



## REVIEW

# Machine learning algorithms: their applications in plant omics and agronomic traits' improvement [version 1; peer review: awaiting peer review]

Itunuoluwa Isewon <sup>1-3</sup>, Oluwabukola Apata<sup>1,2</sup>, Fesobi Oluwamuyiwa<sup>1,2</sup>,  
Olufemi Aromolaran<sup>1,3</sup>, Jelili Oyelade <sup>1-3</sup>

<sup>1</sup>Department of Computer & Information Sciences, Covenant University, Ota, Ogun State, 112233, Nigeria

<sup>2</sup>Covenant Applied Informatics and Communication African Centre of Excellence (CApIC-ACE), Covenant University, Ota, Ogun State, Nigeria

<sup>3</sup>Covenant University Bioinformatics Research (CUBRe), Covenant University, Ota, Ogun State, Nigeria

**V1** First published: 04 Nov 2022, 11:1256  
<https://doi.org/10.12688/f1000research.125425.1>

Latest published: 04 Nov 2022, 11:1256  
<https://doi.org/10.12688/f1000research.125425.1>

## Open Peer Review

**Approval Status** *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

## Abstract

Agronomic traits of plants especially those of economic or aesthetic importance are threatened by climatic and environmental factors such as climate change, biotic, and abiotic stresses. These threats are now being mitigated through the analyses of omics data like genomics, transcriptomics, proteomics, metabolomics, and phenomics. The emergence of high-throughput omics technology has led to an avalanche of plant omics data. Plant research demands novel analytical paradigms to extract and harness large plant omics data for plant improvement effectively and efficiently. Machine learning algorithms are well-suited analytical and computational approaches for the integrative analysis of large unstructured, heterogeneous datasets. This study presents an overview of omics approaches to improve plant agronomic traits and crucial curated plant genomic data sources. Furthermore, we summarize machine learning algorithms and software tools/programming packages used in plant omics research. Lastly, we discuss advancements in machine learning algorithms' applications in improving agronomic traits of economically important plants. Extensive application of machine learning would advance plant omics studies. These advancements would consequently help agricultural scientists improve economically important plants' quality, yield, and tolerance against abiotic and biotic stresses and other plant health-threatening issues.

## Keywords

Agronomic traits, machine learning, multi-omics, plant improvement

**Corresponding author:** Itunuoluwa Isewon ([itunu.isewon@covenantuniversity.edu.ng](mailto:itunu.isewon@covenantuniversity.edu.ng))

**Author roles:** **Isewon I:** Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Apata O:** Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation; **Oluwamuyiwa F:** Data Curation, Methodology, Software, Writing – Original Draft Preparation; **Aromolaran O:** Data Curation, Formal Analysis, Software, Writing – Review & Editing; **Oyelade J:** Conceptualization, Funding Acquisition, Resources, Supervision, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by funding from the World Bank awarded to Covenant Applied Informatics, Communication Africa Centre of Excellence (CApIC-ACE) through the ACE Impact Project (2019 – 2024) and Covenant University Center for Research, Innovation and Discovery (CUCRID)

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2022 Isewon I *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Isewon I, Apata O, Oluwamuyiwa F *et al.* **Machine learning algorithms: their applications in plant omics and agronomic traits' improvement [version 1; peer review: awaiting peer review]** F1000Research 2022, 11:1256  
<https://doi.org/10.12688/f1000research.125425.1>

**First published:** 04 Nov 2022, 11:1256 <https://doi.org/10.12688/f1000research.125425.1>

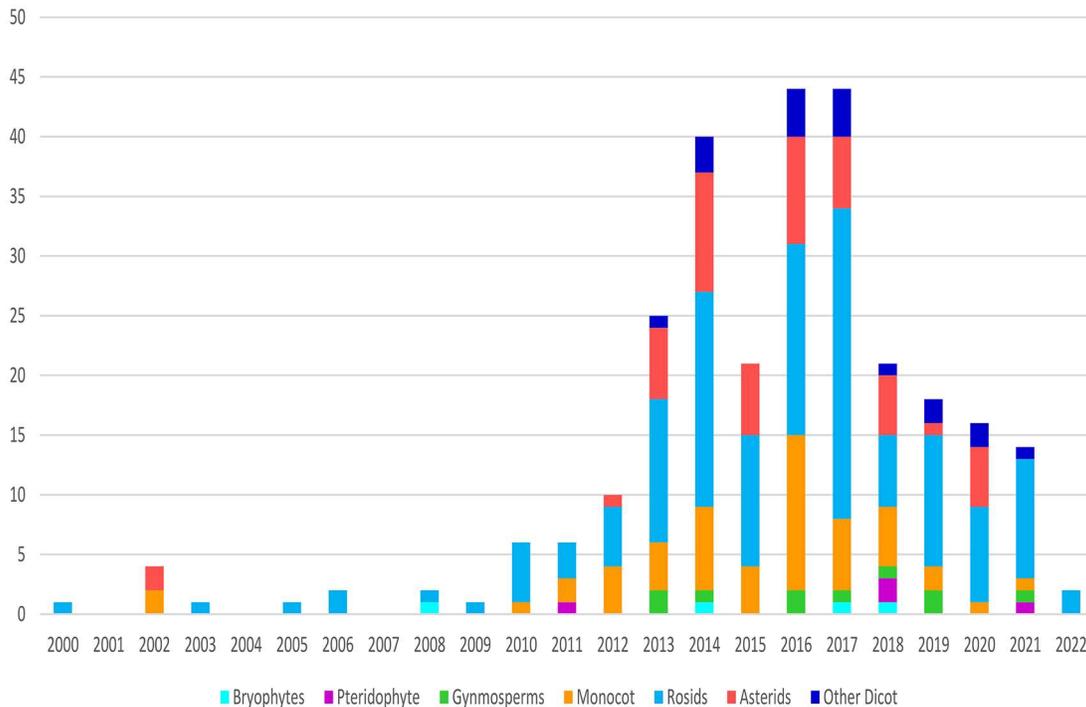
## 1. Introduction

The global agricultural system is threatened by ecological events such as climate change and other environmental stresses.<sup>1,2</sup> These events affect the yield, stability, and quality of production of economically important plants, *i.e.*, medicinal plants, fruit crops, food crops, cereal or grain crops, legume seed crops or pulses, *etc.*<sup>1,3</sup> These challenges are addressed by omics approaches *via* numerous unconventional improvement methodologies.<sup>4,5</sup> Omics approaches involve analysis of the constituents of genome sequence and other macromolecules generated from encoded genomic information. The utilization of omics-derived knowledge and technologies in plant improvement strategies is limited and difficult. Other drawbacks of omics technologies include a lack of data integration and effective phenotype-genotype correlation strategies. As a result, promoting the integration of computational biology and plant genomics to assist plant development is critical.<sup>6,7</sup> This paper gives an overview of omics methodologies for improving plant agronomic traits as well as essential curated plant genomic data sources. We discuss the bioinformatics software, tools, and packages that are utilized in omics-based plant improvement research. We also dissect how machine learning algorithms has been used to improve agronomic features of commercially significant plants, their major contributions and future outlook in plant omics and agronomics.

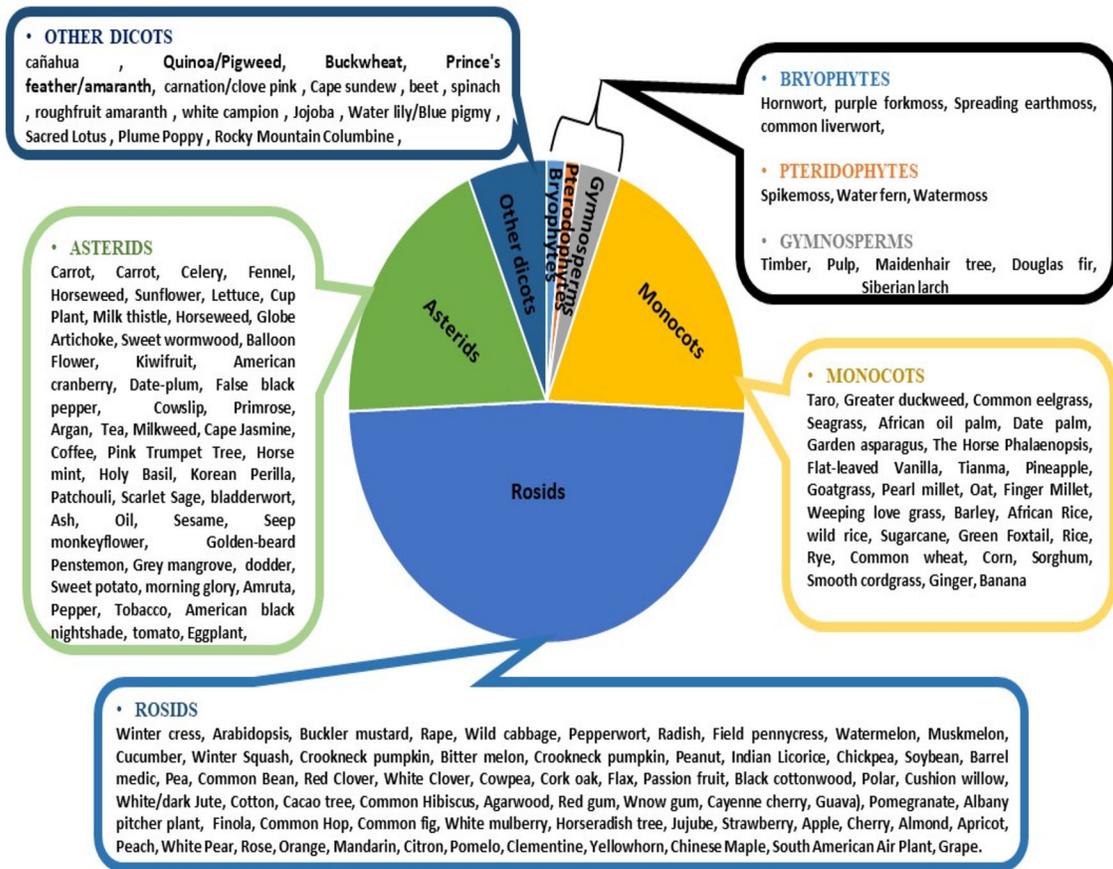
### 1.1 Plant genome sequences and bioinformatics resources

