Plant genome resequencing and population genomics: Current status and future prospects

Bo Song[1,8], Weidong Ning[1,2,8], Di Wei[3], Mengyun Jiang[1,4,5], Kun Zhu[1,4,5], Xingwei Wang[1,4,5], David Edwards[6], Damaris A. Odeny[7] and Shifeng Cheng[1,*]

[1]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

[2]Huazhong Agricultural University, College of Informatics, Hubei Key Laboratory of Agricultural Bioinformatics, Wuhan, Hubei, China

[3]Biotechnology Research Institute, Guangxi Academy of Agricultural Sciences, Nanning 53007, China

[4]State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng 475004, China

[5]Shenzhen Research Institute of Henan University, Shenzhen 518000, China

[6]School of Biological Sciences and Institute of Agriculture, University of Western Australia, Perth, WA, Australia

[7]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) - Eastern and Southern Africa, Nairobi, Kenya

[8]These authors contributed equally to this article.

*Correspondence: Shifeng Cheng (chengshifeng@caas.cn)

https://doi.org/10.1016/j.molp.2023.07.009

ABSTRACT

Advances in DNA sequencing technology have sparked a genomics revolution, driving breakthroughs in plant genetics and crop breeding. Recently, the focus has shifted from cataloging genetic diversity in plants to exploring their functional significance and delivering beneficial alleles for crop improvement. This transformation has been facilitated by the increasing adoption of whole-genome resequencing. In this review, we summarize the current progress of population-based genome resequencing studies and how these studies affect crop breeding. A total of 187 land plants from 163 countries have been resequenced, comprising 54 413 accessions. As part of resequencing efforts 367 traits have been surveyed and 86 genome-wide association studies have been conducted. Economically important crops, particularly cereals, vegetables, and legumes, have dominated the resequencing efforts, leaving a gap in 49 orders, including Lycopodiales, Liliales, Acorales, Austrobaileyales, and Commelinales. The resequenced germplasm is distributed across diverse geographic locations, providing a global perspective on plant genomics. We highlight genes that have been selected during domestication, or associated with agronomic traits, and form a repository of candidate genes for future research and application. Despite the opportunities for cross-species comparative genomics, many population genomic datasets are not accessible, impeding secondary analyses. We call for a more open and collaborative approach to population genomics that promotes data sharing and encourages contribution-based credit policy. The number of plant genome resequencing studies will continue to rise with the decreasing DNA sequencing costs, coupled with advances in analysis and computational technologies. This expansion, in terms of both scale and quality, holds promise for deeper insights into plant trait genetics and breeding design.

Key words: WGRS, WGS, resequencing, genome variation, adaptation

INTRODUCTION

The publication of the *Arabidopsis thaliana* genome in 2000 marked a significant milestone in plant genomics, heralding the beginning of a new era in plant research (Poczai et al., 2014). Since then, the field has witnessed remarkable advances in genomics, resulting in the sequencing and assembly of more than 700 plant genomes.

These advances have revolutionized our understanding of plant biology and facilitated groundbreaking discoveries in trait
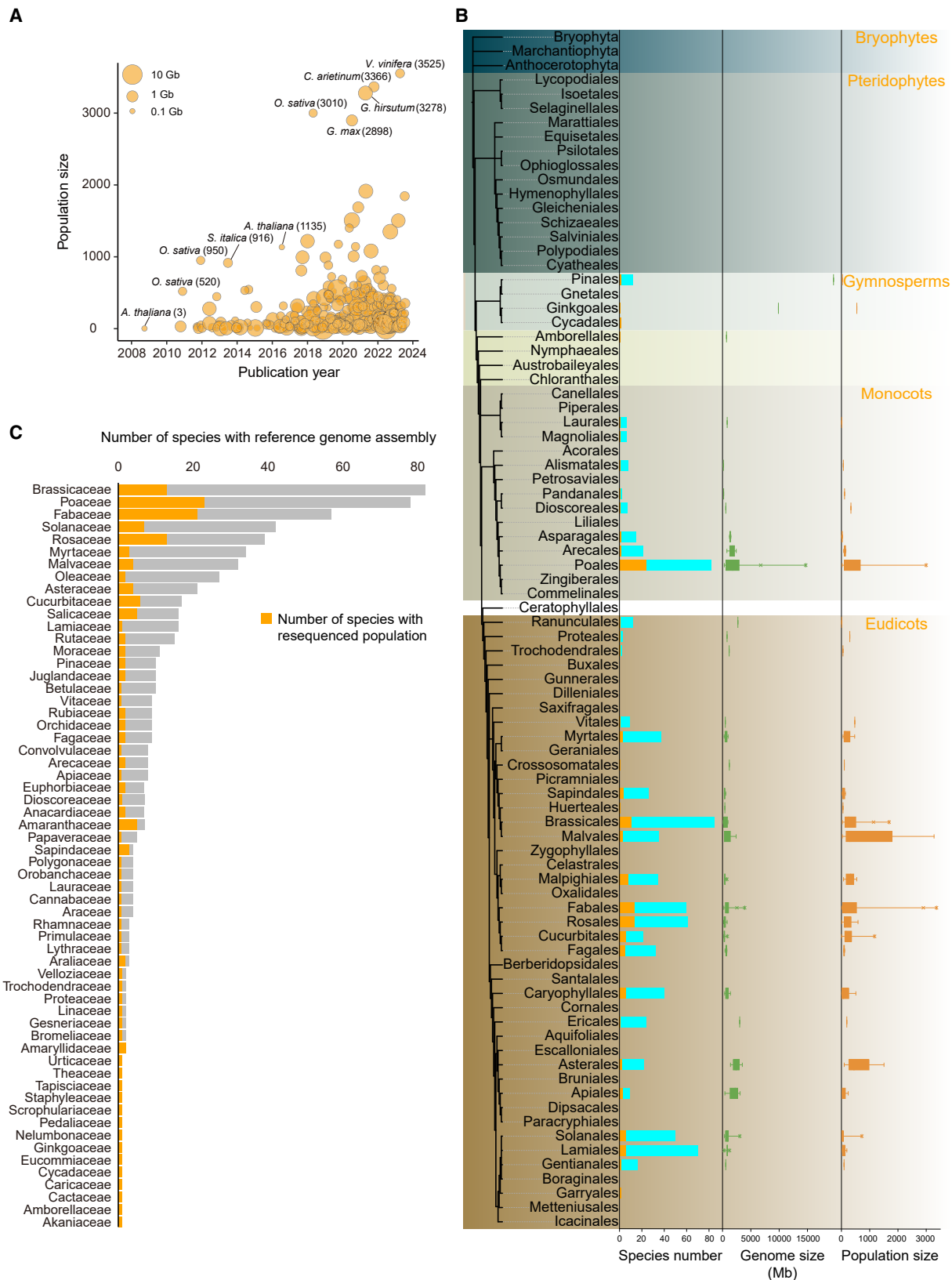
Figure 1. Temporal progression of resequencing and phylogenetic distribution of resequenced plant populations
(A) The changes of the size of resequenced populations. Individual data points are size scaled in accordance with the genome size of corresponding species.

(legend continued on next page)

genetics and crop improvement (Marks et al., 2021; Sun et al., 2021; Kress et al., 2022).

Following the accomplishment of reference genome assemblies, genome resequencing initiatives have explored the genetic diversity in plant populations. Initially, the high cost of DNA sequencing posed challenges for these endeavors. The first study in the 1001 Genomes Project, for example, sequenced only three ecotypes of *A. thaliana*, namely the reference (Col-0) and two others (Bur-0 and Tsu-1), using short-read Illumina technology with coverages ranging from 15- to 20-fold (Ossowski et al., 2008). Despite these limitations, the alignment of these short reads to the reference genome sequence (Col-0) enabled the identification of 823 325 single-nucleotide polymorphisms (SNPs). With decreasing costs, whole-genome resequencing has expanded in scale, enabling characterization of genetic variations within larger populations. For example, the comparison of SNP diversity between wild and domesticated landraces has enabled the characterization of domestication sweeps and the genes associated with domestication (Purugganan and Fuller, 2009; Meyer and Purugganan, 2013; Zhao et al., 2019a; Wei et al., 2021a). Integrating genome resequencing with phenotype measurements, such as crop yield, plant height, and metabolite abundance, has facilitated the detection of loci underlying these traits (Fang et al., 2017; Wu et al., 2019; Alseekh et al., 2021; Uffelmann et al., 2021; Marsh et al., 2023), and plant genome resequencing has greatly enhanced our understanding of crop domestication, plant adaptation, and evolution and the identification of loci conferring valuable traits (Huang et al., 2010; Jia et al., 2013; Abrouk et al., 2020; Mamidi et al., 2020; Kang et al., 2021; Lovell et al., 2021; Wang et al., 2021; Feng et al., 2022).

