MDPI

Review

Multi-Omics Approaches and Resources for Systems-Level Gene Function Prediction in the Plant Kingdom

Muhammad-Redha Abdullah-Zawawi [1,2], **Nisha Govender** [2,*], **Sarahani Harun** [2], **Nor Azlan Nor Muhammad** [2], **Zamri Zainal** [2,3] **and Zeti-Azura Mohamed-Hussein** [2,3]

1   UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur 56000, Malaysia
2   Institute of System Biology (INBIOSIS), Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Malaysia
3   Faculty of Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Malaysia
*   Correspondence: nishag@ukm.edu.my; Tel.: +60-3-8916465

Abstract: In higher plants, the complexity of a system and the components within and among species are rapidly dissected by omics technologies. Multi-omics datasets are integrated to infer and enable a comprehensive understanding of the life processes of organisms of interest. Further, growing open-source datasets coupled with the emergence of high-performance computing and development of computational tools for biological sciences have assisted in silico functional prediction of unknown genes, proteins and metabolites, otherwise known as uncharacterized. The systems biology approach includes data collection and filtration, system modelling, experimentation and the establishment of new hypotheses for experimental validation. Informatics technologies add meaningful sense to the output generated by complex bioinformatics algorithms, which are now freely available in a user-friendly graphical user interface. These resources accentuate gene function prediction at a relatively minimal cost and effort. Herein, we present a comprehensive view of relevant approaches available for system-level gene function prediction in the plant kingdom. Together, the most recent applications and sought-after principles for gene mining are discussed to benefit the plant research community. A realistic tabulation of plant genomic resources is included for a less laborious and accurate candidate gene discovery in basic plant research and improvement strategies.

Keywords: computational approaches; functional genomics; metabolomics; systems biology; transcriptomics; plant breeding

1. Introduction

The plant kingdom is comprised of photosynthetic eukaryotes, mainly green plants. The enormous variations among and within plant populations include the physical forms, reproductive mechanisms, carbon assimilation strategies (photosynthesis metabolisms), growth and development and other factors such as responses against pests and pathogens, stress environments and productivity [1]. Plants are drastically subjected to constant changes that appear invisible to the human eye, otherwise regarded as unknown.

The phenotype accounts for highly flexible differences which result from the genetics (G), environment (E), and genetics by environment interaction (GXE). The deoxyribonucleic acid (DNA) molecule is the central hereditary unit, as the genetic material is passed from one generation to the other. Composed of four different nucleotides (adenine, thymine, cytosine and guanine), DNA carries gene fragments that encode protein molecules, of which protein-encoding genes contribute to a relatively minor portion (2%) of the total genetic material (genome). The major fraction (98%) of the genome is represented by non-coding sequences, which may indirectly participate in the protein-coding gene expression mechanisms and actions. The central dogma of molecular biology maintains genetic integrity at each life cycle via replication (DNA–DNA), reverse-transcription (RNA–DNA), transcription (DNA–RNA) and translation (RNA–protein) [2]. On the other hand, gene regulatory elements

(enhancers and silencers), non-coding RNAs such as microRNA (miRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), long non-coding RNAs (lncRNAs), and Piwi-interacting RNA (piRNA) are explicitly reported to affect gene expression levels, DNA methylation, alternative splicing events, and epigenetics [3,4].

While the study of the entire genetic material of an organism is known as genomics, the landscape of all the elemental genes expressed (transcripts) at a given time/condition is referred as the transcriptome. Transcripts are translated into protein molecules which may undergo further modifications to form small molecules of <15,000 Da (known as metabolites). These catalogues of proteins and metabolites synthesized at a given time/condition are studied in proteomics and metabolomics, respectively. Thus, transcripts, proteins and metabolites are central components driving the complexity of a biological organism. The growing application of various omics technologies has marked a burst of scientific and technological omics-based approaches offering a wealth of plant science information. "Omics" data are either interpreted independently or integrated via multi-omics analysis to understand critical questions in plant-based research [5].

Systems biology approaches (SBA) offer a plethora of virtual modelling systems equipped with *in silico* designs for gene function prediction [6]. Revolutionized by high-throughput omics technologies, SBA offers a vast amount of big data generated at the molecular level [7,8]. In parallel, computational biology has gained importance alongside SBA for dissecting and further improving the biological information of the target organisms *per se* [9,10]. Moving forward, conventional approaches that are dependent on sequence information to predict the putative biological functions (Gene Ontology classification) of a target gene have expanded robustly to accommodate organizational level-function annotations: the structural features of a given sequence, the interaction between the gene product and the cellular entity, and the phenotypic diversity of a population. In recent years, machine learning approaches and deep learning architectures such as feature-based and artificial neural networks (convolutional neural networks (CNNs) and recurrent neural networks) have been massively deployed in plant research [11,12]. The latter was evidently highly advantageous. For example, in *cis*-regulatory element (CRE) prediction, the CNN, in the absence of a priori knowledge on the target location, outperforms conventional k-mer enrichment, expectation maximization and Gibbs sampling methods with a lower false positive rate [13–15].

In this review, we highlight the use of multimodal-omics data and outline the most prominent tools employed for gene function prediction in plant research. Open-source databases available for plant-based omics studies are presented. Further, several plant-related case studies in relation to gene discovery and pathway reconstruction using unknown genes are discussed. Lastly, we emphasize and suggest the importance of integrated multi-modal analyses for gene function prediction and identification in both basic and translation plant research.

2. The Omics-Platform

2.1. Genomics

The development and application of next-generation sequencing (NGS) technologies have revolutionized crop improvement strategies primarily through genome exploration and gene discovery [16,17]. Genomics study infers the function and evolutionary history of plants, and with growing NGS technologies such as Illumina, Pacific Biosciences, Beijing Genomic Institute (BGI), Twist Bioscience, 10XGenomics and Oxford Nanopore, the research output (scientific publication) has significantly increased over the last decade (2012 to 2022) (Figure 1). The NGS technologies are indeed robust tools for genome characterization (genome size and genome ploidy level) and genetic variation identification at the genome and/or population level. Genomic datasets are established by means of comprehensive methods which involves the target species' DNA isolation, sequencing and sequence annotation using bioinformatics tools. Whole-genome sequencing (WGS) requires the entire DNA content of a single organism, while exome sequencing examines the coding DNA

sequences (exons) of a genome. Another technique, namely genotype by sequencing (GBS), is a combinatorial technique that employs restriction enzymes to select single nucleotide polymorphisms (SNP) within a population. Epigenomics targets the gene-regulating components such as DNA methylation [18,19].
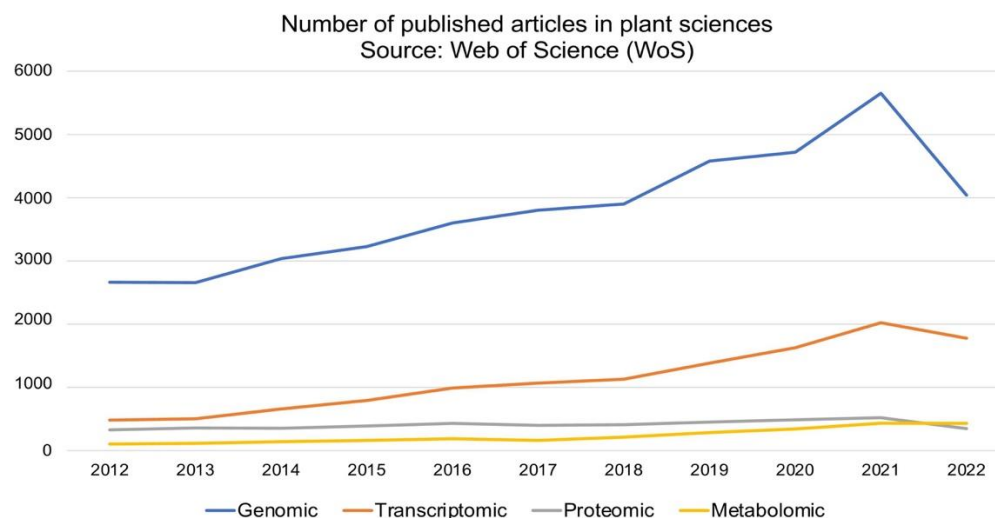
Figure 1. Scholarly omics-related articles published under the plant sciences category from 2012 to 2022. The literature search using Web of Science (https://www.webofknowledge.com) search engine was accessed on 18 September 2022 with Boolean 'or' and the following keywords: genomic, genome, transcriptomic, transcriptome, proteomic, proteome, metabolomic and metabolome.

The decreasing cost of genome sequencing has led to a deluge of plant genome sequences, particularly of agricultural crop sequences [20,21]. Sequencing price varies by the experimental designs and each design considers a myriad of technical features, such as number of reads, read length, methodology and technology. The most used methodologies to generate paired end reads in Illumina are Hiseq (100–250 bp) and Miseq (up to 300 bp). The latter has a low throughput and thus is highly recommended for small genomes <20 Mb. Next, PacBio emerged as a third-generation technology for complex genome sequencing of about 2.5–80 kb. The detection principle is based on the nucleotide excitation of a single molecule, and the technology is subjected to high error rates. The MinION by Oxford Nanopore sequences up to 20 Gb and comes with a low cost, portability features and a high error rate, comparatively much higher than PacBio. Another affordable NGS platform is BGISEQ, a forthcoming technology gaining a foothold in Asia. This technology generates single-end and paired-end reads of about 50–100 bp [22,23]. To date, Illumina remains the best quality read-producing technology. The quality of read profiles generated by Illumina can be evaluated in real time, and poor reads are filtered off using various user-friendly applications as follows: FastQC [24], Cutadapt [25], AdapterRemoval [26], Skewer [27], and Trimmomatic [28].

Plant genome assembly is challenged by the genome size, sequence repetitive nature and ploidy level (autoploid and alloploid). For example, a wheat genome of about 17 Gb features three independent sub-genomes [29]. The genome assembly procedure becomes easiest with the availability of a single allele per locus, although that is not the usual case in most plant genomes. In a systemic comparison between plant and vertebrate genomes using the unbiased kmer-based approach, plant genomes showed higher repeat contents [30].

Upon genome assembly, subsequent genome annotation is required to identify functional elements present along the genome sequence [31]. The genome structural annotation or gene predicting process adds biological meaning to the raw sequences and offers fundamental insights into the biology of the target species. However, the genome annotation process for high-quality genome assemblies is often challenged by the gene density and the

introns abundantly present in a genome. There are three distinct computational algorithms developed for detecting the coding region; ab initio (intrinsic), evidence-based (extrinsic) and genomic sequence comparison. The ab initio gene finding prediction software includes the hidden Markov models (HMM), conditional random field, support vector machine, and neural networks. Integrating the information from both the content sensor and signal sensor [31–35], the content sensor classifies the DNA sequence as coding or non-coding, whilst the signal sensor identifies specific functional regions (donor or acceptor of splice site) throughout the genome [30]. Ab initio gene predictors, for instance, GenePRIMP [36], SnowyOwl [37], CodingQuarry [38], BRAKER1 [39], MAKER2 [40], MAKER-P [41] and Seqping [42], can thus be used as a pipeline to predict a reliable annotation on the newly sequenced genomes.

The evidence-based method exploits a cost-effective approach in the form of transcriptional evidence by expressed sequence tags (ESTs) or complementary DNA (cDNA) [43]. The genomic sequence comparison identifies the relativity of the content sensor to the sequence of other genomic DNA [44]. Among the notable comparative gene-finding predictors, CONTRAST [44] has a higher accuracy in both exon/gene sensitivity and specificity than any previous year predictors; N-SCAN, TWINSCAN [45] and GENSCAN [46]. The ab initio and genomic sequence comparison methods are somehow less convincing than evidence-based due to automatic prediction based on training datasets and have poor quality in algorithms that often result in errors.

Genome sequence data facilitate comparative genomic studies targeted to infer the functions of unknown genes [47,48], enable reconstruction of metabolic pathways [49,50] and advance the understanding of evolutionary relationships between and among species [51]. Genome annotation is generally performed using sequence similarity search whereby annotated genes which encode proteins are matched with known proteins available in open repositories [48,52]. To date, plant genomic information can be retrieved from public databases such as NCBI [52] and Ensembl Plants [53]. Meanwhile, PlantGDB [54], PLAZA [55], Gramene [56] and Phytozome [57].

2.1.1. Genomic-Assisted Gene Discovery for Crop Improvement

Genomics is the key enabler of the five Gs in crop improvement instruments: (i) genome assembly, (ii) germplasm characterization, (iii) gene function identification, (iv) genomic breeding and (v) gene editing [58]. Crops with established genome assemblies are research-friendly, as the ease of computational analyses is becoming highly feasible. Plant genetic resources play a fundamental role in leveraging maximum genetic gain in a breeding program. Genetic variation under the natural setting offers breeders the basis for selection and further exploitation for crop improvement. Genetic diversity of highly valuable agronomic traits such as yield, yield-related traits, and resistance against biotic and abiotic components are amongst the most widely exploited traits for further modifications [59]. Generally, mining desirable genetic variants for subsequent improvement serves as the underlying principle of crop genetic improvement. Population-level characterization of genetic variation includes the identification of deletions, insertions, transversions, copy numbers and single nucleotide polymorphisms (SNPs). A germplasm collection holds a broad genetic diversity; thus, the accurate characterization of a large-scale germplasm remains challenging. Nevertheless, advances in genotyping and phenotyping technologies have revolutionized genomic breeding (GB) approaches.

Early GB methods were developed using markers specifically associated with genes and the quantitative trait loci governing major effects of a trait per se. Such methods were extensively applied in early GB programs: marker-assisted selection (MAS), marker-assisted backcrossing (MABC) and marker-assisted recurrent selection (MARS) [60]. Later, in the quest for genetic gain and enhanced breeding efficiency, new, improved methods emerged: genome-wide association study (GWAS), expression QTL (eQTL), haplotype-based breeding, forward breeding (FB), genomic selection (GS) and speed breeding (SB) [60,61].

2.1.2. Single Cell Sequencing

A single cell is the basic structural and functional unit of living organisms. The formation and function of higher-level tissues and organs are influenced by the various genetic mechanisms along stimuli at the cellular environment. Cell heterogeneity refers to the diverse cell states formed throughout cell growth (genetic and molecular biological changes). With highly specialized structures and functions, the cells of multicellular organisms share identical genetics and sets of genetic instructions in the translation of a functional organism. Single-cell genomics offers the cell-specific landscape information regarding the organisms' genetics, capturing the cell physiology dynamics [62].