Genome sequencing was made possible with the advent of sequencing technologies.<sup>8,9</sup> Complete sequencing of a plant genome was first demonstrated for a model plant named Arabidopsis (*Arabidopsis thaliana*)<sup>10,11</sup> and afterward for rice (*Oryza sativa*). Subsequently, the whole genome of over 250 species in the plant kingdom have been sequenced: bryophytes, pteridophytes, gymnosperms, and angiosperms<sup>12,13</sup> (Figure 1). Angiosperms account for 95% of the sequenced species, most of which are economically important plants or their wild relatives (Figure 2). Food crops like rice, wheat, beans, oat, maize, and soybean are among the sequenced plants, as are ornamental plants like orchid and hibiscus, industrial plants like oilseed, hemp, and spice/herbs like garlic, ginger, turmeric, moringa, artemisia, and neem, which are known for their high therapeutic value.

There have been a variety of databases created to access plant genome datasets.<sup>14,11</sup> The model plant *A. thaliana* genome database launched in 2001 was the premier plant genome database.<sup>10,15</sup> Subsequently, many databases and resources have been developed for plant genomes. The earliest genome databases were essentially archives of genome sequence data. These databases have expanded into genome portals/hubs that combine different genomic data and web servers that



**Figure 1. Published plant genome sequences from 2000 to date.** Most sequenced plants are angiosperms and are subdivided into three groups. Most of the sequenced angiosperms fall under rosids and asterids clades. Other sequenced angiosperms clades are grouped here as other dicots.



**Figure 2. Percentage of sequenced plants with their common names.** 94% of sequenced plants are angiosperms consisting of both monocots and dicots. Percentage of other plants are 1%, 2%, 3% for Pteridophytes, Bryophytes, and Gymnosperms, respectively.

offer online genomics analysis. The availability of annotated plant genome data has led to many discoveries, including genome organization and gene function.<sup>16</sup> These discoveries elucidate the complexity, evolution, and dynamics of plant genomes, contributing to a deeper understanding of plant biology.<sup>6,17</sup> Available genomic information includes cis-elements, gene expression data, protein interactome, transcriptional and post-transcriptional data. These genome databases exist as single species and comprehensive databases as shown in Table 1.

**Table 1. General plant genomics databases and tools.**

Database	Description	Website
AgBase	A unified resource for functional analysis in agriculture	<a href="http://www.agbase.msstate.edu/">http://www.agbase.msstate.edu/</a>
Ensembl Plants	A genome-centric portal for plant species	<a href="http://plants.ensembl.org">http://plants.ensembl.org</a>
AutoSNPdb	An annotated single nucleotide polymorphism database for crop plants	<a href="http://autosnpdb.qfab.org.au/">http://autosnpdb.qfab.org.au/</a>
BarleyBase	An expression profiling database for plant genomics	<a href="http://www.barleybase.org/">http://www.barleybase.org/</a>

**Table 1.** *Continued*

Database	Description	Website
CR-EST	A resource for crop ESTs Search for sequence, classification, clustering and annotation data of crop EST projects	<a href="http://pgrc.ipk-gatersleben.de/cr-est/">http://pgrc.ipk-gatersleben.de/cr-est/</a>
CSRDB	A small RNA integrated database and browser resource for cereals	<a href="http://sundarlab.ucdavis.edu/smrnas/">http://sundarlab.ucdavis.edu/smrnas/</a>
ChromDB	The Chromatin Database	<a href="http://www.chromdb.org/">http://www.chromdb.org/</a>
DRASTICx97INSIGHTS	Querying information in a plant gene expression database	<a href="http://www.drastic.org.uk/">http://www.drastic.org.uk/</a>
FLAGdb++	A Database for the Functional Analysis of the Arabidopsis Genome	<a href="http://urgv.evry.inra.fr/projects/FLAGdb++/HTML/index.shtml">http://urgv.evry.inra.fr/projects/FLAGdb++/HTML/index.shtml</a>
GCP	The Generation Challenge Programme	<a href="http://www.generationcp.org/">http://www.generationcp.org/</a>
GGT	Graphical GenoTypes - Software for visualization and analysis of genetic data	<a href="http://www.plantbreeding.wur.nl/">http://www.plantbreeding.wur.nl/</a>
GabiPD	Integrative Plant Omics Database	<a href="http://www.gabipd.org/">http://www.gabipd.org/</a>
GeneSeqer@PlantGDB	Gene structure prediction in plant genomes - Predict gene structures of plant genomes	<a href="http://www.plantgdb.org/PlantGDB-cgi/GeneSeqer/PlantGDBgs.cgi">http://www.plantgdb.org/PlantGDB-cgi/GeneSeqer/PlantGDBgs.cgi</a>
GrainGenes	The genome database for small-grain crops	<a href="http://wheat.pw.usda.gov/index.shtml">http://wheat.pw.usda.gov/index.shtml</a>
Gramene	A resource for comparative grass genomics	<a href="http://www.gramene.org/">http://www.gramene.org/</a>
MIPS	Analysis and annotation of genome information	<a href="http://mips.gsf.de/">http://mips.gsf.de/</a>
MetaCrop	A detailed database of crop plant metabolism	<a href="http://metacrop.ipk-gatersleben.de/">http://metacrop.ipk-gatersleben.de/</a>
NIASGBdb	National Institute of Agrobiological Sciences Gene Bank DataBase	<a href="http://www.gene.affrc.go.jp/databases_en.php">http://www.gene.affrc.go.jp/databases_en.php</a>
P3DB	Plant Protein Phosphorylation Database	<a href="http://digbio.missouri.edu/p3db/">http://digbio.missouri.edu/p3db/</a>
PHYTOPROT	A Database of Clusters of Plant Proteins	<a href="https://urgi.versailles.inra.fr/phytoprot/">https://urgi.versailles.inra.fr/phytoprot/</a>
PIP	A database of potential intron polymorphism markers	<a href="http://ibi.zju.edu.cn/pgl/pip/">http://ibi.zju.edu.cn/pgl/pip/</a>
PLACE	Plant cis-acting regulatory DNA elements	<a href="http://www.dna.affrc.go.jp/PLACE/">http://www.dna.affrc.go.jp/PLACE/</a>
PLANT-PIs --	A database for protease inhibitors and their genes in higher plants	<a href="http://plantpis.ba.itb.cnr.it/">http://plantpis.ba.itb.cnr.it/</a>
PLecDom	Plant Lectin Domains server	<a href="http://www.nipgr.res.in/plecdom.html">http://www.nipgr.res.in/plecdom.html</a>
PMRD	Plant MicroRNA Database	<a href="http://bioinformatics.cau.edu.cn/PMRD/">http://bioinformatics.cau.edu.cn/PMRD/</a>
PODB	The Plant Organelles Database	<a href="http://podb.nibb.ac.jp/Organellome">http://podb.nibb.ac.jp/Organellome</a>
POGs/PlantRBP	A resource for comparative genomics in plants	<a href="http://cas-pogs.uoregon.edu/#/">http://cas-pogs.uoregon.edu/#/</a>