The construction of whole-genome variation maps has facilitated marker-assisted selection and genomic selection approaches to accelerate crop improvement. The driving force behind these advances is the growth of DNA sequencing data, often coupled with the construction of structured family-based mapping populations and, more recently, the application of machine-learning approaches for trait association (Danilevicz et al., 2021a, 2021b; Bayer et al., 2021; Gill et al., 2022). The development and increasing adoption of long-read DNA sequencing technologies (e.g., PacBio HiFi and Nanopore) have enabled the capturing of more extensive complex structural variations (SVs) and repetitive regions, which has pushed the boundaries of plant genome research.

We review studies of plant genome resequencing and the insights gained in relation to crop domestication and selection, plant-environment interactions, and genome-phenotype associations and provide a repository of candidate genes associated with various traits, which can be used for future research and application. In total, 187 plant species with 54 413 accessions have been resequenced. Although the resequenced species span the entire range of land plants, taxonomic gaps exist and most of the plant genome resequencing studies are focused on crops, particularly cereals and legumes. Future efforts are needed to fill these gaps

to further our understanding of plant genome evolution. The sequenced germplasms were collected globally from locations with a wide span of latitude and altitude and substantial variation of environmental factors, such as temperature, precipitation, and solar radiation, which can potentially facilitate more extensive studies of plant genomic responses to environments. The collation of data from broad communities presents an opportunity to conduct extensive comparative studies of plants. However, most of these genotyping datasets are not accessible. This lack of accessibility has hindered the secondary use of these datasets and led to redundant resequencing of germplasm.

WHOLE-GENOME RESEQUENCING: FROM MODEL ORGANISMS TO TAXONOMIC DIVERSITY

Whole-genome resequencing projects have grown rapidly with the continued reduction in DNA sequencing costs (Figure 1A). To date, genome resequencing has been reported for 187 plant species (Figure 1B; Supplemental Table 1). The number of accessions in each of these resequencing studies reached a peak of 3525 for grapevine (*Vitis vinifera*) accessions (Dong et al., 2023), followed by 3366 chickpea (*Cicer arietinum*) accessions with the completion of the chickpea genome project (Varshney et al., 2021), 3010 *Oryza sativa* accessions (Wang et al., 2018a), and 2898 soybean (*Glycine max*) accessions (Liu et al., 2020a). These large-scale resequencing endeavors have facilitated the generation of nearly complete SNP maps, including rare alleles. Initially, genome resequencing studies focused on model plants and crops with small, diploid genomes, such as *O. sativa* (Huang et al., 2010) and tomato (*Solanum lycopersicum*) (Causse et al., 2013), after the release of respective high-quality reference genome assemblies (Figure 1A; Supplemental Video 1) (Marks et al., 2021; Sun et al., 2021). The resequencing of plants with large and more complex genomes faced the challenges of higher costs and demanding computational requirements. However, despite these limitations, population genomic studies on plants with large and complex genomes emerged soon after the completion of reference genome sequences (Fernandez et al., 2022). For example, after the publication of the hexaploid wheat (*Triticum aestivum*) reference genome in 2018 (Appels et al., 2018), several resequencing studies of wheat were published in 2019 and 2020 (Cheng et al., 2019; Hao et al., 2020; Walkowiak et al., 2020; Zhou et al., 2020). As costs come down, genome and pan-genome references are being constructed for the less mainstream crops (Bayer et al., 2020; Golicz et al., 2020), including sorghum (*Sorghum bicolor* L.) (Ruperao et al., 2021), pigeonpea (*Cajanus cajan* L.) (Zhao et al., 2020), and banana (*Musa* spp.) (Rijzaani et al., 2022). The development of more advanced sequencing techniques (Wenger et al., 2019; Khan et al., 2020; Michael and VanBuren, 2020; Yuan et al., 2020) and assembly algorithms (Koren et al., 2017, 2018; Ruan and Li, 2020; Chen et al., 2021) capable of managing complex genomes, including heterozygous and polyploid genomes (Danilevicz et al.,

---

(B) The phylogenomic distribution and summary of the resequenced plant populations. The stacking histogram delineates the cumulative number of species with reference genome assemblies, with the quantity of resequenced species in each taxonomic order represented by orange bars. Accompanying boxplots depict the range of genome size and population size for resequenced species per order.
(C) Number of species with a reference genome assembly and the species in each family with a resequenced population. The number of resequenced species within each family is designated by orange bars.

2020), is expected to further expand the number of resequenced species and the publication of additional high-quality reference genome assemblies.

While reference genomes have been established for species spanning the phylogenetic tree of land plants, resequencing studies have been more concentrated within a few orders and families (Figures 1B and 1C; Supplemental Tables 2 and 3). Out of the 187 resequenced species, 38% belong to four families, with 12% (23) in Poaceae, 11% (21) in Fabaceae, 7% (13) in Brassicaceae, and 8% (14) in Rosaceae (Figure 1C; Supplemental Table 3). When considering the number of accessions resequenced for each species, it is evident that the scale of genome resequencing studies is strongly influenced by the economic importance of the species, with a significant bias toward crops, particularly those in the Fabaceae and Poaceae families. There are still 49 orders for which no species have been resequenced. Although reference genomes are available for some species in 23 of these orders, a reference genome is still lacking for the remaining 26 orders (Supplemental Table 2). With the decreasing cost of DNA sequencing, recent efforts have expanded beyond crops to include a broader range of species, including gymnosperms (Zhao et al., 2019b; De La Torre et al., 2019; Kress et al., 2022).

Gymnosperms once dominated land on earth but are represented by only four lineages today: Coniferophyta, Cycadophyta, Ginkgophyta, and Gnetophyta, comprising approximately 1000 species (Crang et al., 2018). The genomic datasets obtained from these gymnosperms provide opportunities to enhance our understanding of plant genome evolution and adaptation. For example, the sequencing of 545 *Ginkgo biloba* genomes has provided insights into the population dynamics of this species during the Pleistocene glaciations, revealing multiple rounds of expansion and reduction that correlated with climate change (Zhao et al., 2019b). The exploration of taxonomic diversity can be enriched by studies of medicinal plants that are widely distributed taxonomically. Reference genomes of medicinal plants are rapidly accumulating, and we can expect a similar growth in genome resequencing studies. Ren et al. (2021) recently resequenced 110 accessions of *Cannabis sativa*, leading to the identification of genes associated with the differentiation of drug-type and hemp-type cultivars.

The African Orphan Crops Consortium (AOCC), established in 2011, aims to generate reference genome assemblies for 101 African orphan crops and to resequence 100 accessions for each crop (http://africanorphancrops.org/). Although the focus of the AOCC is on crops, many of these species do not belong to major crop families. Plants targeted by AOCC include species in 28 orders, representing 45 families, in 19 of which no resequencing has been done to date (Hendre et al., 2019). Consequently, the completion of this project would not only advance research and breeding for these orphan crops but also contribute to filling the taxonomic gaps in genome resequencing endeavors.

CO-DOMESTICATION OF LEGUMES AND CEREALS

A total of 54 413 accessions were obtained from 163 countries across all continents except Antarctica (Figure 2A;

Supplemental Table 4). The distribution of germplasm collection centers aligns with the Vavilovian centers of crop plant origins, with eight centers showing exact concordance (Harlan, 1971) (Figure 2A). Additionally, two centers, West Africa and America, have been suggested to be domestication centers for rice in other studies (Cubry et al., 2018). Interestingly, there is a significant overlap in the collection locations of grasses and legumes (Figure 2B), implying co-domestication. The domestication centers and spreading routes can be deduced by analyzing the SNP datasets of populations, and by taking this approach, resequencing initiatives have reported the domestication centers of 28 crops, including six cereals, six legumes, six vegetables, and 10 fruits (Supplemental Table 5). The results reveal that the domestication of cereal and legume crops is often intertwined. For example, rice, millet, and soybean were domesticated in China, while wheat, barley, chickpea, and lentil were domesticated in the Fertile Crescent. Sub-Saharan Africa was the domestication center for crops such as sorghum, pearl millet, and cowpea, whereas maize, common bean, and lima bean were domesticated in Meso-America (Abbo and Gopher, 2017). These domestication centers represent the early agricultural communities where different crops were domesticated. The simultaneous domestication of cereals and legumes could also be attributed to their complementary nutritional profiles. Cereals serve as the primary carbohydrate source but have low protein content (Henchion et al., 2017), whereas legumes are rich in proteins (20%–45%) (Maphosa and Jideani, 2017) and minerals. Moreover, legumes are abundant in essential amino acids such as lysine but low in sulfur-containing amino acids (SCAAs) such as methionine, cystine, and cysteine. On the other hand, cereals are rich in SCAAs but deficient in lysine (Staniak et al., 2014). The combined consumption of cereals and legumes in the domestication centers likely provided balanced nutrition to the inhabitants.

