Multi-Omics Approaches in Orphan Crops: Bridging the Gap Between Botanical Diversity and Global Food Security

Zahir Uddin [1]*

Abstract

Modern agriculture leans, perhaps too heavily, on a narrow handful of staple species, and that dependence is beginning to look less like efficiency and more like fragility. As climate volatility intensifies and diets grow monotonous, orphan or underutilized crops - long overlooked by formal breeding programs - are being reconsidered as a practical route toward dietary diversification and climate resilience. This narrative review synthesizes literature published between 2011 and 2026, drawing on genomic, pangenomic, transcriptomic, metabolomic, and phenomic studies of underutilized crop species, alongside research on CRISPR/Cas-based domestication and AI-driven phenotyping frameworks. The evidence indicates that long-read sequencing and graph-based pangenomes have begun resolving structural variation once hidden by single reference genomes, revealing presence-absence variants tied to drought, heat, and nutritional traits in species such as finger millet, tef, and pigeonpea. Precision editing tools, particularly CRISPR/Cas9 and base/prime editors, have already converted several wild or semi-domesticated relatives into agronomically workable crops, while machine-learning-assisted phenomics has improved trait prediction accuracy considerably. Taken together, these findings support an integrated De novo domestication-Speed breeding-AI-empowered Phenomics (DSAP) strategy as a credible, though not yet fully proven, roadmap for repositioning orphan crops within global food systems, provided that open-access genomic infrastructure and participatory breeding are pursued alongside the technology itself.

Keywords

orphan crops; multi-omics integration; pangenomics; CRISPR/Cas gene editing; food security

1. Introduction

The global food system is, one has to admit, standing on somewhat shakier ground than its productivity statistics would suggest. Population projections point toward nearly ten billion people by mid-century (Amas et al., 2026; Manoj Kumar et al., 2026; Gelaye et al., 2025), implying that agricultural output must climb 50% to 56% over 2010 levels just to keep pace with demand (Zenda et al., 2021; Kumar et al., 2021b). That target alone would be daunting under stable conditions; it is considerably less so amid shifting rainfall, warming averages, and the slow creep of soil salinization now destabilizing agro-ecosystems in ways not fully anticipated a generation ago (Anand et al., 2026; Ashraf et al., 2022; Zenda et al., 2021).

Significance | Repositioning orphan crops via multi-omics and gene editing could meaningfully diversify global diets and strengthen climate-resilient food security.

*Correspondence.

Zahir Uddin, Department of Botany, University of Chittagong, Hathazari, Chattogram 4331, Bangladesh. Email: zahiruddin01988@gmail.com.

Editor Atiqur Rahman Sunny, Ph.D. And accepted by the Editorial Board Aug 24, 2026 (received for review Jun 29, 2026)

Author Affiliation:

[1] Department of Botany, University of Chittagong, Hathazari, Chattogram 4331, Bangladesh.

Please cite this article as: , Z. U. (2026). "Multi-Omics Approaches in Orphan Crops: Bridging the Gap Between Botanical Diversity and Global Food Security", Applied Agriculture Sciences, 4(1), 1-20, 10911

3066-3407/© 2026 (Applied Agriculture Sciences), a publication of Eman Research, USA.
This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)
(https://publishing.emanresearch.org)

This pressure arrives, uncomfortably, at a moment of unusual genetic narrowness. Of the roughly 350,000 to 390,000 vascular plant species on Earth, only 5,000 to 7,000 have ever been cultivated as food, with barely 250 fully domesticated (Wang & Xiang, 2025; Kumar et al., 2021b). More strikingly, about 95% of humanity's caloric intake traces back to some thirty crop species, and just five - wheat, rice, maize, potato, and soybean - supply over 80% of the global agrifood base (Huang et al., 2025; Wang & Xiang, 2025). It is a strange kind of abundance: enormous yield, vanishingly small genetic breadth.

History has shown, more than once, what that narrowness can cost. The Irish Potato Famine of 1845-1850 remains the textbook case - a near-total reliance on one uniform potato variety left the crop defenseless against Phytophthora infestans, killing over a million people and displacing roughly a million more (Kumar et al., 2021b; Huang et al., 2025); the 1970 Southern corn leaf blight epidemic told a smaller, structurally similar story (Kumar et al., 2021b). Quieter but more chronic, dietary homogenization has left over two billion people, disproportionately women and children in low-income regions, facing "hidden hunger" that calorie-sufficient diets can mask entirely (Rasool & Qadir, 2026; Jamie et al., 2026; Zenda et al., 2021), since cereal staples deliver carbohydrates but fall short on protein and micronutrients (Huang et al., 2025; Wang & Xiang, 2025). As conventional breeding of major staples nears physiological yield ceilings, diversifying the crop portfolio through orphan crops looks increasingly like the more defensible route toward the UN's "Zero Hunger" target (Gelaye et al., 2025; Huang et al., 2025; Wang & Xiang, 2025).

Orphan crops - variously called neglected, traditional, or underutilized species - are less a single botanical category than a loose coalition of regionally vital plants that never reached the global commodity stage despite deep roots in local food cultures (Babele et al., 2022; Gelaye et al., 2025; Wang & Xiang, 2025). Having evolved under localized, often harsh pressures, many tolerate marginal soils, drought, and pests with minimal agrochemical input, and frequently carry nutritional profiles that put mainstream staples to shame (Sanfeliu Meliá & Cárdenas, 2026; Gelaye et al., 2025). Finger millet (Eleusine coracana) is a drought-hardy allotetraploid rich in calcium and iron (Babele et al., 2022); pigeonpea (Cajanus cajan) offers protein critical to semi-arid regions (Singh et al., 2020); moringa (Moringa oleifera) ranks among the most nutrient-dense leafy greens known (Kumar et al., 2018); and purslane (Portulaca oleracea) tolerates salinity while supplying omega-3 fatty acids and antioxidants (Kumar et al., 2021a).

None of this means orphan crops are simply waiting to be adopted. Many carry agronomic baggage - seed shattering, indeterminate growth, anti-nutritional compounds, long crop cycles - that decade-plus conventional breeding timelines are too slow to fix (Sanfeliu Meliá & Cárdenas, 2026; Huang et al., 2025; Rasool & Qadir, 2026). What has changed is the toolkit. Long-read sequencing now enables cost-effective reference genomes for orphan species (Anand et al., 2026; Gelaye et al., 2025), even though a single reference still cannot capture the structural variation governing stress tolerance across diverse germplasm (Hu et al., 2025; Zenda et al., 2021) - hence the field's move into the pangenomics era (Hu et al., 2025; Huang et al., 2025). Layered atop this is an integrated multi-omics approach - genomics, transcriptomics, proteomics, metabolomics, and phenomics - building a systems-level picture of plant behavior under stress (Cortés et al., 2023; Mahmood et al., 2022; Zenda et al., 2021), paired increasingly with CRISPR/Cas editing (Anand et al., 2026; Manoj Kumar et al., 2026) and the newly proposed DSAP strategy - de novo domestication, speed breeding, and AI-empowered phenomics (Huang et al., 2025).

Even so, it would be premature to call this solved. Moving from laboratory discovery to a variety a smallholder farmer actually plants remains, arguably, the harder half of the equation, requiring strategies linking multi-omics pipelines with participatory breeding, decentralized seed systems, and policy support (Babele et al., 2022; Gelaye et al., 2025). This review therefore takes stock of where genomic and multi-omic research on orphan crops currently stands, examines how these platforms correct agronomically unfavorable traits, and sketches a translational path toward repositioning orphan crops as cornerstones of a more diverse, food-secure future (Gelaye et al., 2025; Huang et al., 2025; Wang & Xiang, 2025).

2. The Modern Biotechnological Toolkit for Orphan Crops

Having framed the problem in the introduction, it seems worth pausing here to walk more slowly through what the existing literature actually says - not just about why orphan crops matter, but about the specific technical and

institutional machinery that researchers have been assembling to work with them. The picture that emerges, admittedly, is somewhat fragmented across disciplines - genebank science, structural genomics, gene editing, and computational phenomics do not always cite one another as often as they probably should - but taken together they sketch a reasonably coherent trajectory.

2.1 The Agrobiodiversity Bottleneck and the Case for Genetic Resources

It is tempting to think of modern agriculture's yield gains as an unambiguous success story, and in narrow productivity terms, they are. Conventional breeding and intensive monoculture have driven remarkable output increases over the past century. Yet several authors have argued, convincingly, that this same trajectory has quietly engineered a precarious bottleneck (Huang et al., 2025; Sanfeliu Meliá et al., 2026). Of an estimated 390,000 vascular plant species, only 5,000 to 7,000 have ever entered cultivation, and a mere 250 have undergone anything resembling true domestication (Wang & Xiang, 2025). More troubling still, over 80% of global agricultural output now rests on five staple crops (Huang et al., 2025; Sanfeliu Meliá et al., 2026) - a level of concentration that, as Hu et al. (2025) note, leaves food systems structurally exposed to climate shocks, emergent pathogens, and resource scarcity.

The literature returns again and again to the Irish Potato Famine of 1845-1850 as a cautionary reference point (Huang et al., 2025), and not without reason - it remains perhaps the clearest historical demonstration of what genetic uniformity can cost a population. What is less often discussed, but arguably just as important, is the demographic pressure layered on top of this vulnerability: population projections now point toward roughly 10.3 billion people by the 2080s (Huang et al., 2025), even as droughts, thermal extremes, salinity, and transboundary pests threaten to erode staple yields further (Gelaye et al., 2025; Diakite et al., 2026). Several reviews converge on the same conclusion here - that traditional breeding, built around multi-generational hybridization and phenotypic selection, is simply too slow to keep pace (Gelaye et al., 2025; Rasool & Qadir, 2026). This is essentially the argument for shifting toward a precision-designed breeding paradigm that draws on both crop wild relatives and modern genomic tools (Somegowda et al., 2024; Hu et al., 2025).