**Table 1.** *Continued*

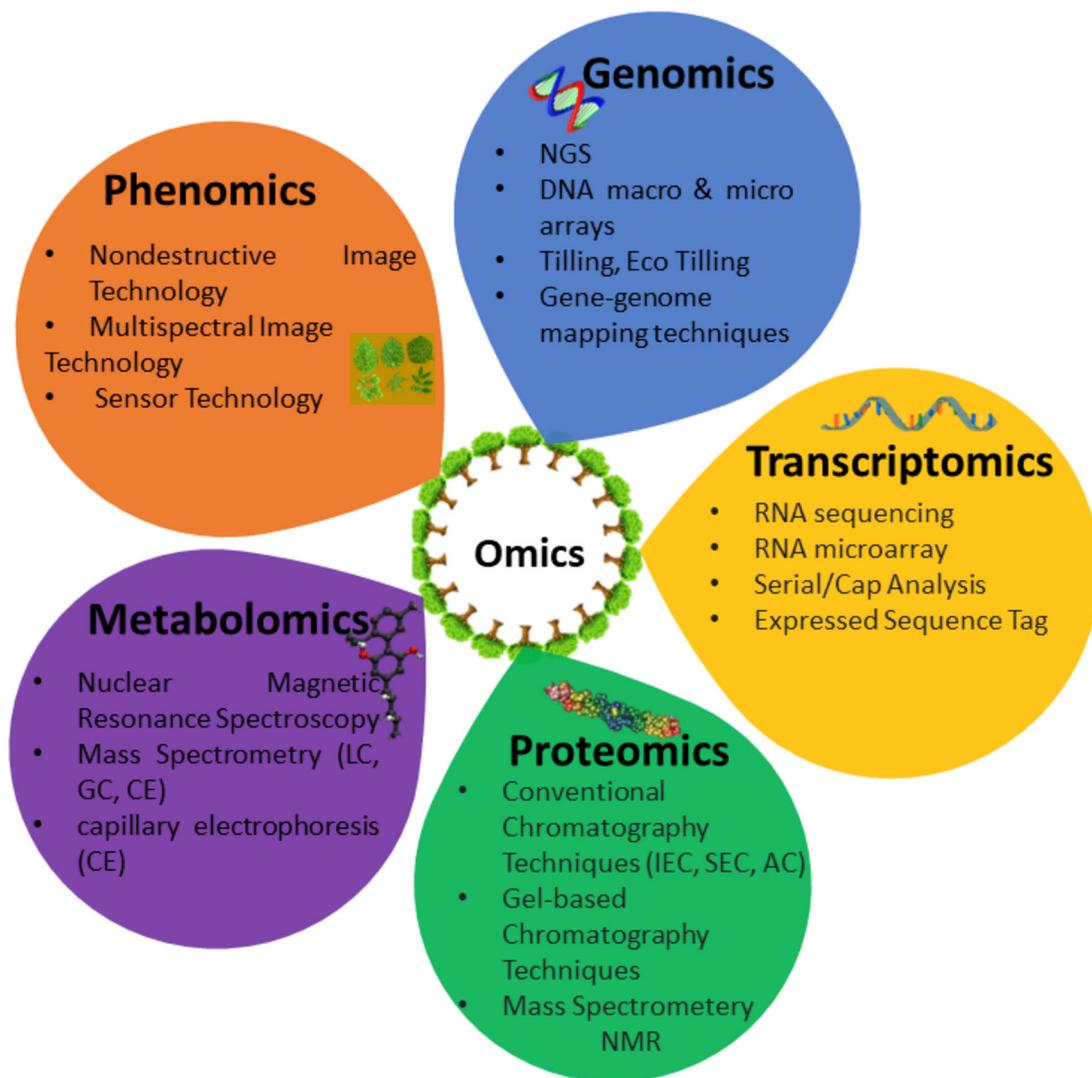
<b>Database</b>	<b>Description</b>	<b>Website</b>
PREP Suite	Predictive RNA Editor for Plants	<a href="http://prep.unl.edu/">http://prep.unl.edu/</a>
PRGDB	Plant Resistance Genes DataBase	<a href="http://prgdb.cbm.fvg.it/">http://prgdb.cbm.fvg.it/</a>
PathoPlantxae	A platform for microarray expression data to analyze co-regulated genes involved in plant defense responses	<a href="http://www.pathoplant.de/">http://www.pathoplant.de/</a>
Phytome	A platform for plant comparative genomics	<a href="http://www.phytome.org/">http://www.phytome.org/</a>
Plant MPSS databases	Signature-based transcriptional resources for analyses of mRNA and small RNA	<a href="http://mpss.udel.edu/">http://mpss.udel.edu/</a>
Plant snoRNA database	Search for comprehensive information on small nucleolar RNAs in plants	<a href="http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/home">http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/home</a>
PlantCARE	A database of plant cis-acting elements	<a href="http://bioinformatics.psb.ugent.be/webtools/plantcare/html/">http://bioinformatics.psb.ugent.be/webtools/plantcare/html/</a>
PlantGDB	Plant Genome Database and Analysis tools	<a href="http://www.plantgdb.org/">http://www.plantgdb.org/</a>
PlnTFDB	Database of Plant Transcription Factor	<a href="http://www.softberry.com/berry.phtml?topic=plantprom&amp;group=data&amp;subgroup=plantprom">http://www.softberry.com/berry.phtml?topic=plantprom&amp;group=data&amp;subgroup=plantprom</a>
PlantTribes	A gene and gene family resource for comparative genomics in plants	<a href="http://planttfdb.cbi.pku.edu.cn/">http://planttfdb.cbi.pku.edu.cn/</a>
PlantTFDB	Plant Transcription Factor Databases	<a href="http://fgp.huck.psu.edu/tribe.html">http://fgp.huck.psu.edu/tribe.html</a>
PmiRKB	Plant MicroRNA Knowledge Base - Find information about plant microRNAs	<a href="http://bis.zju.edu.cn/pmirkb/">http://bis.zju.edu.cn/pmirkb/</a>
SALAD	Surveyed contained motif Alignment diagram and the Associating Dendrogram	<a href="http://salad.dna.affrc.go.jp/salad/en/">http://salad.dna.affrc.go.jp/salad/en/</a>
The Adaptive Evolution Database (TAED)	A phylogeny based comparative genomics tools	<a href="http://www.bioinfo.no/tools/TAED">http://www.bioinfo.no/tools/TAED</a>
The Plant DNA C-values Database	Search for information on plant DNA C-values and genome sizes.	<a href="http://data.kew.org/cvalues/homepage.html">http://data.kew.org/cvalues/homepage.html</a>
The PlantsP Functional Genomics Database	Search for information on plant kinases and phosphatases	<a href="http://plantsp.sdsc.edu/">http://plantsp.sdsc.edu/</a>
The PlantsT Functional Genomics Databases	Search for genes and proteins involved in plant membrane transportation	<a href="http://plantst.sdsc.edu/">http://plantst.sdsc.edu/</a>
The TIGR Plant Repeat Databases	A Collective Resource for the Identification of Repetitive Sequences in Plant	<a href="http://www.tigr.org/tdb/e2k1/plant.repeats/index.shtml">http://www.tigr.org/tdb/e2k1/plant.repeats/index.shtml</a>
The TIGR Plant Transcript Assemblies database	Search for plant EST and cDNA sequences from this comprehensive collection	<a href="http://plantta.tigr.org/">http://plantta.tigr.org/</a>
TropGENE-DB	A Multi-Tropical Crop Information System	<a href="http://tropgenedb.cirad.fr/">http://tropgenedb.cirad.fr/</a>

**Table 1.** Continued

Database	Description	Website
UK CropNet	A collection of databases and bioinformatics resources for crop plant genomics	<a href="http://ukcrop.net/">http://ukcrop.net/</a>
openSputnik x97	A database for annotated expressed sequences tags information and comparative plant genomics analysis	<a href="http://sputnik.btk.fi/">http://sputnik.btk.fi/</a>

### 1.2 Plant omics technologies

Plant studies involving the analysis of biological macromolecules are collectively termed plant omics. Omics is a broad field of study encompassing subfields like genomics, transcriptomics, proteomics, metabolomics, phenomics, glycomics, lipidomic, *etc.* Plant genomics involves studying the compositions, organizations, functions, and structures of genetic materials (DNA/RNA) and molecular genetic networks of interactions in the plant genome.<sup>17,18</sup> While genome structure



**Figure 3. Important plant omics branches and their major techniques.** Representation of major omics approaches in plant molecular studies and the methods utilized in conventional analysis of plant omics datasets.

and organization are studied in structural genomics,<sup>13,19</sup> functional genomics investigates the functions, interaction, and regulation of gene and gene products.<sup>5,20</sup>

Plant functional genomics is a goldmine in agronomic traits improvement. It incorporates other omics approaches like transcriptomics, proteomics, metabolomics, phenomics, *etc.*<sup>16,21</sup> (Figure 3). Other aspects of genomics are epigenomics, mutagenomics, and pangenomics.<sup>22</sup> Epigenetic changes, such as histone modifications, small RNA and DNA methylations occurring at the genomic phase, are dissected within epigenomics. Mutagenomics is used to explore mutation events mediating modified genotype and phenotype in mutant species. Pangenomics studies the whole set of genomic sequences present in the entire population of a species. It also explores dispensable genomes that are individual specific or partially shared. Mutagenomics and pangenomics are new omics techniques in crop sciences.<sup>22,23</sup>

Plant transcriptomics involves investigating the control of plant metabolite production processes at the RNA level. Transcript-level gene expression control regulates the whole plant's development and growth.<sup>24</sup> Plant proteomics explores the structural and functional features of proteins in a living organism. It encompasses studies on plants' typical morphological and physiological properties.<sup>16,22</sup> The role of proteins in controlling the plant metabolic processes is also studied in plant proteomics, especially in medicinal plants.<sup>25</sup>

Plant metabolomics involves profiling primary and secondary metabolites in plants.<sup>26</sup> Metabolic data are useful when developing metabolic correlation networks. These networks can aid comparative analysis of cellular compartments such as carbon and nitrogen transport and partitioning in plants.<sup>27</sup> In addition, the molecular and cellular regulation of different enzymatic processes can also be investigated.<sup>25</sup> Plant phenomics involves the systematic study of phenotypes such as plant composition, growth, and production analysis. This study can be conducted both in controlled environments and in the field. Field phenomics involves the measurement of phenotypes that exist under both cultivated and natural conditions. Studies in controlled environments involve glasshouses, growth chambers, and other systems where growth conditions can be manipulated.<sup>28</sup> These multi-omics approaches have emerged successful for plant research, including agronomic traits improvement over the last few decades. Agronomic traits are desirable plants' genetic or phenotypic features, *i.e.*, quality traits, disease resistance, pest resistance, insecticide tolerance, temperature, drought, and other adverse environmental factors tolerance traits. Quality traits encompass morphological features like plant height, seed weight; physiological features like chlorophyll content and photosynthetic rate<sup>29</sup>; economic features like improved crop yield, processing, and storage; pharmaceutical and industrial features like the elimination of toxins and allergen, increased nutritional or dietary value and increased medicinal values.<sup>30</sup> Recent research suggests that when multi-omics technologies are integrated, they can be better harnessed to improve genetic development, crop breeding science, plant stress resistance, and other agronomic traits.<sup>22,23</sup>