GENE SELECTION DURING DOMESTICATION

Selection for traits such as low seed shattering and enhanced inflorescence structures in grain crops during domestication leads to reduced nucleotide diversity (Gross and Olsen, 2010). Comparative analysis of wild and cultivated populations allows for the detection of selection signals during domestication and the identification of genes for breeding improved varieties (Tirnaz et al., 2022). Huang et al. (2012a) identified 60, 62, and 55 domestication-related loci in *indica*, *japonica*, and mixed rice populations, respectively, encompassing known domestication genes linked to seed shattering (*sh4*, *qSH1*), hull color (*Bh4*), pericarp color (*Rc*), tiller angle (*PROG1*), grain width (*qSW5*), and grain quality (*Waxy*). *PROG1* also exhibited evidence of strong selection during the domestication of African rice (*O. glaberrima*), indicating convergence of domestication between Asia and Africa (Cubry et al., 2018). Domestication of other cereal crops, such as wheat (Cheng et al., 2019; Zhou et al., 2020), maize (Hufford et al., 2012), *Coix* (Liu et al., 2020b), pearl millet (Varshney et al., 2017), foxtail millet (Jia et al., 2013), and fonio millet (Abrouk et al., 2020), also involved selection of genes involved in plant architecture, spike morphology, seed shattering, grain width, or flowering.
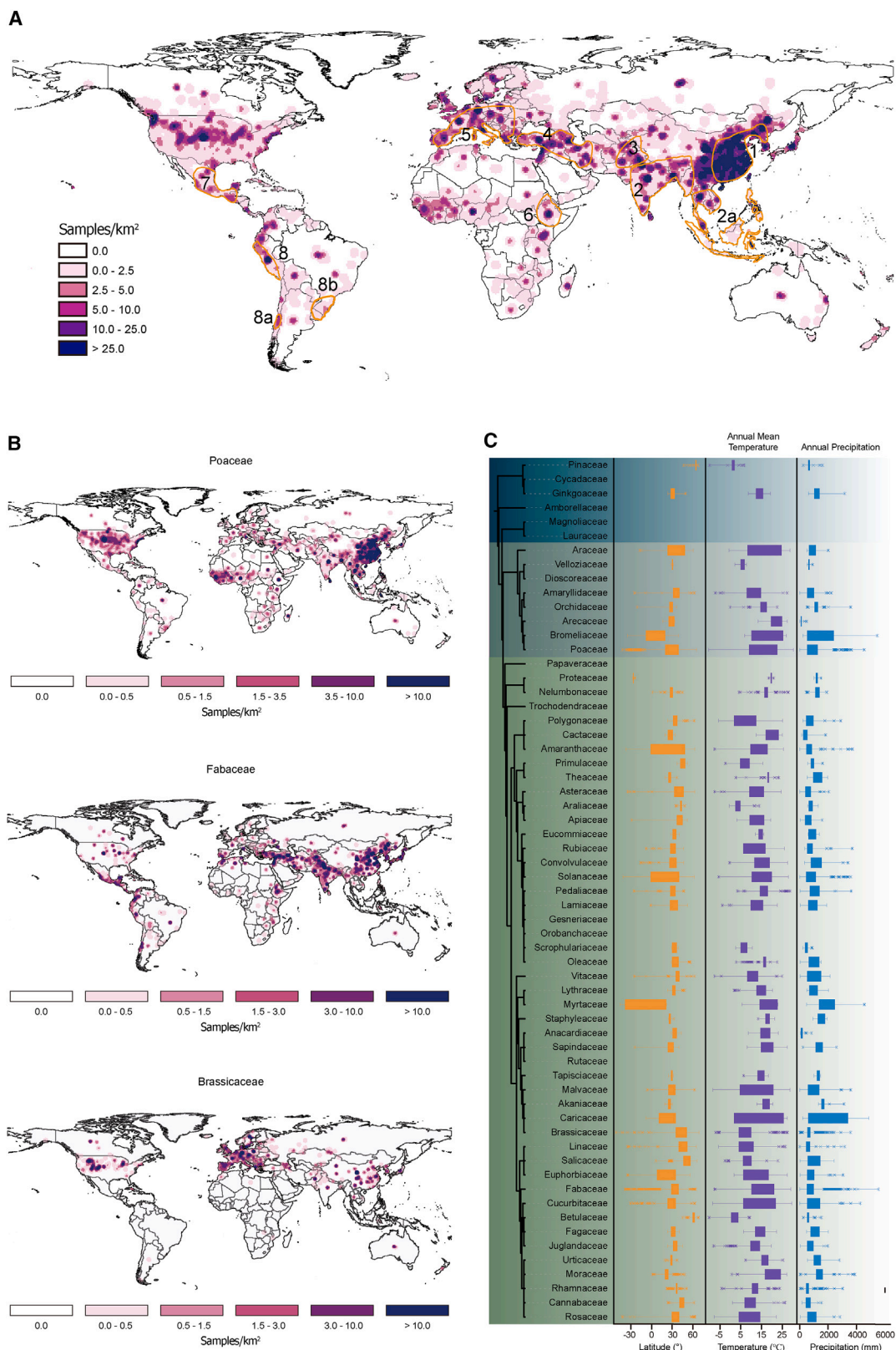
Figure 2. Geographic distribution and environmental context of germplasm collection sites for population resequencing.
(A) Global distribution of all collected germplasms, with numbered orange-circled regions denoting Vavilov centers of origin.
(B) Geographical distribution of the top three families (Poaceae, Fabaceae, and Brassicaceae) with the most resequenced germplasms.
(C) Latitude, average temperature, and precipitation variables at the locations of collected, resequenced germplasms. Additional environmental variables are provided in Supplemental Table 4.

Genes selected in wheat included those controlling non-brittle rachis and shattering (e.g., *TtBtr1-A* and *TtBtr1-B*) as well as those involved in circadian rhythm pathways and flowering (Cheng et al., 2019; Zhou et al., 2020). Approximately 7.6% of the maize genome experienced selection during domestication, and these regions included genes potentially associated with flowering, yield, nitrogen mobilization, seed germination, and gibberellin biosynthesis (Hufford et al., 2012). Selection in legume genomes primarily targeted genes involved in pod dehiscence, seed size, seed development and germination, stress responses, and flowering (Zhou et al., 2015; Varshney et al., 2019; Garcia et al., 2021). For example, the selected regions detected in the soybean genome overlapped with quantitative trait loci (QTLs) harboring genes responsible for pod dehiscence (Zhou et al., 2015). A significant loss of gene content and average gene number was also observed during domestication in soybean in a study of >1000 resequenced lines from the US Department of Agriculture (USDA) soybean collection (Bayer et al., 2022b). These data have been reused to define characteristics of gene families affecting traits of agronomic importance such as protein content (Marsh et al., 2022) and oil production (Derbyshire et al., 2023). Domestication sweeps in fruit and vegetable crops favored genes related to taste, flavor, nutrition, and disease resistance (Shang et al., 2014; Duan et al., 2017; Zhu et al., 2018; Guo et al., 2019). Peach domestication is associated with sweeps related to sugar and acid content (Li et al., 2019a). Domestication in leafy vegetables favored regions associated with leaf traits such as leaf surface texture, leaf base shape, and petiole color. For example, selected regions in the spinach genome contain genes involved in cutin assembly, and a selected region in lettuce was found to overlap with a QTL controlling lobed leaf morphology (Wei et al., 2021a; Cai et al., 2021).

Genes selected during domestication are targets for further crop improvement, especially because of their consistent overlap with QTLs for desirable agronomic traits (Huang et al., 2012a; Zhao et al., 2019a; Wei et al., 2021a; Garcia et al., 2021), some of which have undergone secondary selection in breeding programs (Mace et al., 2013; Li et al., 2019a; Varshney et al., 2021). However, functional characterization is still lacking for many genes within these regions, warranting further investigation.

GENOMICS RESPONSES TO ENVIRONMENTAL ADAPTATION

With the increasing effect of climate change on agriculture, there is an urgent need to breed crops adapted to new and more variable climates (Abberton et al., 2016; Anderson et al., 2020). Understanding species-wide diversity enhances the identification of suitable adaptive alleles (Marsh et al., 2021; Petereit et al., 2022). Resequencing datasets that span a wide range of habitats can provide valuable insights into understanding how genomes respond to changing environments. Figure 2A presents an example of a resequencing dataset with collection sites that vary in latitude (68°N to 52°S) and altitude (−29 to 5482 m) (Figure 2A; Supplemental Video 1). To illustrate the environmental variation, we examined climatic data from Worldclim (https://www.worldclim.org/), which showed a range of annual average temperatures (−10.5°C–30.4°C), precipitation

(0–5561 mm), and solar radiation (7367–22 789 kJ m^{-2} day^{-1}) across these locations (Figure 2C; Supplemental Table 4). This rich geographic and bioclimatic diversity has facilitated the investigation of genomic responses to different environments and the identification of adaptive genes (Exposito-Alonso et al., 2019; Li et al., 2020; Mamidi et al., 2020; Ji et al., 2021).

Flowering time

Flowering time is an important phenotype associated with plant adaptation. The 1001 Genomes Project uncovered significant associations between the flowering time of *Arabidopsis* and genes involved in vernalization and flowering regulation. Notable examples include *VERNALIZATION INSENSITIVE3* (*VIN3*), *FLOWERING LOCUS C* (*FLC*), *FLOWERING LOCUS T* (*FT*), *SHORT VEGETATIVE PHASE* (*SVP*), and *DELAY OF GERMINATION 1* (*DOG1*). In *Brassica napus*, 28 MADS-box genes, including *FLC*, were found to be associated with variation of flowering time (Song et al., 2020). Similarly, *SENSITIVITY TO RED LIGHT REDUCED 1* (*SRR1*) and *VIN3* were found to be associated with flowering time in *B. juncea* (Kang et al., 2021), while, in sunflower, genomic regions harboring *FT*, *FLOWERING LOCUS M* (*FLM*), and *EARLY FLOWERING7* (*ELF7*) were associated with flowering (Todesco et al., 2020). Several of these genes are conserved across different plant species, reflecting conserved mechanisms that regulate flowering time and adaptation to environments.

