAED): a phylogeny based tool for comparative genomics. *Nucleic acids research*, 2005. **33**(suppl_1): p. D495-D497.
60. Hartmann, S., et al., Phytome: a platform for plant comparative genomics. *Nucleic Acids Research*, 2006. **34**(suppl_1): p. D724-D730.
61. Vandepoele, K., A guide to the PLAZA 3.0 plant comparative genomic database, in *Plant Genomics Databases*. 2017, Springer. p. 183-200.
62. Zheng, H.-Q., et al., EXPath tool—a system for comprehensively analyzing regulatory pathways and coexpression networks from high-throughput transcriptome data. *DNA Research*, 2017. **24**(4): p. 371-375.
63. Botstein, D., et al., Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*, 1980. **32**(3): p. 314.
64. Phillips, R.L. and I.K. Vasil, *DNA-based markers in plants*. Vol. 6. 2013: Springer Science & Business Media.
65. Takeya, M., et al., NIASGBdb: NIAS Genebank databases for genetic resources and plant disease information. *Nucleic acids research*, 2010. **39**(suppl_1): p. D1108-D1113.
66. Button, D.K., et al., DRASTIC—INSIGHTS: querying information in a plant gene expression database. *Nucleic acids research*, 2006. **34**(suppl_1): p. D712-D716.
67. Sanseverino, W., et al., PRGdb 2.0: towards a community-based database model for the analysis of R-genes in plants. *Nucleic acids research*, 2012. **41**(D1): p. D1167-D1171.
68. Naika, M., et al., STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and

- stress-responsive genes in Arabidopsis and rice. *Plant and Cell Physiology*, 2013. **54**(2): p. e8-e8.
69. Sandhu, M., et al., RiceMetaSys for salt and drought stress responsive genes in rice: a web interface for crop improvement. *BMC bioinformatics*, 2017. **18**(1): p. 432.
 70. Priya, P. and M. Jain, RiceSRTFDB: a database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. Database, 2013. **2013**.
 71. Alter, S., et al., DroughtDB: an expert-curated compilation of plant drought stress genes and their homologs in nine species. Database, 2015. **2015**.
 72. Zhang, D., et al., CSTDB: A Crop Stress-tolerance Gene and Protein Database Integrated by Convolutional Neural Networks. *bioRxiv*, 2018: p. 456343.
 73. Berz, J., et al., HEATSTER: a database and web server for identification and classification of heat stress transcription factors in plants. *Bioinformatics and biology insights*, 2019. **13**: p. 1177932218821365.
 74. Kumar, S.A., et al., PSPDB: plant stress protein database. *Plant molecular biology reporter*, 2014. **32**(4): p. 940-942.
 75. Mansueto, L., et al., SNP-Seek II: A resource for allele mining and analysis of big genomic data in *Oryza sativa*. *Current Plant Biology*, 2016. **7-8**: p. 16-25.
 76. Rafalski, J.A., Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Science*, 2002. **162**(3): p. 329-333.
 77. Wilkinson, P.A., et al., CerealsDB 2.0: an integrated resource for plant breeders and scientists. *BMC bioinformatics*, 2012. **13**(1): p. 219.
 78. Rudd, S., H. Schoof, and K. Mayer, PlantMarkers—a database of predicted molecular markers from plants. *Nucleic Acids Research*, 2005. **33**(suppl_1): p. D628-D632.
 79. Gonzalez-Ibeas, D., et al., MELOGEN: an EST database for melon functional genomics. *BMC Genomics*, 2007. **8**(1): p. 306.
 80. Jorrín, J.V., A.M. Maldonado, and M.A. Castillejo, Plant proteome analysis: a 2006 update. *Proteomics*, 2007. **7**(16): p. 2947-2962.
 81. Amme, S., et al., Proteome analysis of cold stress response in *Arabidopsis thaliana* using DIGE-technology. *Journal of experimental botany*, 2006. **57**(7): p. 1537-1546.
 82. Leo, F.D., et al., PLANT-PIs: a database for plant protease inhibitors and their genes. *Nucleic acids research*, 2002. **30**(1): p. 347-348.
 83. Walker, N.S., N. Stiffler, and A. Barkan, POGs/PlantRBP: a resource for comparative genomics in plants. *Nucleic acids research*, 2007. **35**(suppl_1): p. D852-D856.
 84. Gendler, K., T. Paulsen, and C. Napoli, ChromDB: the chromatin database. *Nucleic acids research*, 2008. **36**(suppl_1): p. D298-D302.
 85. Mewes, H.-W., et al., MIPS: a database for genomes and protein sequences. *Nucleic acids research*, 2002. **30**(1): p. 31-34.
 86. Yao, Q., et al., P3DB 3.0: from plant phosphorylation sites to protein networks. *Nucleic acids research*, 2014. **42**(D1): p. D1206-D1213.
 87. Mohseni-Zadeh, S., et al., PHYTOPROT: a database of clusters of plant proteins. *Nucleic acids research*, 2004. **32**(suppl_1): p. D351-D353.
 88. Mihara, M., T. Itoh, and T. Izawa, SALAD database: a motif-based database of protein annotations for plant comparative genomics. *Nucleic acids research*, 2010. **38**(suppl_1): p. D835-D842.
 89. Kushwaha, S.K., et al., NBSPred: a support vector machine-based high-throughput pipeline for plant resistance protein NBSLR prediction. *Bioinformatics*, 2016. **32**(8): p. 1223-1225.
 90. Chandra, N.R., et al., Lectindb: a plant lectin database. *Glycobiology*, 2006. **16**(10): p. 938-946.
 91. Gribskov, M., et al., PlantsP: a functional genomics database for plant phosphorylation. *Nucleic acids research*, 2001. **29**(1): p. 111-113.
 92. Tchieu, J.H., et al., The PlantsP and PlantsT functional genomics databases. *Nucleic acids research*, 2003. **31**(1): p. 342-344.