The discovery of cell-specific transcription, tissue-specific spatial gene expression, the role of cell localization, the binding and activity of transcription factors, and the chromatin and cis-regulatory signatures of a system of interest is now feasible with growing commercial and specialized equipment systems catered toward resolving cell-specific activities. The chromatin accessibility profiling methods such as the DNase 1 hypersensitive site sequencing and assay for transposase-accessible chromatin sequencing (ATAC-seq) measure the chromatin accessibility for plant regulatory DNA across population-level species [63]. The disadvantages of these methods include a tendency to mask the cell-specific and rare events of a target tissue. Alternatively, improved high-cost systems such as the single-cell ATAC seq assays (integrated co-encapsulation or barcoding of individual cells) perform sequencing at the single-cell level [64]. In transcriptional profiling using the scRNA-seq method, the following strategies are most frequently employed: (i) fluorescence activated sorting (FACS), (ii) isolation of nuclei tagged in individual cell types (INTACT) and (iii) laser capture microdissection (LCM). Both FACS and INTACT have restricted use on selected plant species only, whereas the LCM offers a broader application range on a vast number of plant species. In general, these methods lack markers corresponding to the different differentiation states of the cell types [65].

The establishment of the Plant Cell Atlas in 2019 officially marked the trajectory of single-cell studies performed by the plant research community. Comprehensive high-resolution plant cell information (nucleic acids, proteins and metabolites) is built and shared among the scientific community [66]. Single-cell RNA sequencing (scRNA-seq) resolves cell-to-cell heterogeneity using high-throughput technologies: Drop-Seq, Chromium, Seq-well, SMART-seq 3 and iCell8 [67]. These methods offer a variety of features, which account for the following factors: (1) the target mRNA region (5′, 3′ or full length), (2) the number of cells, (3) the cell preparation technique (droplets, cell sorting and nanowells), (4) unique molecular identifiers (UMIs)—the mRNA molecule label, (5) cell size, and (6) method availability. In numerous previous studies, scRNA-seq applied on numerous tissues (Arabidopsis, rice, peanut, maize) revealed high heterogeneity, highlighting the expression signatures of cell types and development trajectories [68]. In the conventional RNA-seq method, the bulk information (average gene expression of the sample) is obtained, whereas the scRNA-seq technique consists of pools of information, each corresponding to the different types of cells present in the sample. The cell preparation is rendered as the utmost challenge to obtaining a decent result with accurate interpretations. Optimizing the protoplast isolation is vital, considering the following factors in a typical plant cell: cell density, cell wall thickness, digestion efficacy (influenced by cuticle, lignin, suberin and other deposition), enzyme type and requirement and enzyme digestion time [67,69].

2.1.3. Genome-Wide Association Study (GWAS)

Amongst these methods, GS is the most preferred tool for breeding programs, as the method does not rely on diagnostic markers entirely and the selection is made on the breeding lines evaluated according to genomic-estimated breeding values (GEBV) generated from the genomic-wide marker data sets. Genomic selection (GS) gathers the additive effects of all the genes governing the genetic variance of a given trait. With each independent gene imparting a relatively small effect, the number of genes controlling a single trait may stretch from hundreds to thousands [60,70]. Using a genome-wide

marker and phenotype information, the GS method establishes the association between markers and phenotypes from an observed population. A GS analysis was first performed following Fisher's infinite model, and soon was extended to the genomic best linear unbiased prediction (GBLUP) model. The latter accommodates GXE interactions and thus offers a more accurate prediction [61,71]. Later, the Markov chain, Monte Carlo and Bayesian modelling methods were developed to include non-additive genetic effects such as adverse environmental conditions. In the GS method, machine learning builds a training/reference population of individuals with information of interest (genotype and phenotype) to train prediction models on the test population or selection candidates. The prediction accuracy is affected by training set population size, density/number of the genome-wide markers and the heritability of the trait of interest [72].

Genomics, together with advanced-level genomic tools, open-source genome resources and powerful technologies, have accelerated crop breeding through rapid trait discovery techniques. Proposed 15 years ago, genomics-assisted breeding (GAB) has now expedited a broad range of breeding programs for resistance enhancement against diseases and tolerance improvement against abiotic factors such as submergence, salinity and drought. In rice, the "Improved Samba Mahsuri", a GAB product, carries the Xa21, xa13, xa5 and xa38 genes governing the bacterial blight (BB) disease (causal pathogen, *Xanthomonas oryzae*) along with Pi-2 and Pi-54, blast disease (causal pathogen, *Magnoporthe oryzae*) resistance genes [73–75].

2.1.4. Pan-Genomics

There are about 390 thousand land plant species, and their genomes are highly complicated (highly repetitive DNA content, polyploidy and heterozygosity) and diverse (genome size varying from 60 Mb to 150 Gb). Plant genome changes arise from evolutionary forces that shaped plant speciation and evolution. Pan-genomics, a subset of plant genomic research, is highly suitable for plant species with extensive genetic diversity at the population level. Pan-genomes have been developed for important agricultural crops and model plants such as rice, Arabidopsis, barley, soybean, maize, wheat, tomato, etc. [76]. The key principles of pan-genomics include the comparison of high-quality genomes to provide insights into the collection of core and dispensable genes in a species population. Generally, a single genome or a small number of genomes do not make a good sample in pan-genome construction. Integration of many high-quality genomes is important to obtain comprehensive genetic information of the target population [77]. Genes are designated as the basic units defining a pan-genome. Pan-genome studies are most useful in understanding plants with a wide spectrum of genetic diversity and gene pools. In brief, the pan-genome strategy first establishes a target population of highly diverse individuals. A good selection of representative individuals in the population is reflected by phenotypic diversity, as determined by the phylogenetic relationship among the individuals of the population. Next, a high-quality genome assembly method for long reads is employed using automatic annotation pipelines. The construction approaches available for pan-genome analyses includes the de novo assembly (detects variant types and classifies genes into core and dispensable), iterative assembly (based on a single reference genome), and graph-based assembly strategy (utilizes graphs from a reference genome to represent the diversity and variations). Comprehensive tools and pipelines popularly employed in pan-genome analyses were exhaustively described by Li et al., 2022 [78].

2.2. Transcriptomics

A transcriptome is an atlas of RNA transcripts of a tissue, cell or defined specific condition [79]. Using the genome information, a transcriptome is "read" to obtain a comprehensive description of the genes expressed at a given time point. The mapping and quantification of the transcriptional activity are central to transcriptome studies. In the modern era, the transcriptomes are produced either by the microarray [80] or RNA-sequencing (RNA-seq) technology [81]. The latter is preferred by the plant research community due

to higher precision in capturing lowly expressed RNAs and isoforms [81]. Comparatively, the RNA-seq technology detects a greater percentage of novel transcripts than the microarray [82,83]. In most transcriptome data analyses, the raw count data are subjected to differentially expressed genes (DEGs) analysis, co-expression network construction and other techniques such as alternative splicing and isoform analysis [84,85]. Both DEG and network analyses are used extensively to discover genes underpinning various biological processes such as plant defense response [86], regulation [87], water stress *JAZ1* in *G. arboreum* [88], desiccation tolerance and drought (such as LEA) in *A. thaliana* seeds [89], cellulose synthase in secondary cell wall synthesis [90] and cell wall-related genes in *A. thaliana* [91].

In 2002, the Gene Expression Omnibus (GEO) repository was first established as an open repository for gene expression data obtained from various platforms such as microarrays, serial analysis of gene expression (SAGE) and other sequence-based data [92]. Since then, the number of open-source gene expression data repositories for various plant species and condition-specific has been on the rise: The Arabidopsis Information Resource (TAIR) [93], TRAVA [94], RiceXPro [95], Transcriptome Encyclopedia of Rice (TENOR) [96], Barley Gene Expression Database (Bex-db) [97], and Plant Stress RNA-Seq Nexus (PSRN) [98] (Table 1).

Transcriptome data relate to the prediction of genome-scale reconstruction from previous studies: the starch biosynthesis of *Manihot esculenta* [99], the light and temperature acclimation in *Arabidopsis thaliana* [100], and the biosynthesis of biotic stress-regulated pathways (i.e., tryptophan, auxin and serotonin) in *Oryza sativa* [101]. High and low levels of mRNA transcription have improved the understanding of the response outcome in the genome, especially those mechanistic associations between the cellular trade-offs and epistatic gene interactions [102,103].

Table 1. Plant omics databases, as accessed on 24 August 2022.

Omics Type	Database	Organism	URL	References
Genomics	Plant Genome Database (PlantGDB)	Plants	http://www.plantgdb.org	[54]
	Plant Genome DataBase Japan (PGDBj)	Plants	http://pgdbj.jp/?ln=en	[104]
	National Center for Biotechnology Information (NCBI)	Various	https://www.ncbi.nlm.nih.gov	[52]
	Ensembl Plants	Plants	http://plants.ensembl.org/	[53]
	Phytozome	Plants	https://phytozome.jgi.doe.gov	[57]
	PLAZA	Plants	https://bioinformatics.psb.ugent.be/plaza/	[55]
	Plant Genome and Systems Biology (PGSB PlantsDB)	Plants	http://pgsb.helmholtz-muenchen.de/plant/plantsdb.jsp	[105]
	Chloroplast Genome Database (ChloroplastDB)	Plants	http://chloroplast.cbio.psu.edu/	[106]
	The Solanaceae Genomics Resource (Spud DB)	Potato	http://solanaceae.plantbiology.msu.edu	[107]
	Melon Genome Database (Melonomics)	Melon	https://www.melonomics.net/	[108]
	Maize Genetics and Genomics Database (MaizeGDB)	Maize	https://www.maizegdb.org	[109]
	Rice Annotation Project Database (RAP-DB)	Rice	https://rapdb.dna.affrc.go.jp	[110]
	Rice Genome Annotation Project (RGAP)	Rice	http://rice.plantbiology.msu.edu	[111]
	GrainGenes	Wheat, Barley, rye, oat	http://wheat.pw.usda.gov/GG3/	[112]
	SoyBase	Soy	Soybase.org	[113]
	Genome Database for Rosaceae (GDR)	Rosaceae plants	https://www.rosaceae.org/	[114]
	Brassica Database (BRAD)	Brassica plants	http://brassicadb.org/brad/	[115]

Table 1. *Cont.*

Omics Type	Database	Organism	URL	References
Transcriptomics	Gene Expression Omnibus (GEO)	Various	https://www.ncbi.nlm.nih.gov/geo/	[92]
	AgriSeqDB	Plants	https://expression.latrobe.edu.au/agriseqdb	[116]
	The Bio-Analytic Resource for Plant Biology (BAR)	Plants	http://bar.utoronto.ca	[117]
and	The Arabidopsis Information Resource (TAIR)	Arabidopsis	https://www.arabidopsis.org	[93]
	Transcriptome Variation Analysis (TRAVA)	Arabidopsis	http://travadb.org	[94]
	The Rice Expression Profile Database (RiceXPro)	Rice	https://ricexpro.dna.affrc.go.jp	[95]
	Transcriptome Encycloperdia of Rice (TENOR)	Rice	http://tenor.dna.affrc.go.jp/	[96]
	Barley Gene Expression Database (Bex-db)	Barley	http://barleyflc.dna.affrc.go.jp/hvdb/	[97]
	Plant Stress RNA-seq Nexus (PSRN)	Plants	http://syslab5.nchu.edu.tw	[98]
	Plant microRNA database (PMRD)	Plants	http://bioinformatics.cau.edu.cn/PMRD/	[118]
Interactomics	STRING	Various	https://string-db.org	[119]
	Database of Interacting Proteins (DIP)	Various	http://dip.doe-mbi.ucla.edu	[120]
	Protein–Protein Interaction Database for Maize (PPIM)	Maize	http://comp-sysbio.org/ppim	[121]
	IntAct	Various	https://www.ebi.ac.uk/intact/	[122]
	Oryza sativa Protein–Protein Interaction Network (PRIN)	Rice	http://bis.zju.edu.cn/prin/	[123]
	Biomolecular Interaction Network Database (BIND)	Various	http://bind.ca	[124]
	The Biological General Repository for Interaction Datasets (BioGRID)	Various	https://thebiogrid.org	[125]
	Arabidopsis thaliana Protein Interaction Network (AtPIN)	Arabidopsis	https://atpin.bioinfoguy.net	[126]
	PlaPPISite	Plants	http://zzdlab.com/plappisite/index.php	[127]
	3D interacting domains (3did)	Various	https://3did.irbbarcelona.org	[128]
	Molecular INTeraction database (MINT)	Various	http://mint.bio.uniroma2.it/mint/	[129]
	ATTED-II	Plants	http://atted.jp/	[130]
	CressExpress	Arabidopsis	http://cressexpress.org/	[131]
	Arabidopsis Network (AraNet)	Arabidopsis	http://www.inetbio.org/aranet/	[132]
	Co-expressed Biological Processes (CoP)	Plants	http://webs2.kazusa.or.jp/kagiana/cop0911/	[133]
	EXPath	Plants	http://expath.itps.ncku.edu.tw/	[134]
	Plant Omics Data Center (PODC)	Plants	http://bioinf.mind.meiji.ac.jp/podc/	[135]
	Plant Netwrok (PlaNet)	Plants	http://aranet.mpimp-golm.mpg.de/	[136]
	OryzaExpress	Rice	http://plantomics.mind.meiji.ac.jp/OryzaExpress/	[137]
	PlantExpress	Rice, Arabidopsis	http://plantomics.mind.meiji.ac.jp/PlantExpress/	[138]
	Rice Functionally Related Gene Expression Network Database (RiceFREND)	Rice	http://ricefrend.dna.affrc.go.jp/	[139]
	Vitis vinifera Co-expression Database (VTCdb)	Grape	http://vtcdb.adelaide.edu.au/	[140]
	GeneMania	Various	http://genemania.org/	[141]
	A Comprehensive Systems-Biology Database (CSB.DB)	Various	http://www.csbdb.de/csbdb/home/databases.html	[142]
	RapaNet	Brassica	http://bioinfo.mju.ac.kr/arraynet/brassica300k/query/	[143]
	Rice Expression Database (RED)	Rice	http://expression.ic4r.org	[144]
	PhytoNet	Various	www.gene2function.de	[145]
	CoNekT	Plants	https://conekt.sbs.ntu.edu.sg	[146]
	CoCoCoNet	Plants	https://milton.cshl.edu/CoCoCoNet	[147]

Transcriptome-Wide Association Studies: Prediction of Genes Governing Complex traits

Global transcriptional activity measured by the transcriptome-wide association studies (TWAS) offers a fundamental understanding of the spatiotemporal regulation of transcription events in plants [148]. Transcription causes variation, often observed as a collection of events resulting from altered coding sequences. Both mRNA and protein expression are spatial and temporal targets for selecting variations caused by the coding sequences. TWAS unravel endophenotype or variation that is predominantly caused by genetic factors. Such a feature is highly valuable for prioritizing candidate genes governing complex agronomic traits. TWAS was recently proposed as a powerful tool to predict trait-associated gene

expression based on GWAS summary data [149]. TWAS, in combination with GWAS, increases the power of detection of unknown genes and offers a selection of prioritized causal genes [150,151].