Temperature and precipitation

Genes associated with response to abiotic stress are valuable resources for crop breeding and a subject of extensive investigation. For example, the gene *HEAT-INTOLERANT1* (*HIT1*), which confers heat stress tolerance, was found to be associated with several temperature-related climatic variables in a study on sunflower (Todesco et al., 2020). In *Setaria viridis*, 140 loci near or within genes related to organellar genome maintenance, protein breakdown, gibberellic acid homeostasis, and fucose metabolism were associated with adaptation to extreme temperature and precipitation (Mamidi et al., 2020). Analysis of fonio millet revealed that genomic loci associated with heat stress also harbored genes related to hormone biosynthesis, metabolic processes, shoot apical meristem maintenance, and root growth and development, while a locus associated with drought response included a homolog of *Arabidopsis WSD1*, which is known to enhance drought tolerance through wax accumulation (Abrouk et al., 2020). In a study in soybean, SNPs within genes encoding several protein kinases, and a gene encoding a 14-3-3 protein, and a gene encoding a C-repeat-binding factor were associated with temperature response, and homologs of these genes in *Arabidopsis* are known for their involvement in cold and freezing stress tolerance (Wang et al., 2022). Similarly, a walnut genome-environment association study identified three temperature-associated genomic regions containing genes related to cold tolerance, along with a precipitation-associated region harboring a cluster of nucleotide-binding site leucine-rich repeat (NBS-LRR) genes and a gene with a GRAM domain known to confer drought and salt tolerance in rice (Zheng et al., 2020a; Ji et al., 2021).

By revisiting the genotyping dataset of the 1001 *Arabidopsis* genomes project, Ferrero-Serrano and Assmann (2019) found that variation in the heterotrimeric G-protein subunit AGG3 was

associated with cold tolerance, and this was experimentally confirmed by observing the cold-induced leaf senescence in single and triple *agg* mutants. Certain genes involved in flowering have also been reported to overlap with regions responsive to temperature or precipitation. For example, a walnut gene homologous to *Arabidopsis TERMINAL FLOWER 1* (*TFL1*) was found in a temperature-associated region; a soybean homolog of the *PHYTOCHROME E* gene was located within a block associated with temperature; and a flowering-related gene, *GmELF4*, was located near a region associated with precipitation response (Li et al., 2020). When comparing genes regulating flowering with those associated with temperature adaptation, it is apparent that different plants employ distinct elements to cope with changing environments.

Solar radiation

Investigating adaptation to variation in solar radiation often involves the study of accessions collected from high-altitude regions, such as the Qinghai-Tibetan plateau. In the case of qingke, a Tibetan variety of barley, several genomic regions were found to be associated with high-altitude adaptation (Zeng et al., 2018). In a study of Tibetan poplar, 74 SNPs were found to be associated with altitude, while 82 SNPs were associated with solar radiation. Notably, an SNP located within a gene encoding a putative receptor-like protein kinase was found to be associated with both variables (Zheng et al., 2020b). In wheat, nearly 2000 genomic regions harboring 3847 genes were identified as being under selection in a high-altitude environment. The authors implicated a gene homologous to *Arabidopsis ELONGATED HYPOCOTYL 5* (*HY5*) as playing an important role in mediating wheat's adaptation to the plateau. Interestingly, over 86% of Tibetan wheat accessions carry one haplotype (AT haplotype) of *HY5*, whereas most accessions from low-altitude environments carry an alternative GC haplotype (Guo et al., 2020a). In walnut, a gene encoding leucoanthocyanidin dioxygenase (*LDOX*), which is involved in flavonoid biosynthesis and known to combat reactive oxygen species resulting from ultraviolet radiation, was located within a genomic region associated with adaptation to high altitude (Ji et al., 2021), and this gene was positively selected in walnut accessions collected from Tibet. In a genome-environment association study of *Arabidopsis*, the *SPA1-RELATED 2* (*SPA2*) gene, which is known to respond to various light conditions and to inhibit de-etiolation of *Arabidopsis* seedlings, was found to be significantly associated with radiation intensity during summer (Ferrero-Serrano and Assmann, 2019). These findings highlight the genetic mechanisms underlying plant adaptation to solar radiation and the importance of specific genes in enabling plants to thrive under varying light conditions.

Soil fertility

Adaptation to variation in soil fertility has also been investigated. In a study of *A. lyrata*, seven loci potentially associated with serpentine adaptation were identified by comparing the nucleotide diversity between populations collected from serpentine and granitic soils (Turner et al., 2010). In sunflower, three genomic regions on chromosomes 9, 11, and 14 were found to be associated with soil characteristics (Todesco et al., 2020). Furthermore, Liu et al. (2021) discovered an SNP on rice chromosome 6 related to tillering in response to nitrogen, which
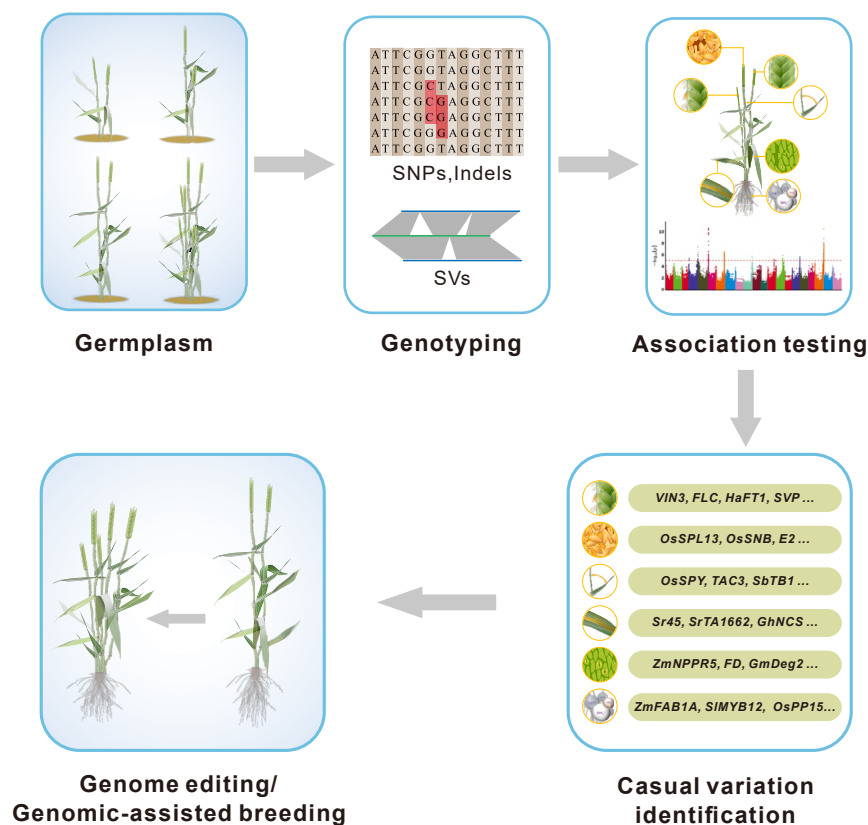
serves as a proxy for nitrogen-use efficiency (NUE). This finding was made by sequencing accessions collected from locations with varying soil nitrogen. These studies shed light on the genetic mechanisms underlying plant adaptation to soil fertility, particularly in relation to specific soil types and nutrient availability.

GENOME-TRAIT ASSOCIATION STUDIES

Advances in DNA resequencing have enabled rapid discovery of SNPs in crops, facilitating the identification of genes or loci associated with a variety of important traits, thereby enhancing our understanding of these crops and accelerating their breeding (Figure 3) (Huang and Han, 2014). Through genome-wide association studies (GWASs), regions associated with specific traits have been identified. In addition to loci associated with climatic adaptation, several genomic regions have been reported to be related to field traits, such as NUE, yield components, and metabolite content. A total of 8332 loci conferring 367 phenotypes were identified from 38 crops, among which 6573 loci were associated with 257 field traits (Supplemental Table 6). Four major categories of phenotypes were studied, namely yield, plant architecture, tolerance to biotic and abiotic stresses, and metabolite diversity. Supplemental Table 7 provides a list of genes related to these traits that were identified in previous GWAS analysis. Although most of these genes have not yet been experimentally validated, they hold potential as candidates for future research and application in breeding.