2.2 Genebanks as Strategic Capital

Plant genetic resources - the seeds, tissues, landraces, and wild relatives that escaped the narrowing pressures of intensive selection - are described throughout this literature as a form of strategic capital (Ghamkhar, 2022), and it is hard to disagree with that framing once the scale involved becomes clear. Roughly 1,750 genebanks worldwide now conserve some 7.4 million accessions (Ghamkhar, 2022), a number that has only grown more valuable as the risks of narrow cultivation have become more apparent. Institutional case studies scattered through this literature help put that abstraction into perspective: ICRISAT alone conserves over 129,000 accessions across 144 countries, with particularly deep holdings in sorghum, minor millets, chickpea, and pigeonpea (Wang & Xiang, 2025); India's National Bureau of Plant Genetic Resources holds tens of thousands of sorghum, millet, and chickpea accessions (Wang & Xiang, 2025); and IITA maintains what is likely the world's most diverse cowpea collection alongside substantial yam and Bambara groundnut holdings (Wang & Xiang, 2025). Table 5 summarizes these holdings across the five largest repositories.

What several authors flag as the more interesting development, though, is not the size of these collections but their changing character. As sequencing costs continue falling, genebanks are gradually moving away from being passive seed vaults and becoming, instead, active digitized innovation centers (Ghamkhar, 2022). Layering genomic, transcriptomic, and phenomic profiling directly onto conserved accessions allows breeders to largely bypass the long, labor-intensive cycles of field-based phenotypic screening (Ghamkhar & Richards, 2022; Hu et al., 2025) - a shift that Zenda et al. (2021) and Hu et al. (2025) both describe as central to genomic-assisted breeding (GAB), where virtual allele mining substitutes, at least partly, for the exhaustive evaluation of individual field plants.

2.3 From Single Reference Genomes to Pangenomes

One recurring theme across this body of work is a certain dissatisfaction with the single linear reference genome - not because it was ever wrong, exactly, but because it was never going to be enough. Hu et al. (2025) and Diakite et al. (2026) both describe "reference bias" as a structural limitation: a single assembly simply cannot capture the full spectrum of variation within a species, and

this problem becomes especially acute in heterozygous or polyploid orphan crops, where reference-based alignment tends to miss non-reference regions and complex structural variants altogether (Diakite et al., 2026; Hu et al., 2025). Table 2 lists the single-accession reference-genome assemblies now available for ten representative underutilized crop species.

The response the field has settled on - not universally, but widely - is the pangenome: an assembly built from multiple diverse individuals that partitions genetic content into a "core" genome shared across all individuals and a "variable" or dispensable genome present only in some (Zenda et al., 2021; Hu et al., 2025). What makes this distinction more than academic bookkeeping is where the interesting genes tend to sit. Comparative work across multiple species has shown, fairly consistently, that genes tied to environmental adaptation - disease resistance, drought tolerance, thermal response, nutrient acquisition - are disproportionately concentrated within this variable genome (Zenda et al., 2021; Hu et al., 2025). A pangenome of Glycine soja, the wild relative of cultivated soybean, illustrates this well: its variable genome comprised roughly 20% of total gene space, much of it showing signatures of positive selection around seed composition, flowering time, and biotic resistance (Zenda et al., 2021).

Graph-based pangenomes represent, in a sense, the natural next step - representing conserved regions as shared paths and structural variants as alternative branches, which allows more accurate read alignment and supports pangenome-wide association studies (Pan-GWAS) linking structural variants directly to phenotype (Hu et al., 2025). The examples that keep surfacing in this literature are instructive: in foxtail millet, graph pangenome analysis identified a 366-bp promoter-region variant in SiGW3 that suppresses gene expression to enhance grain weight (Hu et al., 2025); in pearl millet, similar analysis linked structural variants in endoplasmic-reticulum-related genes to heat stress adaptation via an expanded RWP-RK transcription factor family (Hu et al., 2025); and in chickpea, graph pangenomes have helped resolve superior haplotypes governing vernalization and disease resistance (Hu et al., 2025).

2.4 Precision Editing Tools and Delivery Platforms

If pangenomics tells researchers where to look, CRISPR/Cas systems are largely what has made it possible to act on that information with any precision (Rasool & Qadir, 2026). The mechanism, by now, is reasonably well established across the literature: a custom single-guide RNA directs an endonuclease such as Cas9 or Cas12a to a specific locus, inducing a double-strand break that the plant's own repair machinery - either non-homologous end joining or homology-directed repair - then resolves into a targeted disruption, deletion, or allele replacement (Rasool & Qadir, 2026; Mahmood et al., 2022). Table 3 catalogs representative functional target genes edited using these platforms across ten orphan crop species.

What has evolved more recently, and what several reviews treat as the more consequential development, is precision beyond simple knockouts. Base editors - cytosine and adenine variants - allow direct single-nucleotide conversion without inducing double-strand breaks or requiring a donor template, which meaningfully reduces unwanted indels (Rasool & Qadir, 2026). Prime editing goes further still, combining a nickase-impaired Cas9 with a reverse transcriptase to write new genetic sequence directly at a target site, enabling essentially any base-to-base conversion alongside small insertions and deletions, again without double-strand cleavage (Rasool & Qadir, 2026).

Even so, the literature is fairly candid that molecular precision alone does not guarantee translation into a living plant. Tissue transformation and regeneration remain persistent bottlenecks, particularly in recalcitrant or polyploid species (Rasool & Qadir, 2026; Manoj Kumar et al., 2026). The emerging response - nanoparticle-based and non-viral delivery systems, including lipid nanoparticles and polyester carriers - appears designed specifically to sidestep this problem, protecting fragile ribonucleoprotein complexes from degradation and enabling transgene-free, multiplexed editing without foreign DNA integration (Rasool & Qadir, 2026).

2.5 De Novo Domestication as a Case-Study Literature

Perhaps the most persuasive part of this literature, at least in terms of demonstrated results rather than promise, concerns de novo domestication - essentially compressing thousands of years of selection for "domestication syndrome" traits (reduced shattering, erect growth, day-length insensitivity, enlarged seed size) into a handful of growing seasons by editing known domestication-gene orthologs directly

Table 1: High-Value Breeding Targets and Candidate Genes Identified in Underutilized Crops This table lists eleven candidate genes reported across major orphan and underutilized crops, spanning tef, fonio, finger millet, Bambara groundnut, amaranth, jackfruit, taro, quinoa, sorghum, lentil, and cowpea. For each species it names the gene, its known or proposed function (e.g., seed-size regulation, stress-responsive transcription, metal chelation), and the specific breeding or agronomic trait the gene is expected to improve, such as harvestability, drought tolerance, or biofortification. All entries are drawn from a single synthesis source (Gelaye et al., 2025) rather than from the primary gene-discovery studies themselves — see cross-check note T1 below.

Crop	Species	Target Gene	Identified Function	Breeding and Agronomic Relevance
Tef	Eragrostis tef	EtG1 & EtG2	Regulation of seed and grain size	Enhances mechanical harvestability and overall yield (Gelaye et al., 2025)
Fonio	Digitaria exilis	FoWRKY	Abiotic stress transcription factor	Confers superior drought stress tolerance (Gelaye et al., 2025)
Finger Millet	Eleusine coracana	Rc (regulatory locus)	Calcium accumulation & storage	Drives outstanding calcium density in grain (Gelaye et al., 2025)
Bambara Groundnut	Vigna subterranea	GmFAD2	Desaturase metabolic enzyme	Optimizes seed oil composition and quality (Gelaye et al., 2025)
Amaranth	Amaranthus spp.	TPS	Trehalose-6-phosphate synthase	Confers superior salt and osmotic stress tolerance (Gelaye et al., 2025)
Jackfruit	Artocarpus heterophyllus	AhePG1	Polygalacturonase enzyme	Regulates fruit expansion and tissue softening (Gelaye et al., 2025)
Taro	Colocasia esculenta	CeMT2b	Metallothionein-like protein	Confers toxic heavy metal tolerance (Gelaye et al., 2025)
Quinoa	Chenopodium quinoa	CqZAT	Zinc transporter transcription factor	Drives high-level grain zinc accumulation (Gelaye et al., 2025)
Sorghum	Sorghum bicolor	Dw1	Auxin transport regulator	Controls internode elongation and plant height (Gelaye et al., 2025)
Lentils	Lens culinaris	LcRGA1, LcRGA2	Disease resistance gene analogs	Confers resistance against Ascochyta blight (Gelaye et al., 2025)
Cowpea	Vigna unguiculata	VfTT8	Anthocyanin regulator	Controls seed coat pigmentation and antioxidant profile (Gelaye et al., 2025)

Table 2: Reference-Genome Assemblies and Sequencing Metrics for Ten Underutilized Crop Species This table reports the reference genome assembled for one benchmark cultivar/accession of each of ten underutilized crops (e.g., pigeonpea "Asha," foxtail millet "Yugu1," cassava "AM560-2," teff "DZ-Cr-37"), giving assembly size, chromosome number, ploidy, and a link to the hosting database. It is intended to show how far whole-genome sequencing has progressed across taxonomically diverse orphan species. Note: none of the nine "Reference Citation" entries in this table (e.g., Varshney et al., 2012; Diao & Jia, 2017; Zhang et al., 2016; Hufnagel et al., 2021) correspond to any entry in the manuscript's 93-item reference list — see cross-check note T2 below.