## 2. Major areas of application of omics technologies for agronomic traits improvement

### 2.1 Genomics-assisted pre-breeding

Genomics-assisted pre-breeding is a genetic manipulation strategy to improve agronomic traits of interest in plants at the DNA level.<sup>31</sup> Genomics-assisted pre-breeding approaches positively contribute to the efficiency of diseases and climate-resilient crop development.<sup>32–34</sup> Crop breeding across the globe has relied on a series of phenotypic selection and crossing before the genomic era to generate superior crop genotypes.<sup>35</sup> Genome sequence availability has paved the way for identifying all genes and genetic variants associated with agronomics traits.<sup>36,37</sup> Besides, it has made it possible to assess genotype level changes incurred during breeding processes.<sup>38</sup> Plant breeders have utilized genomics and bioinformatics in gene-level resolution of agronomic variation using quantitative trait loci (QTL) mapping<sup>39–41</sup> and genome-wide association studies (GWAS).<sup>42,43</sup> For instance, studies have recently been conducted to develop multiple stress-adaptable rice species that are disease and climate resilient using genomics-assisted breeding techniques such as quantitative trait locus (QTL), gene/markers-phenotype association and phenotype selection.<sup>44–46</sup> Pea breeding projects used genetic marker-trait associations to boost valued yield and market-preferred agronomic traits.<sup>47,48</sup> Miedaner *et al.*<sup>43</sup> used high-density genotype arrays and comprehensive phenotyping of the same species population across diverse conditions, locations, and seasons in genomic selection and population mapping to speed the breeding of disease resistance traits in maize, small-grain cereals, and wheat. Hu *et al.*<sup>39</sup> harnessed genomic selection (GS) and genome-wide genetic variants to prevent reiterated phenotyping in breeding cycles. These studies indicate new breeding techniques such as speed breeding and genomic selection to boost genetic and trait improvements. However, the lack of robust phenotypic data limits the efficient utilization of available genomic information and technologies in genomics-assisted breeding.

### 2.2 Evolution and crop diversity

Variation in gene content among individuals within the same species is caused by genetic variation ranging from single-nucleotide polymorphisms to substantial structural variants (SVs). Due to human and natural selection acts, this variation offers the raw material on which evolution occurs.<sup>49,50</sup> Deviation in agricultural plants' phenotypic and genetic

characteristics is referred to as crop diversity.<sup>36</sup> The understanding of crop diversity is enhanced by plant genomics at both species and gene levels.<sup>51</sup> According to recent research, a single reference genome is insufficient to capture a species' entire genetic diversity landscape. Pan-genome analysis provides a platform for evaluating a species' genetic diversity by looking at its whole genomic repertoire. Pan-genomic studies have shed new light on the landscape of diversity and improvement of major crops such as *Brachypodium distachyon*,<sup>52</sup> *Brassica* Spp.,<sup>53–55</sup> maize,<sup>56</sup> rice,<sup>57,58</sup> soybean,<sup>59</sup> wheat<sup>60</sup> etc. Evolution in plant diversity is correlated with and relatively predictable by heterogenous biotic and abiotic environmental stress induced by global climate change. These stresses, in turn, affect crop yield and crop-growing seasons.<sup>61</sup> A study on natural plant populations shows that the organization and evolution of plant populations' diversity at all genomic regions is nonrandom at the molecular and organismal level.<sup>62,63</sup> Therefore, plants can evolve under climatic gradients resulting in clinal adaptation. Hence, the breeding of climate resilience crops can be facilitated by understanding the genomic basis of clinal adaptation in crop species.<sup>64</sup>

### 2.3 Abiotic and biotic stresses

Biotic stresses are instigated by living organisms such as insects, parasitic plant nematodes, diseases, or weeds in production agriculture.<sup>65</sup> Genomics approaches to biotic stress include ribonucleic acid interference (RNAi) silencing and transgenesis. RNA interference (RNAi) silencing is employed against viruses and some fungi, while transgenesis has been exploited to develop resistance against some fungi, for example, *Fusarium* head blight.<sup>66</sup> Genome-wide identification and expression analysis in legume crops also revealed the role of small RNA biogenesis mediators in biotic stress response regulation.<sup>42</sup>

Abiotic stresses, such as low or high temperatures, heavy metals, insufficient or excessive water, high salinity, ultraviolet radiation, are hostile to developing plants, resulting in significant wane in crop yield worldwide.<sup>67</sup> According to Kumar *et al.*,<sup>28</sup> knowledge of plants' response to abiotic stresses can be enhanced by integrating information generated from metabolomics and proteomics with genomics data. Sustenance of yield in crops threatened by abiotic stresses is a significant challenge in breeding resilient crop varieties.<sup>68</sup> A study on *Brassica oleracea* shows that heat stress transcription factors are integral to signal transduction pathways functioning in response to environmental stresses and are suggested to contribute significantly to various stress responses.<sup>62</sup> Heat stress transcription factors genes were identified in the *in silico* analysis of *B. oleracea*. The identified genes may be exploited in developing crop varieties resilient to global climate change.<sup>1,62</sup>

### 2.4 Population studies

Population genomics is employed to study adaptation and speciation. Population genomics datasets are used in GWAS to detect the genes responsible for adaptive phenotypic variations of large plant population samples.<sup>35,69</sup> For instance, Bamba *et al.* identified specific adaptation loci in a GWA study and unveiled the molecular basis of genetic trade-offs. It also showed that ecological fitness could be predicted by polygenic effects of several loci associated with local climate.<sup>35</sup> Medicinal plants' diversity is of exceptional interest because of their ethnomedicine role. GWAS studies on adaptive genotypic and phenotypic variation provide a framework to assess the diversity of medicinal plant application across different cultures and infer modifications in plant use over time.<sup>70,71</sup> Other genomic approaches such as genomic selection, nested association mapping, genetic diversity, and allele mining have been integrated into crop improvement programs to address the genetic issues associated with maize productivity and nutritional contents.<sup>72</sup>

Plant omics studies have greatly helped our understanding and interpretation of plant responses to ecological influences and their contribution to key developmental processes important for crop yield and food quality. However, there are still some problems, such as a lack of data integration and robust techniques for phenotype-genotype correlation. Also, the use of omics-derived knowledge and tools in plant improvement strategies is limited and difficult. As a result, there is a pressing need to promote the integration of computational biology and plant genomics to benefit plant improvement.<sup>6,7</sup>

## 3. Applications of machine learning in plant omics and agronomics

### 3.1 Machine learning algorithms and resources

Machine learning (ML) is a computer science field that utilizes algorithms to learn and capture the characteristics of target patterns of complex datasets.<sup>73</sup> Machine learning algorithms are generally classified into the following categories: supervised, semi-supervised, unsupervised, reinforcement, and deep learning.<sup>74</sup> A supervised ML algorithm is trained using a labeled dataset. It learns to respond more accurately based on these training sets by comparing its output with the given input.<sup>67,73</sup> Semi-supervised algorithms provide a tool that harnesses the potential of both supervised and unsupervised learning. These algorithms are ideally adapted for model building and can be used for classification, regression, and prediction.<sup>75</sup> Unsupervised learning is all about identifying unexplained existing patterns from the data to generate pattern rules. Unsupervised learning is a learning approach focused on statistics and thus applied to the issue of discovering a hidden structure in unlabeled data.<sup>7,74,76</sup>

**Table 2. Existing machine learning tools for plant omics data analysis.**

Application area	Developed tools	URL	Algorithms	Selected features
Plants Mitochondrially Localized Proteins Prediction	MU-LOC <sup>78</sup>	<a href="http://mu-loc.org/">http://mu-loc.org/</a>	SVM and DNN	gene co-expression information, protein position weight matrix, amino acid compositions, and N-terminal sequence information,
plant resistance protein NBSLRR prediction	NBSPred <sup>79</sup>	<a href="http://soilecology.biol.lu.se/nbs/">http://soilecology.biol.lu.se/nbs/</a>	SVM	R-protein and non-R-protein sequences attribute like sequence domain and compositional frequencies
Ribosomal proteins (RPs) prediction	RAMA <sup>80</sup>	<a href="http://inctipp.bioagro.ufv.br:8080/Rama">http://inctipp.bioagro.ufv.br:8080/Rama</a> .	MLP, RF, and NB	Amino acid side chains attributes
Plant disease resistance proteins prediction	DRPPP <sup>81</sup>	<a href="http://14.139.240.55/NGS/download.php">http://14.139.240.55/NGS/download.php</a>	SVM, MLP, and RF	genomic sequence (satellite DNAs)
Geminiviruses Gene and genera classification	Fangorn Forest (F2) <sup>82</sup>	<a href="http://www.geminivirus.org:8080/geminivirusdw/discoveryGeminivirus.jsp">www.geminivirus.org:8080/geminivirusdw/discoveryGeminivirus.jsp</a> .	SVM, MLP, and RF	genomic sequence (satellite DNAs)
Transcriptomes for stress responses in Arabidopsis	mIDNA <sup>83</sup>	<a href="http://www.plantcell.org/content/26/2/520.short#def-8">www.plantcell.org/content/26/2/520.short#def-8</a>	RF with PSOL algorithm, SVM, and NN	patterns of 32 known stress-related gene expression traits and the complementary expression characteristic

Reinforcement learning is considered an intermediate form of learning as the algorithm is only provided with an answer that tells whether the output is correct or not.<sup>75</sup> Deep learning is built on artificial neural networks (ANN). The algorithms extract higher-level features from the raw input using multiple layers of neural networks. Learning of the algorithm can be unsupervised, semi-supervised, or supervised.<sup>73</sup> Machine learning approaches provide unique techniques for integrating and analyzing omics data, allowing for the improvement of crops and other economically important plants. Some machine learning algorithms have been used to developed tools specifically for plant omics analysis. Table 2 highlights the existing machine learning tools for plant omics analysis. Machine learning algorithms have a broad range of applications in plant genomics. These algorithms play vital roles in genome assembly, iterative gene regulatory network inference, and the identification of true SNPs in polyploid plants.<sup>77</sup>