Yield

Yield-related traits hold great economic importance and have therefore been extensively studied. Many important loci related to yield have been identified through GWAS. For example, a QTL region associated with soybean seed weight has been mapped to the *qSW* locus (Zhou et al., 2015). Si et al. (2016) and Yu et al. (2017) identified genes encoding a plant-specific transcription factor, *OsSPL13*, and APETALA2/ethylene-responsive element binding protein 125 that control rice grain length and enhance yield. Ma et al. (2019) identified another gene controlling rice grain size, *OsSNB*, where a mutation can lead to an increase of grain length, width, and weight. Transcription factors have been found to play important roles in crop yield traits. For example, an MYB transcription factor was associated with rice grain weight (Huang et al., 2012b), while a region containing two auxin response factor-encoding genes was linked to storage root weight in cassava (Hu et al., 2021). Additionally, ARF18 and ribosome recycling factor genes were found to be associated with seed weight and silique length of rapeseed (Hu et al., 2022). In tartary buckwheat (*Fagopyrum tataricum*), genes encoding an AP2 transcription factor and a tryptophan aminotransferase-related protein were identified as candidates influencing seed weight and seed size, respectively (Zhang et al., 2021a). Similarly, an AP2 transcription factor has been associated with cotton yield traits, and an AP2/ETHYLENE response factor-encoding gene was identified as conferring an increase of lint yield, the number of bolls per plant, and a decrease of seed index (Fang et al., 2017). WRKY, bZIP, and NAC transcription factors were found to be associated with the seed size of castor bean (Cai et al., 2021). Genes involved in ubiquitination have also been associated with yield traits. A region harboring a gene encoding a ubiquitin-conjugating

in rapeseed. In sorghum, the *SbTB1* gene was identified as a candidate related to lateral branch number (Wu et al., 2022), while a GWAS analysis in *Populus* led to the identification of *GAUT9* as a gene influencing leaf area (Chhetri et al., 2019).

Tolerance to abiotic stress

The ability of plants to tolerate various abiotic stresses, such as extreme temperature, salt, and drought, is important for their growth and adaptation. GWAS has proved to be valuable in identifying loci and candidate genes associated with tolerance to these stresses. For example, GWAS revealed three cotton loci containing genes from the G-type lectin receptor kinase family that are related to high-temperature responses (Ma et al., 2021a). In rice, GWAS identified a gene encoding Stress-Associated Protein 16 (*OsSAP16*), which controls low-temperature germination (Wang et al., 2018b). Additionally, GWAS enabled the identification of a gene, *NIP1;1*, in *Arabidopsis* that confers sensitivity to toxic H_2O_2 (Sadhukhan et al., 2017). Several transporters associated with salt tolerance were identified through GWAS in different plants, including *OsHKT1;1*, *OSHKT1;4*, *OsHKT1;5*, *OsHKT2;1*, *OsMOLYBDATE TRANSPORTER1;1*, *OsHAK5*, *OsHAK6*, and *OsZmHAK4* (Meyer et al., 2016; Campbell et al., 2017; Yang et al., 2018; Zhang et al., 2019a; Hartley et al., 2020). However, identifying genes associated with these traits can be laborious, highlighting the need for more efficient methods. One such method is a non-destructive phenotyping pipeline that offers a means of automatically monitoring drought response in crops. In a rice study by Guo et al. (2018), 51 image-based traits (i-traits) related to drought tolerance (DT) were collected using a non-destructive phenotyping facility, and subsequently used for GWAS. A total of 443 loci associated with DT were identified, 93% of which overlapped with previously reported DT-related loci. Furthermore, a new gene (*OsPP15*) associated with DT was discovered and experimentally validated. The same technology was applied to maize to identify loci for DT by monitoring phenotypes under normal and drought-stress conditions for 98 days (Wu et al., 2021). Consequently, 1592 loci containing 2318 candidate genes were detected, including previously known drought-responsive genes, two of which (*ZmcPGM2* and *ZmFAB1A*) have been experimentally validated. These studies demonstrate the value of non-destructive phenotyping performed in combination with GWAS for the identification of genes and loci associated with tolerance to abiotic stress in crops.

Resistance to biotic stress

Both major and minor genes have been implicated in the response to biotic stress. The major disease resistance

enzyme E2 was associated with yield-related traits in soybean and flax seed (Fang et al., 2017; Guo et al., 2020b). Studies have found that cytochrome genes are frequently associated with seed weight in *B. juncea* (Kang et al., 2021) and common bean (Wu et al., 2020).

Plant architecture

Plant architecture is determined by various complex traits, including plant height, branch structure, tiller angle, and leaf and panicle morphology (Cheng, 2022). The application of plant genome resequencing has expedited the identification of several genes governing plant architecture. In rice, loci related to tiller angle were detected by GWAS, leading to the identification of *TAC3* as a candidate causal gene (Dong et al., 2016) responsible for increased tiller angle. Similarly, three basic helix-loop-helix (*bHLH*) family genes, *OsbHLH153*, *OsbHLH173*, and *OsbHLH174*, were found to be associated with leaf angle in rice (Dong et al., 2018). Genes with pleiotropic effects have also been reported, including a GATA zinc-finger-type transcription factor associated with both plant height and panicle length, as well as *OsNAL1*, which is associated with panicle number and leaf blade width (Yano et al., 2016). In rice, a GWAS facilitated the identification of an *N*-acetyl glucosamine transferase-encoding gene (*OsSPY*) that controls rice architecture (Yano et al., 2019). Some regions associated with tiller angle and leaf angle in different studies were found to be overlapping (Dong et al., 2018). Resequencing in spinach led to the identification of a *FAR1*-related gene that was located within a locus associated with petiole width, while a beta-tubulin gene was proposed as a candidate for a leaf surface texture trait (Cai et al., 2021). Hu et al. (2022) reported *BnPIN2*, which encodes an auxin efflux carrier, to be associated with leaf angle

(R) genes often have a classic NBS-LRR structure. In a GWAS for rice blast resistance, associated SNPs were reported to co-localize with an NBS-LRR gene (Xiao et al., 2021). Similar reports have been made in spinach for white rust disease and downy mildew (Cai et al., 2021; Shi et al., 2022). Consequently, the R gene enrichment and sequencing (RenSeq) method was developed for NBS-LRR gene sequencing (Jupe et al., 2013) and has been successfully applied in several different plants, including tomato (Jupe et al., 2013), potato (Witek et al., 2016; Chen et al., 2018), *Arabidopsis* (Van de Weyer et al., 2019), wheat (Arora et al., 2019; Strachan et al., 2019), and lettuce (Parra et al., 2021). When combined with mutagenesis and kmer-GWAS, RenSeq enables rapid identification of genes conferring resistance to different diseases. For example, the stem rust resistance genes *Sr22* and *Sr45* were identified in hexaploid wheat (Steuernagel et al., 2016), while *Sr33*, *Sr45*, *Sr46*, and *SrTA1662* were identified in *Aegilops tauschii* (Arora et al., 2019), the D genome donor of bread wheat. Other genes associated with disease resistance in other species include *Ga11G2352*, a gene encoding a glutathione *S*-transferase and conferring resistance to *Fusarium* wilt disease in cotton (Du et al., 2018). Another gene, *GhNCS*, encoding *S*-norcoclaurine synthase, was identified as a candidate associated with susceptibility to *Verticillium* wilt of cotton (Ma et al., 2021b). In maize, *ZmFBL41*, a gene encoding an F-box protein, was associated with resistance to maize banded leaf and sheath blight caused by *Rhizoctonia solani* (Li et al., 2019b). An F-box protein-encoding gene in *Medicago truncatula* was also shown to be involved in disease resistance, with SNPs within its promoter region reported to be associated with resistance to *Aphanomyces euteiches* (Bonhomme et al., 2014). Using GWAS, Li et al. (2017) identified a natural allele of a C2H2-type transcription factor that conferred broad-spectrum resistance to rice blast.

Metabolites

Metabolite-based GWAS (mGWAS) has increased in popularity with the ability to profile diverse metabolites coupled with crop resequencing, and it has enabled the identification of loci associated with flavor, nutrition, and abundance of bioactive substrates. mGWAS has been particularly useful in studying leafy vegetables, fruit crops, and medicinal plants. The first mGWAS study was conducted in *A. thaliana*, in which 43 metabolites including the contents of individual glucosinolate (GSL) secondary metabolites and their ratios, were investigated, leading to the identification of two loci conferring variation in GSLs (Chan et al., 2010). Since then, mGWAS has been widely applied in the study of horticultural crops, including apple (*Malus*) (Bilbrey et al., 2021) and tomato (*S. lycopersicum*) (Zhu et al., 2018).