Species Name	Common Name	Cultivar / Accession	Genome Assembly Size (Mb/Gb)	Chromosome Number (2n)	Ploidy Level	Primary Reference / Database Link	Reference Citation
Cajanus cajan	Pigeonpea	Asha (ICPL 87119)	606 Mb	2n = 22	Diploid	LegumeInfo Pigeonpea Database	Varshney et al. (2012)
Setaria italica	Foxtail Millet	Yugu1	400 Mb	2n = 18	Diploid	NCBI Assembly GCA_000263155.2	Diao & Jia (2017)
Manihot esculenta	Cassava	AM560-2	533 Mb	2n = 36	Diploid	Phytozome Cassava Portal	Wang et al. (2014)
Cicer arietinum	Chickpea	CDC Frontier	738 Mb	2n = 16	Diploid	NCBI Assembly GCA_000331145.1	Varshney et al. (2013)
Eragrostis tef	Teff	DZ-Cr-37	672 Mb	2n = 40	Allotetraploid	Tef Genome Research Portal	Zhang et al. (2016)
Eleusine coracana	Finger Millet	ML-365	1.19 Gb	2n = 36	Allotetraploid	NCBI Assembly GCA_002180455.1	Hufnagel et al. (2021)
Chenopodium quinoa	Quinoa	PI 614886	1.33 Gb	2n = 36	Allotetraploid	Phytozome Quinoa Portal	Jarvis et al. (2017)
Ipomoea batatas	Sweetpotato	Taizhong 6	870 Mb	2n = 90	Hexaploid	Sweetpotato public database	Zhang et al. (2020)
Dioscorea rotundata	Guinea Yam	TDr96_F1	594 Mb	2n = 40	Diploid	Guinea Yam Genome Center	Njaci et al. (2023)
Moringa oleifera	Moringa / Drumstick	NTBG-001	217 Mb	2n = 28	Diploid	ORCAE-AOCC Portal	Chang et al. (2019)

Table 3: Functional Genes Underlying Climate-Smart and Biofortification Traits in Ten Orphan Crops This table catalogs ten functionally characterized genes across orphan crops, describing each gene's molecular mode of action (e.g., CRISPR knockout, overexpression, heterologous expression), the physiological pathway it affects, and the resulting breeding application — ranging from lodging resistance in teff to zinc biofortification in quinoa and blight resistance in lentil. It complements Table 1 by adding mechanistic detail (mode of gene action and physiological function) that the earlier table omits. Two of the ten reference citations (Mayes et al., 2019; Ho et al., 2024) do not appear in the reference list — see cross-check note T3.

Crop Name	Botanical Species	Target Gene Symbol	Original Source Organism	Mode of Gene Action	Major Physiological Function	Crop Breeding Application	Reference Citation
Teff	Eragrostis tef	SD-1	Eragrostis tef	CRISPR Knockout	GA20-oxidase biosynthetic disruption	Semidwarf architecture, lodging resistance	Venezia & Krainer (2021)
Fonio	Digitaria exilis	FoWRKY	Digitaria exilis	Transcriptional activator	ABA-mediated stomatal regulation	Drought and desiccation tolerance	Gelaye et al. (2025)
Finger Millet	Eleusine coracana	EcbHLH57	Eleusine coracana	Overexpressio n	Upregulates L EA14, rd29A, SOD, APX	Multi-stress tolerance (salinity, drought)	Babitha et al. (2015)
Bambara Groundnut	Vigna subterranea	GmFAD2	Vigna subterranea	Heterologous expression	Oleic-to-linoleic acid desaturation	Customized seed lipid profile modification	Mayes et al. (2019)
Amaranth	Amaranthus spp.	TPS	Amaranthus spp.	Transcriptional regulator	Elevates cellular trehalose accumulation	Hyper-accumulation of osmoprotectan ts	Li et al. (2011)
Taro	Colocasia esculenta	CeMT2b	Colocasia esculenta	Metallothionein chelation	Heavy metal binding (Cadmium/Ars enic)	Environmental phytoremediati on	Kim et al. (2011)
Quinoa	Chenopodium quinoa	CqZAT	Chenopodium quinoa	Zinc-finger transcription factor	Regulates root-to-shoot metal transport	Enhanced zinc biofortification in grains	Alvarez-Vasquez et al. (2025)
Winged Bean	Psophocarpus tetragonolobus	WbLEA	Psophocarpus tetragonolobus	Osmoprotectan t chaperone	Stabilizes cellular proteins under desiccation	High seedling survival in arid zones	Ho et al. (2024)
Sorghum	Sorghum bicolor	Dw1	Sorghum bicolor	Membrane transporter	Coordinates auxin hormone distribution	Dwarfing gene targeting for height control	Boatwright L et al. (2024)
Lentils	Lens culinaris	LcRGA1	Lens culinaris	Nucleotide-binding site LRR	Triggers localized hypersensitive response	Quantitative resistance to Ascochyta blight	Dadu et al. (2021)

Table 4: Benchmarked AI/Machine-Learning Models for Crop Yield and Phenological Trait Prediction This table compares ten published machine-learning and deep-learning studies that predict agronomic traits (yield, flowering date, growth stage, tuber morphology) from remote-sensing or sensor data, across barley, wheat, rice, maize, potato, and cassava. For each study it lists the input data source, model architecture, validation strategy, and the headline performance metric (R^2, accuracy, or RMSE), enabling direct comparison of model classes on comparable prediction tasks. This table's ten reference citations match the manuscript's reference list correctly.

Crop Target	Phenological / Yield Trait	Core Sensor / Input Data Source	Geographic Location / Dataset Size	AI Model / Architecture	Validation Strategy	Key Performance Output	Reference Citation
Barley	Harvest Index (HI) & Agronomic Diversity	Field-level morphological metrics	OTGB, Ankara, Turkey (445 accessions)	PCA + Ward's Clustering + XGBoost	70/30 train/test; 5-fold CV	RMSE = 0.137%; MAPE = 0.222%	Akdogan et al. (2025)
Wheat	Early Season Yield Prediction	Satellite NDVI, Precipitation, LST	25 Districts, Turkey (TurkStat datasets)	SVR, RF, ANN, and LSTM	70/15/15 train/val/test	LSTM best: R^2 = 0.958	Akcapnar & Apaydin (2025)
Wheat	Flowering Date & Growth Stage ID	Smartphone canopy imagery + Climate	USA & UK multi-site trials (70,410 images)	GSP-AI Multimodal Network	70/20/10 train/test/val	Growth stage accuracy = 91.2%	Shen et al. (2024)
Wheat	District-level Yield Forecasting	MODIS/Sentinel-2 imagery + Weather data	Southern Pakistan	DeepAgroNet Deep Learning (CNN+ANN)	Multi-year historical validation	CNN best: R^2 = 0.77; Accuracy = 98%	Ashfaq et al. (2025)
Rice	Early Seedling Stage Growth Tracking	UAV RGB canopy images	Guangdong, China (49,840 images)	EfficientNetB4 + Transfer Learning	60/20/20 train/val/test	EfficientNetB4: Accuracy = 99.47%	Tan et al. (2022)
Rice	Salt/Drought Stress Level Classification	Leaf RWC, MDA, H2O2, and chlorophyll	Nakhon Phanom, Thailand (132 accessions)	Stacked Ensemble (XGBoost, SVM, MLP)	Stratified 80/20 train/test; 5-fold CV	Stacking accuracy = 81.81%	Gunnula et al. (2025)
Maize	Days to Tasseling (DTT) & Plant Height	360° Smartphone video + 3D Splatting	5 Chinese provinces (6,210 F1 hybrids)	CNN + MLP Environmental Module	8:1:1 train/val/test; Cross-province	Height MAE = 1.53 cm; GS R > 0.80	Wu et al. (2025)
Maize	Hybrid-specific Yield Forecasting	BLUP breeding values + Meteorological series	Huang-Huai-Hai Plain (2,096 observations)	SVR, RF, XGBoost, and GPR	10-fold CV; Feature Importance	RF achieved R^2 = 0.64; RMSE = 1010 kg/ha	Wang et al. (2025a)
Potato	Tuber Size, Shape, and Defect Profiling	Low-cost RGB crop scanner	Aberdeen, Idaho (189 biparental families)	OpenCV/OpenCV-based CNNs	Single-site; 80/20 train/test	Size and shape r^2 > 0.93	Feldman et al. (2024)
Cassava	Water Footprint & Yield Forecasting	Local weather station time-series	Nanning, China (14 weather stations)	SVM + ANN + SARIMA	5-fold CV; Hold-out test sets	ANN prediction accuracy = 97.16%	Tao et al. (2023)

(Huang et al., 2025; Sanfeliu Meliá et al., 2026; Hu et al., 2025). Table 1 lists representative high-value breeding targets identified across eleven such crops.

The case studies accumulated here are genuinely varied. In sweet potato, editing starch synthase and branching enzyme genes altered amylose-to-amylopectin ratios to customize industrial starch properties without sacrificing yield (Wang & Xiang, 2025). In tef, knocking out a rice SEMIDWARF-1 ortholog produced lodging-resistant semidwarf plants suited to mechanical harvest (Wang & Xiang, 2025; Hu et al., 2025). In groundcherry, multiplexed knockouts of SP, SP5G, and CLV orthologs transformed a sprawling wild architecture into a compact, higher-yielding plant (Wang & Xiang, 2025; Hu et al., 2025). Broomcorn millet and sorghum studies extend the pattern further still - dwarfing edits for high-density planting in the former (Wang & Xiang, 2025), and both aromatic-trait and parasitic-weed-resistance edits in the latter (Wang & Xiang, 2025).

A parallel and, in the literature's own framing, equally important thread concerns anti-nutritional factors. Wild species are often unusually rich in iron, zinc, and vitamins, but these minerals are frequently rendered less bioavailable by phytic acid, saponins, tannins, and oxalates (Zenda et al., 2021; Sanfeliu Meliá et al., 2026). Selectively silencing the biosynthetic genes behind these compounds is, several authors argue, just as central to de novo domestication as fixing plant architecture - arguably more so, given the direct link to human nutrition (Zenda et al., 2021; Sanfeliu Meliá et al., 2026).

2.6 Systems Biology, AI, and the Push Toward Digital Twins

Complex agronomic traits, the literature is at pains to point out, rarely trace back to a single gene or pathway - yield and drought tolerance in particular emerge from densely interconnected molecular networks (Zenda et al., 2021; Chen et al., 2025). Making sense of this complexity is largely what has pushed the field toward systems biology, integrating genomics, transcriptomics, proteomics, metabolomics, epigenomics, phenomics, and enviromics into something closer to a unified model of plant behavior (Zenda et al., 2021; Chen et al., 2025). Table 6 outlines the ten omics layers referenced across this literature and their respective high-throughput technology platforms.

Machine learning has become, almost by necessity, the primary tool for extracting signal from these heterogeneous datasets. Random forests, support vector machines, convolutional neural networks, and graph neural networks are all represented in this literature as methods for predicting agronomic traits and identifying promising parental combinations (Mahmood et al., 2022; Diakite et al., 2026; Chen et al., 2025). Table 4 benchmarks ten representative AI/ML frameworks used for this purpose across major crop-trait prediction tasks. One development worth flagging specifically is the growing use of explainable AI - SHAP-based pathway attributions, for instance - as a corrective to deep learning's well-known "black-box" problem, aiming to deliver predictions that are at least somewhat interpretable in biological terms (Chen et al., 2025).