2.3. Phenome

For the past decade, plant phenomics has made significant strides with the advancement of imaging and sensor technologies in measuring a wide range of traits or phenotypic variations in response to environmental factors or genetic modifications [152]. Phenomic data aid in the understanding of the pathways that link genotypes to phenotypes and determine the underlying causes of complex events in crop yields and diseases [153]. Gathering relevant phenotypic data across multiple organizational levels is a key step in phenomics, which aims to characterize the full range of phenotypes that can be expected from a given genome. Therefore, plant phenotyping can be stratified as per resolution and dimensionality (from molecular to entire plant) and environments (from lab to field settings) [154]. The phenotyping method has become an outstanding tool for integrating knowledge into producing high-performance cultivars, particularly for breeders seeking to develop higher tolerant cultivars against abiotic and biotic challenges.

Handling high-dimensional phenomics data necessitates advanced computational methods. In plants, both quantitative trait locus (QTL) mapping and high-throughput phenotyping (HTP) are being utilized to identify the underlying genes responsible for the desired phenotypes. The development of NGS techniques has facilitated rapid and cost-effective access to a vast amount of genomic data, allowing QTL mapping-based marker-assisted selection (MAS) to be conducted. QTL mapping relies heavily on high-quality phenomics data. Near-infrared reflectance spectroscopy (NIRS) data composed of phenomics information have been used as predictors to compare its predictive ability with marker data [155]. The phenomics study via NIRS has been shown to achieve promising predictive abilities in crops, including soybean [156], maize [157] and sugarcane [158]. HTP, on the other hand, relies on automated trait analysis in producing phenotypic data, such as imaging techniques. This technique uses computational image-analysis tools to parse images or videos of traits such as root architecture, height, morphology, and photosynthetic status to extract the latter information [159].

2.4. Epigenetic Modification

Chromatin is a complex structure consisting of DNA and histone proteins that are susceptible to epigenetic mechanisms, such as DNA methylation, histone tail modifications and methylation mediated by small RNA (e.g., miRNA, piRNA and/or snRNA). Epigenomics is a dynamic process that alters gene regulation activities that cause plant morphology and development to become abnormal due to environmental factors such as biotic and abiotic stress. DNA methylation patterns vary greatly between plant species. Genes and transposable elements (TEs) in angiosperms are typically methylated at CHG and CHH (H = A, C, or T) nucleotides, whereas CG methylation is highly abundant in animals.

Epigenomic studies using high-throughput sequencing methods, such as methylation arrays, chromatin immunoprecipitation sequencing (ChIP-Seq), assay for transposase-accessible chromatin sequencing (ATAC-Seq), reduced-representation bisulfite sequencing (RRBS-Seq), methylated DNA immunoprecipitation sequencing (MeDIP-Seq), and bisulfite sequencing (BS-Seq), have made it feasible to investigate the roles of epigenetic mechanisms and regulatory pathways in plants at a genome-wide scale [160]. Methylation arrays were the first epigenetic technologies developed to study DNA-methylated CpG islands characterized by the presence of cytosine-guanine sequences. However, the use of methylation arrays in plant studies is still limited compared to other sequencing methods and in mammals [161]. Bisulfite sequencing is widely regarded as the gold standard for detecting 5-methyl-cytosine (5mC) due to its ability to sequence the genome at a base-pair level. Other methods, such as MeDIP-Seq and RRBS-Seq, only examine the preselected genomic regions based on the prevalence of CpG content or methylation [162]. Meanwhile, ChIP-Seq

is a powerful method used to study the interaction between transcription factors and DNA binding sites and provides additional information about epigenetic modification based on the chromatin structure or histone changes [163]. To date, more than 11,000 ChIP-Seq data series have been deposited in the Gene Expression Omnibus (GEO) database. Others such as the ATAC-Seq database recorded a total of 1880 data series [164]. By using ATAC-Seq, the chromatin accessibility with DNA methylation changes can be determined using hyperactive Tn5 transposase that cleaves the DNA and then inserts sequencing adapters into open chromatin regions [165].

Epigenomic technologies are widely used to identify genes that underpin various functions. For instance, ChIP-Seq analysis was used by Li and colleagues to identify genes involved in the activation and repression of gene regulation in response to abiotic stress. Additionally, ATAC-Seq profiling of accessible chromatin was carried out to investigate the transcriptional regulatory landscape of plant genomes, which appears to be conserved across the root tips of plant species during development [166,167]. Using BS-Seq, Li et al. (2020) [168] discovered that methylated cytosines (mCG) contributed to the difference in methylation levels in drought stress, revealing extensive DNA methylation changes in response to drought. The MeDIP-Seq profiling of olive development also showed differential DNA methylation in secondary metabolism, which is responsible for the quality of olive oil. This finding provides an insight into the significance of the methylation status of olives during the ripening process [169].

2.4.1. Interactomics

Interactomics, the study of interactions between functional elements within an organism is revolutionizing genetic research. The systemic dissection of functional genes riding the phenotype of interest is analyzed by genetic models or networks within and between genetic layers. For example, the genome-wide protein–protein interaction analysis may represent two distinct genetic layers, namely the proteomics and transcriptomics. An interactome study exploits large datasets generated by multi-omics technologies to improve the predictive power in understanding the role of functional elements of a complex biological system [170]. It offers valuable information on the associations between functional elements across multiple biological processes. The functional elements of an interaction network are represented as nodes, whilst the relationships between the nodes are edges. The edges are constructed based on the correlation measurements derived algorithmically from quantitative omics-datasets [171,172].

A rapid research pace in co-expression network construction and analysis using transcriptomes (RNA-seq and microarray generated datasets) may have arisen from the increasingly growing open-source databases. Databases of co-expression datasets include the ATTED-II [130], AraNet [132], GeneMANIA [141] and others, as listed in Table 1. In higher plants, co-expression network analyses have successfully dissected gene function prediction in glucosinolate biosynthesis [173], cell wall biosynthesis [174], transcriptional regulation of hormone biosynthesis [175] and Arabidopsis aliphatic glucosinolate biosynthetic pathway [176].

Protein–protein interaction (PPI) dissects the physical interactions between a group of proteins, ultimately imparting a global understanding of the functional mechanisms of a proteome landscape, domain interactions and motif and site association of complexes [177]. PPI datasets are generated by means of in vivo, in vitro and in silico methods [178]. The in vivo methods such as yeast two-hybrid (Y2H) [179], split ubiquitin system (SUS) [180] and bimolecular fluorescence complementation (BiFC) [181] test the physical interactions between two proteins. Meanwhile, in vitro methods such as affinity purification mass spectrometry (AP-MS) [182] and protein microarrays [183] quantify PPIs and protein activities. Both the in vivo and in vitro methods serve as evidence in PPI network construction. In addition, various computational methods have been developed for PPI prediction: interolog mapping, gene/domain-fusion inference, domain/motif-domain transfer, gene co-expression network and machine learning approaches.

2.4.2. Resources for Plant Protein–Protein Interactions

The number of plant species-specific experimentally validated or predicted PPIs has been growing with the development of new info-centric databases [127]. As such, the PlaPPISite houses comprehensive and high coverage interactomes of 13 different species (zzdlab.com/plappisite/index.php) [127]. The Interacting Proteins (DIP) [120] and 3D interacting domains (3did) [128] databases integrate information from the Protein Data Bank (PDB) for the identification of protein interaction sites [183]. Concerning plant PPI data, the Biological General Repository for Interaction Datasets (BioGRID) [126], Molecular Interaction database (MINT) [129], Biomolecular Interaction Network Database (BIND) [124], Functional Protein Association Networks (STRING) [119] and *Arabidopsis thaliana* Protein Interaction Network (AtPIN) [126] are rendered the most widely employed databases in plant functional studies.

2.4.3. Integrated Multi-Layer Omics Data for Functional Studies in Plant

The functional aspects governing phenotypic diversity are cumulatively driven by the distinct layers of the central dogma. Genetics research, along with integrated multi-omics approaches, has made a major leap forward in gene function prediction and identification. The use of multi-omics datasets established from a single experiment is essential for significant characterization and identification of gene/protein/biomolecules and their putative roles in the biological pathways and processes. The omics-to-interactome relationship using multi-layer omics modules is shown in Figure 2. Integrative omics recruits at least two or more distinct genetic layers, often established from omics technologies.
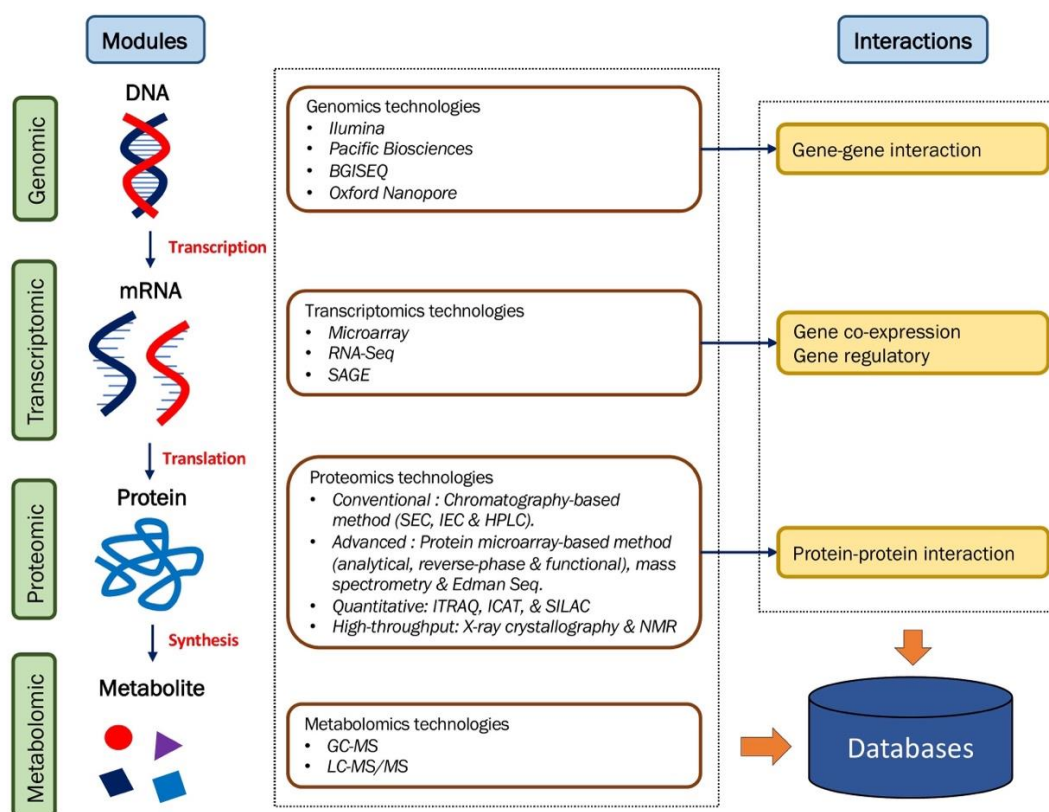
Figure 2. Type of omics datasets. Omics datasets can be divided into two categories: modules and interactions. Module data indicate molecular sequences of biological systems; DNA (genome) transcribed into mRNA (transcriptome), later translated into proteins (proteome), and lastly synthesized into metabolite (metabolome). Interaction data, known as interactomes, represent the relationships of module data generated from respective platforms. The omics resources of the omics technologies and network interactions can be downloaded from the databases.

- Genomics-transcriptomics

Co-expression networks using transcriptome-based analysis have facilitated the characterization of unknown gene functions within the metabolic pathways [171]. A combinatorial analysis of sequence similarity and co-expression identifies conserved co-expression networks across various crop species [184]. With the recent development of co-expression network databases such as PlaNet [136], PhytoNet [145], CoNekT [146], and CoCoCoNet [148], it is now possible to gain insights into the potential causal effect of gene interactions on a trait of interest. In a study by Liu et al. [185], the gene regulatory network (GRN) constructed from transcriptome data elucidated the relationship between transcription factors and target genes via direct interaction. In another study, GRN was utilized to investigate the genome-wide transcriptional response of fruit development [186]. The hub genes (group of tightly associated genes) identified from the co-expression network correlated with TFs, suggesting potential regulatory mechanisms involved in fruit development metabolism [185,186].

- Transcriptomics-proteomics

The integration of genome and transcriptome data may demonstrate the abundance of protein; however, it does not compulsorily correlate with the corresponding mRNA levels [187,188]. This may likely occur when there is a protein synthesis delay at the regulation and post-translational modification process [189], along with other factors such as the density of the ribosomal subunit [190] and physical characteristics of the transcript [191].

In a study which investigated maize leaf development, the correlation between the mRNA and protein abundance was relatively weak during the leaf transition from heterotrophic to autotrophic cells compared to later stages of development [192]. In another study, both proteome and transcriptome data were integrated to understand tomato pericarp ripening [193]. The post-translational mechanism occurred during ripening when the protein abundance and mRNA levels showed a weak correlation, in contrast to the early stage of tomato ripening [193]. Integrative methods are effectively deployed to understand plant responses toward biotic and abiotic stresses. Peng and colleagues suggested that several cotton stress-responsive proteins (gigantae protein, α-crystalline heat shock protein, and β-1-pyrroline-5-carboxylate synthetase) regulate the alternative splicing events as the mRNA levels were significantly correlated with protein abundance under salt-stress condition [194]. The alternative splicing event allows the translation of spliced mRNAs (from a single gene) into multiple proteins [195].

2.5. Candidate Gene Mining in the Context of Pathway Reconstruction

In plant biology research, gene function identification is primarily challenging as the study requires a large-scale dimension [196]. In *Arabidopsis thaliana*, approximately 27,500 genes that encode proteins were reported in 2013 by The Arabidopsis Information Resource (TAIR), and this number was expected to increase with time. In the same year, 30% of genes were reported to be experimentally validated, compared to only 11% in 2007 [197,198]. Within 50 years of Arabidopsis research, more than 50,000 publications have been released and stored in the TAIR database for data curation, the annotation of newly discovered genes and metabolic pathway refinement [199]. Candidate gene mining in higher plants is much more challenging compared to bacteria due to tissue-level complexity at the cellular level and the lack of functional information about existing annotated gene functions [43].

Challenges in Cellular Pathway Reconstruction

Plants synthesize more than a million different types of metabolites [200]. Cellular pathways such as the metabolic, biochemical, and signal transduction of plant function influence the system-level behavior, growth and development processes. Incomplete metabolic pathways from weak annotation necessitate pathway reconstruction [44,201].

The identification of candidate genes in an incomplete metabolic pathway may result in unknown proteins. These unknown proteins could represent a missing enzymatic reaction underpinning a dead-end metabolite. The first step in pathway reconstruction begins with the identification of orthologous genes. Orthologous genes in different species arise from speciation events. Since a common ancestor holds orthologous genes, they retain a similar gene function. Orthologous domains/proteins are retrievable in Clusters of Eukaryotic Orthologous Groups (KOG) via the Clusters of Orthologous Groups (COG) (https://www.ncbi.nlm.nih.gov) [202] and WU-BLAST2 server (dove.embl-heidelberg.de/Blast2), as accessed on 19 August 2022 [203].