Sugar content has also been widely studied, resulting in the identification of sucrose synthase and raffinose synthase as being associated with watermelon flesh sweetness (Guo et al., 2019). Loci associated with sucrose and acid content, as well as 12 aromatic compounds, were identified in grapevine (Liang et al., 2019), while loci associated with soluble sugar contents, malate, and citrate accumulation (Liao et al., 2021) have been identified in apple. Flavor and taste in vegetables and fruits are often linked to volatile compounds. In a tomato study, 251 loci were found through GWAS to be associated with 20 traits, including 4 non-volatile and 15 volatile flavor compounds (Tieman et al., 2017). In a study in blueberry, several SNPs were reported to be associated with volatile compounds, including linalool, *D*-limonene, and eucalyptol (Ferrão et al., 2020). The abundance of stone cells in pear has been associated with 12 loci, with several genes implicated in lignin regulation (Zhang et al., 2021b). Comprehensive metabolic profiling of 442 accessions of *S. lycopersicum*, coupled with genomic association analysis, led to the identification of 3526 loci associated with 514 metabolites, including 351 uncharacterized ones (Zhu et al., 2018). In the same study, analysis of the association between genomic variants and transcriptome variation was performed, resulting in the identification of 2566 *cis*- and 93 587 *trans*-eQTLs for 2566 and 2461 genes, respectively. Additionally, genomic variants have been associated with other molecules, including proteins (Zhang et al., 2019b; Liang et al., 2019; Liu et al., 2019; Sun et al., 2019) and minerals (Qin et al., 2015; Hindu et al., 2018). Although only a few loci have been identified thus far to be associated with small molecules (Supplemental Table 6), ongoing research investment in medicinal plants rich in secondary compounds with potential pharmaceutical or industrial value will likely lead to an expansion of this list.

GENOMIC-ASSISTED BREEDING

The identification of genomic variants through genome resequencing studies provides valuable resources for breeding and crop improvement. These DNA sequence variants can be used to introduce specific alleles, predict hybrid performance, and design more effective breeding strategies. For example, the introduction of haplotype H of *OsTCP19* (*OsTCP*-H) into two elite rice cultivars increased tiller number under low- and moderate-nitrogen conditions (Liu et al., 2021). Resequencing more than 3000 accessions of chickpea, Varshney et al. (2021) identified 24 consistent and stable superior haplotypes absent in 80% of breeding lines and assisted in the selection of 56 donor lines for use in breeding. Resequencing data can be helpful in the prediction of hybrid performance. A model trained to predict grain yield in pearl millet identified 170 promising hybrid combinations, with 11 showing improved performance and 159 combinations showing potential for future improvement (Varshney et al., 2017). Similarly, models trained using GWAS-derived SNPs enabled the prediction of white rust resistance in spinach, achieving a correlation higher than 0.75 with only 40 SNPs (Shi et al., 2022). In a recent maize study, environmental variables were integrated into models to accurately predict maize yield across Europe under different conditions (Millet et al., 2019). Wei et al. (2021b) developed RiceNavi, an *in silico* breeding system composed of three major modules: RiceNavi-QTNpick, -Sim, and -Sample-Select. This system enables sample selection, breeding design, and optimization *in silico*, allowing for precise introgression of target genotypes in a short period while reducing linkage drag. While the power and efficiency of RiceNavi has been demonstrated in rice, more sophisticated systems for other crops are still required. Plant genome resequencing data provide fundamental resources with considerable value for training performance prediction models and developing advanced *in silico* breeding systems that can significantly accelerate crop breeding programs.

CONCLUDING REMARKS AND PERSPECTIVES

A new era of plant population genomics

The decreasing cost of high-throughput DNA sequencing has led to an explosion of large-scale population genomic studies. The read-length limitations of early next-generation sequencing (NGS) platforms (Slatko et al., 2018) encouraged the heavy reliance of population studies mainly on SNP variants. Although some earlier studies employed methods for identifying SVs (Liu et al., 2020a; Mamidi et al., 2020; Todesco et al., 2020), it was challenging to accurately detect SVs from NGS datasets (Ho et al., 2020; Yuan et al., 2021). The availability of more advanced DNA sequencing technologies currently offers a solution for the characterization of SVs. The long-read DNA sequencing techniques from PacBio high fidelity (HiFi) and Oxford Nanopore GridION have been used to construct pan-genomes for several crops (Khan et al., 2020), including soybean (Liu et al., 2020a), wheat (Walkowiak et al., 2020; Bayer et al., 2022a), rice (Qin et al., 2021), barley (Jayakodi et al., 2020), tomato (Zhou et al., 2022), and potato (Tang et al., 2022). These studies mark the beginning of a new era of genome research. Long-read genome resequencing has provided insights into the evolution and roles of SVs, which cannot be adequately resolved using short NGS reads. For example, Walkowiak et al. (2020) detected several presence/absence variants and copy number variants of genes, along with large-scale SVs, including introgressions, translocations, and shifts of centromeres across wheat genomes. In a soybean resequencing study, the genomes of 26 representative soybean individuals were assembled *de novo*, revealing fusion genes and large-scale SVs. Approximately 1000 of the SVs were reported to result in changes in gene expression levels associated with agronomic traits (Liu et al., 2020a).
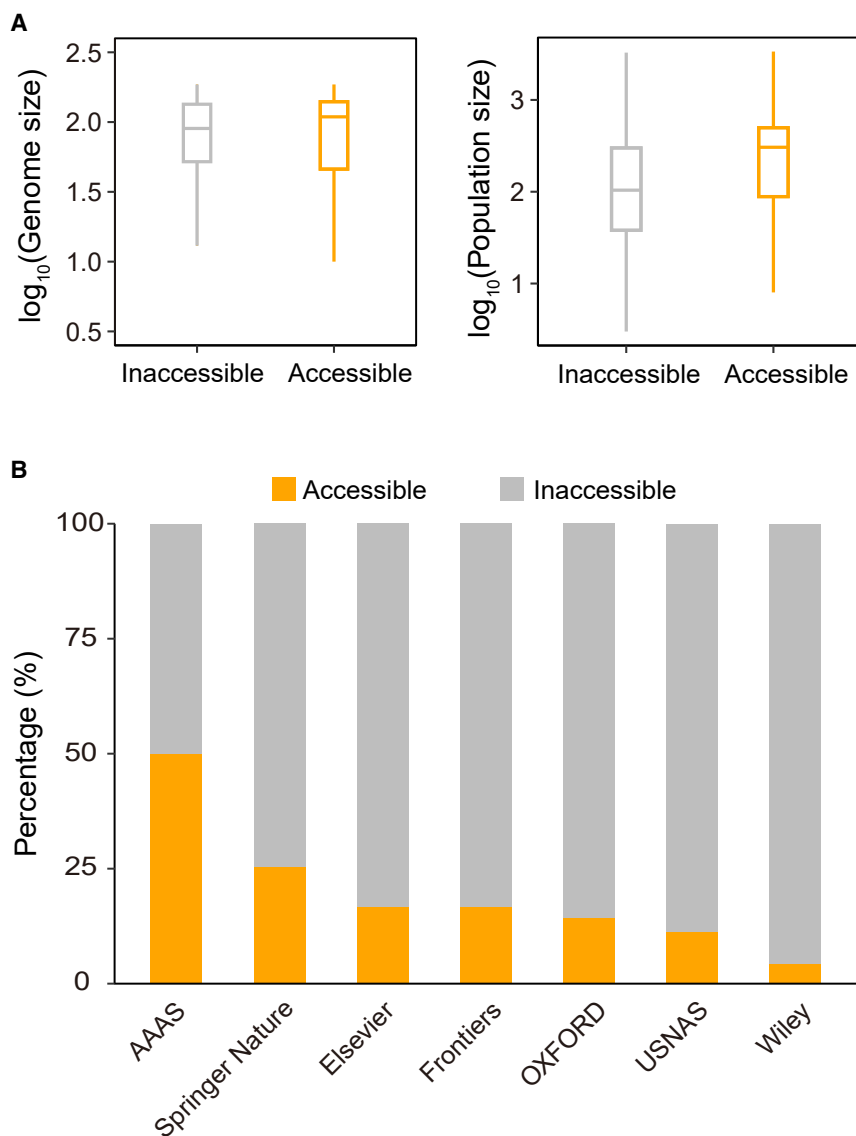
The application of long-read DNA sequencing has enabled the construction of graph-based pan-genomes that can replace single reference genomes in plant genome studies (Bayer et al., 2020; Zanini et al., 2022), capturing heritability that is often missed in linear references (Edwards and Batley, 2022)(Zhou et al., 2022). As variant calling relies on a reference genome, it is important to reevaluate population resequencing data once a graph pan-genome reference becomes available. One approach is to map NGS reads to the graph-genome reference and call the variants. Several tools have been developed for this purpose, including GraphTyper (Eggertsson et al., 2017), GraphAligner (Rautiainen and Marschall, 2020), and HISAT2 (Kim et al., 2019). Although graph-based mapping and variant calling are generally slower and require more memory compared to linear reference genome analysis (Eizenga et al., 2020), a newly developed tool, Giraffe, can accurately map NGS reads to the graph-genome and call variants, including SNPs, insertions and deletions, and SVs, at a speed comparable to that of a single linear genome (Sirén et al., 2020). An alternative solution involves converting single genomes to graph pan-genomes, which is more efficient and may require less computational resources. By doing this, SNPs identified from single genomes can be relocated to the pan-genome coordinate system. However, this approach may not be suitable for SVs that are difficult to detect using short NGS reads. In a recent human study, long-read sequencing was used to resequence more

than 3000 individuals (Beyter et al., 2021). While the high cost of long-read sequencing currently limits its population-scale application in plant studies, decreasing costs will increasingly make long-read sequencing the standard for population studies. Variants called from long reads are more capable of resolving SVs and insertions and deletions in highly repetitive genomes, and in some cases facilitate the direct detection of methylated nucleotides, such as *N*6-methyladenine (m6A), 5-methylcytosine (5mC), and 5-hydroxymethylcytosine (5hmC) (Flusberg et al., 2010; Tse et al., 2021; Ni et al., 2023). Epigenetic variation in genomes can result in substantial phenotypic changes (Ong-Abdullah et al., 2015; Song et al., 2017; Han et al., 2018; Ji et al., 2018; Wibowo et al., 2018), and the application of long-read techniques can potentially identify epigenetically modified alleles conferring traits that cannot be solely explained by DNA sequence variants. These emerging techniques hold promise for plant population studies, although their application at large scale is currently limited by affordability for many research groups.