### 3.2 Precision plant breeding

Precision breeding is a genetic engineering technique that involves reproducing organisms of the same species together to preserve desirable characteristics and create a stronger hybrid.<sup>84</sup> Traditional statistical methods mainly used in plant breeding strategies are ineffective in plant data analysis because of the non-deterministic and nonlinear nature of plant features attributed to environment, genotype, and interaction.<sup>85</sup> Machine learning has enabled effective plant phenotyping and data mining for patterns such as genotype and trait correlation.<sup>39</sup> It has also been successful in genomic selection. Genomic selection is a critical method used in selecting plant species with genetic gains of interest in plant breeding. Applying different ML algorithms in building GS models has produced robust and accurate prediction.<sup>86,87</sup> Multilayer neural networks (NNs) have been used in genetic value prediction in plant breeding. NN models are efficient in predicting genetic value, regardless of the population size, heritability, or coefficient of variation. Thus, the ANN is promising for genetic value prediction in unbiased experiments.<sup>88</sup> Multilayer NNs have been utilized to select bean genotypes with highly stable phenotypes using 13 genotypes of common beans between 2002 and 2006. The integration of this model in plant breeding has enabled precise genetic value prediction and selection.<sup>89</sup>

Also, deep learning generated a robust prediction accuracy in grain yield compared to the conventional linear statistical methods used in traditional plant breeding when analyzing multiple traits with mixed ordinal, continuous, and binary phenotype data. Univariate and multivariate deep learning models' predictive performance was assessed using the Durum wheat (*Triticum turgidum* var. durum Desf.) dataset. Deep learning model performance shows that it has a

promising potential to be a successful model for accurate genomic prediction in plant breeding.<sup>90</sup> The flexibility of machine learning algorithms makes them a viable alternative to traditional parametric methods for predicting categorical and continuous responses in genomic selection.<sup>91</sup>

An ensemble of RF and SVM was implemented to improve genotype-phenotype classification using manually derived root trait datasets. The combined model accurately identified the most distinguishing root traits and corresponding cultivar differentiation. The model's performance demonstrates the potential of ML approaches in unbiased cultivars classification and trait selection.<sup>92</sup>

Additionally, predictive models have aided the integration of additive and dominance effects in GWAS and have enhanced the prediction of complex agronomic traits in polyploid plant species. For instance, a study revealed the feasibility of genome-wide prediction of potato agronomic traits despite being an autotetraploid food crop. It also shows that GWA prediction is viable in selecting breeding values in elite germplasm with substantial non-additive genetic variance.<sup>93</sup>

### 3.3 Phenomics

Plant phenomics is a systematic study of plant phenotypes.<sup>28</sup> In recent years, plant field phenotyping has gotten a lot of attention with the possibility of crop fields' high-throughput analysis.<sup>94</sup> The application of machine learning methods and the various technological developments for image analysis have improved quantitative crop traits assessment.<sup>92,94,95</sup> For instance, CNN-based detection and analysis of wheat spikes using wheat field trials images captured over one planting season achieved an average accuracy of 88 to 94% across diverse groups of test images. CNN's high-performance accuracy shows that it is a robust model for genome-based selection and prediction in plant breeding.<sup>96</sup> Also, the RF algorithm was used in plant image segmentation involving the acquisition and processing of several plant images samples.<sup>97</sup> The predictions made by the model enabled the discovery of various parameters relevant to plant growth.<sup>94</sup>

### 3.4 Stress resilience phenotyping

ML learning has been exploited in identifying favorable agronomic traits, including abiotic and biotic stress resistance. ML algorithms are integrated into conventional statistical methodologies to optimize the accuracy of plants stresses prediction and detection.<sup>96</sup> For instance, 25,000 soybean leaflets images exposed to varied diseases and nutritional perturbation were used to develop a convolutional neural network (CNN), which can infer the image features of the disease types and dietary deficiencies at high resolution. The prediction accuracy of the ML framework was very close to that of human expert diagnosis. Other plants' induced stresses can also be identified, classified, and quantified using the model. The model can also be adapted to identify, classify, and quantify the induced stresses in other plants.<sup>95</sup>

Random forest has been used to predict metabolite and transcript markers in drought tolerance prediction using experimental drought-stressed plant field trial datasets. The low error rate recorded in the model shows that the model could be considered as an alternative model for accurate prediction and identification of molecular markers.<sup>98</sup> RF was used to identify suitable features combination for phenotypic traits prediction using data derived from various agro-management treatment experiments. This approach achieved optimal prediction accuracy and improved plant breeding strategies by enabling maximal allocation of stress management resources.<sup>99</sup> Sanz-Carbonell and colleagues used deep sequencing and computational approaches such as PCA and Clustering analysis to infer the biotic and abiotic stress responses regulatory network mediated by miRNA. 24 miRNAs were used in this study, all of which are known to alter expression significantly under stressful conditions. The prediction generated inference that target genes of miRNAs down-regulated under stress conditions contribute to plant response to stress, whereas miRNAs that are up-regulated control genes associated with growth and development.<sup>67</sup> Soybean fields were screened for tolerance to soybean iron chlorosis deficiency (abiotic stress in soybean) using linear discriminant analysis (LDA) and SVM. The phenotypic data obtained from soybean fields were used in model training and predicting soybeans' iron chlorosis deficiency stress. The ML application has helped evaluate the severity of real-time stress in the soybean sector.<sup>95</sup>

### 3.5 Plant-pathogen interaction and diseases prediction

Plant diseases and pests pose a significant threat to agriculture. Early identification of plant diseases and pests would aid in developing effective treatment strategies while economic losses are mitigated.<sup>100</sup> Diverse ML approaches for precise disease recognition and prediction have been implemented in plant populations.<sup>101,102</sup> Neural networks (NNs) have achieved impressive results in plant disease prediction using image classification. A deep convolutional network was implemented in leaf image classification model for disease recognition. The developed model showed a high predictive performance in distinguishing plant leaves from their surroundings and recognized 13 plant diseases types on healthy leaves.<sup>103</sup> In another approach, a heterogeneous ensemble of deep-learning-based neural network models was used in detecting tomato plants diseases and pests using images collected on-site by imaging devices of varying resolutions.

The ensemble model successfully handled image complexity in the plant's surrounding area and recognized nine different diseases and pests.<sup>100,104</sup> Therefore, deep CNNs are promising in automatic classification and detection of diseases traits from leaf images. In addition, CNN has shown optimal performance when implemented in plant–pathogen interaction, pest, and disease recognition in some studies. These studies include prediction of pests and diseases occurrence in cotton<sup>105</sup>; rice plant diseases and pests recognition<sup>106</sup>; rice blast disease prediction<sup>107,108</sup>; image-based potato tuber disease detection<sup>109</sup> and so on. The CNN model is a high-performing method for detecting plant diseases, and it can be implemented and optimized for practical applications.

SVM has also been used for weather-based rice blast prediction and has proven suitable for plant disease forecasting with incredible predictive accuracy. A world-first SVM-based web server for rice blast prediction was developed. Plant scientists and farmers have benefited from this tool, especially in their decision-making.<sup>110</sup> In the pixel-wise quantification and identification of powdery mildew diseased barley tissue, SVM classification was used to establish hypersensitive response spots using multispectral imaging of diseased barley plants. SVM application enabled precise automatic identification of barley interaction with powdery mildew.<sup>111</sup>

Recently, a data-driven ML approach named ApoplastP was proposed. RF classifier is the base algorithm for ApoplastP and has shown high performance in predicting protein localization in plant apoplast. At first, differences in the constituents of apoplastic and intracellular plant proteins were unknown. However, the advent of ApoplastP enabled the exploration of differences in the composition of plant proteins. The plant apoplast is integral to plant–pathogen interactions, transport, and intercellular signaling. Hence, integrating and optimizing machine learning algorithms in apoplastic localization prediction will aid functional studies and help predict whether an effector will localize to the apoplast or enter the plant cells.<sup>112</sup>

Also, RF has been implemented to build an inter-species protein–protein interaction (PPI) prediction model using *Arabidopsis*–pathogen PPI data acquired both experimentally and from PPI public databases-UniProt. A critical assessment of the model performance showed that random forest integration with linear statistical methods using sequence information and network attributes as model features resulted in substantial and robust improvement in performance.<sup>24</sup>

In addition, RF classification has been used to exploit protein biomarkers' potential in precision breeding using biomarkers generated assays of 104 potatoes (*Solanum tuberosum*) peptides. These peptides were selected using diagonal linear discriminant analysis, bagging, principal component analysis (PCA), and SVM and then classified with RF classifiers. The ML algorithms' application helped identify *Phytophthora infestans* resistance in leaves, tubers and its effect on plant yield using potato leaf secretome data.<sup>109</sup> Early disease detection enables farmers to use timely and targeted crop protection strategies. With the use of ML, researchers have improved the accuracy of object detection and recognition systems dramatically.

### 3.6 Challenges and future outlook

ML applications in plant genomics and agronomics have majorly contributed to efficient breeding of crops with desirable agronomic traits, plant phenotyping, genetic trait prediction, and precise disease prediction such as in rice, soybeans, maize, beans, *etc.*<sup>72,95,106,110</sup> However, several limitations still exist. Firstly, the black-box nature of some sophisticated ML algorithms inhibits interpretation. The plant research community is more interested and fascinated with the biological implications of the prediction than the accuracy of the predictive model. Hence, there is a need for further processing and careful interpretation of the predictive model output using conforming biological knowledge. Additionally, the dimensionality of omics datasets poses challenges such as multicollinearity, overfitting, and sparsity which are difficult to avoid. Though contemporary machine learning methods and the huge sample size can partially alleviate these problems, the model's accuracy can be significantly enhanced by using different fine-tuning, augmentation, and optimization techniques.<sup>107</sup> Data integration from various sources is necessary for GS-assisted breeding and other trait improvement approaches.<sup>113</sup> Simultaneous analysis of multiple omics datasets can advance our understanding of complex biological phenomena.<sup>78,79</sup> Another challenge is the limited and inconsistent information on plant–pathogen interaction phenotypic information. The ML models used in plant disease recognition can be extended by enriching the plant disease database with plant–pathogen interaction phenotype data. Developing more robust classification algorithms with an expanded number of diseases classes will improve plant disease recognition and forecasting.<sup>103,105,108,106</sup> Finally, a comprehensive plant database must be constructed to facilitate comparative studies and promote research collaborations on critical plant science problems.