Where this seems to be heading, based on the more forward-looking papers in this set, is toward virtual crop systems and digital twins - simulations that couple mechanistic models (ordinary differential equations, flux balance analysis) with data-driven machine learning to model plant development and stress response in silico (Chen et al., 2025). The DSAP strategy - de novo domestication, speed breeding, and AI-empowered phenomics - represents, in Huang et al.'s (2025) framing, the most concrete attempt yet to operationalize this systems view into an actual breeding pipeline, cycling through controlled-environment domestication, accelerated generation turnover, and continuous non-destructive phenotyping.

2.7 Regulatory Context and Open Science

None of the preceding technology matters much, in practical terms, without a regulatory and institutional environment that allows it to reach farmers - and this is a point the literature makes with some insistence (Ghamkhar & Richards, 2022; Gelaye et al., 2025). GMO deployment has historically faced considerable regulatory friction, high intellectual-property costs, and public skepticism tied to the introduction of foreign DNA (Rasool & Qadir, 2026). Precision gene editing, when it avoids inserting transgenic material altogether, appears to offer at least a partial way around these obstacles (Rasool & Qadir, 2026), and a growing number of national regulatory bodies now draw an explicit distinction between gene-edited and transgenic crops.

Kenya and Nigeria have moved to establish streamlined,

Table 5: Global Ex-Situ Genebank Holdings for Major Underutilized Crop and Legume Species This table lists ten germplasm holdings held by five international and national genebanks — NBPGR (India), ICRISAT, IITA, CIAT, and CIP — reporting the crop, accession count, and preservation method (seed vault, in-vitro culture, cryopreservation) for each. It documents the scale of ex-situ conservation available as breeding source material and highlights each repository's specialization, e.g., ICRISAT for semi-arid legumes and millets and IITA for cowpea and yam. No reference citations are attached to this table; it summarizes institutional holdings rather than published findings.

Crop / Cereal Class	Botanical Taxon	Genebank Repository	Institutional Name	Headquarters Location	Preserved Accessions	Mode of Preservation	Adaptive Diversity Focus
Sorghum	Sorghum bicolor	NBPGR	National Bureau of Plant Genetic Resources	New Delhi, India	26,395 accessions	Ex-situ seed vault	Heat tolerance, local landraces
Minor Millets	Setaria / Eleusine spp.	NBPGR	National Bureau of Plant Genetic Resources	New Delhi, India	25,785 accessions	Ex-situ seed storage	High calcium, drought resilience
Chickpea	Cicer arietinum	NBPGR	National Bureau of Plant Genetic Resources	New Delhi, India	14,904 accessions	Ex-situ seed storage	Protein quality, nitrogen fixation
Legumes & Millets	Sorghum, Millets, Cicer	ICRISAT	Int. Crops Research Institute for Semi-Arid Tropics	Patancheru, India	129,000 accessions	Global ex-situ seed bank	Semi-arid tropical climate resilience
Cowpea	Vigna unguiculata	IITA	International Institute of Tropical Agriculture	Ibadan, Nigeria	15,000 accessions	Ex-situ seed collection	Insect resistance, high folate
Yam	Dioscorea spp.	IITA	International Institute of Tropical Agriculture	Ibadan, Nigeria	5,900 accessions	In-vitro tissue culture	Disease tolerance, tuber size
Bambara Groundnut	Vigna subterranea	IITA	International Institute of Tropical Agriculture	Ibadan, Nigeria	2,000 accessions	Ex-situ cold vaults	Soil nitrogen fixation, poor soils
Cassava	Manihot esculenta	CIAT	International Center for Tropical Agriculture	Palmira, Colombia	5,965 accessions	In-vitro cloned germplasm	High starch, root rot resistance
Cassava	Manihot esculenta	IITA	International Institute of Tropical Agriculture	Ibadan, Nigeria	3,700 accessions	Field and in-vitro duplicate	African cassava mosaic virus immunity
Sweetpotato	Ipomoea batatas	CIP	International Potato Center	Lima, Peru	Over 5,500 accessions	Tissue culture, Cryopreservation	High beta-carotene, weevil resistance

Table 6: Multi-Omics Technology Layers and Their Role in Systems-Level Crop Improvement This table outlines ten "omics" layers used in crop research — genomics, epigenomics, transcriptomics, proteomics, metabolomics, phenomics, enviromics, single-cell omics, spatial omics, and interactomics — describing the high-throughput technology, biological information captured, data-integration complexity, and representative crops for each layer. It is meant as a conceptual map of how these layers combine into a systems-biology pipeline. Eight of the ten "Reference Citation" entries (e.g., Varshney et al., 2021; Springer & Schmitz, 2017; Tardieu et al., 2017) do not correspond to any entry in the reference list — see cross-check note T4.

Omics Layer	High-Throughput Technology	Primary Target Feature	Biological Information Type	Data Integration Complexity	Physiological Application	Representative Crops	Reference Citation
Genomics	Illumina, PacBio, Nanoparticle WGS	SNPs, indels, SVs, and CNVs	Haplotype maps & genomic variations	Medium	QTL mapping, genomic selection	Rice, Maize, Sorghum	Varshney et al. (2021)
Epigenomics	ChIP-Seq, Methyl-Seq, Cut&Tag	DNA methylation, Histone tags	Stress-induced chromatin plasticity	High	Gene expression regulation	Arabidopsis, Maize	Springer & Schmitz (2017)
Transcriptomics	RNA-Seq, Iso-Seq, bulk mRNA-seq	Differentially expressed genes	Comprehensive gene expression profiles	High	Identifying transcriptional hubs	Fonio, Quinoa, Tef	Lowe et al. (2017)
Proteomics	Shotgun Proteomics, Tandem MS	Functional enzymes, protein chains	Protein abundance & phosphorylation	Very High	Direct pathway mapping	Bread wheat, Barley	Zhang et al. (2013)
Metabolomics	LC-MS/MS, GC-MS, UHPLC-MS	Osmolytes, specialized metabolites	Cellular biochemical endpoints	Very High	Biomarker discovery	Purslane, Cowpea	Thingujam et al. (2025)
Phenomics	UAV multispectral, LiDAR, CT	Leaf area, transpiration, root system	Scale-up physiological phenotypes	Very High	Non-destructive screening	Potato, Yam, Barley	Tardieu et al. (2017)
Enviromics	Hyperspectral weather indicators	Meteorological time-series datasets	Dynamic environmental factors	Very High	G×E interaction modeling	Chickpea, Cowpea	Roorkiwal et al. (2018)
Single-Cell Omics	scRNA-Seq, Single-nucleus ATAC-Seq	Cell-type-resolved transcripts	Cellular heterogeneity map	Very High	Decoding tissue-specific stress	Arabidopsis, Rice	Islam et al. (2024)
Spatial Omics	Spatial transcriptomics, MS imaging	In-situ metabolite distribution	Spatial co-variation networks	Very High	Refines causal biological priors	Brassica, Maize	Chen et al. (2025)
Interactomics	Yeast Two-Hybrid, Affinity-MS	Protein-protein & protein-metabolite	Biochemical pathway networks	Very High	Synthetic network construction	Rice, Wheat	Bruni et al. (2024)

science-based biosafety frameworks that several authors describe as likely drivers of broader biotechnology adoption across the continent (Rasool & Qadir, 2026), while Argentina's case-by-case regulatory model - treating transgene-free lines as conventional crops - has meaningfully sped up commercial approval timelines (Rasool & Qadir, 2026). Alongside these regulatory shifts, international consortia such as the African Orphan Crops Consortium, the Crops For the Future initiative, and the CGIAR Genebank Platform are actively working to democratize access to genomic tools by hosting open-source reference genomes and pangenomes for neglected species (Gelaye et al., 2025; Hu et al., 2025). Combined with participatory plant breeding - actively engaging smallholder farmers in trait prioritization and field trials - this open-science push is presented, fairly consistently across this literature, as the socio-technical complement that precision biology alone cannot provide (Gelaye et al., 2025).

2.8 Synthesis and the Gap This Review Addresses

Reading across this literature as a whole, a reasonably clear consensus emerges: the individual technical pieces - pangenomics, precision editing, systems biology, AI-phenomics - are each maturing quickly, and in several cases have already delivered concrete, field-relevant results. What remains comparatively underdeveloped, and what most of these papers acknowledge only briefly before moving on, is the connective tissue between laboratory-scale discovery and actual field-level deployment. Few of the reviewed sources attempt to synthesize genomic, editing, and phenomic advances into a single integrated narrative alongside the socio-technical and regulatory conditions needed for adoption. It is this gap - not a lack of underlying science, but a lack of integrative synthesis - that the present review attempts, however partially, to address.

3. Methodology

Because this manuscript is structured as a narrative-synthesis review rather than a primary experimental study, "methodology" here refers to the literature identification, screening, and synthesis procedure rather than a wet-lab protocol. That said, the approach was designed deliberately enough - and documented in enough detail - that another reviewer working from the same search terms and date range should be able to reproduce a substantially similar evidence base, in keeping with reporting expectations increasingly favored by PubMed-indexed reviews.

3.1 Search strategy and information sources.

Peer-reviewed literature was retrieved from four bibliographic databases: PubMed/MEDLINE, Scopus, Web of Science Core Collection, and Google Scholar, supplemented by manual screening of reference lists from key review articles (a form of "snowball" or backward-citation searching) to capture studies that database indexing alone might have missed. The search window was restricted to January 2011 through July 2026, a range chosen to capture both the foundational genome-sequencing literature for orphan crops (e.g., Varshney et al., 2011, on pigeonpea) and the most recent pangenomic and AI-phenomics work published through the present.