Greater data accessibility is needed

Sharing data with the research community has proved to be beneficial, with many outstanding studies conducted through the compilation and secondary analysis of public data providing insights into genome evolution and transcription regulation (Ramírez-González et al., 2018; Zhang et al., 2023). Population genomic datasets have significant potential for reuse by other researchers for diverse purposes, such as meta-GWAS (Zhao et al., 2019c) and the identification of unannotated open reading frames (Jiang et al., 2022). However, the secondary analyses of population genomic datasets are relatively uncommon, partly due to the difficulty in accessing these datasets, which can be more challenging than accessing genome or transcriptome datasets. Despite most journals requiring or encouraging the sharing of genomic data, access to genotyping data in many studies remains restricted. A survey of data-sharing policies in journals revealed that approximately 36% of the journals require sharing of genomic data (Vasilevsky et al., 2017), with 12% mandating data sharing as a condition of publication. For example, the Nature Portfolio of journals requires data availability as a condition of publication, and the submission of genetic polymorphism data to public repositories is mandatory (https://www.nature.com/ng/editorial-policies/reporting-standards). However, in some publications within these journals, either no data access is provided or the access links are invalid. In our analysis of currently published plant genome resequencing studies, only 21% (74 of 356) of papers published between 2008 and 2023 provided access to the variant calling format (VCF) file, and, even among those, only 82% of the links were still valid (Supplemental Table 1).

To understand the poor accessibility of genotyping datasets, we compared studies with available genotyping datasets to those without in terms of the population and genome sizes they studied. Surprisingly, datasets derived from larger genomes and larger populations were more accessible than small datasets (Figure 4A), despite the potential difficulty in uploading large datasets. This counterintuitive conclusion suggests that the challenge of transferring large data is not a major factor preventing the sharing of genotyping datasets. Two reasons

standardized format (VCF format) as suggested by many journals or required by the database. To better facilitate further analyses of these data, we propose to include the following information in the submitted datasets: (1) the description of species, such as the name and taxonomy; (2) the description of genomes, including the ploidy and the size of the genome, as well as the version of the reference genome; (3) the tool and parameters used for SNP calling and filtering; (4) detailed information about the samples, such as collection locations and the breeding status (wild species, landraces, or elite cultivars); and (5) the raw SNP datasets in case future users need to refilter the SNPs with their customized parameters.

In summary, although the quantity and quality of population resequencing data are rapidly increasing, the analysis of existing datasets remains inadequate. More comprehensive analyses and deeper mining of these datasets are necessary for future studies. The accumulation of large-scale plant population genomic datasets has provided an unprecedented opportunity to further our understanding of plant genome evolution and adaptation. A future direction for plant genomic studies could involve performing meta-analysis using these datasets to gain further insights into plant genome evolution and response to environmental change. The broad distribution of the resequenced plant genomes across the Tree of Life (Figure 1B) facilitates the analysis of intra- and inter-species evolution of plant genomes, which could be further accelerated by the ongoing accumulation of relevant datasets. Secondary analyses of publicly available genomic datasets hold great promise for generating novel insights and hypotheses related to plant genome evolution. To facilitate such analyses, barriers to their access should be removed by encouraging more open sharing. This is not only essential for reproducing published analysis but also beneficial for reducing redundant work and fostering secondary analyses through data compilation.

could potentially explain this observation. First, most large-scale population studies are international collaborations involving many institutes from different countries, such as the 1001 Genomes Project (Alonso-Blanco et al., 2016) and the 3000 Rice Genomes Project (Wang et al., 2018a), where data generating and sharing is one of the major goals of these projects. Second, large-scale population studies have a greater chance of being published in top-tier journals that require data sharing as a condition of publication. Indeed, our data show a strong correlation between data accessibility and publishers. Datasets published in AAAS and Springer Nature journals are more accessible than datasets from other journals (Figure 4B). We noticed that many of these datasets, particularly old datasets, became inaccessible only because of the discontinued maintenance of the sites where the datasets were deposited. Therefore, it is necessary to submit the datasets to a permanent database, such as European Variation Archive (https://www.ebi.ac.uk/eva/?Home), China National GeneBank Database (https://db.cngb.org/), and the Genome Variation Map of the China National Center for Bioinformation (https://ngdc.cncb.ac.cn/gvm/home) with a

SUPPLEMENTAL INFORMATION
Supplemental information is available at *Molecular Plant Online*.

FUNDING

AUTHOR CONTRIBUTIONS

C.S. and B.S. conceived this work;. W.N., D.W., M.J., X.W., and K.Z. performed literature search and data collection;. W.N., B.S., and M.J. participated in the visualization of data;. B.S., W.N., C.S., D.E., and D.O. wrote and revised the manuscript.

ACKNOWLEDGMENTS

REFERENCES

Abberton, M., Batley, J., Bentley, A., Bryant, J., Cai, H., Cockram, J., Costa de Oliveira, A., Cseke, L.J., Dempewolf, H., De Pace, C., et al. (2016). Global agricultural intensification during climate change: A role for genomics. Plant Biotechnol. J. **14**:1095–1098.

Abbo, S., and Gopher, A. (2017). Near Eastern Plant Domestication: A History of Thought. Trends Plant Sci. **22**:491–511.

Abrouk, M., Ahmed, H.I., Cubry, P., Šimoníková, D., Cauet, S., Pailles, Y., Bettgenhaeuser, J., Gapa, L., Scarcelli, N., Couderc, M., et al. (2020). Fonio millet genome unlocks African orphan crop diversity for agriculture in a changing climate. Nat. Commun. **11**:4488.

Alonso-Blanco, C., Andrade, J., Becker, C., et al. (2016). 1,135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. Cell **166**:481–491.

Alseekh, S., Kostova, D., Bulut, M., and Fernie, A.R. (2021). Genome-wide association studies: assessing trait characteristics in model and crop plants. Cell. Mol. Life Sci. **78**:5743–5754.

Anderson, R., Bayer, P.E., and Edwards, D. (2020). Climate change and the need for agricultural adaptation. Curr. Opin. Plant Biol. **56**:197–202.

Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., Pozniak, C.J., Choulet, F., Distelfeld, A., Poland, J., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science **361**:eaar7191.

Arora, S., Steuernagel, B., Gaurav, K., Chandramohan, S., Long, Y., Matny, O., Johnson, R., Enk, J., Periyannan, S., Singh, N., et al. (2019). Resistance gene cloning from a wild crop relative by sequence capture and association genetics. Nat. Biotechnol. **37**:139–143.

Bayer, P.E., Golicz, A.A., Scheben, A., Batley, J., and Edwards, D. (2020). Plant pan-genomes are the new reference. Nat. Plants **6**:914–920.

Bayer, P.E., Petereit, J., Danilevicz, M.F., Anderson, R., Batley, J., and Edwards, D. (2021). The application of pangenomics and machine learning in genomic selection in plants. Plant Genome **14**:e20112.

Bayer, P.E., Petereit, J., Durant, É., Monat, C., Rouard, M., Hu, H., Chapman, B., Li, C., Cheng, S., Batley, J., and Edwards, D. (2022a). Wheat Panache: A pangenome graph database representing presence–absence variation across sixteen bread wheat genomes. Plant Genome **15**:e20221.

Bayer, P.E., Valliyodan, B., Hu, H., Marsh, J.I., Yuan, Y., Vuong, T.D., Patil, G., Song, Q., Batley, J., Varshney, R.K., et al. (2022b). Sequencing the USDA core soybean collection reveals gene loss during domestication and breeding. Plant Genome **15**:e20109.

Beyter, D., Ingimundardottir, H., Oddsson, A., Eggertsson, H.P., Bjornsson, E., Jonsson, H., Atlason, B.A., Kristmundsdottir, S., Mehringer, S., Hardarson, M.T., et al. (2021). Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. Nat. Genet. **53**:779–786.

Bilbrey, E.A., Williamson, K., Hatzakis, E., Miller, D.D., Fresnedo-Ramírez, J., and Cooperstone, J.L. (2021). Integrating genomics and multiplatform metabolomics enables metabolite quantitative trait loci detection in breeding-relevant apple germplasm. New Phytol. **232**:1944–1958.

Bonhomme, M., André, O., Badis, Y., Ronfort, J., Burgarella, C., Chantret, N., Prosperi, J.M., Briskine, R., Mudge, J., Debéllé, F., et al. (2014). High-density genome-wide association mapping implicates an F-box encoding gene in Medicago truncatula resistance to Aphanomyces euteiches. New Phytol. **201**:1328–1342.