3. Guilt-by-Association (GBA), a Method for Gene Discovery

A gene co-expression network (GCN) is a powerful tool to uncover unknown genes based on correlation values computed (gene expression data) among a series of experimental samples/conditions [204]. A candidate gene is assumed to co-function with a partnering gene in the event of correlation, an association measure (Figure 3). The GCN is built by calculating the correlation of mRNA expression levels across samples. The transcripts are represented as nodes connected by either weighted or unweighted correlation values, represented as edges. Unweighted edges imply a binary graph, whilst weighted edges score the different strengths of the edges of a completely connected graph.
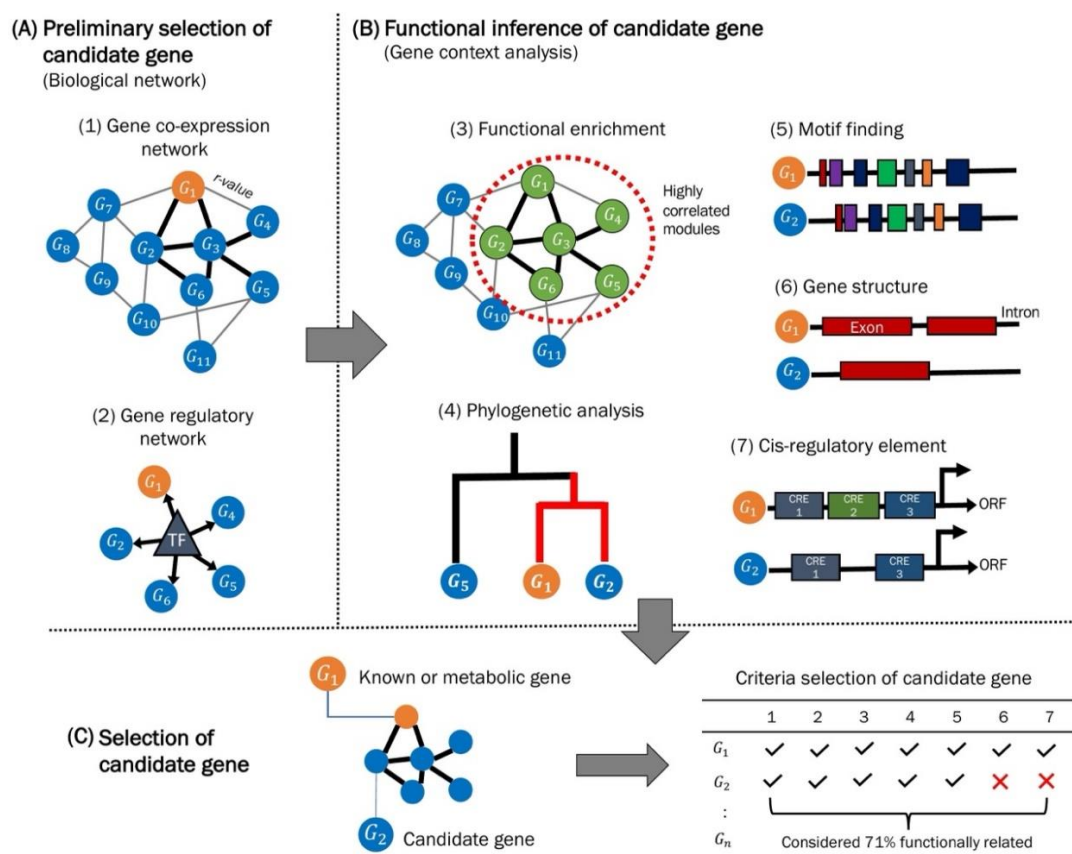
Figure 3. Guilt-by-association techniques for candidate gene discovery. (**A**) Preliminary selection of candidate gene using the biological network, in which the co-functional information of candidate gene with known gene (i.e., metabolic or other functional genes) can be extracted from gene correlation and coregulation/regulatory network (steps 1–2). (**B**) Co-functional information can be inferred by gene context analyses (steps 3–7). Candidate genes will be observed based on enriched in a similar function (step 3), clustered in a monophyletic group (step 4), shared similar distributions of motifs (step 5) and exon/intron structure (step 6), and lastly consist similar CREs (step 7). (**C**) The ranking of the high confidence candidate gene will be observed based on the predicted co-function similarity.

In 2000, the "guilt-by-association (GBA)" principle was proposed to unravel the gene function of uncharacterized or hypothetical targets within a functional network [205]. Assuming that two interacting genes or proteins are hypothetically bound to a similar or related cellular function [205,206], the GBA assesses for biological information of a co-expression network such as functional links between genes: plants [207], yeast, and bacteria [208]. Gene co-expression and co-regulation have become a standard technique to identify the function of unknown genes in metabolic pathways [176]. By using mRNA data from RNA-seq/microarray technologies, genes with similar expression profiles are hypothetically presumed to be regulated by a similar transcription factor [209,210].

According to Hansen et al. [207], two genes with similar features (sequence, structure, and expression pattern) may share a similar function. Gene context analysis and gene network study are commonly known as GBA. Genomic evidence using comparative genomics and gene co-expression networks infer the participation of the candidate genes in a similar or related pathway by identifying the possible association with genes of known functions [48,207,211]. According to Osterman and Overbeek [48], the gene context technique ranks candidate genes through multiple assignments. Thus, candidate genes with highly similar contexts are measured as high-confidence genes with potential functional association with known genes [212–215].

Basically, GCN construction involves three key steps: (i) input data comprising an expression matrix (m = gene across n = conditions) vector with n = dimension (Table 1), (ii) similarity measurement/association measure and gene similarity matrix, and (iii) the threshold value (cutoff correlation value). The association measures are calculated using Pearson's correlation coefficient (PCC), Spearman's correlation coefficient (SCC) and others, dependent on the dataset distribution. The gene interaction of GCN is defined as the correlation between genes [211–213]. Correlation values that meet the threshold criteria assume significant interactions [212].

The threshold value selection criteria vary for unweighted and weighted GCN [214]. There is no rule of thumb applied for setting the threshold values. Although a soft threshold value (nearing zero) is considered less significant, it compensates for the robustness of a weighted GCN [215]. On the flip side, important genes might be missed out from the network with a highly stringent threshold selection [216]. A hard threshold (r = 0.8 to 1.0) has been shown to be more relevant in studies inferring biological relationships. The validity of the biological information computed based on the GO functional similarity measure increases at r > 0.8 [215]. GCN has been widely applied to Arabidopsis for the identification of genes corresponding to cell wall biosynthetic [90], fatty acid chain [217], photorespiration [218], immune response [219] and other metabolic pathways. In others, a random threshold correlation value was applied in GCN construction: r = 0.7 in GCN of biotic and abiotic attack [220,221] and r = 0.83 for leaf development [222].

In weighted GCN, the strength of the interaction is reflected by the score distribution (0 to 1) [214]. Contrarily, in unweighted GCN, the interaction score is computed by binary values, whereby 0 represents no correlation, and 1 indicates the presence of correlation [214]. The WGCNA [223] and webCEMiTool [224] are freely available computational resources available for weighted GCN construction. Others that feature differential GCN construction include dcanr [225], Ebcoexpress [226], MODA [227], DICER [228], and DiffCoEx [229] and CoExNet [230] for unweighted GCN. The differential GCN infers the causal regulatory changes between sample groups of different conditions [87]. For example, the comparative co-expression of mRNA and lncRNA in *Cleistogenes songorica* under water-deficient conditions identified differentially expressed mRNAs and lncRNA of common TFs families. The function of lncRNAs was identified as drought stress regulation via interaction with miRNAs and protein-coding genes [231].

Recently, another method in GCN, comparative GCN analysis, incorporated gene homology and co-expressed gene information for functional prediction in different plant species [148]. Comparative GCN can be executed by predicting conserved interaction between homolog genes from two or more species. Conserved genes with similar co-

expression profiles showed significant biological similarities and differences in Arabidopsis and maize [232]. Obviously, a gene that integrates both the sequence similarity and co-expression profile information provides a better prediction accuracy than independent single-information analysis. The integration of homology gene and correlated gene expression allows useful information on candidate genes to be obtained from the conserved gene modules. The functional annotation could be drawn relatively from model plants to the crops of interest [233,234]. There are various web servers and applications available for co-expression and comparative plant studies; EXPath [134], Plant Network (PlaNet) [136], RED [144], PhytoNet [145], and CoCoCoNet [147]. These tools combine and compare the conserved GCN among plant species, ultimately aiding gene function prediction.

4. Predictive Modelling, Artificial Intelligence and Machine Learning Based Methods

The 'big data' era in plant sciences offers massive omics-datasets that are extremely large, noisy and heterogeneous in nature. Gene, protein and metabolite prediction using phenotypic datasets from various genotypes under adverse environmental conditions increases the call for scientific approaches that could effectively handle big data with parallel integration of multi-modality phenomics, metabolomics, genomics, proteomics, transcriptomics, etc. [235]. In this context, artificial intelligence (AI) and machine learning (ML) fit perfectly to support the decision-making processes in various plant research areas while accommodating diverse and fragmented datasets: the prediction of genome regions favorable for genetic modifications, modelling the genotype–environment interactions, the dissection of complex plant traits, and the prediction of genome crossover regions. ML comprises algorithms that learn to perform a required task using a given dataset. There are two distinctive types of machine learning: (i) supervised learning, where output prediction is dependent on the input data (training data), and (ii) unsupervised learning, which identifies patterns in an unlabeled dataset [236]. The most common unsupervised methods used in plant research include principal component analysis (PCA), clustering and Autoencoder. The PCA method corrects for data variability by linear transformation of the variables. The clustering method clusters the data observations based on the similarity features. The Autoencoder utilizes artificial neural networks to perform reconstruction using compressed input data to minimize the differences in the original dataset [237,238]. In plant research, the supervised method is much preferred compared to the unsupervised method; nevertheless, the selection of ML methods is largely influenced by data availability and the objective of the analysis [239].

5. Conclusions

Viridiplantae is estimated to consist of about several hundred thousand species. With tremendous advances in sequencing technologies and computational tools, genome sequencing and assembly have emerged as important strategies for decoding genetic information of plant species. Undeniably, plant species with decoded genetic information are better placed for manipulation and subsequent improvement in breeding programs. Important crops, primarily food crops such as rice, wheat, sunflower, soybean and many others, have been the species of interest in high-throughput next-generation sequencing (NGS) technology. Leading barriers in the success of elucidating the plant genetic landscape includes the large and inherent complexity of plant genomes attributed by polyploidy, phenotypic variation and heterozygosity factors observed in repetitive sequences, transposable elements (TE), tandem arrays, and ribosomal gene clusters. To date, only <1000 draft plant assemblies have been constructed using the NGS platforms. Nevertheless, new methods are being robustly developed to enhance specificity against the research biological question. Optimized computational algorithms, computational power and sequencing technologies are increasingly catered toward answering specific research questions. The ultimate challenge in gene function prediction involves employing that most appropriate technological tools feasible to the experimenter. With climate change on the chart of global issue, food security requires serious attention in the realm of an ever-growing human

population. Plant breeding is the utmost fundamental strategy in crop yield improvement. Conventional breeding programs are being replaced by rapid high-throughput breeding approaches, ultimately to gain better resolution in effective breeding programs. Gene function prediction and identification is a pre-requisite step that informs the design of a plant breeding method. Modern biological research provides comprehensive insights into system-level variation using collated multi-omics tools and integrative system biology approaches. What is the concerted pool of genes, proteins and metabolites underpinning a complex trait? The ease of dissecting the research questions posed here becomes much less with integrative omics analyses which favor high-confidence predictions.

Author Contributions: Conceptualization, M.-R.A.-Z.; methodology, M.-R.A.-Z.; validation, N.G., S.H., N.A.N.M., Z.Z. and Z.-A.M.-H.; formal analysis, M.-R.A.-Z.; writing—original draft preparation, M.-R.A.-Z.; writing—review and editing, N.G., S.H., N.A.N.M., Z.Z. and Z.-A.M.-H.; supervision, Z.-A.M.-H.; project administration, S.H., N.A.N.M., Z.Z. and Z.-A.M.-H.; funding acquisition, Z.-A.M.-H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Parry, M.A.J.; Reynolds, M.; Salvucci, M.E.; Raines, C.; Andralojc, P.J.; Zhu, X.-G.; Price, D.G.; Condon, A.G.; Furbank, R.T. Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. *J. Exp. Bot.* **2011**, *62*, 453–467. [CrossRef]
2. Pramanik, D.; Shelake, R.M.; Kim, M.J.; Kim, J.-Y. CRISPR-Mediated Engineering across the Central Dogma in Plant Biology for Basic Research and Crop Improvement. *Mol. Plant* **2021**, *14*, 127–150. [CrossRef] [PubMed]
3. Zhang, P.; Wu, W.; Chen, Q.; Chen, M. Non-Coding RNAs and their Integrated Networks. *J. Integr. Bioinform.* **2019**, *16*, 20190027. [CrossRef]
4. Yu, Y.; Zhang, Y.; Chen, X.; Chen, Y. Plant Noncoding RNAs: Hidden Players in Development and Stress Responses. *Annu. Rev. Cell Dev. Biol.* **2019**, *35*, 407–431. [CrossRef] [PubMed]
5. Qian, Y.; Huang, S.-S.C. Improving plant gene regulatory network inference by integrative analysis of multi-omics and high resolution data sets. *Curr. Opin. Syst. Biol.* **2022**, *22*, 8–15. [CrossRef]
6. Herrgård, M.J.; Swainston, N.; Dobson, P.; Dunn, W.B.; Arga, K.Y.; Arvas, M.; Blüthgen, N.; Borger, S.; Costenoble, R.; Heinemann, M.; et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* **2008**, *26*, 1155–1160. [CrossRef]
7. Raikhel, N.V.; Coruzzi, G.M. Achieving the in silico plant. Systems biology and the future of plant biological research. *Plant Physiol.* **2003**, *132*, 404–409.
8. Santos, F.; Boele, J.; Teusink, B. A Practical Guide to Genome-Scale Metabolic Models and Their Analysis. *Methods Enzymol.* **2011**, *500*, 509–532. [CrossRef] [PubMed]
9. Feist, A.M.; Henry, C.S.; Reed, J.L.; Krummenacker, M.; Joyce, A.R.; Karp, P.D.; Broadbelt, L.J.; Hatzimanikatis, V.; Palsson, B. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **2007**, *3*, 121. [CrossRef]
10. McCloskey, D.; Palsson, B.; Feist, A.M. Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol.* **2013**, *9*, 661. [CrossRef] [PubMed]
11. Mahood, E.H.; Kruse, L.H.; Moghe, G.D. Machine learning: A powerful tool for gene function prediction in plants. *Appl. Plant Sci.* **2020**, *8*, e11376. [CrossRef] [PubMed]
12. Mahmoud, M.; Gobet, N.; Cruz-Dávalos, D.I.; Mounier, N.; Dessimoz, C.; Sedlazeck, F.J. Structural variant calling: The long and the short of it. *Genome Biol.* **2019**, *20*, 1–14. [CrossRef] [PubMed]
13. Weirauch, M.T.; Cote, A.; Norel, R.; Annala, M.; Zhao, Y.; Riley, T.R.; Saez-Rodriguez, J.; Cokelaer, T.; Vedenko, A.; Talukder, S.; et al. Evaluation of methods for modeling transcription factor sequence specificity. *Nat. Biotechnol.* **2013**, *31*, 126–134. [CrossRef]
14. Vu, T.T.D.; Jung, J. Protein function prediction with gene ontology: From traditional to deep learning models. *PeerJ* **2021**, *9*, e12019. [CrossRef]
15. Metzker, M.L. Sequencing technologies—the next generation. *Nat. Rev. Genet.* **2010**, *11*, 31–46. [CrossRef]