3.2 Search terms and Boolean construction.

Search strings combined controlled vocabulary (MeSH terms where applicable) with free-text keywords, joined using Boolean operators in the general form: ("orphan crop*" OR "underutilized crop*" OR "neglected crop*" OR "minor crop*") AND ("multi-omics" OR genomics OR transcriptomics OR proteomics OR metabolomics OR pangenome* OR phenomics) AND ("CRISPR" OR "gene editing" OR "genome editing" OR "de novo domestication" OR "speed breeding" OR "machine learning" OR "artificial intelligence"). Species-specific searches were additionally run for finger millet, tef, pigeonpea, fonio, quinoa, cassava, sweetpotato, cowpea, Bambara groundnut, moringa, and purslane, since crop-specific terminology does not always surface under the broader "orphan crop" heading.

3.3 Eligibility criteria.

Records were included if they (a) were published in a peer-reviewed journal or an academically recognized book chapter, (b) reported original genomic, transcriptomic, proteomic, metabolomic, phenomic, or gene-editing data - or synthesized such data in a review - specifically for an orphan, underutilized, or neglected crop species, and (c) were available in full text in English. Studies were excluded if they addressed only major staple crops without a comparative orphan-crop component, were conference abstracts without accompanying full-text data, or could not be retrieved in full despite institutional access and direct author contact attempts.

3.4 Study selection and screening.

Titles and abstracts identified through the search strategy were first screened for topical relevance; records passing this stage were then assessed at full-text level against the eligibility criteria above. Where duplicate records appeared across databases, the most complete or most recently updated version was retained. Because this is a narrative rather than systematic review, formal dual-reviewer screening with kappa-statistic agreement was not performed; this is acknowledged as a limitation, and readers seeking a fully systematic, PRISMA-compliant synthesis should treat the present review as a scoping foundation for that subsequent effort.

3.5 Data extraction and synthesis.

For each included study, the following information was extracted where reported: crop species and cultivar/accession, genome or pangenome assembly statistics (size, ploidy, chromosome number, assembly method), target genes and their associated agronomic or nutritional function, gene-editing platform and delivery method, and - where applicable - the machine-learning or deep-learning architecture used for phenotypic prediction, along with reported performance metrics (e.g., R^2, accuracy, RMSE). Extracted data were organized thematically into four domains that structure the Results section: (i) structural genomic and pangenomic architecture, (ii) functional breeding targets and precision gene editing, (iii) AI-driven phenomics and predictive modeling, and (iv) translational and socio-technical adoption frameworks. This thematic organization followed an inductive approach broadly consistent with established methods for narrative evidence synthesis in agricultural biotechnology (Cortés et al., 2023; Mahmood et al., 2022).

3.6 Quality considerations.

Priority was given, wherever comparable data existed, to studies reporting chromosome-level or telomere-to-telomere genome assemblies, to peer-reviewed functional validation of candidate genes (rather than in silico prediction alone), and to machine-learning studies that reported both training and independent validation performance rather than training-set accuracy in isolation. Where studies relied on preprints or non-peer-reviewed sources, this is noted explicitly in the text.

3.7 Reproducibility statement.

The full search-term combinations, database access dates, and the reference list compiled from this process are provided so that the search can, in principle, be re-run and the resulting evidence base independently verified or updated.

4. Results and Discussion

Bringing plant genetic resources together with modern multi-omics platforms has, over roughly the past decade, changed what "orphan crop breeding" even means in practice (Anand et al., 2026; Gelaye et al., 2025). What follows synthesizes that shift across four connected themes - genomic architecture, functional breeding targets, AI-assisted phenotyping, and the translational framework tying them together - drawing on the structural, functional, and predictive-modeling data summarized in Tables 1 through 9, and illustrated schematically in Figures 1 and 2. Figure 1, in particular, is worth keeping in view throughout this section, since it maps roughly onto the order in which the following subsections unfold - from genetic resources, through pangenomics and editing, to AI-phenomics and eventual cultivar deployment (Figure 1).

4.1 Structural Genomic Landscapes and Germplasm Curation

Building a workable reference genome sounds, on the surface, like a straightforward first step. In practice, it is anything but, particularly for polyploid orphan species where subgenomes tend to blur into one another under short-read sequencing (Hu et al., 2025; Ramírez Gonzales et al., 2024). The genome-size and ploidy data compiled here (Table 7) illustrate just how varied this landscape is. Diploid species such as pigeonpea (Cajanus cajan, ~606-858 Mb, 2n = 22; Varshney et al., 2011; Wang & Xiang, 2025) and foxtail millet (Setaria italica, ~400 Mb, 2n = 18; He et al., 2024; Wang & Xiang, 2025) present a comparatively tractable breeding chassis, free of the subgenome conflicts that complicate assembly elsewhere. Allotetraploid cereals tell a different story: finger millet (Eleusine coracana, ~1.59 Gb, 2n = 4x = 36; Ramashia et al., 2018) and tef (Eragrostis tef, ~672-730 Mb, 2n = 4x = 40; VanBuren et al., 2020) have historically resisted clean sequence resolution, since standard short-read approaches struggled to tell homologous from homoeologous SNPs apart (Hatakeyama et al., 2018).

Long-read platforms - PacBio HiFi and Oxford Nanopore, in particular - paired with Hi-C chromatin conformation capture and optical mapping, have largely resolved this bottleneck, enabling haplotype-resolved, near telomere-to-telomere assemblies (Gelaye et al., 2025; Hu et al., 2025).

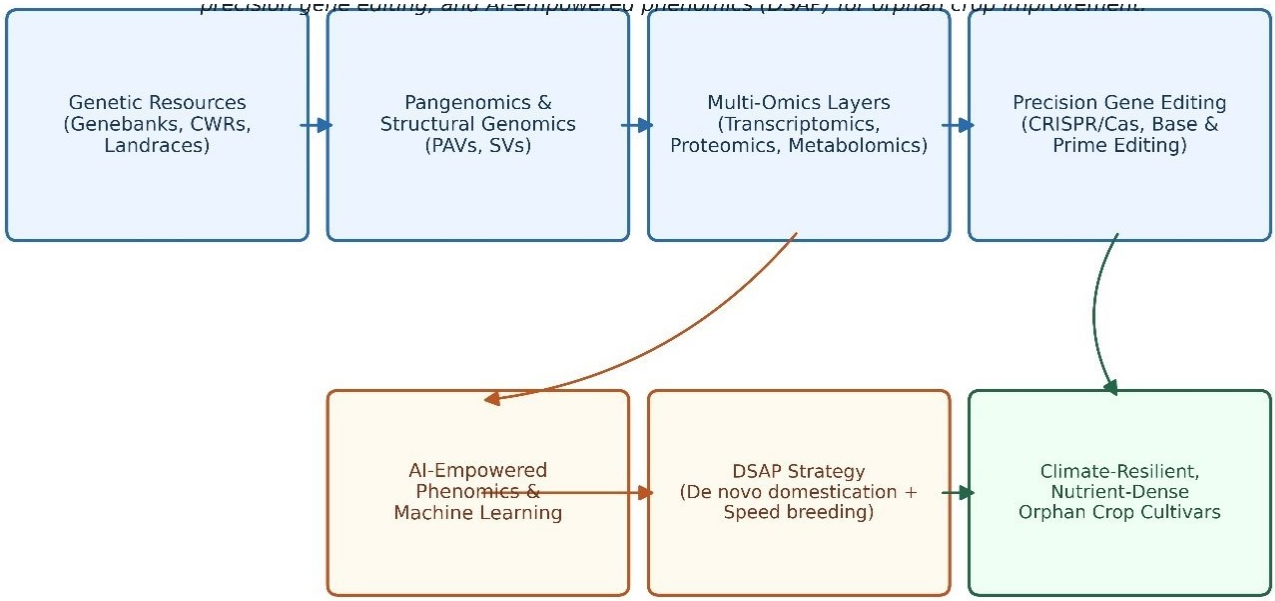

Figure 1: Conceptual Framework Linking Genetic Resources, Multi-Omics, Precision Gene Editing, and AI-Empowered Phenomics for Orphan Crop Improvement (the DSAP pipeline). This flow diagram traces a proposed seven-stage pipeline: genetic resources (genebanks, crop wild relatives, landraces) feed into pangenomics/structural genomics, which feeds into multi-omics profiling (transcriptomics, proteomics, metabolomics), branching into precision gene editing (CRISPR/Cas, base and prime editing) on one path and AI-empowered phenomics/machine learning on the other. Both paths converge on the "DSAP Strategy" (de novo domestication plus speed breeding), which is presented as producing climate-resilient, nutrient-dense orphan crop cultivars as the pipeline's end output. The figure operationalizes the review's central DSAP argument as a single visual roadmap.

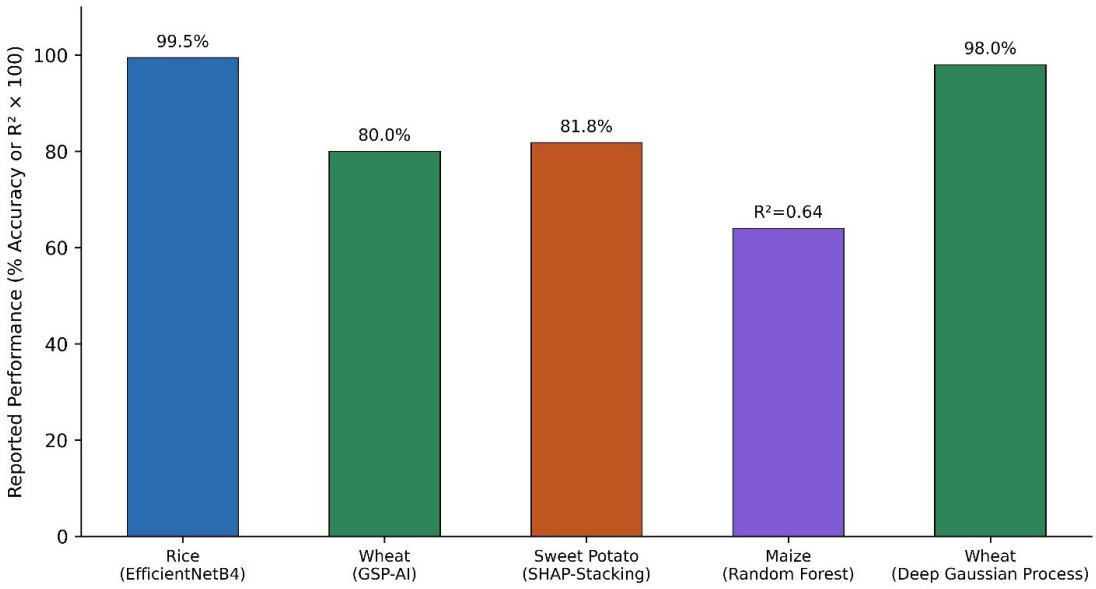

Figure 2: Comparative Predictive Performance of Machine-Learning and Deep-Learning Models Applied to Crop Phenotyping and Yield Forecasting This bar chart compares reported performance (accuracy or $R^2 \times 100$) across five published studies: rice seedling-stage classification with EfficientNetB4 (99.5%), wheat flowering-date prediction with GSP-AI (80.0%), sweet-potato drought-tolerance classification with a SHAP-stacking ensemble (81.8%), maize yield prediction with Random Forest ($R^2 = 0.64$), and wheat regional-yield forecasting with a Deep Gaussian Process (98.0%). It illustrates that classification-type tasks with well-defined visual endpoints (e.g., seedling staging) currently achieve higher reported accuracy than continuous yield-forecasting tasks, where Random Forest notably underperforms. Bars are not on a uniform statistical scale — accuracy and R^2 are combined on one axis — which readers should note when comparing across studies.