## 4. Conclusion

Machine learning has shown tremendous promise in studying enormous high-dimensional data sets, although it is still limited in plant molecular studies application. An in-depth understanding of ML models will stimulate ML

implementation for plant biological data analysis. As sequenced plant genome data continues to accumulate, ML will accelerate all plant genomic research fields, including identifying genes associated with biotic and abiotic stress resistance and other genes with significant functions, understanding gene regulation mechanisms, exploring plant genome genetic framework, and estimating breeding values. These advancements would help agricultural researchers improve the quality and yield of crops with stronger tolerance to abiotic and biotic stress and other plant health-threatening issues.

## Data available statement

### Extended data

OSF: Extended data for “Machine learning algorithms: their applications in plant omics and agronomic traits’ improvement”, <https://doi.org/10.17605/OSF.IO/TE6GC>.<sup>14</sup>

Files included:

Supplementary Table 1. Published sequenced plant genomes. Hundreds of plant genomes have been sequenced and published since 2000. The statistics for each genome are taken from the publication, despite several model plants having significant updates to genome assemblies and gene counts. NA, data not available in publication; Mb, megabases; kb, kilobases.

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

## References

- Mousavi-Derazmahalleh M, Bayer PE, Hane JK, *et al.*: **Adapting legume crops to climate change using genomic approaches.** *Plant Cell Environ.* 2019; **42**: 6–19.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- United Nations Convention to Combat Desertification (UNCCD): Sustainable Development Knowledge Platform: n.d. (accessed August 24, 2020).  
[Reference Source](#)
- Zhang Q, Xu M, Xia X, *et al.*: **Crop genetics research in Asia: improving food security and nutrition.** *Theor. Appl. Genet.* 2020; **133**: 1339–1344.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature.* 2000; **408**: 796–815.  
[Publisher Full Text](#)
- Michael TP, VanBuren R: **Progress, challenges and the future of crop genomes.** *Curr. Opin. Plant Biol.* 2015; **24**: 71–81.  
[Publisher Full Text](#)
- Borevitz JO, Ecker JR: **PLANT GENOMICS: The Third Wave.** *Annu. Rev. Genom. Hum. Genet.* 2004; **5**: 443–477.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Abadi S, Yan WX, Amar D, *et al.*: **A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action.** *PLoS Comput. Biol.* 2017; **13**: e1005807.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Marri PR, Ye L, Jia Y, *et al.*: **Advances in Sequencing and Resequencing in Crop Plants.** Varshney RK, Pandey MK, Chitikineni A, editors. *Plant Genetics and Molecular Biology.* Cham: Springer International Publishing; 2018; pp. 11–35.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Akhtar MS, Alaraidh IA, Hakeem KR: **Experimental Approaches for Genome Sequencing.** Hakeem KR, Shaik NA, Banaganapalli B, *et al.*, editors. *Essentials of Bioinformatics, Volume III: In Silico Life Sciences: Agriculture.* Cham: Springer International Publishing; 2019; pp. 159–165.  
[Publisher Full Text](#)
- Huala E, Dickerman AW, Garcia-Hernandez M, *et al.*: **The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant.** *Nucleic Acids Res.* 2001; **29**: 102–105.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hu S, Li G, Yang J, *et al.*: **Aquatic Plant Genomics: Advances, Applications, and Prospects.** *Int. J. Genomics.* 2017; **2017**: 1–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Pryer KM, Schneider H, Zimmer EA, *et al.*: **Ann Banks, Deciding among green plants for whole genome studies.** *Trends Plant Sci.* 2002; **7**: 550–554.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cooper L, Jaiswal P: **The Plant Ontology: A Tool for Plant Genomics.** Edwards D, editor. *Plant Bioinformatics.* New York, NY: Springer New York; 2016; pp. 89–114.  
[Publisher Full Text](#)
- Isewon IM, Apata OR, Oluwamuyiwa FA, *et al.*: **Machine Learning Algorithms: Their Applications in Plant Omics and Agronomic Traits Improvement.** 2022.  
[Publisher Full Text](#)
- Chen F, Dong W, Zhang J, *et al.*: **The Sequenced Angiosperm Genomes and Genome Databases.** *Front. Plant Sci.* 2018; **9**.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rai A, Yamazaki M, Saito K: **A new era in plant functional genomics.** *Curr. Opin. Syst. Biol.* 2019; **15**: 58–67.  
[Publisher Full Text](#)
- Abdurakhmonov IY: **Genomics Era for Plants and Crop Species – Advances Made and Needed Tasks Ahead.** Abdurakhmonov IY, editor. *Plant Genomics.* InTech; 2016.  
[Publisher Full Text](#)
- Shrivastava P, Kumar R: **Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation.** *Saudi J. Biol. Sci.* 2015; **22**: 123–131.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Terry N, Rouzé P, Montagu MV: **Plant genomics.** *FEBS Lett.* 1999; **452**: 3–6.  
[Publisher Full Text](#)
- Exposito-Alonso M, Drost H-G, Burbano HA, *et al.*: **The Earth BioGenome project: opportunities and challenges for plant genomics and conservation.** *Plant J.* 2020; **102**: 222–229.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Holtorf H, Guitton M-C, Reski R: **Plant functional genomics.** *Naturwissenschaften.* 2002; **89**: 235–249.  
[Publisher Full Text](#)
- Yang Y, Saand MA, Huang L, *et al.*: **Applications of Multi-Omics Technologies for Crop Improvement.** *Front. Plant Sci.* 2021; **12**:

1846.  
[PubMed Abstract](#) | [Publisher Full Text](#)
23. Muthamilarasan M, Singh NK, Prasad M: **Multi-omics approaches for strategic improvement of stress tolerance in underutilized crop species: A climate change perspective.** *Adv. Genet.* 2019; **103**: 1–38.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  24. Yang S, Li H, He H, *et al.*: **Critical assessment and performance improvement of plant-pathogen protein-protein interaction prediction methods.** *Brief. Bioinformatics.* 2019; **20**: 274–287.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  25. Babar MM, Zaidi N-SS, Pothineni VR, *et al.*: **Application of Bioinformatics and System Biology in Medicinal Plant Studies.** Hakeem KR, Malik A, Vardar-Sukan F, *et al.*, editors. *Plant Bioinformatics: Decoding the Phyta.* Cham: Springer International Publishing; 2017; pp. 375–393.  
[Publisher Full Text](#)
  26. Wolfender J-L, Rudaz S, Hae Choi Y, *et al.*: **Plant Metabolomics: From Holistic Data to Relevant Biomarkers.** *Curr. Med. Chem.* 2013; **20**: 1056–1090.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  27. Kempinski C, Jiang Z, Bell S, *et al.*: **Metabolic Engineering of Higher Plants and Algae for Isoprenoid Production.** Schrader J, Bohlmann J, editors. *Biotechnology of Isoprenoids.* Cham: Springer International Publishing; 2015; pp. 161–199.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  28. Kumar J, Pratap A, Kumar S: **Plant Phenomics: An Overview.** Kumar J, Pratap A, Kumar S, editors. *Phenomics in Crop Plants: Trends.* Springer India, New Delhi: Options and Limitations; 2015; pp. 1–10.  
[Publisher Full Text](#)
  29. Tshikunde NM, Mashilo J, Shimelis H, *et al.*: **Agronomic and Physiological Traits, and Associated Quantitative Trait Loci (QTL) Affecting Yield Response in Wheat (*Triticum aestivum* L.): A Review.** *Front. Plant Sci.* 2019; **10**.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  30. Fan G, Liu X, Sun S, *et al.*: **The Chromosome Level Genome and Genome-wide Association Study for the Agronomic Traits of Panax Notoginseng.** *iScience.* 2020; **23**: 101538.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  31. Varshney RK, Bohra A, Yu J, *et al.*: **Designing Future Crops: Genomics-Assisted Breeding Comes of Age.** *Trends Plant Sci.* 2021; **26**: 631–649.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  32. Varshney RK, Singh VK, Kumar A, *et al.*: **Can genomics deliver climate-change ready crops?** *Curr. Opin. Plant Biol.* 2018; **45**: 205–211.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  33. Thudi M, Palakurthi R, Schnable JC, *et al.*: **Genomic resources in plant breeding for sustainable agriculture.** *J. Plant Physiol.* 2021; **257**: 153351.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  34. Zenda T, Liu S, Dong A, *et al.*: **Advances in Cereal Crop Genomics for Resilience under Climate Change.** *Life (Basel, Switzerland).* 2021; **11**.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  35. Bamba M, Kawaguchi YW, Tsuchimatsu T: **Plant adaptation and speciation studied by population genomic approaches.** *Develop. Growth Differ.* 2019; **61**: 12–24.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  36. Gedil M, Menkir A: **An Integrated Molecular and Conventional Breeding Scheme for Enhancing Genetic Gain in Maize in Africa.** *Front. Plant Sci.* 2019; **10**: 1430.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  37. Jha UC, Sharma KD, Nayyar H, *et al.*: **Breeding and Genomics Interventions for Developing Ascochyta Blight Resistant Grain Legumes.** *Int. J. Mol. Sci.* 2022; **23**: 2217.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  38. Kole C, Muthamilarasan M, Henry R, *et al.*: **Application of genomics-assisted breeding for generation of climate resilient crops: progress and prospects.** *Front. Plant Sci.* 2015; **6**.  
[Publisher Full Text](#)
  39. Hu H, Scheben A, Edwards D: **Advances in Integrating Genomics and Bioinformatics in the Plant Breeding Pipeline.** *Agriculture.* 2018; **8**: 75.  
[Publisher Full Text](#)
  40. Singh RK, Prasad A, Muthamilarasan M, *et al.*: **Breeding and biotechnological interventions for trait improvement: status and prospects.** *Planta.* 2020; **252**: 54.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  41. Jia B, Conner RL, Penner WC, *et al.*: **Quantitative Trait Locus Mapping of Marsh Spot Disease Resistance in Cranberry Common Bean (*Phaseolus vulgaris* L.).** *Int. J. Mol. Sci.* 2022; **23**: 7639.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  42. Garg V, Agarwal G, Pazhamala LT, *et al.*: **Genome-Wide Identification, Characterization, and Expression Analysis of Small RNA Biogenesis Purveyors Reveal Their Role in Regulation of Biotic Stress Responses in Three Legume Crops.** *Front. Plant Sci.* 2017; **8**.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  43. Barreto FZ, Rosa JRBF, Balsalobre TWA, *et al.*: **A genome-wide association study identified loci for yield component traits in sugarcane (*Saccharum spp.*).** *PLoS One.* 2019; **14**: e0219843.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  44. Sandhu N, Yadav S, Catolos M, *et al.*: **Developing Climate-Resilient, Direct-Seeded, Adapted Multiple-Stress-Tolerant Rice Applying Genomics-Assisted Breeding.** *Front. Plant Sci.* 2021; **12**.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  45. He Z, Xin Y, Wang C, *et al.*: **Genomics-Assisted Improvement of Super High-Yield Hybrid Rice Variety "Super 1000" for Resistance to Bacterial Blight and Blast Diseases.** *Front. Plant Sci.* 2022; **13**: 881244.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  46. Sandhu N, Singh J, Singh G, *et al.*: **Development and validation of a novel core set of KASP markers for the traits improving grain yield and adaptability of rice under direct-seeded cultivation conditions.** *Genomics.* 2022; **114**: 110269.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  47. Bohra A, Saxena KB, Varshney RK, *et al.*: **Genomics-assisted breeding for pigeonpea improvement.** *Theor. Appl. Genet.* 2020; **133**: 1721–1737.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  48. Mohanty JK, Jha UC, Dixit GP, *et al.*: **Harnessing the hidden allelic diversity of wild Cicer to accelerate genomics-assisted chickpea crop improvement.** *Mol. Biol. Rep.* 2022; **49**: 5697–5715.  
[Publisher Full Text](#)
  49. Miedaner T, Boeven ALG-C, Gaikpa DS, *et al.*: **Genomics-Assisted Breeding for Quantitative Disease Resistances in Small-Grain Cereals and Maize.** *Int. J. Mol. Sci.* 2020; **21**: E9717.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  50. Tao Y, Zhao X, Mace E, *et al.*: **Exploring and Exploiting Pan-genomics for Crop Improvement.** *Mol. Plant.* 2019; **12**: 156–169.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  51. Muthamilarasan M, Theriappan P, Prasad M: **Recent advances in crop genomics for ensuring food security.** *Curr. Sci.* 2013; **105**: 155–158.
  52. Gordon SP, Contreras-Moreira B, Woods DP, *et al.*: **Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure.** *Nat. Commun.* 2017; **8**: 2184.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  53. Lin K, Zhang N, Severing EI, *et al.*: **Beyond genomic variation - comparison and functional annotation of three Brassica rapagenomes: a turnip, a rapid cycling and a Chinese cabbage.** *BMC Genomics.* 2014; **15**: 250.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  54. Golicz AA, Bayer PE, Barker GC, *et al.*: **The pan-genome of an agronomically important crop plant Brassica oleracea.** *Nat. Commun.* 2016; **7**: 13390.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  55. Hurgobin B, Golicz AA, Bayer PE, *et al.*: **Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid Brassica napus.** *Plant Biotechnol. J.* 2018; **16**: 1265–1274.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  56. Hirsch CN, Foerster JM, Johnson JM, *et al.*: **Insights into the Maize Pan-Genome and Pan-Transcriptome.** *Plant Cell.* 2014; **26**: 121–135.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  57. Wang W, Mauleon R, Hu Z, *et al.*: **Genomic variation in 3,010 diverse accessions of Asian cultivated rice.** *Nature.* 2018; **557**: 43–49.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  58. Zhao Q, Feng Q, Lu H, *et al.*: **Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice.** *Nat. Genet.* 2018; **50**: 278–284.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  59. Li Y, Zhou G, Ma J, *et al.*: **De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits.** *Nat. Biotechnol.* 2014; **32**: 1045–1052.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  60. Montenegro JD, Golicz AA, Bayer PE, *et al.*: **The pan-genome of hexaploid bread wheat.** *Plant J.* 2017; **90**: 1007–1013.  
[PubMed Abstract](#) | [Publisher Full Text](#)