16. Shendure, J.; Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145. [CrossRef]
17. Singh, S.; Rao, A.; Mishra, P.; Yadav, A.K.; Maurya, R.; Kaur, S.; Tandon, G. Bioinformatics in Next-Generation Genome Se-quencing. In *Current Trends in Bioinformatics: An Insight*; Wadhwa, G., Shanmughavel, P., Singh, A., Bellare, J., Eds.; Springer: Singapore, 2018; pp. 27–38.
18. Kühner, S.; van Noort, V.; Betts, M.J.; Leo-Macias, A.; Batisse, C.; Rode, M.; Yamada, T.; Maier, T.; Bader, S.; Beltran-Alvarez, P.; et al. Proteome Organization in a Genome-Reduced Bacterium. *Science* **2009**, *326*, 1235–1240. [CrossRef]
19. Edwards, D.; Batley, J. Plant bioinformatics: From genome to phenome. *Trends Biotechnol.* **2004**, *22*, 232–237. [CrossRef] [PubMed]
20. Bolger, M.E.; Weisshaar, B.; Scholz, U.; Stein, N.; Usadel, B.; Mayer, K.F. Plant genome sequencing—Applications for crop improvement. *Curr. Opin. Biotechnol.* **2014**, *26*, 31–37. [CrossRef]
21. Cao, Y.; Fanning, S.; Proos, S.; Jordan, K.; Srikumar, S. A Review on the Applications of Next Generation Sequencing Tech-nologies as Applied to Food-Related Microbiome Studies. *Front. Microbiol.* **2017**, *8*, 1829. [CrossRef]
22. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [CrossRef] [PubMed]
23. Haynes, E.; Jimenez, E.; Pardo, M.A.; Helyar, S.J. The future of NGS (Next Generation Sequencing) analysis in testing food authenticity. *Food Control* **2019**, *101*, 134–143. [CrossRef]
24. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: http://www.bioinformatics. babraham.ac.uk/projects/fastqc (accessed on 28 July 2022).
25. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* **2011**, *17*, 10–12. [CrossRef]
26. Schubert, M.; Lindgreen, S.; Orlando, L. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **2016**, *9*, 88. [CrossRef] [PubMed]
27. Jiang, H.; Lei, R.; Ding, S.-W.; Zhu, S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinform.* **2014**, *15*, 182. [CrossRef]
28. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
29. International Wheat Genome Sequencing Consortium (IWGSC). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **2014**, *345*, 1251788. [CrossRef]
30. Jiao, W.B.; Schneeberger, K. Chromosome-level assemblies of multiple Arabidopsis genomes reveal hotspots of rear-rangements with altered evolutionary dynamics. *Nat. Commun.* **2020**, *11*, 989. [CrossRef]
31. Abril, J.F.; Castellano Hereza, S. Genome Annotation. In *Encyclopedia of Bioinformatics and Computational Biology*; Ranganathan, S., Gribskov, M., Schönbach, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 195–209.
32. Liu, Y.; Guo, J.; Hu, G.; Zhu, H. Gene prediction in metagenomic fragments based on the SVM algorithm. *BMC Bioinform.* **2013**, *14*, S12. [CrossRef]
33. Scalzitti, N.; Jeannin-Girardon, A.; Collet, P.; Poch, O.; Thompson, J.D. A benchmark study of ab initio gene prediction methods in diverse eukaryotic organisms. *BMC Genom.* **2020**, *21*, 293. [CrossRef]
34. Wang, Z.; Chen, Y.; Li, Y. A Brief Review of Computational Gene Prediction Methods. *Genom. Proteom. Bioinform.* **2004**, *2*, 216–221. [CrossRef]
35. Huang, Y.; Chen, S.-Y.; Deng, F. Well-characterized sequence features of eukaryote genomes and implications for ab initio gene prediction. *Comput. Struct. Biotechnol. J.* **2016**, *14*, 298–303. [CrossRef] [PubMed]
36. Pati, A.; Ivanova, N.N.; Mikhailova, N.; Ovchinnikova, G.; Hooper, S.D.; Lykidis, A.; Kyrpides, N.C. GenePRIMP: A gene prediction improvement pipeline for prokaryotic genomes. *Nat. Chem. Biol.* **2010**, *7*, 455–457. [CrossRef] [PubMed]
37. Reid, I.; O'Toole, N.; Zabaneh, O.; Nourzadeh, R.; Dahdouli, M.; Abdellateef, M.; Gordon, P.M.; Soh, J.; Butler, G.; Sensen, C.W.; et al. SnowyOwl: Accurate prediction of fungal genes by using RNA-Seq and homology information to select among ab initio models. *BMC Bioinform.* **2014**, *15*, 229. [CrossRef]
38. Testa, A.C.; Hane, J.K.; Ellwood, S.R.; Oliver, R.P. CodingQuarry: Highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genom.* **2015**, *16*, 170. [CrossRef]
39. Hoff, K.J.; Lange, S.; Lomsadze, A.; Borodovsky, M.; Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **2016**, *32*, 767–769. [CrossRef]
40. Holt, C.; Yandell, M. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **2011**, *12*, 491. [CrossRef]
41. Campbell, M.S.; Law, M.; Holt, C.; Stein, J.C.; Moghe, G.; Hufnagel, D.; Lei, J.; Achawanantakun, R.; Jiao, D.; Lawrence, C.J.; et al. MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. *Plant Physiol.* **2013**, *164*, 513–524. [CrossRef]
42. Chan, K.-L.; Rosli, R.; Tatarinova, T.V.; Hogan, M.; Firdaus-Raih, M.; Low, E.-T.L. Seqping: Gene prediction pipeline for plant genomes using self-training gene models and transcriptomic data. *BMC Bioinform.* **2017**, *18*, 1–7. [CrossRef]
43. Liang, C.; Mao, L.; Ware, D.; Stein, L. Evidence-based gene predictions in plant genomes. *Genome Res.* **2009**, *19*, 1912–1923. [CrossRef]
44. Flicek, P. Gene prediction: Compare and CONTRAST. *Genome Biol.* **2007**, *8*, 233. [CrossRef] [PubMed]
45. Van Baren, M.J.; Koebbe, B.C.; Brent, M.R. Using N-SCAN or TWINSCAN to predict gene structures in genomic DNA se-quences. *Curr. Protoc. Bioinform.* **2007**, *20*, 4–8. [CrossRef] [PubMed]

46. Richmond, T. Identification of complete gene structures in genomic DNA. *Genome Biol.* **2000**, *1*, reports222. [CrossRef]
47. Seaver, S.M.D.; Henry, C.S.; Hanson, A.D. Frontiers in metabolic reconstruction and modeling of plant genomes. *J. Exp. Bot.* **2012**, *63*, 2247–2258. [CrossRef]
48. Osterman, A.; Overbeek, R. Missing genes in metabolic pathways: A comparative genomics approach. *Curr. Opin. Chem. Biol.* **2003**, *7*, 238–251. [CrossRef]
49. De Oliveira Dal'Molin, C.G.; Nielsen, L.K. Plant genome-scale metabolic reconstruction and modelling. *Curr. Opin. Biotechnol.* **2013**, *24*, 271–277. [CrossRef]
50. Pont, C.; Wagner, S.; Kremer, A.; Orlando, L.; Plomion, C.; Salse, J. Paleogenomics: Reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biol.* **2019**, *20*, 29. [CrossRef]
51. Rai, A.; Saito, K. Omics data input for metabolic modeling. *Curr. Opin. Biotechnol.* **2016**, *37*, 127–134. [CrossRef]
52. Pruitt, K.D.; Tatusova, T.; Maglott, D.R. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **2005**, *33*, D501–D504. [CrossRef]
53. Bolser, D.; Staines, D.M.; Pritchard, E.; Kersey, P. Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. *Methods Mol. Biol.* **2016**, *1374*, 115–140. [CrossRef]
54. Dong, Q.; Schlueter, S.D.; Brendel, V. PlantGDB, plant genome database and analysis tools. *Nucleic Acids Res.* **2004**, *32*, 354D–359D. [CrossRef] [PubMed]
55. Proost, S.; Van Bel, M.; Vaneechoutte, D.; Van De Peer, Y.; Inzé, D.; Mueller-Roeber, B.; Vandepoele, K. PLAZA 3.0: An access point for plant comparative genomics. *Nucleic Acids Res.* **2015**, *43*, D974–D981. [CrossRef] [PubMed]
56. Liang, C.; Jaiswal, P.; Hebbard, C.; Avraham, S.; Buckler, E.S.; Casstevens, T.; Hurwitz, B.; McCouch, S.; Ni, J.; Pujar, A.; et al. Gramene: A growing plant comparative genomics resource. *Nucleic Acids Res.* **2006**, *36*, D947–D953. [CrossRef]
57. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [CrossRef] [PubMed]
58. Varshney, R.K.; Sinha, P.; Singh, V.K.; Kumar, A.; Zhang, Q.; Bennetzen, J.L. 5Gs for crop genetic improvement. *Curr. Opin. Plant Biol.* **2020**, *56*, 190–196. [CrossRef]
59. Varshney, R.K.; Pandey, M.K.; Bohra, A.; Singh, V.K.; Thudi, M.; Saxena, R.K. Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor. Appl. Genet.* **2019**, *132*, 797–816. [CrossRef]
60. Varshney, R.K.; Bohra, A.; Roorkiwal, M.; Barmukh, R.; Cowling, W.A.; Chitikineni, A.; Lam, H.-M.; Hickey, L.T.; Croser, J.S.; Bayer, P.E.; et al. Fast-forward breeding for a food-secure world. *Trends Genet.* **2021**, *37*, 1124–1136. [CrossRef]
61. Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; de los Campos, G.; Burgueño, J.; González-Camacho, J.M.; Pérez-Elizalde, S.; Beyene, Y.; et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Sci.* **2017**, *22*, 961–975. [CrossRef]
62. Kundaje, A.; Meuleman, W.; Ernst, J.; Bilenky, M.; Yen, A.; Heravi-Moussavi, A.; Kheradpour, P.; Zhang, Z.; Wang, J.; Ziller, M.J.; et al. Integrative analysis of 111 reference human epigenomes. *Nature* **2015**, *518*, 317–330. [CrossRef]
63. Minnoye, L.; Marinov, G.K.; Krausgruber, T.; Pan, L.; Marand, A.P.; Secchia, S.; Greenleaf, W.J.; Furlong, E.E.M.; Zhao, K.; Schmitz, R.J.; et al. Chromatin accessibility profiling methods. *Nat. Rev. Methods Prim.* **2021**, *1*, 58. [CrossRef]
64. Moore, J.E.; Purcaro, M.J.; Pratt, H.E.; Epstein, C.B.; Shoresh, N.; Adrian, J.; Kawli, T.; Davis, C.A.; Dobin, A.; Kaul, R.; et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **2020**, *583*, 699–710. [CrossRef]
65. Tu, X.; Marand, A.P.; Schmitz, R.J.; Zhong, S. A combinatorial indexing strategy for low-cost epigenomic profiling of plant single cells. *Plant Comm.* **2022**, *3*, 100308. [CrossRef] [PubMed]
66. Rodriguez-Villalon, A.; Brady, S.M. Single cell RNA sequencing and its promise in reconstructing plant vascular cell lineages. *Curr. Opin. Plant Biol.* **2019**, *48*, 47–56. [CrossRef] [PubMed]
67. Rhee, S.Y.; Birnbaum, K.D.; Ehrhardt, D.W. Towards building a plant cell atlas. *Trends Plant Sci.* **2019**, *24*, 303–310. [CrossRef] [PubMed]
68. Denyer, T.; Timmermans, M.C. Crafting a blueprint for single-cell RNA sequencing. *Trends Plant Sci.* **2022**, *27*, 92–103. [CrossRef]
69. Giacomello, S. A new era for plant science: Spatial single-cell transcriptomics. *Curr. Opin. Plant Biol.* **2021**, *60*, 102041. [CrossRef]
70. Li, X.; Zhang, X.; Gao, S.; Cui, F.; Chen, W.; Fan, L.; Qi, Y. Single-cell RNA sequencing reveals the landscape of maize root tips and assists in identification of cell type-specific nitrate-response genes. *Crop J.* **2022**. [CrossRef]
71. He, T.; Li, C. Harness the power of genomic selection and the potential of germplasm in crop breeding for global food security in the era with rapid climate change. *Crop J.* **2020**, *8*, 688–700. [CrossRef]
72. Cuevas, J.; Crossa, J.; Montesinos-López, O.A.; Burgueño, J.; Pérez-Rodríguez, P.; de Los Campos, G. Bayesian Genomic Pre-diction with Genotype × Environment Interaction Kernel Models. *G3 Genes Genomes Genet.* **2017**, *7*, 41–53.
73. Mulesa, T.H.; Westengen, O.T. Against the grain? A historical institutional analysis of access governance of plant genetic resources for food and agriculture in Ethiopia. *J. World Intellect. Prop.* **2020**, *23*, 82–120. [CrossRef]
74. Yugander, A.; Sundaram, R.M.; Singh, K.; Ladhalakshmi, D.; Rao, L.V.S.; Madhav, M.S.; Badri, J.; Prasad, M.S.; Laha, G.S. Incorporation of the novel bacterial blight resistance gene Xa38 into the genetic background of elite rice variety Improved Samba Mahsuri. *PLoS ONE* **2018**, *13*, e0198260. [CrossRef] [PubMed]
75. Ratna Madhavi, K.; Rambabu, R.; Abhilash Kumar, V.; Vijay Kumar, S.; Aruna, J.; Ramesh, S.; Sundaram, R.M.; Laha, G.S.; Sheshu Madhav, M.; Prasad, M.S. Marker assisted introgression of blast (Pi-2 and Pi-54) genes into the genetic background of elite, bacterial blight resistant indica rice variety, Improved Samba Mahsuri. *Euphytica* **2016**, *212*, 331–342. [CrossRef]

76. Huang, C.; Chen, Z.; Liang, C. Oryza pan-genomics: A new foundation for future rice research and improvement. *Crop J.* **2021**, *9*, 622–632. [CrossRef]

77. Li, H.; Wang, S.; Chai, S.; Yang, Z.; Zhang, Q.; Xin, H.; Xu, Y.; Lin, S.; Chen, X.; Yao, Z.; et al. Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nat. Commun.* **2022**, *13*, 682. [CrossRef] [PubMed]

78. Li, W.; Liu, J.; Zhang, H.; Liu, Z.; Wang, Y.; Xing, L.; He, Q.; Du, H. Plant pan-genomics: Recent advances, new challenges, and roads ahead. *J. Genet. Genom.* **2022**. [CrossRef]

79. Lowe, R.; Shirley, N.; Bleackley, M.; Dolan, S.; Shafee, T. Transcriptomics technologies. *PLoS Comput. Biol.* **2017**, *13*, e1005457. [CrossRef]

80. Govindarajan, R.; Duraiyan, J.; Kaliyappan, K.; Palanisamy, M. Microarray and its applications. *J. Pharm. Bioallied. Sci.* **2012**, *4*, S310–S312.

81. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [CrossRef]

82. Zhao, S.; Fung-Leung, W.-P.; Bittner, A.; Ngo, K.; Liu, X. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS ONE* **2014**, *9*, e78644. [CrossRef] [PubMed]

83. Wilhelm, B.T.; Landry, J.-R. RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **2009**, *48*, 249–257. [CrossRef]

84. Costa-Silva, J.; Domingues, D.; Martins Lopes, F. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* **2017**, *12*, e0190152. [CrossRef]

85. Yeung, K.Y.; Medvedovic, M.; Bumgarner, R.E. From co-expression to co-regulation: How many microarray experiments do we need? *Genome Biol.* **2004**, *5*, R48. [CrossRef] [PubMed]

86. Ederli, L.; Dawe, A.; Pasqualini, S.; Quaglia, M.; Xiong, L.; Gehring, C. Arabidopsis flower specific defense gene expression patterns affect resistance to pathogens. *Front. Plant Sci.* **2015**, *6*, 79. [CrossRef] [PubMed]

87. Inoue, M.; Horimoto, K. Relationship between regulatory pattern of gene expression level and gene function. *PLoS ONE* **2017**, *12*, e0177430. [CrossRef] [PubMed]

88. You, Q.; Zhang, L.; Yi, X.; Zhang, K.; Yao, D.; Zhang, X.; Wang, Q.; Zhao, X.; Ling, Y.; Xu, W.; et al. Co-expression network analyses identify functional modules associated with development and stress response in Gossypium arboreum. *Sci. Rep.* **2016**, *6*, 38436. [CrossRef]

89. Costa, M.C.D.; Righetti, K.; Nijveen, H.; Yazdanpanah, F.; Ligterink, W.; Buitink, J.; Hilhorst, H.W.M. A gene co-expression network predicts functional genes controlling the re-establishment of desiccation tolerance in germinated Arabidopsis thaliana seeds. *Planta* **2015**, *242*, 435–449. [CrossRef]

90. Ruprecht, C.; Mutwil, M.; Saxe, F.; Eder, M.; Nikoloski, Z.; Persson, S. Large-Scale Co-Expression Approach to Dissect Secondary Cell Wall Formation Across Plant Species. *Front. Plant Sci.* **2011**, *2*, 23. [CrossRef]

91. Wang, S.; Yin, Y.; Ma, Q.; Tang, X.; Hao, D.; Xu, Y. Genome-scale identification of cell-wall related genes in Arabidopsis based on co-expression network analysis. *BMC Plant Biol.* **2012**, *12*, 138. [CrossRef] [PubMed]

92. Barrett, T.; Edgar, R. Gene Expression Omnibus: Microarray Data Storage, Submission, Retrieval, and Analysis. *Methods Enzymol.* **2006**, *411*, 352–369. [CrossRef] [PubMed]

93. Huala, E.; Dickerman, A.W.; Garcia-Hernandez, M.; Weems, D.; Reiser, L.; LaFond, F.; Hanley, D.; Kiphart, D.; Zhuang, M.; Huang, W.; et al. The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information re-trieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* **2001**, *29*, 102–105. [CrossRef] [PubMed]

94. Klepikova, A.V.; Kulakovskiy, I.V.; Kasianov, A.S.; Logacheva, M.D.; Penin, A.A. An update to database TraVA: Organ-specific cold stress response in Arabidopsis thaliana. *BMC Plant Biol.* **2019**, *19*, 29–40. [CrossRef] [PubMed]

95. Sato, Y.; Antonio, B.A.; Namiki, N.; Takehisa, H.; Minami, H.; Kamatsuki, K.; Sugimoto, K.; Shimizu, Y.; Hirochika, H.; Nagamura, Y. RiceXPro: A platform for monitoring gene expression in japonica rice grown under natural field conditions. *Nucleic Acids Res.* **2011**, *39*, D1141–D1148. [CrossRef] [PubMed]

96. Kawahara, Y.; Oono, Y.; Wakimoto, H.; Ogata, J.; Kanamori, H.; Sasaki, H.; Mori, S.; Matsumoto, T.; Itoh, T. TENOR: Database for Comprehensive mRNA-Seq Experiments in Rice. *Plant Cell Physiol.* **2016**, *57*, e7. [CrossRef]

97. Tanaka, T.; Sakai, H.; Fujii, N.; Kobayashi, F.; Nakamura, S.; Itoh, T.; Matsumoto, T.; Wu, J. bex-db: Bioinformatics workbench for comprehensive analysis of barley-expressed genes. *Breed. Sci.* **2013**, *63*, 430–434. [CrossRef]

98. Li, J.-R.; Liu, C.-C.; Sun, C.-H.; Chen, Y.-T. Plant stress RNA-seq Nexus: A stress-specific transcriptome database in plant cells. *BMC Genom.* **2018**, *19*, 966. [CrossRef]

99. Saithong, T.; Rongsirikul, O.; Kalapanulak, S.; Chiewchankaset, P.; Siriwat, W.; Netrphan, S.; Suksangpanomrung, M.; Meechai, A.; Cheevadhanarak, S. Starch biosynthesis in cassava: A genome-based pathway reconstruction and its exploitation in data integration. *BMC Syst. Biol.* **2013**, *7*, 75. [CrossRef]

100. Töpfer, N.; Caldana, C.; Grimbs, S.; Willmitzer, L.; Fernie, A.R.; Nikoloski, Z. Integration of Genome-Scale Modeling and Transcript Profiling Reveals Metabolic Pathways Underlying Light and Temperature Acclimation in *Arabidopsis. Plant Cell* **2013**, *25*, 1197–1211. [CrossRef]

101. Dharmawardhana, P.; Ren, L.; Amarasinghe, V.; Monaco, M.; Thomason, J.; Ravenscroft, D.; McCouch, S.; Ware, D.; Jaiswal, P. A genome scale metabolic network for rice and accompanying analysis of tryptophan, auxin and serotonin biosynthesis regulation under biotic stress. *Rice* **2013**, *6*, 15. [CrossRef]

102. Assefa, T.; Otyama, P.I.; Brown, A.V.; Kalberer, S.R.; Kulkarni, R.S.; Cannon, S.B. Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. *BMC Genom.* **2019**, *20*, 527. [CrossRef]

103. Weiße, A.Y.; Oyarzún, D.A.; Danos, V.; Swain, P.S. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E1038–E1047. [CrossRef]

104. Nakaya, A.; Ichihara, H.; Asamizu, E.; Shirasawa, S.; Nakamura, Y.; Tabata, S.; Hirakawa, H. Plant Genome DataBase Japan (PGDBj). *Methods Mol. Biol.* **2017**, *1533*, 45–77. [CrossRef] [PubMed]

105. Spannagl, M.; Nussbaumer, T.; Bader, K.C.; Martis, M.M.; Seidel, M.; Kugler, K.G.; Gundlach, H.; Mayer, K.F. PGSB PlantsDB: Updates to the database framework for comparative plant genome research. *Nucleic Acids Res.* **2016**, *44*, D1141–D1147. [CrossRef]

106. Cui, L.; Veeraraghavan, N.; Richter, A.; Wall, K.; Jansen, R.K.; Leebens-Mack, J.; Makalowska, I.; dePamphilis, C.W. Chloro-plastDB: The Chloroplast Genome Database. *Nucleic Acids Res.* **2006**, *34*, D692–D696. [CrossRef] [PubMed]

107. Hirsch, C.; Hamilton, J.; Childs, K.; Cepela, J.; Crisovan, E.; Vaillancourt, B.; Hirsch, C.N.; Habermann, M.; Neal, B.; Buell, C.R. Spud DB: A Resource for Mining Sequences, Genotypes, and Phenotypes to Accelerate Potato Breeding. *Plant Genome* **2014**, *7*, plantgenome2013-10. [CrossRef]

108. Ruggieri, V.; Alexiou, K.; Morata, J.; Argyris, J.; Pujol, M.; Yano, R.; Nonaka, S.; Ezura, H.; Latrasse, D.; Boualem, A.; et al. An improved assembly and annotation of the melon (*Cucumis melo* L.) reference genome. *Sci. Rep.* **2018**, *8*, 8088. [CrossRef] [PubMed]

109. Portwood, J.L., II; Woodhouse, M.R.; Cannon, E.K.; Gardiner, J.M.; Harper, L.C.; Schaeffer, M.L.; Walsh, J.R.; Sen, T.Z.; Cho, K.T.; Schott, D.A.; et al. MaizeGDB 2018: The maize multi-genome genetics and genomics database. *Nucleic Acids Res.* **2019**, *47*, D1146–D1154. [CrossRef] [PubMed]

110. Sakai, H.; Lee, S.S.; Tanaka, T.; Numa, H.; Kim, J.; Kawahara, Y.; Wakimoto, H.; Yang, C.-C.; Iwamoto, M.; Abe, T.; et al. Rice Annotation Project Database (RAP-DB): An Integrative and Interactive Database for Rice Genomics. *Plant Cell Physiol.* **2013**, *54*, e6. [CrossRef]

111. Kawahara, Y.; de la Bastide, M.; Hamilton, J.P.; Kanamori, H.; McCombie, W.R.; Ouyang, S.; Schwartz, D.C.; Tanaka, T.; Wu, J.; Zhou, S.; et al. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 4. [CrossRef]

112. Yao, E.; Blake, V.C.; Cooper, L.; Wight, C.P.; Michel, S.; Cagirici, H.B.; Lazo, G.R.; Birkett, C.L.; Waring, D.J.; Jannink, J.-L.; et al. GrainGenes: A data-rich repository for small grains genetics and genomics. *Database* **2022**, *2022*, baac034. [CrossRef]

113. Brown, A.V.; Conners, S.I.; Huang, W.; Wilkey, A.P.; Grant, D.; Weeks, N.T.; Cannon, S.B.; Graham, M.A.; Nelson, R.T. A new decade and new data at SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* **2021**, *49*, D1496–D1501. [CrossRef]

114. Jung, S.; Staton, M.; Lee, T.; Blenda, A.; Svancara, R.; Abbott, A.; Main, D. GDR (Genome Database for Rosaceae): Integrated web-database for Rosaceae genomics and genetics data. *Nucleic Acids Res.* **2008**, *36*, D1034–D1040. [CrossRef] [PubMed]

115. Chen, H.; Wang, T.; He, X.; Cai, X.; Lin, R.; Liang, J.; Wu, J.; King, G.; Wang, X. BRAD V3.0: An upgraded Brassicaceae database. *Nucleic Acids Res.* **2022**, *50*, D1432–D1441. [CrossRef] [PubMed]

116. Robinson, A.J.; Tamiru, M.; Salby, R.; Bolitho, C.; Williams, A.; Huggard, S.; Fisch, E.; Unsworth, K.; Whelan, J.; Lewsey, M.G. AgriSeqDB: An online RNA-Seq database for functional studies of agriculturally relevant plant species. *BMC Plant Biol.* **2018**, *18*, 200. [CrossRef] [PubMed]

117. Waese, J.; Provart, N.J. The Bio-Analytic Resource for Plant Biology. In *Plant Genomics Databases. Methods in Molecular Biology*; Van Dijk, A., Ed.; Humana Press: New York, NJ, USA, 2017; Volume 1533, pp. 119–148. [CrossRef]

118. Zhang, Z.; Yu, J.; Li, D.; Zhang, Z.; Liu, F.; Zhou, X.; Wang, T.; Ling, Y.; Su, Z. PMRD: Plant microRNA database. *Nucleic Acids Res.* **2010**, *38*, D806–D813. [CrossRef]

119. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [CrossRef]

120. Xenarios, I.; Rice, D.W.; Salwinski, L.; Baron, M.K.; Marcotte, E.M.; Eisenberg, D. DIP: The database of interacting proteins. *Nucleic Acids Res.* **2000**, *28*, 289–291. [CrossRef]

121. Zhu, G.; Wu, A.; Xu, X.-J.; Xiao, P.-P.; Lu, L.; Liu, J.; Cao, Y.; Chen, L.; Wu, J.; Zhao, X.-M. PPIM: A Protein-Protein Interaction Database for Maize. *Plant Physiol.* **2016**, *170*, 618–626. [CrossRef]

122. del Toro, N.; Shrivastava, A.; Ragueneau, E.; Meldal, B.; Combe, C.; Barrera, E.; Perfetto, L.; How, K.; Ratan, P.; Shirodkar, G.; et al. The IntAct database: Efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* **2022**, *50*, D648–D653. [CrossRef]

123. Gu, H.; Zhu, P.; Jiao, Y.; Meng, Y.; Chen, M. PRIN: A predicted rice interactome network. *BMC Bioinform.* **2011**, *12*, 161. [CrossRef]

124. Bader, G.; Donaldson, I.; Wolting, C.; Ouellette, B.F.F.; Pawson, T.; Hogue, C.W.V. BIND–The Biomolecular Interaction Network Database. *Nucleic Acids Res.* **2003**, *31*, 248–250. [CrossRef]

125. Chatr-Aryamontri, A.; Breitkreutz, B.-J.; Oughtred, R.; Boucher, L.; Heinicke, S.; Chen, D.; Stark, C.; Breitkreutz, A.; Kolas, N.; O'Donnell, L.; et al. The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* **2014**, *43*, D470–D478. [CrossRef] [PubMed]

126. Brandão, M.M.; Dantas, L.L.; Silva-Filho, M.C. AtPIN: Arabidopsis thaliana Protein Interaction Network. *BMC Bioinform.* **2009**, *10*, 454. [CrossRef] [PubMed]

127. Yang, X.; Yang, S.; Qi, H.; Wang, T.; Li, H.; Zhang, Z. PlaPPISite: A comprehensive resource for plant protein-protein interaction sites. *BMC Plant Biol.* **2020**, *20*, 61. [CrossRef]

128. Stein, A.; Russell, R.B.; Aloy, P. 3did: Interacting protein domains of known three-dimensional structure. *Nucleic Acids Res.* **2005**, *33*, D413–D417. [CrossRef] [PubMed]

129. Chatr-Aryamontri, A.; Ceol, A.; Palazzi, L.M.; Nardelli, G.; Schneider, M.V.; Castagnoli, L.; Cesareni, G. MINT: The Molecular INTeraction database. *Nucleic Acids Res.* **2007**, *35*, D572–D574. [CrossRef] [PubMed]