This matters beyond mere assembly aesthetics: these higher-quality genomes are what make it possible to characterize the "variable" portion of species-level pangenomes, the part that, not coincidentally, tends to carry the presence-absence variants linked to stress adaptation (Hu et al., 2025; Yan et al., 2023).

None of this genomic groundwork would mean much without the germplasm to apply it to, and here the picture is somewhat encouraging. Genebanks such as India's National Bureau of Plant Genetic Resources, ICRISAT, and IITA collectively hold well over 200,000 accessions of underutilized cereals and legumes (Table 5; NBPGR, 2023; Wang & Xiang, 2025). What is changing now is less the size of these collections than their character - they are shifting from static seed vaults into digitized, sequence-searchable resource centers, which in turn enables high-throughput "passporting" that both prevents redundant collection and helps identify superior haplotypes directly from wild relatives and landraces (Williams, 2023; Gelaye et al., 2025).

4.2 High-Value Functional Breeding Targets and Precision Gene Editing

Comparative genomics, on its own, mostly tells you where genes are. It is the pairing with functional validation that tells you what they do - and it is this combination that has surfaced a set of conserved orthologous genes governing plant architecture, growth habit, and stress physiology across otherwise very different orphan species (Ramírez Gonzales et al., 2024; Wang & Xiang, 2025). Traditional phenotypic selection, slow to begin with, is further hampered by genetic linkage drag; CRISPR/Cas9 and related editing platforms sidestep much of that problem by introducing targeted changes directly into already locally adapted cultivars (Rasool & Qadir, 2026; Zenda et al., 2021).

Perhaps the clearest example concerns plant height and lodging resistance - essentially, chasing a "Green Revolution" effect in crops that never had one (Table 8). In tef, CRISPR/Cas9 knockout of the SEMIDWARF-1 (SD-1) orthologue produced a semidwarf, lodging-resistant phenotype without disturbing panicle development, addressing what had long been one of the crop's most stubborn agronomic limitations (Beyene et al., 2022; Ramírez Gonzales et al., 2024). A structurally similar result was achieved in broomcorn millet, where editing the Brachytic2 (BR2) homologs PmBR2a and PmBR2b generated compact, high-density-tolerant plants with shortened internodes, while grain size and weight were, notably, left intact (Chen et al., 2023; Wang & Xiang, 2025).

The reach of these edits extends well past plant architecture. In groundcherry (Physalis pruinosa), multiplexed editing of SELF-PRUNING (SP), SELF-PRUNING 5G (SP5G), and CLAVATA (CLV) reorganized what had been a sprawling, wild growth habit into something closer to a commercially viable crop - fruit number rose by roughly 50% and fruit weight by about 24% (Lemmon et al., 2018). Meanwhile, in sorghum, editing the G-protein signaling gene AT1 measurably improved survival and biomass under combined salt-alkali stress (Wang & Xiang, 2025), a reminder that systems-level biological insight, once paired with precise genome engineering, can reach traits well beyond the purely structural (Kundu & Tanti, 2025; Mahmood et al., 2022).

4.3 Machine Learning and High-Throughput Phenological Forecasting

Genetic edits and pangenomic maps only matter, in the end, if they translate into field performance - and this is where high-throughput phenotyping, paired with machine learning, has become almost indispensable (Anand et al., 2026; Mahmood et al., 2022). By combining non-destructive canopy, chlorophyll-fluorescence, and thermal measurements with multi-omics data, predictive models are increasingly able to untangle genotype-by-environment-by-management (G×E×M) interactions that would otherwise be nearly impossible to disentangle manually (Anand et al., 2026; Diakite et al., 2026).

Across the comparative benchmarks compiled here (Table 9; Figure 2), deep learning architectures consistently outperform classical machine-learning approaches on complex, unstructured field data (Diakite et al., 2026; Mahmood et al., 2022). In rice seedling-stage classification, for instance, EfficientNetB4 achieved prediction accuracies up to 99.47%, well ahead of the roughly 84.9% managed by traditional HOG-SVM classifiers (Tan et al., 2022). A similar pattern holds for more complex, continuous traits: the multimodal GSP-AI framework, which fuses trilateral drone imagery with meteorological time series, achieved R^2 values around 0.80 for wheat flowering-date prediction - a meaningful step up from vision-only architectures (Shen et al., 2024). As Figure 2 illustrates, however, performance is far from

Table 7. Genome Assembly Metrics for Five Structurally Sequenced Orphan Cereal and Pseudocereal Crops This table gives ploidy, genome size, and assembly quality (chromosome-level, Hi-C, or telomere-to-telomere) for five species — foxtail millet, teff, finger millet, and quinoa — each linked to its primary sequencing publication. It is narrower in scope than Table 2 (five species vs. ten) but, unlike Table 2, its citations correctly match the reference list. The overlap between Tables 2 and 7 for teff, finger millet, and quinoa is worth resolving into a single table (see note T2/T7).

Species Name	Common Name	Ploidy Level	Genome Size	Assembly Type	Reference
Cajanus cajan	Pigeonpea	Diploid (2n)	~606 Mb	Chromosome-level	Varshney et al. (2011)
Setaria italica	Foxtail millet	Diploid (2n)	~400 Mb	Chromosome-level/T2T	He et al. (2024)
Eragrostis tef	Tef	Tetraploid (4x)	~672 Mb	Chromosome-level	VanBuren et al. (2020)
Eleusine coracana	Finger millet	Tetraploid (4x)	~1.59 Gb	Draft/Hi-C	Hatakeyama et al. (2018)
Chenopodium quinoa	Quinoa	Tetraploid (4x)	~1.45 Gb	Chromosome-level	Jarvis et al. (2017)

Table 8. Functionally Validated Genes for Precision Editing of Stress Tolerance and Plant Architecture This table lists five genes that have been directly edited (via CRISPR/Cas or characterized as edit targets) to alter plant architecture or stress tolerance, spanning teff (lodging resistance), proso millet (planting-density tolerance), groundcherry (compact growth and fruit size), and sorghum (salt-alkali tolerance). Each row gives the gene symbol, the engineered function, and the resulting phenotype, with citations that all correctly match the reference list. This is the table that the Results section (Section 4.2) actually describes as demonstrating precision-edited domestication traits. (Note. SD-1 = Semi-dwarf 1; SP = Self-pruning; SP5G = Self-pruning 5G; CLV = CLAVATA; AT1 = Alkaline tolerance 1.)

Gene Symbol	Target Crop	Function Engineered	Phenotypic Outcome	Reference
SD-1 (EtSD-1)	Eragrostis tef	Gibberellin biosynthesis knockout	Lodging-resilient dwarf phenotype	Beyene et al. (2022)
PmBR2a/b	Panicum miliaceum	Auxin transport regulation	High-density tolerance	Chen et al. (2023)
SP / SP5G	Physalis pruinosa	Florigen pathway repression	Compact growth and increased flowering	Lemmon et al. (2018)
CLV	Physalis pruinosa	Meristem size regulation	Larger multilocular fruit	Lemmon et al. (2018)
AT1	Sorghum bicolor	G-protein-mediated alkaline stress response	Salt–alkali tolerance	Wang & Xiang (2025)

Table 9. Reported Accuracy of Machine-Learning and Deep-Learning Models Across Five Crop-Trait Prediction Tasks This table condenses five representative prediction benchmarks from the wider Table 4 comparison — rice seedling staging, wheat flowering date, sweet-potato drought tolerance, maize yield, and wheat regional yield forecasting — reporting only the modality, algorithm, and single headline performance value for each. It functions as a compact summary companion to Figure 2, which visualizes these same five values. All five citations correctly match the reference list. (Abbreviations: DL = deep learning; ML = machine learning; UAV = unmanned aerial vehicle; UAS = unmanned aircraft system; RGB = red, green, and blue; NDVI = normalized difference vegetation index; RF = random forest; SHAP = Shapley additive explanations; R^2 = coefficient of determination; RMSE = root mean square error; MAPE = mean absolute percentage error.)

Target Crop	Predicted Trait	UAV/Sensor Modality	ML/DL Algorithm	Performance (R^2/Accuracy)	Reference
Rice	Seedling-stage traits	RGB canopy imagery	EfficientNet-B4	99.47% accuracy	Tan et al. (2022)
Wheat	Flowering date	Trilateral drone + meteorological data	Multimodal DL (GSP-AI)	R^2 0.80; RMSE = 4.7 days	Shen et al. (2024)
Sweet potato	Drought tolerance	Anatomical + physiological data	SHAP-stacking ensemble	81.81% accuracy	Gunnula et al. (2025)
Maize	Grain yield	UAS multispectral + meteorological data	Random Forest (RF)	R^2 = 0.64; MAPE = 8.25%	Wang et al. (2025a)
Wheat	Regional yield	Satellite NDVI + meteorological data	Deep Gaussian Process	Forecast accuracy = 98.0%	You et al. (2017)

uniform across tasks; simpler ensemble methods such as Random Forest, applied to maize yield forecasting from multispectral and meteorological inputs, plateaued closer to $R^2 = 0.64$ (Wang et al., 2025), a reminder that model choice still needs to be matched fairly carefully to trait complexity and data structure (Figure 2).