93. Su, B., et al., PlantMP: a database for moonlighting plant proteins. Database, 2019. **2019**.
94. Mani, M., et al., MoonProt: a database for proteins that are known to moonlight. Nucleic acids research, 2015. **43**(D1): p. D277-D282.
95. Franco-Serrano, L., et al., MultitaskProtDB-II: an update of a database of multitasking/moonlighting proteins. Nucleic acids research, 2018. **46**(D1): p. D645-D648.
96. Jin, J., et al., PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic acids research, 2016: p. gkw982.
97. Szcześniak, M.W., et al., ERISdb: a database of plant splice sites and splicing signals. Plant and Cell Physiology, 2013. **54**(2): p. e10-e10.
98. Mathelier, A., et al., JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic acids research, 2016. **44**(D1): p. D110-D115.
99. Bard, J.B. and S.Y. Rhee, Ontologies in biology: design, applications and future challenges. Nature Reviews Genetics, 2004. **5**(3): p. 213-222.
100. Ashburner, M., et al., Gene ontology: tool for the unification of biology. Nature genetics, 2000. **25**(1): p. 25-29.
101. Mao, X., et al., Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics, 2005. **21**(19): p. 3787-3793.
102. Consortium, G.O., The Gene Ontology (GO) database and informatics resource. Nucleic acids research, 2004. **32**(suppl_1): p. D258-D261.
103. Künne, C., et al., CR-EST: a resource for crop ESTs. Nucleic acids research, 2005. **33**(suppl_1): p. D619-D621.
104. Johnson, C., et al., CSRDB: a small RNA integrated database and browser resource for cereals. Nucleic Acids Res, 2007. **35**(Database issue): p. D829-33.
105. Schlueter, S.D., Q. Dong, and V. Brendel, GeneSeqer@ PlantGDB: gene structure prediction in plant genomes. Nucleic acids research, 2003. **31**(13): p. 3597-3600.
106. Schreiber, F., et al., MetaCrop 2.0: managing and exploring information about crop plant metabolism. Nucleic acids research, 2012. **40**(D1): p. D1173-D1177.
107. Higo, K., et al., PLACE: a database of plant cis-acting regulatory DNA elements. Nucleic acids research, 1998. **26**(1): p. 358-359.
108. Childs, K.L., et al., The TIGR plant transcript assemblies database. Nucleic Acids Research, 2007. **35**(suppl_1): p. D846-D851.
109. Cooper, L., et al., The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. Nucleic acids research, 2018. **46**(D1): p. D1168-D1180.
110. Ge, S.X., D. Jung, and R. Yao, ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics, 2020. **36**(8): p. 2628-2629.
111. Zeng, X., et al., CMAUP: a database of collective molecular activities of useful plants. Nucleic acids research, 2019. **47**(D1): p. D1118-D1127.
112. Kriventseva, E.V., et al., OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic acids research, 2019. **47**(D1): p. D807-D811.
113. Rayson, S., et al., A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in Arabidopsis thaliana NMD mutants. PloS one, 2012. **7**(2): p. e31917.
114. Urban, M., et al., PHI-base: a new interface and further additions for the multi-species pathogen–host interactions database. Nucleic acids research, 2017. **45**(D1): p. D604-D610.
115. Bülow, L., M. Schindler, and R. Hehl, PathoPlant@: a platform for microarray expression data to analyze co-regulated genes involved in plant defense responses. Nucleic acids research, 2007. **35**(suppl_1): p. D841-D845.
116. Mano, S., et al., The Plant Organelles Database 2 (PODB2): an updated resource containing movie data of plant organelle dynamics. Plant and cell physiology, 2011. **52**(2): p. 244-253.
117. Fernandez-Pozo, N., et al., The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. Nucleic acids research, 2015. **43**(D1): p. D1036-D1041.

118. Ruiz, M., et al., TropGENE-DB, a multi-tropical crop information system. *Nucleic acids research*, 2004. **32**(suppl_1): p. D364-D367.
119. Miettinen, K., et al., The TriForC database: a comprehensive up-to-date resource of plant triterpene biosynthesis. *Nucleic acids research*, 2018. **46**(D1): p. D586-D594.