130. Aoki, Y.; Okamura, Y.; Tadaka, S.; Kinoshita, K.; Obayashi, T. ATTED-II in 2016: A Plant Coexpression Database Towards Lineage-Specific Coexpression. *Plant Cell Physiol.* **2016**, *57*, e5. [CrossRef] [PubMed]

131. Srinivasasainagendra, V.; Page, G.P.; Mehta, T.; Coulibaly, I.; Loraine, A.E. CressExpress: A Tool for Large-Scale Mining of Expression Data from Arabidopsis. *Plant Physiol.* **2008**, *147*, 1004–1016. [CrossRef]

132. Lee, T.; Yang, S.; Kim, E.; Ko, Y.; Hwang, S.; Shin, J.; Shim, J.E.; Shim, H.; Kim, H.; Kim, C.; et al. AraNet v2: An improved database of co-functional gene networks for the study of Arabidopsis thaliana and 27 other nonmodel plant species. *Nucleic Acids Res.* **2015**, *43*, D996–D1002. [CrossRef]

133. Ogata, Y.; Suzuki, H.; Sakurai, N.; Shibata, D. CoP: A database for characterizing co-expressed gene modules with biological information in plants. *Bioinformatics* **2010**, *26*, 1267–1268. [CrossRef]

134. Chien, C.-H.; Chow, C.-N.; Wu, N.-Y.; Chiang-Hsieh, Y.-F.; Hou, P.-F.; Chang, W.-C. EXPath: A database of comparative expression analysis inferring metabolic pathways for plants. *BMC Genom.* **2015**, *16*, S6. [CrossRef]

135. Ohyanagi, H.; Takano, T.; Terashima, S.; Kobayashi, M.; Kanno, M.; Morimoto, K.; Kanegae, H.; Sasaki, Y.; Saito, M.; Asano, S.; et al. Plant Omics Data Center: An integrated web repository for interspecies gene expression networks with NLP-based cu-ration. *Plant Cell Physiol.* **2015**, *56*, e9. [CrossRef]

136. Mutwil, M.; Klie, S.; Tohge, T.; Giorgi, F.; Wilkins, O.; Campbell, M.; Fernie, A.R.; Usadel, B.; Nikoloski, Z.; Persson, S. PlaNet: Combined Sequence and Expression Comparisons across Plant Networks Derived from Seven Species. *Plant Cell* **2011**, *23*, 895–910. [CrossRef] [PubMed]

137. Hamada, K.; Hongo, K.; Suwabe, K.; Shimizu, A.; Nagayama, T.; Abe, R.; Kikuchi, S.; Yamamoto, N.; Fujii, T.; Yokoyama, K.; et al. OryzaExpress: An Integrated Database of Gene Expression Networks and Omics Annotations in Rice. *Plant Cell Physiol.* **2011**, *52*, 220–229. [CrossRef]

138. Kudo, T.; Terashima, S.; Takaki, Y.; Tomita, K.; Saito, M.; Kanno, M.; Yokoyama, K.; Yano, K. PlantExpress: A Database Inte-grating OryzaExpress and ArthaExpress for Single-species and Cross-species Gene Expression Network Analyses with Mi-croarray-Based Transcriptome Data. *Plant Cell Physiol.* **2017**, *58*, e1. [CrossRef] [PubMed]

139. Sato, Y.; Namiki, N.; Takehisa, H.; Kamatsuki, K.; Minami, H.; Ikawa, H.; Ohyanagi, H.; Sugimoto, K.; Itoh, J.-I.; Antonio, B.A.; et al. RiceFREND: A platform for retrieving coexpressed gene networks in rice. *Nucleic Acids Res.* **2013**, *41*, D1214–D1221. [CrossRef] [PubMed]

140. Wong, D.C.; Sweetman, C.; Drew, D.P.; Ford, C.M. VTCdb: A gene co-expression database for the crop species *Vitis vinifera* (grapevine). *BMC Genom.* **2013**, *14*, 882. [CrossRef] [PubMed]

141. Warde-Farley, D.; Donaldson, S.L.; Comes, O.; Zuberi, K.; Badrawi, R.; Chao, P.; Franz, M.; Grouios, C.; Kazi, F.; Lopes, C.T.; et al. The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **2010**, *38*, W214–W220. [CrossRef]

142. Steinhauser, D.; Usadel, B.; Luedemann, A.; Thimm, O.; Kopka, J. CSB.DB: A comprehensive systems-biology database. *Bioinfor-matics* **2004**, *20*, 3647–3651. [CrossRef] [PubMed]

143. Kim, J.; Jun, K.M.; Kim, J.S.; Chae, S.; Pahk, Y.-M.; Lee, T.-H.; Sohn, S.-I.; Lee, S.I.; Lim, M.-H.; Kim, C.-K.; et al. RapaNet: A Web Tool for the Co-Expression Analysis of *Brassica rapa* Genes. *Evol. Bioinform.* **2017**, *13*. [CrossRef] [PubMed]

144. Xia, L.; Zou, D.; Sang, J.; Xu, X.; Yin, H.; Li, M.; Wu, S.; Hu, S.; Hao, L.; Zhang, Z. Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. *J. Genet. Genom.* **2017**, *44*, 235–241. [CrossRef]

145. Ferrari, C.; Proost, S.; Ruprecht, C.; Mutwil, M. PhytoNet: Comparative co-expression network analyses across phytoplankton and land plants. *Nucleic Acids Res.* **2018**, *46*, W76–W83. [CrossRef] [PubMed]

146. Proost, S.; Mutwil, M. CoNekT: An open-source framework for comparative genomic and transcriptomic network analyses. *Nucleic Acids Res.* **2018**, *46*, W133–W140. [CrossRef] [PubMed]

147. Lee, J.; Shah, M.; Ballouz, S.; Crow, M.; Gillis, J. CoCoCoNet: Conserved and comparative co-expression across a diverse set of species. *Nucleic Acids Res.* **2020**, *48*, W566–W571. [CrossRef]

148. Tang, S.; Zhao, H.; Lu, S.; Yu, L.; Zhang, G.; Zhang, Y.; Yang, Q.Y.; Zhou, Y.; Wang, X.; Ma, W.; et al. Genome- and transcrip-tome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. *Mol. Plant* **2021**, *14*, 470–487. [CrossRef] [PubMed]

149. Gusev, A.; Ko, A.; Shi, H.; Bhatia, G.; Chung, W.; Penninx, B.W.J.H.; Jansen, R.; de Geus, E.J.C.; Boomsma, D.I.; Wright, F.A.; et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **2016**, *48*, 245–252. [CrossRef]

150. Kremling, K.A.G.; Diepenbrock, C.H.; Gore, M.A.; Buckler, E.S.; Bandillo, N.B. Transcriptome-Wide Association Supplements Genome-Wide Association in *Zea mays*. *G3 Genes Genomes Genet.* **2019**, *9*, 3023–3033. [CrossRef] [PubMed]

151. Wu, D.; Li, X.; Tanaka, R.; Wood, J.C.; Tibbs-Cortes, L.E.; Magallanes-Lundback, M.; Bornowski, N.; Hamilton, J.P.; Vaillancourt, B.; Diepenbrock, C.H.; et al. Combining GWAS and TWAS to identify candidate causal genes for tocochromanol levels in maize grain. *Genetics* **2022**, *221*. [CrossRef]

152. Perez-Sanz, F.; Navarro, P.J.; Egea-Cortines, M. Plant phenomics: An overview of image acquisition technologies and image data analysis algorithms. *GigaScience* **2017**, *6*, gix092. [CrossRef] [PubMed]

153. Houle, D.; Govindaraju, D.R.; Omholt, S. Phenomics: The next challenge. *Nat. Rev. Genet.* **2010**, *11*, 855–866. [CrossRef]
154. Lobos, G.A.; Camargo, A.V.; del Pozo, A.; Araus, J.L.; Ortiz, R.; Doonan, J.H. Editorial: Plant Phenotyping and Phenomics for Plant Breeding. *Front. Plant Sci.* **2017**, *8*, 2181. [CrossRef]
155. Zhu, X.; Maurer, H.P.; Jenz, M.; Hahn, V.; Ruckelshausen, A.; Leiser, W.L.; Würschum, T. The performance of phenomic selection depends on the genetic architecture of the target trait. *Theor. Appl. Genet.* **2021**, *135*, 653–665. [CrossRef]
156. Parmley, K.; Nagasubramanian, K.; Sarkar, S.; Ganapathysubramanian, B.; Singh, A.K. Development of Optimized Phenomic Predictors for Efficient Plant Breeding Decisions Using Phenomic-Assisted Selection in Soybean. *Plant Phenomics* **2019**, *2019*, 5809404. [CrossRef]
157. Lane, H.M.; Murray, S.C.; Montesinos-Lopez, A.; Crossa, J.; Rooney, D.K.; Barrero-Farfan, I.D.; De la Fuente, G.N.; Morgan, C.L.S. Phenomic selection and prediction of maize grain yield from near-infrered reflectance spectroscopy of kernels. *Plant Phenome. J.* **2020**, *3*, e0117737. [CrossRef]
158. Gonçalves, M.T.V.; Morota, G.; Costa, P.M.dA.; Vidigal, P.M.P.; Barbosa, M.H.P.; Peternelli, L.A. Near-infrared spec-troscopy outperforms genomics for predicting sugarcane feedstock quality traits. *PLoS ONE* **2021**, *16*, e0236853. [CrossRef]
159. Li, Z.; Sillanpää, M.J. Dynamic Quantitative Trait Locus Analysis of Plant Phenomic Data. *Trends Plant Sci.* **2015**, *20*, 822–833. [CrossRef] [PubMed]
160. Kumar, S. Epigenomics of Plant Responses to Environmental Stress. *Epigenomes* **2018**, *2*, 6. [CrossRef]
161. Arneson, A.; Haghani, A.; Thompson, M.J.; Pellegrini, M.; Bin Kwon, S.; Vu, H.; Maciejewski, E.; Yao, M.; Li, C.Z.; Lu, A.T.; et al. A mammalian methylation array for profiling methylation levels at conserved sequences. *Nat. Commun.* **2022**, *13*, 783. [CrossRef]
162. Grehl, C.; Wagner, M.; Lemnian, I.; Glaser, B.; Grosse, I. performance of mapping approaches for whole genome bisulfite sequencing data in crop plants. *Front. Plant Sci.* **2020**, *11*, 176. [CrossRef] [PubMed]
163. Kaufmann, K.; Muino, J.M.; Østerås, M.; Farinelli, L.; Krajewski, P.; Angenent, G.C. Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nat. Protoc.* **2010**, *5*, 457–472. [CrossRef]
164. Tollefsbol, T.O. Advances in epigenetic technology. *Methods Mol. Biol.* **2011**, *791*, 1–10. [CrossRef] [PubMed]
165. Bajic, M.; Maher, K.A.; Deal, R.B. Identification of Open Chromatin Regions in Plant Genomes Using ATAC-Seq. *Plant Chromatin Dyn.* **2017**, *1675*, 183–201. [CrossRef]
166. Li, B.; Carey, M.; Workman, J.L. The Role of Chromatin during Transcription. *Cell* **2007**, *128*, 707–719. [CrossRef]
167. Li, R.; Hu, F.; Li, B.; Zhang, Y.; Chen, M.; Fan, T.; Wang, T. Whole genome bisulfite sequencing methylome analysis of mulberry (*Morus alba*) reveals epigenome modifications in response to drought stress. *Sci. Rep.* **2020**, *10*, 232. [CrossRef]
168. Maher, K.A.; Bajic, M.; Kajala, K.; Reynoso, M.; Pauluzzi, G.; West, D.A.; Zumstein, K.; Woodhouse, M.; Bubb, K.L.; Dorrity, M.W.; et al. Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules. *Plant Cell* **2018**, *30*, 15–36. [CrossRef]
169. Badad, O.; Lahssassi, N.; Zaid, N.; El-Baze, A.; Zaid, Y.; Meksem, J.; Lightfoot, D.A.; Tombuloglu, H.; Zaid, E.H.; Unver, T.; et al. Genome-wide MeDIP-Seq profiling of wild and cultivated olives trees suggests DNA methylation fin-gerprint on the sensory quality of olive oil. *Plants* **2021**, *10*, 1405. [CrossRef] [PubMed]
170. Hawe, J.; Theis, F.J.; Heinig, M. Inferring Interaction Networks from Multi-Omics Data. *Front. Genet.* **2019**, *10*, 535. [CrossRef]
171. Schaefer, R.J.; Michno, J.-M.; Myers, C.L. Unraveling gene function in agricultural species using gene co-expression networks. *Biochim. Biophys. Acta Gene Regul. Mech.* **2017**, *1860*, 53–63. [CrossRef] [PubMed]
172. Provart, N.J. Correlation networks visualization. *Front. Plant Sci.* **2012**, *3*, 240. [CrossRef]
173. Chen, Y.; Yan, X.; Chen, S. Bioinformatic analysis of molecular network of glucosinolate biosynthesis. *Comput. Biol. Chem.* **2011**, *35*, 10–18. [CrossRef] [PubMed]
174. Yang, X.; Ye, C.-Y.; Bisaria, A.; Tuskan, G.A.; Kalluri, U.C. Identification of candidate genes in Arabidopsis and Populus cell wall biosynthesis using text-mining, co-expression network analysis and comparative genomics. *Plant Sci.* **2011**, *181*, 675–687. [CrossRef] [PubMed]
175. Van Verk, M.C.; Bol, J.F.; Linthorst, H.J. Prospecting for Genes involved in transcriptional regulation of plant defenses, a bioinformatics approach. *BMC Plant Biol.* **2011**, *11*, 88. [CrossRef] [PubMed]
176. Ashari, K.-S.; Abdullah-Zawawi, M.-R.; Harun, S.; Mohamed-Hussein, Z.-A. Reconstruction of the Transcriptional Regulatory Network in Arabidopsis thaliana Aliphatic Glucosinolate Biosynthetic Pathway. *Sains Malays.* **2018**, *47*, 2993–3002. [CrossRef]
177. De Las Rivas, J.; Fontanillo, C. Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. *PLoS Comput. Biol.* **2010**, *6*, e1000807. [CrossRef] [PubMed]
178. Sheth, B.P.; Thaker, V.S. Plant systems biology: Insights, advances and challenges. *Planta* **2014**, *240*, 33–54. [CrossRef] [PubMed]
179. Miller, J.; Stagljar, I. Using the Yeast Two-Hybrid System to Identify Interacting Proteins. *Methods Mol. Biol.* **2004**, *261*, 247–262. [CrossRef] [PubMed]
180. Grefen, C.; Lalonde, S.; Obrdlik, P. Split-ubiquitin system for identifying protein-protein interactions in membrane and full-length proteins. *Curr. Protoc. Neurosci.* **2007**, *41*, 5–27. [CrossRef]
181. Kerppola, T.K. Bimolecular Fluorescence Complementation (BiFC) Analysis as a Probe of Protein Interactions in Living Cells. *Annu. Rev. Biophys.* **2008**, *37*, 465–487. [CrossRef]
182. Morris, J.H.; Knudsen, G.M.; Verschueren, E.; Johnson, J.R.; Cimermancic, P.; Greninger, A.L.; Pico, A.R. Affinity purifica-tion-mass spectrometry and network analysis to understand protein-protein interactions. *Nat. Protoc.* **2014**, *9*, 2539–2554. [CrossRef]