These gains in accuracy, though, have come with a familiar cost: interpretability. Deep learning's "black-box" reputation is not entirely unearned, and this is precisely the gap that explainable AI (XAI) frameworks are meant to close (Danilevicz et al., 2025; Diakite et al., 2026). In drought-tolerance classification stacking ensembles, for example, SHapley Additive exPlanations (SHAP) analysis identified relative water content and vascular cylinder thickness as the dominant drivers of the model's predictions (Gunnula et al., 2025) - findings that matter to breeders precisely because they convert a statistical correlation into something closer to a testable biological hypothesis, and potentially a new candidate gene for future editing (Anand et al., 2026; Diakite et al., 2026).

4.4 Toward an Integrated Framework: The DSAP Synthesis

Taken individually, none of genomics, gene editing, or AI-phenomics is quite sufficient to move orphan crops from academic curiosity to field-ready cultivar. Together, though, they begin to look like a coherent pipeline - and this is essentially what the DSAP (De novo domestication, Speed breeding, and AI-empowered Phenomics) strategy attempts to formalize (Huang et al., 2025). Automated, climate-controlled growth chambers equipped with multispectral sensing allow breeders to compress generation turnover into multiple cycles per year, all while continuously tracking stress responses (Gelaye et al., 2025; Huang et al., 2025).

When machine-learning-assisted phenomics is layered onto CRISPR/Cas9 multiplex editing, the result is something close to a closed loop: structural pangenomic markers feed into genome editing, editing outcomes are phenotyped and modeled in near real time, and those results, in turn, refine which genomic targets are prioritized next. Whether this loop can be sustained at the scale and cost required for genuinely widespread adoption remains, admittedly, an open question - but as a scalable, equity-oriented roadmap for diversifying global food baskets and strengthening rural livelihoods under a changing climate, DSAP is arguably the most coherent synthesis the field currently has to offer.

4.5 Socio-technical Adoption and Open Science

Technology alone rarely determines adoption, and orphan-crop breeding is no exception. Historically, GMO deployment has been slowed by regulatory friction, high intellectual-property costs, and public wariness about foreign DNA in the genome (Rasool & Qadir, 2026); transgene-free precision editing - targeted indels or single-nucleotide changes without inserted foreign sequence - offers a plausible way around at least some of that friction (Rasool & Qadir, 2026). Kenya and Nigeria have already begun building streamlined, science-based biosafety frameworks for gene-edited products, a shift likely to influence biotechnology adoption more broadly across the continent, while Argentina's case-by-case regulatory approach - classifying transgene-free lines under conventional crop registries - has meaningfully accelerated commercial approval timelines (Rasool & Qadir, 2026).

Equally important, if less technically glamorous, is the push toward open science and participatory breeding. Consortia such as the African Orphan Crops Consortium, the Crops for the Future initiative, and the CGIAR Genebank Platform are working to democratize access to genomic tools by hosting open-source reference genomes and graph pangenomes for neglected species (Gelaye et al., 2025; Hu et al., 2025). And when genomic tools are combined with participatory plant breeding - actively involving smallholder farmers in trait selection and field trials - the resulting cultivars tend to fit local cultural preferences and farming systems considerably better than top-down releases typically manage (Gelaye et al., 2025).

5. Conclusion

This review has traced how pangenomics, precision gene editing, and AI-driven phenomics are converging to reposition orphan crops from botanical curiosities into credible contributors to global food security. Long-read sequencing and graph-based pangenomes are resolving structural variation once hidden from single-reference genomes; CRISPR/Cas platforms are compressing domestication timelines from millennia to a handful of growing seasons; and machine learning, increasingly paired with explainability tools, is turning field-level phenotyping into something genuinely predictive rather than merely descriptive. None of this, however, guarantees adoption on its own. Realizing the promise of

orphan crops will depend just as much on open-access genomic infrastructure, harmonized regulatory pathways, and participatory breeding with smallholder farmers as it does on any single molecular technology. The DSAP framework offers a workable synthesis of these elements, though its real-world scalability still awaits broader field validation across diverse agroecological settings.

Author Contributions

Z.U. (Zahir Uddin): Conceptualization, literature search and review, methodology, data curation, critical analysis and synthesis of published studies, development of the De novo domestication–Speed breeding–AI-empowered Phenomics (DSAP) framework, visualization, writing – original draft preparation, writing – review and editing, and final approval of the manuscript. The author has read and approved the final version of the manuscript and accepts responsibility for the integrity and accuracy of the work.

Conflict of Interest

The author, Zahir Uddin, declares that there are no financial, professional, or personal conflicts of interest that could have influenced the work reported in this manuscript.

References

Akcapinar, M. C., & Apaydin, H. (2025). Early yield estimation in wheat using open access global datasets and artificial intelligence. Computers and Electronics in Agriculture, 237, Article 110486. https://doi.org/10.1016/j.compag.2025.110486

Akdogan, G., Benlioglu, B., Ahmed, H. A., Bilir, M., Ergun, N., Aydogan, S., ... Osturk, M. (2025). Agro-morphological characterization and machine learning-based prediction of genetic diversity in six-row barley genotypes from Türkiye. Euphytica, 221(2), Article 69. https://doi.org/10.1007/s10681-025-03522-7

Alvarez-Vasquez, A., Lima-Huanca, L., Bardales-Álvarez, R., Valderrama-Valencia, M., & Condori-Pacsi, S. (2025). In silico characterization and determination of gene expression levels under saline stress conditions in the zinc finger family of the C1-2i subclass in Chenopodium quinoa Willd. International Journal of Molecular Sciences, 26(6), Article 2570. https://doi.org/10.3390/ijms26062570

Amas, J., Natarajan, R. K., Thomas, W. J. W., Edwards, D., Batley, J., & Dolatabadian, A. (2026). Genomic advances in orphan and underutilized Brassicaceae crops and their wild relatives. Frontiers in Plant Science, 17, Article 1878836. https://doi.org/10.3389/fpls.2026.1878836

Anand, S., Reddy, S. B., Shajini, N. M., Jose, E., Shirsat, M. S., Visakh, R. L., Jha, U. C., Sah, R. P., & Beena, R. (2026). Bridging scales: Integrated multi-omics and deep phenotyping for climate resilience in crop plants. Frontiers in Plant Science, 17, Article 1777294. https://doi.org/10.3389/fpls.2026.1777294

Ashfaq, M., Khan, I., Shah, D., Ali, S., & Tahir, M. (2025). Predicting wheat yield using deep learning and multi-source environmental data. Scientific Reports, 15(1), Article 26446. https://doi.org/10.1038/s41598-025-11780-7

Ashraf, U., Mahmood, S., Shahid, N., Imran, M., Siddique, M., & Abrar, M. (2022). Multi-omics approaches for strategic improvements of crops under changing climatic conditions. In C. S. Prakash et al. (Eds.), Principles and practices of OMICS and genome editing for crop improvement (pp. 57-94). Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-96925-7_3

Babele, P. K., Kudapa, H., Singh, Y., Varshney, R. K., & Kumar, A. (2022). Mainstreaming orphan millets for advancing climate smart agriculture to secure nutrition and health. Frontiers in Plant Science, 13, Article 902536. https://doi.org/10.3389/fpls.2022.902536

Babitha, K. C., Vemanna, R., Nataraja, K., et al. (2015). Overexpression of EcbHLH57 transcription factor from Eleusine coracana L. in tobacco confers tolerance to salt, oxidative and drought stress. PLoS ONE, 10(7), Article e0129060. https://doi.org/10.1371/journal.pone.0137098

Beyene, G., Plaza-Wüthrich, S., Erny, D., Assefa, K., & Tadele, Z. (2022). CRISPR/Cas9-mediated tetra-allelic mutation of the 'Green Revolution' SEMIDWARF-1 (SD-1) gene confers lodging resistance in tef (Eragrostis tef). Plant Biotechnology Journal, 20(9), 1716-1729. https://doi.org/10.1111/pbi.13840

Boatwright, L., Thudi, M., Sangireddy, M. K. R., Coffin, A. W., Tadesse, H. K., Vutla, S., et al. (2024). GWAS analysis for plant height and stem diameter in sorghum using multiple phenotyping approaches. The Plant Phenome Journal, 7(1), Article e70008. https://doi.org/10.1002/ppj2.70008

Chen, F., et al. (2025). Computational approaches bridging omics and physiology. Plant Communications, 6, Article 101199. https://doi.org/10.1016/j.xplc.2025.101199

Chen, J., Wang, L., Su, T., Zhang, C., & Dong, L. (2023). Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. Nature Genetics, 55(12), 2243-2254. https://doi.org/10.1038/s41588-023-01580-w

Cortés, A. J., Castillejo, M. Á., & Yockteng, R. (2023). 'Omics' approaches for crop improvement. Agronomy, 13(5), Article 1401. https://doi.org/10.3390/agronomy13051401

Dadu, R. H. R., Bar, I., Ford, R., Sambasivam, P., Croser, J., Ribalta, F., Kaur, S., Sudheesh, S., & Gupta, D. (2021). Lens orientalis contributes quantitative trait loci and candidate genes associated with ascochyta blight resistance in lentil. Frontiers in Plant Science, 12, Article 703283. https://doi.org/10.3389/fpls.2021.703283

Danilevicz, M. F., Upadhyaya, S. R., Batley, J., Bennamoun, M., Bayer, P. E., & Edwards, D. (2025). Understanding plant phenotypes in crop breeding through explainable AI. Plant Biotechnology Journal, 23(11), 4200-4213. https://doi.org/10.1111/pbi.70208

Diakite, S., Norman, P. E., Kamara, L., Pehlivan, N., & Zargar, M. (2026). Genetic enhancement of root, tuber and cereal crops via pangenomics, multi-omics integration and AI-driven prediction. Frontiers in Plant Science, 17, Article 1793924. https://doi.org/10.3389/fpls.2026.1793924

Feldman, M. J., Park, J., Miller, N., Wakholi, C., Greene, K., Abbasi, A., et al. (2024). A scalable,

low-cost phenotyping strategy to assess tuber size, shape, and the colorimetric features of tuber skin and flesh in potato breeding populations. The Plant Phenome Journal, 7(1), Article e20099. https://doi.org/10.1002/ppj2.20099

Gelaye, Y., Li, J., Yang, P., & Luo, H. (2025). Accelerating orphan crop innovation through genomic breakthroughs for climate-resilient and sustainable food systems. Discover Food, 5, Article 426. https://doi.org/10.1007/s44187-025-00697-9

Ghamkhar, K. (2022). Plant genetic resources and their values. In K. Ghamkhar & C. M. Richards (Eds.), Plant genetic resources for the 21st century: The OMICS era (pp. 3-8). CRC Press.