61. Neeta L, Golicz A, Singh M, *et al.*: **Genome-wide analysis of the Hsf gene family in Brassica oleracea and a comparative analysis of the Hsf gene family in B. oleracea, B. rapa and B. napus.** *Funct. Integr. Genomics.* 2019; **19**: 515–531.  
[PubMed Abstract](#) | [Publisher Full Text](#)
62. Ndjondjop MN, Alachiotis N, Pavlidis P, *et al.*: **Comparisons of molecular diversity indices, selective sweeps and population structure of African rice with its wild progenitor and Asian rice.** *Theor. Appl. Genet.* 2019; **132**: 1145–1158.  
[PubMed Abstract](#) | [Publisher Full Text](#)
63. Nevo E: **Ecological genomics of natural plant populations: the Israeli perspective.** *Methods Mol. Biol.* 2009; **513**: 321–344.  
[PubMed Abstract](#) | [Publisher Full Text](#)
64. Olatoye MO, Hu Z, Maina F, *et al.*: **Genomic Signatures of Adaptation to a Precipitation Gradient in Nigerian Sorghum.** *G3 (Bethesda).* 2018; **8**: 3269–3281.  
[PubMed Abstract](#) | [Publisher Full Text](#)
65. Juliana P, Poland J, Huerta-Espino J, *et al.*: **Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics.** *Nat. Genet.* 2019; **51**: 1530–1539.  
[PubMed Abstract](#) | [Publisher Full Text](#)
66. Riccroch AE: **What Will Be the Benefits of Biotech Wheat for European Agriculture?** Bhalla PL, Singh MB, editors. *Wheat Biotechnology: Methods and Protocols.* New York, NY: Springer; 2017; pp. 25–35.  
[Publisher Full Text](#)
67. Sanz-Carbonell A, Marques MC, Bustamante A, *et al.*: **Inferring the regulatory network of the miRNA-mediated response to biotic and abiotic stress in melon.** *BMC Plant Biol.* 2019; **19**: 78.  
[PubMed Abstract](#) | [Publisher Full Text](#)
68. Mochida K, Shinozaki K: **Genomics and Bioinformatics Resources for Crop Improvement.** *Plant Cell Physiol.* 2010; **51**: 497–523.  
[PubMed Abstract](#) | [Publisher Full Text](#)
69. Katuramu DN, Branham SE, Levi A, *et al.*: **Genome-Wide Association Analysis of Resistance to Pseudoperonospora cubensis in Citron Watermelon.** *Plant Dis.* 2022; **106**: 1952–1958.  
[PubMed Abstract](#) | [Publisher Full Text](#)
70. Teixidor-Toneu I, Jordan FM, Hawkins JA: **Comparative phylogenetic methods and the cultural evolution of medicinal plant use.** *Nat Plants.* 2018; **4**: 754–761.  
[Publisher Full Text](#)
71. Guidini R, Jahani M, Huang K, *et al.*: **Genome Wide Association Mapping in Sunflower (*Helianthus annuus* L.) reveals Common Loci and Putative Candidate Genes for resistance to Diaporthe gulyae and D. helianthi causing Phomopsis Stem Canker.** *Plant Dis.* 2022.  
[Publisher Full Text](#)
72. Chakradhar T, Hindu V, Reddy PS: **Genomic-based-breeding tools for tropical maize improvement.** *Genetica.* 2017; **145**: 525–539.  
[PubMed Abstract](#) | [Publisher Full Text](#)
73. Sun S, Wang C, Ding H, *et al.*: **Machine learning and its applications in plant molecular studies.** *Brief. Funct. Genomics.* 2020; **19**: 40–48.  
[Publisher Full Text](#)
74. Alzubi J, Nayyar A, Kumar A: **Machine Learning from Theory to Algorithms: An Overview.** *J. Phys. Conf. Ser.* 2018; **1142**: 012012.  
[Publisher Full Text](#)
75. Carbonell JG, Michalski RS, Mitchell TM: **1 - AN OVERVIEW OF MACHINE LEARNING.** Michalski RS, Carbonell JG, Mitchell TM, editors. *Machine Learning.* San Francisco (CA): Morgan Kaufmann; 1983; pp. 3–23.  
[Publisher Full Text](#)
76. Matukumalli LK, Grefenstette JJ, Hyten DL, *et al.*: **Application of machine learning in SNP discovery.** *BMC Bioinformatics.* 2006; **7**: 4.  
[PubMed Abstract](#) | [Publisher Full Text](#)
77. Esposito S, Carputo D, Cardì T, *et al.*: **Applications and Trends of Machine Learning in Genomics and Phenomics for Next-Generation Breeding.** *Plants.* 2020; **9**: 34.  
[PubMed Abstract](#) | [Publisher Full Text](#)
78. Zhang N, Rao RSP, Salvato F, *et al.*: **MU-LOC: A Machine-Learning Method for Predicting Mitochondrially Localized Proteins in Plants.** *Front. Plant Sci.* 2018; **9**.  
[PubMed Abstract](#) | [Publisher Full Text](#)
79. Kushwaha SK, Chauhan P, Hedlund K, *et al.*: **NBSPred: a support vector machine-based high-throughput pipeline for plant resistance protein NBSLR prediction.** *Bioinformatics.* 2016; **32**: 1223–1225.  
[PubMed Abstract](#) | [Publisher Full Text](#)
80. Carvalho TFM, Silva JCF, Calil IP, *et al.*: **Rama: a machine learning approach for ribosomal protein prediction in plants.** *Sci. Rep.* 2017; **7**: 16273.  
[PubMed Abstract](#) | [Publisher Full Text](#)
81. Pal T, Jaiswal V, Chauhan R: **DRPPP: A machine learning based tool for prediction of disease resistance proteins in plants.** *Comput. Biol. Med.* 2016; **78**: 42–48.  
[PubMed Abstract](#) | [Publisher Full Text](#)
82. Silva JC, Carvalho T, Fontes E, *et al.*: **Fangorn Forest (F2): a machine learning approach to classify genes and genera in the family Geminiviridae.** *BMC Bioinformatics.* 2017; **18**: 431.  
[PubMed Abstract](#) | [Publisher Full Text](#)
83. Ma C, Xin M, Feldmann KA, *et al.*: **Machine Learning-Based Differential Network Analysis: A Study of Stress-Responsive Transcriptomes in Arabidopsis.** *Plant Cell.* 2014; **26**: 520–537.  
[PubMed Abstract](#) | [Publisher Full Text](#)
84. van Dijk ADJ, Kootstra G, Kruijer W, *et al.*: **Machine learning in plant science and plant breeding.** *IScience.* 2021; **24**: 101890.  
[PubMed Abstract](#) | [Publisher Full Text](#)
85. Niaziyan M, Niedbala G: **Machine Learning for Plant Breeding and Biotechnology.** *Agriculture.* 2020; **10**: 436.  
[Publisher Full Text](#)
86. Guzzetta G, Jurman G, Furlanello C: **A machine learning pipeline for quantitative phenotype prediction from genotype data.** *BMC Bioinformatics.* 2010; **11**: S3.  
[PubMed Abstract](#) | [Publisher Full Text](#)
87. Crossa J, Pérez-Rodríguez P, Cuevas J, *et al.*: **Genomic Selection in Plant Breeding: Methods, Models, and Perspectives.** *Trends Plant Sci.* 2017; **22**: 961–975.  
[PubMed Abstract](#) | [Publisher Full Text](#)
88. Peixoto LA, Bhering LL, Cruz CD: **Artificial neural networks reveal efficiency in genetic value prediction.** *Genet. Mol. Res.* 2015; **14**: 6796–6807.  
[PubMed Abstract](#) | [Publisher Full Text](#)
89. Corrêa AM, Teodoro PE, Gonçalves MC, *et al.*: **Artificial intelligence in the selection of common bean genotypes with high phenotypic stability.** *Genet. Mol. Res.* 2016; **15**.  
[Publisher Full Text](#)
90. Montesinos-López OA, Montesinos-López A, Tuberosa R, *et al.*: **Multi-Trait, Multi-Environment Genomic Prediction of Durum Wheat With Genomic Best Linear Unbiased Predictor and Deep Learning Methods.** *Front. Plant Sci.* 2019; **10**.  
[Publisher Full Text](#)
91. González-Camacho JM, Ornella L, Pérez-Rodríguez P, *et al.*: **Applications of Machine Learning Methods to Genomic Selection in Breeding Wheat for Rust Resistance.** *Plant Genome.* 2018; **11**: 170104.  
[PubMed Abstract](#) | [Publisher Full Text](#)
92. Zhao BG, Rewald B: **Phenotyping: Using Machine Learning for Improved Pairwise Genotype Classification Based on Root Traits.** *Front. Plant Sci.* 2016; **7**.  
[Publisher Full Text](#)
93. Endelman JB, Carley CAS, Bethke PC, *et al.*: **Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato.** *Genetics.* 2018; **209**: 77–87.  
[PubMed Abstract](#) | [Publisher Full Text](#)
94. Selvaraj MG, Valderrama M, Guzman D, *et al.*: **Machine learning for high-throughput field phenotyping and image processing provides insight into the association of above and below-ground traits in cassava (Manihot esculenta Crantz).** *Plant Methods.* 2020; **16**: 87.  
[PubMed Abstract](#) | [Publisher Full Text](#)
95. Dobbels AA, Lorenz AJ: **Soybean iron deficiency chlorosis high throughput phenotyping using an unmanned aircraft system.** *Plant Methods.* 2019; **15**: 97.  
[PubMed Abstract](#) | [Publisher Full Text](#)
96. Hasan MM, Chopin JP, Laga H, *et al.*: **Detection and analysis of wheat spikes using Convolutional Neural Networks.** *Plant Methods.* 2018; **14**: 100.  
[PubMed Abstract](#) | [Publisher Full Text](#)
97. de Carvalho RRB, Marmolejo Cortes DF, Sousa MBE, *et al.*: **Image-based phenotyping of cassava roots for diversity studies and carotenoids prediction.** *PLoS One.* 2022; **17**: e0263326.  
[PubMed Abstract](#) | [Publisher Full Text](#)
98. Sprenger H, Erban A, Seddig S, *et al.*: **Metabolite and transcript markers for the prediction of potato drought tolerance.** *Plant Biotechnol. J.* 2018; **16**: 939–950.  
[PubMed Abstract](#) | [Publisher Full Text](#)
99. Parmley KA, Higgins RH, Ganapathysubramanian B, *et al.*: **Machine Learning Approach for Prescriptive Plant Breeding.** *Sci. Rep.* 2019; **9**: 17132.  
[PubMed Abstract](#) | [Publisher Full Text](#)
100. Fuentes A, Yoon S, Kim SC, *et al.*: **A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests**

- Recognition.** *Sensors (Basel)*. 2017; **17**.  
[PubMed Abstract](#) | [Publisher Full Text](#)
101. Danilevicz MF, Gill M, Anderson R, *et al.*: **Plant Genotype to Phenotype Prediction Using Machine Learning.** *Front. Genet.* 2022; **13**: 822173.  
[PubMed Abstract](#) | [Publisher Full Text](#)
102. Simón D, Borsani O, Filippi CV: **RFPDR: a random forest approach for plant disease resistance protein prediction.** *PeerJ.* 2022; **10**: e11683.  
[PubMed Abstract](#) | [Publisher Full Text](#)
103. Sladojevic S, Arsenovic M, Anderla A, *et al.*: **Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification.** *Comput. Intell. Neurosci.* 2016; **2016**: 1–11.  
[PubMed Abstract](#) | [Publisher Full Text](#)
104. Wang Q, Qi F, Sun M, *et al.*: **Identification of Tomato Disease Types and Detection of Infected Areas Based on Deep Convolutional Neural Networks and Object Detection Techniques.** *Comput. Intell. Neurosci.* 2019; **2019**: 9142715–9142753.  
[PubMed Abstract](#) | [Publisher Full Text](#)
105. Xiao Q, Li W, Kai Y, *et al.*: **Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network.** *BMC Bioinformatics.* 2019; **20**: 688.  
[PubMed Abstract](#) | [Publisher Full Text](#)
106. Li D, Wang R, Xie C, *et al.*: **A Recognition Method for Rice Plant Diseases and Pests Video Detection Based on Deep Convolutional Neural Network.** *Sensors (Basel)*. 2020; **20**.  
[PubMed Abstract](#) | [Publisher Full Text](#)
107. Liang W-J, Zhang H, Zhang G-F, *et al.*: **Rice Blast Disease Recognition Using a Deep Convolutional Neural Network.** *Sci. Rep.* 2019; **9**: 2869.  
[PubMed Abstract](#) | [Publisher Full Text](#)
108. Nettleton DF, Katsantonis D, Kalaitzidis A, *et al.*: **Predicting rice blast disease: machine learning versus process-based models.** *BMC Bioinformatics.* 2019; **20**: 514.  
[PubMed Abstract](#) | [Publisher Full Text](#)
109. Oppenheim D, Shani G, Erlich O, *et al.*: **Using Deep Learning for Image-Based Potato Tuber Disease Detection.** *Phytopathology.* 2019; **109**: 1083–1087.  
[Publisher Full Text](#)
110. Kaundal R, Kapoor AS, Raghava GPS: **Machine learning techniques in disease forecasting: a case study on rice blast prediction.** *BMC Bioinformatics.* 2006; **7**: 485.  
[PubMed Abstract](#) | [Publisher Full Text](#)
111. Kuska MT, Behmann J, Großkinsky DK, *et al.*: **Screening of Barley Resistance Against Powdery Mildew by Simultaneous High-Throughput Enzyme Activity Signature Profiling and Multispectral Imaging.** *Front. Plant Sci.* 2018; **9**: 1074.  
[Publisher Full Text](#)
112. Sperschneider J, Dodds PN, Singh KB, *et al.*: **ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning.** *New Phytol.* 2018; **217**: 1764–1778.  
[PubMed Abstract](#) | [Publisher Full Text](#)
113. Montesinos-López OA, Montesinos-López A, Pérez-Rodríguez P, *et al.*: **A review of deep learning applications for genomic selection.** *BMC Genomics.* 2021; **22**: 19.  
[PubMed Abstract](#) | [Publisher Full Text](#)

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**