183. Paul, C.; Rho, H.; Neiswinger, J.; Zhu, H. Characterization of Protein–Protein Interactions Using Protein Microarrays. *Cold Spring Harb. Protoc.* **2016**, *2016*, prot087965. [CrossRef]

184. Harun, S.; Rohani, E.R.; Ohme-Takagi, M.; Goh, H.-H.; Mohamed-Hussein, Z.-A. ADAP is a possible negative regulator of glucosinolate biosynthesis in Arabidopsis thaliana based on clustering and gene expression analyses. *J. Plant Res.* **2021**, *134*, 327–339. [CrossRef]

185. Liu, L.-Y.D.; Hsiao, Y.-C.; Chen, H.-C.; Yang, Y.-W.; Chang, M.-C. Construction of gene causal regulatory networks using microarray data with the coefficient of intrinsic dependence. *Bot. Stud.* **2019**, *60*, 22. [CrossRef] [PubMed]

186. Mounet, F.; Moing, A.; Garcia, V.; Petit, J.; Maucourt, M.; Deborde, C.; Bernillon, S.; Le Gall, G.; Colquhoun, I.; Defernez, M.; et al. Gene and Metabolite Regulatory Network Analysis of Early Developing Fruit Tissues Highlights New Candidate Genes for the Control of Tomato Fruit Composition and Development. *Plant Physiol.* **2009**, *149*, 1505–1528. [CrossRef] [PubMed]

187. Cox, J.; Mann, M. Is proteomics the new genomics? *Cell* **2007**, *130*, 395–398. [CrossRef]

188. Aslam, B.; Basit, M.; Nisar, M.A.; Khurshid, M.; Rasool, M.H. Proteomics: Technologies and Their Applications. *J. Chromatogr. Sci.* **2017**, *55*, 182–196. [CrossRef] [PubMed]

189. Gedeon, T.; Bokes, P. Delayed Protein Synthesis Reduces the Correlation between mRNA and Protein Fluctuations. *Biophys. J.* **2012**, *103*, 377–385. [CrossRef] [PubMed]

190. Riba, A.; Di Nanni, N.; Mittal, N.; Arhné, E.; Schmidt, A.; Zavolan, M. Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 15023–15032. [CrossRef] [PubMed]

191. Neymotin, B.; Ettorre, V.; Gresham, D. Multiple Transcript Properties Related to Translation Affect mRNA Degradation Rates in *Saccharomyces cerevisiae*. *G3 Genes Genomes Genet.* **2016**, *6*, 3475–3483. [CrossRef]

192. Ponnala, L.; Wang, Y.; Sun, Q.; Van Wijk, K.J. Correlation of mRNA and protein abundance in the developing maize leaf. *Plant J.* **2014**, *78*, 424–440. [CrossRef]

193. Osorio, S.; Alba, R.; Damasceno, C.M.; Lopez-Casado, G.; Lohse, M.; Zanor, M.I.; Tohge, T.; Usadel, B.; Rose, J.K.; Fei, Z.; et al. Systems biology of tomato fruit development: Combined transcript, protein, and metabolite analysis of tomato transcription factor (nor, rin) and ethylene receptor (Nr) mutants reveals novel regulatory interactions. *Plant Physiol.* **2011**, *157*, 405–425. [CrossRef]

194. Peng, Z.; He, S.; Gong, W.; Xu, F.; Pan, Z.; Jia, Y.; Geng, X.; Du, X. Integration of proteomic and transcriptomic profiles reveals multiple levels of genetic regulation of salt tolerance in cotton. *BMC Plant Biol.* **2018**, *18*, 128. [CrossRef]

195. Syed, N.H.; Kalyna, M.; Marquez, Y.; Barta, A.; Brown, J.W. Alternative splicing in plants–coming of age. *Trends Plant Sci.* **2012**, *17*, 616–623. [CrossRef] [PubMed]

196. Rhee, S.Y.; Mutwil, M. Towards revealing the functions of all genes in plants. *Trends Plant Sci.* **2014**, *19*, 212–221. [CrossRef] [PubMed]

197. Lamesch, P.; Berardini, T.; Li, D.; Swarbreck, D.; Wilks, C.; Sasidharan, R.; Muller, R.; Dreher, K.; Alexander, D.L.; Garcia-Hernandez, M.; et al. The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **2012**, *40*, D1202–D1210. [CrossRef]

198. Saito, K.; Hirai, M.Y.; Yonekura-Sakakibara, K. Decoding genes with coexpression networks and metabolomics—'Majority report by precogs'. *Trends Plant Sci.* **2008**, *13*, 36–43. [CrossRef] [PubMed]

199. Provart, N.J.; Alonso, J.; Assmann, S.M.; Bergmann, D.; Brady, S.; Brkljacic, J.; Browse, J.; Chapple, C.; Colot, V.; Cutler, S.R.; et al. 50 years of Arabidopsis research: Highlights and future directions. *New Phytol.* **2015**, *209*, 921–944. [CrossRef] [PubMed]

200. Tan, D.-X.; Reiter, R.J. An evolutionary view of melatonin synthesis and metabolism related to its biological functions in plants. *J. Exp. Bot.* **2020**, *71*, 4677–4689. [CrossRef] [PubMed]

201. Schnoes, A.M.; Brown, S.D.; Dodevski, I.; Babbitt, P.C. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLoS Comput. Biol.* **2009**, *5*, e1000605. [CrossRef]

202. Galperin, M.Y.; Wolf, Y.I.; Makarova, K.S.; Alvarez, R.V.; Landsman, D.; Koonin, E.V. COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **2021**, *49*, D274–D281. [CrossRef]

203. Lopez, R.; Silventoinen, V.; Robinson, S.; Kibria, A.; Gish, W. WU-Blast2 server at the European Bioinformatics Institute. *Nucleic Acids Res.* **2003**, *31*, 3795–3798. [CrossRef]

204. Medema, M.H.; Osbourn, A. Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways. *Nat. Prod. Rep.* **2016**, *33*, 951–962. [CrossRef]

205. Oliver, S.G. Guilt-by-association goes global. *Nature* **2000**, *403*, 601–602. [CrossRef] [PubMed]

206. Aravind, L. Guilt by Association: Contextual Information in Genome Analysis. *Genome Res.* **2000**, *10*, 1074–1077. [CrossRef]

207. Hansen, B.O.; Vaid, N.; Musialak-Lange, M.; Janowski, M.; Mutwil, M. Elucidating gene function and function evolution through comparison of co-expression networks of plants. *Front. Plant Sci.* **2014**, *5*, 394. [CrossRef] [PubMed]

208. Luo, F.; Yang, Y.; Zhong, J.; Gao, H.; Khan, L.; Thompson, D.K.; Zhou, J. Constructing gene co-expression networks and pre-dicting functions of unknown genes by random matrix theory. *BMC Bioinform.* **2007**, *8*, 299. [CrossRef]

209. Allocco, D.J.; Kohane, I.S.; Butte, A.J. Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinform.* **2004**, *5*, 18. [CrossRef] [PubMed]

210. Bassel, G.; Gaudinier, A.; Brady, S.; Hennig, L.; Rhee, S.Y.; De Smet, I. Systems Analysis of Plant Functional, Transcriptional, Physical Interaction, and Metabolic Networks. *Plant Cell* **2012**, *24*, 3859–3875. [CrossRef] [PubMed]

211. Van Dam, S.; Võsa, U.; van der Graaf, A.; Franke, L.; de Magalhães, J.P. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform.* **2018**, *19*, 575–592. [CrossRef]

212. López-Kleine, L.; Leal, L.; López, C. Biostatistical approaches for the reconstruction of gene co-expression networks based on transcriptomic data. *Briefings Funct. Genom.* **2013**, *12*, 457–467. [CrossRef]
213. Mahanta, P.; Ahmed, H.A.; Bhattacharyya, D.K.; Kalita, J.K. An effective method for network module extraction from mi-croarray data. *BMC Bioinform.* **2012**, *13*, S4. [CrossRef] [PubMed]
214. Couto, C.M.V.; Comin, C.H.; Costa, L.D.F. Effects of threshold on the topology of gene co-expression networks. *Mol. BioSyst.* **2017**, *13*, 2024–2035. [CrossRef] [PubMed]
215. Borate, B.R.; Chesler, E.J.; Langston, M.A.; Saxton, A.M.; Voy, B.H. Comparison of threshold selection methods for microarray gene co-expression matrices. *BMC Res. Notes* **2009**, *2*, 240. [CrossRef]
216. Perkins, A.D.; Langston, M.A. Threshold selection in gene co-expression networks using spectral graph theory techniques. *BMC Bioinform.* **2009**, *10* (Suppl. S11), S4. [CrossRef] [PubMed]
217. Han, X.; Yin, L.; Xue, H. Co-expression Analysis Identifies CRC and AP1 the Regulator of Arabidopsis Fatty Acid Biosynthesis. *J. Integr. Plant Biol.* **2012**, *54*, 486–499. [CrossRef] [PubMed]
218. Bordych, C.; Eisenhut, M.; Pick, T.; Kuelahoglu, C.; Weber, A.P.M. Co-expression analysis as tool for the discovery of transport proteins in photorespiration. *Plant Biol.* **2013**, *15*, 686–693. [CrossRef]
219. Leal, L.G.; Lopez, C.; López-Kleine, L. Construction and comparison of gene co-expression networks shows complex plant immune responses. *PeerJ* **2014**, *2*, e610. [CrossRef]
220. Barah, P.; Jayavelu, N.D.; Sowdhamini, R.; Shameer, K.; Bones, A.M. Transcriptional regulatory networks in Ara-bidopsis thaliana during single and combined stresses. *Nucleic Acids Res.* **2016**, *44*, 3147–3164. [CrossRef] [PubMed]
221. Barah, P.; Winge, P.; Kuśnierczyk, A.; Tran, D.H.; Bones, A.M. Molecular Signatures in Arabidopsis thaliana in Response to Insect Attack and Bacterial Infection. *PLoS ONE* **2013**, *8*, e58987. [CrossRef] [PubMed]
222. Chang, Y.-M.; Lin, H.-H.; Liu, W.-Y.; Yu, C.-P.; Chen, H.-J.; Wartini, P.P.; Kao, Y.-Y.; Wu, Y.-H.; Lin, J.-J.; Lu, M.-Y.J.; et al. Compara-tive transcriptomics method to infer gene coexpression networks and its applications to maize and rice leaf transcriptomes. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 3091–3099. [CrossRef]
223. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef]
224. Cardozo, L.E.; Russo, P.S.T.; Gomes-Correia, B.; Araujo-Pereira, M.; Sepúlveda-Hermosilla, G.; Maracaja-Coutinho, V.; Nakaya, H.I. webCEMiTool: Co-expression Modular Analysis Made Easy. *Front. Genet.* **2019**, *10*, 146. [CrossRef] [PubMed]
225. Bhuva, D.D.; Cursons, J.; Smyth, G.K.; Davis, M.J. Differential co-expression-based detection of conditional relationships in transcriptional data: Comparative analysis and application to breast cancer. *Genome Biol.* **2019**, *20*, 236. [CrossRef] [PubMed]
226. Dawson, J.A.; Ye, S.; Kendziorski, C. R/EBcoexpress: An empirical Bayesian framework for discovering differential co-expression. *Bioinformatics* **2012**, *28*, 1939–1940. [CrossRef]
227. Li, D.; Brown, J.B.; Orsini, L.; Pan, Z.; Hu, G.; He, S. MODA: MOdule Differential Analysis for Weighted Gene Co-Expression Network. *arXiv* **2016**, arXiv:1605.04739.
228. Amar, D.; Safer, H.; Shamir, R. Dissection of Regulatory Networks that Are Altered in Disease via Differential Co-expression. *PLoS Comput. Biol.* **2013**, *9*, e1002955. [CrossRef]
229. Tesson, B.M.; Breitling, R.; Jansen, R.C. DiffCoEx: A simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinform.* **2010**, *11*, 497. [CrossRef]
230. Henao, J.D. Coexnet: An R Package to Build CO-EXpression NETworks from Microarray Data; Version 1.8.0.; View on Bio-conductor; 2019. Available online: https://rdrr.io/bioc/coexnet/ (accessed on 15 April 2022).
231. Yan, Q.; Wu, F.; Yan, Z.; Li, J.; Ma, T.; Zhang, Y.; Zhao, Y.; Wang, Y.; Zhang, J. Differential co-expression networks of long non-coding RNAs and mRNAs in Cleistogenes songorica under water stress and during recovery. *BMC Plant Biol.* **2019**, *19*, 23. [CrossRef] [PubMed]
232. Ma, S.; Ding, Z.; Li, P. Maize network analysis revealed gene modules involved in development, nutrients utilization, metabolism, and stress response. *BMC Plant Biol.* **2017**, *17*, 131. [CrossRef]
233. Liu, N.; Cheng, F.; Zhong, Y.; Guo, X. Comparative transcriptome and co-expression network analysis of carpel quantitative variation in Paeonia rockii. *BMC Genom.* **2019**, *20*, 683. [CrossRef] [PubMed]
234. McLoughlin, F.; Augustine, R.C.; Marshall, R.S.; Li, F.; Kirkpatrick, L.D.; Otegui, M.S.; Vierstra, R.D. Maize multi-omics reveal roles for autophagic recycling in proteome remodelling and lipid turnover. *Nat. Plants* **2018**, *4*, 1056–1070. [CrossRef]
235. Wang, H.; Wang, S.; Zhang, Y.; Bi, S.; Zhu, X. A brief review of machine learning methods for RNA methylation sites prediction. *Methods* **2022**, *203*, 399–421. [CrossRef]
236. Banerjee, D.; Jindra, M.A.; Linot, A.J.; Pfleger, B.F.; Maranas, C.D. EnZymClass: Substrate specificity prediction tool of plant acyl-ACP thioesterases based on ensemble learning. *Curr. Res. Biotechnol.* **2022**, *4*, 1–9. [CrossRef]
237. Sampaio, M.; Rocha, M.; Dias, O. Exploring synergies between plant metabolic modelling and machine learning. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 1885–1900. [CrossRef] [PubMed]
238. Campos, T.L.; Korhonen, P.K.; Hofmann, A.; Gasser, R.B.; Young, N.D. Harnessing model organism genomics to un-derpin the machine learning-based prediction of essential genes in eukaryotes-Biotechnological implications. *Biotechnol. Adv.* **2021**, *54*, 107822. [CrossRef] [PubMed]
239. Van Dijk, A.D.J.; Kootstra, G.; Kruijer, W.; de Ridder, D. Machine learning in plant science and plant breeding. *iScience* **2021**, *24*, 101890. [CrossRef] [PubMed]