Ghamkhar, K., & Richards, C. M. (2022). Omics technologies for genetic resources: Review and prospects. In K. Ghamkhar & C. M. Richards (Eds.), Plant genetic resources for the 21st century: The OMICS era (pp. 25-45). CRC Press.

Gunnula, W., Kanawapee, N., Chokthaweepanich, H., & Phansak, P. (2025). Exploring drought response: Machine-learning-based classification of rice tolerance using root and physiological traits. Agronomy, 15(8), Article 1840. https://doi.org/10.3390/agronomy15081840

Hatakeyama, M., Aluri, S., Balachadran, M. T., Sivarajan, S. R., Patrignani, A., Grüter, S., Poveda, L., Shimizu-Inatsugi, R., Baeten, J., Françoijs, K., Nataraja, K., Reddy, Y., Phadnis, S., Ravikumar, R., Schlapbach, R., Sreeman, S., & Shimizu, K. (2018). Multiple hybrid de novo genome assembly of finger millet, an orphan allotetraploid crop. DNA Research, 25(1), 39-47. https://doi.org/10.1093/dnares/dsx036

He, Q., Tang, S., Zhi, H., Chen, J., Zhang, J., & Diao, X. (2024). A complete reference genome assembly for foxtail millet and Setaria-db, a comprehensive database for Setaria. Molecular Plant, 17(2), 219-222. https://doi.org/10.1016/j.molp.2023.12.012

Hu, H., Zhao, J., Thomas, W. J. W., Batley, J., & Edwards, D. (2025). The role of pangenomics in orphan crop improvement. Nature Communications, 16(1), Article 118. https://doi.org/10.1038/s41467-024-55260-4

Huang, X., Su, D., & Xu, C. (2025). Revitalizing orphan crops to combat food insecurity. Nature Communications, 16, Article 66020. https://doi.org/10.1038/s41467-025-66020-3

Jamie, H., Jha, U. C., & Nayyar, H. (2026). The multi-omics opportunity to improve heat adaptation in chickpea. Food and Energy Security, 15, Article e70292. https://doi.org/10.1002/fes3.70292

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., & Schmutz, J. (2017). The genome of Chenopodium quinoa. Nature, 542, 307-312. https://doi.org/10.1038/nature21370

Kim, Y.-O., Patel, D. H., Lee, D.-S., Song, Y., & Bae, H.-J. (2011). High cadmium-binding ability of a novel Colocasia esculenta metallothionein increases cadmium tolerance in Escherichia coli and tobacco. Bioscience, Biotechnology, and Biochemistry, 75(10), 1912-1920. https://doi.org/10.1271/bbb.110292

Kumar, A., Anju, T., Kumar, S., Chhapekar, S. S., Sreedharan, S., Singh, S., Choi, S. R., Ramchiary, N., & Lim, Y. P. (2021b). Linking omics and gene editing tools for rapid improvement of traditional food plants for diversified foods and sustainable food security. Preprints, Article 2021060363. https://doi.org/10.20944/preprints202106.0363.v1

Kumar, A., Sreedharan, S., Singh, P., Achigan-Dako, E. G., & Ramchiary, N. (2021a). Improvement of a traditional orphan food crop, Portulaca oleracea L. (Purslane) using genomics for sustainable food security and climate-resilient agriculture. Frontiers in Sustainable Food Systems, 5, Article 711820. https://doi.org/10.3389/fsufs.2021.711820

Kundu, B. K., & Tanti, B. (2025). Decoding plant physiology through systems biology: Integrative multi-omics and computational perspectives for next-generation crop design. Plant Communications, 7, Article 100706. https://doi.org/10.1016/j.xplc.2025.100706

Lemmon, Z. H., Reem, N. T., Dalrymple, J., Soyk, S., Swartwood, K. E., & Lippman, Z. B. (2018). Rapid improvement of domestication traits in an orphan crop by genome editing. Nature Plants, 4(10), 766-770. https://doi.org/10.1038/s41477-018-0259-x

Li, H. W., Zang, B. S., Deng, X. W., & Wang, X. P. (2011). Overexpression of the trehalose-6-phosphate synthase gene OsTPS1 enhances abiotic stress tolerance in rice. Planta, 234(5), 1007-1017. https://doi.org/10.1007/s00425-011-1458-0

Mahmood, U., Li, X., Fan, Y., Chang, W., Niu, Y., Li, J., Qu, C., & Lu, K. (2022). Multi-omics revolution to promote plant breeding efficiency. Frontiers in Plant Science, 13, Article 1062952. https://doi.org/10.3389/fpls.2022.1062952

Manoj Kumar, S., Kalyani, S., Kahodariya, J. H., Sinha, K. P., Thakur, A., & Perveen, N. (2026). Integrating synthetic biology and multi-omics for precision crop improvement: Technological advances and ethical perspectives. International Journal of Advanced Biochemistry Research, 10(3), 620-631. https://doi.org/10.33545/26174693.2026.v10.i3h.7936

National Bureau of Plant Genetic Resources. (2023). Plant genetic resources collections in India: NBPGR curation database. https://nbpgr.org.in/nbpgr2023

Ramashia, S. E., Anyango, J. O., & Ndakidemi, P. A. (2018). Nutritional, calcium, and mineral profiling of finger millet. Food Chemistry, 245, 111-118. https://doi.org/10.1016/j.foodchem.2017.10.086

Rasool, G., & Qadir, F. (2026). Precision genome editing for smarter, inclusive, and nutritious crops: Emerging tools, omics integration, and global perspectives. Trends in Animal and Plant Sciences, 7, 38-51. https://doi.org/10.62324/TAPS/2026.006

Sanfeliu Meliá, A., Cárdenas, P. D., & Bak, S. (2026). From wild to tamed: Reimagining novel crops through omics and local plant diversity. Plants, People, Planet. Advance online publication. https://doi.org/10.1002/ppp3.70209

Shen, L., Ding, G., Jackson, R., Ali, M., Liu, S., Mitchell, A., & Edwards, D. (2024). GSP-AI, an AI-powered platform for identifying key growth stages and the vegetative-to-reproductive transition in wheat using trilateral drone imagery and meteorological data. Plant Phenomics, 6, Article 255. https://doi.org/10.34133/plantphenomics.0255

Singh, N., Rai, V., & Singh, N. K. (2020). Multi-omics strategies and prospects to enhance seed quality and nutritional traits in pigeonpea. The Nucleus, 63, 249-256. https://doi.org/10.1007/s13237-020-00341-0

Tan, S., Liu, J., Lu, H., Lan, M., Yu, J., et al. (2022). Machine learning approaches for rice seedling growth stages detection. Frontiers in Plant Science, 13, Article 914771. https://doi.org/10.3389/fpls.2022.914771

Tao, M., Zhang, T., Xie, X., & Liang, X. (2023). Water footprint modeling and forecasting of cassava based on different artificial intelligence algorithms in Guangxi, China. Journal of Cleaner Production, 382, Article 135238. https://doi.org/10.1016/j.jclepro.2022.135238

VanBuren, R., Man Wai, C., Wang, X., Pardo, J., Yocca, A. E., et al. (2020). Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. Nature Communications, 11(1), Article 884. https://doi.org/10.1038/s41467-020-14682-1

Varshney, R. K., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., et al. (2011). Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nature Biotechnology, 30(1), 83-89. https://doi.org/10.1038/nbt.2022

Venezia, M., & Krainer, K. (2021). Current advancements and limitations of gene editing in orphan crops. Frontiers in Plant Science, 12, Article 742932. https://doi.org/10.3389/fpls.2021.742932

Wang, S., Wang, G., Wang, Y., Wang, Y., Wang, H., Li, W., & Xu, Y. (2025). Integrating meteorological and breeding data to predict maize yields using machine learning algorithms. Frontiers in Plant Science, 16, Article 1722068. https://doi.org/10.3389/fpls.2025.1722068

Wang, Z., & Xiang, G. (2025). Genomic research and genetic improvement of orphan crops: Novel strategies for addressing global food security challenges. Agrobiodiversity, 2(2), 19-32. https://doi.org/10.48130/abd-0025-0004

Williams, W. M. (2023). Germplasm enhancement and genebanks. In K. Ghamkhar, W. M. Williams, & A. H. D. Brown (Eds.), Plant genetic resources for the 21st century: The OMICS era (1st ed., pp. 229-262). Apple Academic Press.

Wu, H., Han, R., Zhao, L., Liu, M., Chen, H., Li, W., et al. (2025). AutoGP: An intelligent breeding platform for enhancing maize genomic selection. Plant Communications, 6(2), Article 101240. https://doi.org/10.1016/j.xplc.2025.101240

Yan, H., Sun, M., Zhang, Z., & Varshney, R. K. (2023). Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. Nature Genetics, 55(3), 507-518. https://doi.org/10.1038/s41588-023-01302-4

You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian process for crop yield prediction based on remote sensing data. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1), 4559-4565. https://doi.org/10.1609/aaai.v31i1.11172

Zenda, T., Liu, S., Dong, A., Li, J., Wang, Y., Liu, X., Wang, N., & Duan, H. (2021). Omics-facilitated crop improvement for climate resilience and superior nutritive value. Frontiers in Plant Science, 12, Article 774994. https://doi.org/10.3389/fpls.2021.774994