Cai, X., Sun, X., Xu, C., Sun, H., Wang, X., Ge, C., Zhang, Z., Wang, Q., Fei, Z., Jiao, C., and Wang, Q. (2021). Genomic analyses provide insights into spinach domestication and the genetic basis of agronomic traits. Nat. Commun. **12**:7246.

Campbell, M.T., Bandillo, N., Al Shiblawi, F.R.A., et al. (2017). Allelic variants of OsHKT1; 1 underlie the divergence between indica and japonica subspecies of rice (Oryza sativa) for root sodium content. PLoS Genet **13**, e1006823.

Causse, M., Desplat, N., Pascual, L., Le Paslier, M.C., Sauvage, C., Bauchet, G., Bérard, A., Bounon, R., Tchoumakov, M., Brunel, D., and Bouchet, J.P. (2013). Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. BMC Genom. **14**:791.

Chan, E.K.F., Rowe, H.C., and Kliebenstein, D.J. (2010). Understanding the evolution of defense metabolites in Arabidopsis thaliana using genome-wide association mapping. Genetics **185**:991–1007.

Chen, X., Lewandowska, D., Armstrong, M.R., Baker, K., Lim, T.Y., Bayer, M., Harrower, B., McLean, K., Jupe, F., Witek, K., et al. (2018). Identification and rapid mapping of a gene conferring broad-spectrum late blight resistance in the diploid potato species Solanum verrucosum through DNA capture technologies. Theor. Appl. Genet. **131**:1287–1297.

Chen, Y., Nie, F., Xie, S.Q., Zheng, Y.F., Dai, Q., Bray, T., Wang, Y.X., Xing, J.F., Huang, Z.J., Wang, D.P., et al. (2021). Efficient assembly of nanopore reads via highly accurate and intact error correction. Nat. Commun. **12**:60.

Cheng, S. (2022). Gregor Mendel: The father of genetics who opened a biological world full of wonders. Mol. Plant **15**:1641–1645.

Cheng, H., Liu, J., Wen, J., Nie, X., Xu, L., Chen, N., Li, Z., Wang, Q., Zheng, Z., Li, M., et al. (2019). Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. Genome Biol. **20**:136.

Chhetri, H.B., Macaya-Sanz, D., Kainer, D., Biswal, A.K., Evans, L.M., Chen, J.G., Collins, C., Hunt, K., Mohanty, S.S., Rosenstiel, T., et al. (2019). Multitrait genome-wide association analysis of Populus trichocarpa identifies key polymorphisms controlling morphological and physiological traits. New Phytol. **223**:293–309.

Crang, R., Lyons-Sobaski, S., and Wise, R. (2018). Plant anatomy: a concept-based approach to the structure of seed plants (Springer).

Cubry, P., Tranchant-Dubreuil, C., Thuillet, A.C., Monat, C., Ndjiondjop, M.N., Labadie, K., Cruaud, C., Engelen, S., Scarcelli, N., Rhoné, B., et al. (2018). The Rise and Fall of African Rice

Cultivation Revealed by Analysis of 246 New Genomes. Curr. Biol. **28**:2274–2282.e6.

Danilevicz, M.F., Tay Fernandez, C.G., Marsh, J.I., Bayer, P.E., and Edwards, D. (2020). Plant pangenomics: approaches, applications and advancements. Curr. Opin. Plant Biol. **54**:18–25.

Danilevicz, M.F., Bayer, P.E., Nestor, B.J., Bennamoun, M., and Edwards, D. (2021a). Resources for image-based high-throughput phenotyping in crops and data sharing challenges. Plant Physiol. **187**:699–715.

Danilevicz, M.F., Bayer, P.E., Boussaid, F., Bennamoun, M., and Edwards, D. (2021b). Maize yield prediction at an early developmental stage using multispectral images and genotype data for preliminary hybrid selection. Remote Sens. **13**:3976.

De La Torre, A.R., Puiu, D., Crepeau, M.W., Stevens, K., Salzberg, S.L., Langley, C.H., and Neale, D.B. (2019). Genomic architecture of complex traits in loblolly pine. New Phytol. **221**:1789–1801.

Derbyshire, M.C., Marsh, J., Tirnaz, S., et al. (2023). Diversity of fatty acid biosynthesis genes across the soybean pangenome. Plant Genome, e20334.

Dong, H., Zhao, H., Xie, W., Han, Z., Li, G., Yao, W., Bai, X., Hu, Y., Guo, Z., Lu, K., et al. (2016). A Novel Tiller Angle Gene, TAC3, together with TAC1 and D2 Largely Determine the Natural Variation of Tiller Angle in Rice Cultivars. PLoS Genet. **12**:e1006412.

Dong, H., Zhao, H., Li, S., Han, Z., Hu, G., Liu, C., Yang, G., Wang, G., Xie, W., and Xing, Y. (2018). Genome-wide association studies reveal that members of bHLH subfamily 16 share a conserved function in regulating flag leaf angle in rice (Oryza sativa). PLoS Genet. **14**:e1007323.

Dong, Y., Duan, S., Xia, Q., et al. (2023). Dual domestications and origin of traits in grapevine evolution. Sci **379**:892–901.

Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., Li, N., Zhang, X., Sun, J., Liu, M., et al. (2018). Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nat. Genet. **50**:796–802.

Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., Wang, X., Jiao, C., Legall, N., Mao, L., et al. (2017). Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. Nat. Commun. **8**:249.

Edwards, D., and Batley, J. (2022). Graph pangenomes find missing heritability. Nat. Genet. **54**:919–920.

Eggertsson, H.P., Jonsson, H., Kristmundsdottir, S., Hjartarson, E., Kehr, B., Masson, G., Zink, F., Hjorleifsson, K.E., Jonasdottir, A., Jonasdottir, A., et al. (2017). Graphtyper enables population-scale genotyping using pangenome graphs. Nat. Genet. **49**:1654–1660.

Eizenga, J.M., Novak, A.M., Sibbesen, J.A., Heumos, S., Ghaffaari, A., Hickey, G., Chang, X., Seaman, J.D., Rounthwaite, R., Ebler, J., et al. (2020). Pangenome Graphs. Annu. Rev. Genomics Hum. Genet. **21**:139–162.

Exposito-Alonso, M., 500 Genomes Field Experiment Team, Exposito-Alonso, M., Gómez Rodríguez, R., Barragán, C., Capovilla, G., Chae, E., Devos, J., Dogan, E.S., Friedemann, C., et al. (2019). Natural selection on the Arabidopsis thaliana genome in present and future climates. Nature **573**:126–129.

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., Hu, G., Zhou, Z., Yu, H., Zhang, M., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. Genome Biol. **18**:161.

Feng, C., Wang, X., Wu, S., Ning, W., Song, B., Yan, J., and Cheng, S. (2022). HAPPE: A Tool for Population Haplotype Analysis and Visualization in Editable Excel Tables. Front. Plant Sci. **13**:1–7.

Tay Fernandez, C.G., Nestor, B.J., Danilevicz, M.F., Gill, M., Petereit, J., Bayer, P.E., Finnegan, P.M., Batley, J., and Edwards, D. (2022). Pangenomes as a Resource to Accelerate Breeding of Under-Utilised Crop Species. Int. J. Mol. Sci. **23**:2671.

Ferrão, L.F.V., Johnson, T.S., Benevenuto, J., Edger, P.P., Colquhoun, T.A., and Munoz, P.R. (2020). Genome-wide association of volatiles reveals candidate loci for blueberry flavor. New Phytol. **226**:1725–1737.

Ferrero-Serrano, Á., and Assmann, S.M. (2019). Phenotypic and genome-wide association with the local environment of Arabidopsis. Nat. Ecol. Evol. **3**:274–285.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat. Methods **7**:461–465.

Garcia, T., Duitama, J., Zullo, S.S., Gil, J., Ariani, A., Dohle, S., Palkovic, A., Skeen, P., Bermudez-Santana, C.I., Debouck, D.G., et al. (2021). Comprehensive genomic resources related to domestication and crop improvement traits in Lima bean. Nat. Commun. **12**:702.

Gill, M., Anderson, R., Hu, H., Bennamoun, M., Petereit, J., Valliyodan, B., Nguyen, H.T., Batley, J., Bayer, P.E., and Edwards, D. (2022). Machine learning models outperform deep learning models, provide interpretation and facilitate feature selection for soybean trait prediction. BMC Plant Biol. **22**:180.

Golicz, A.A., Bayer, P.E., Bhalla, P.L., Batley, J., and Edwards, D. (2020). Pangenomics Comes of Age: From Bacteria to Plant and Animal Applications. Trends Genet. **36**:132–145.

Gross, B.L., and Olsen, K.M. (2010). Genetic perspectives on crop domestication. Trends Plant Sci. **15**:529–537.

Guo, Z., Yang, W., Chang, Y., et al. (2018). Genome-wide association studies of image traits reveal genetic architecture of drought resistance in rice. Mol. Plant **11**:789–805.

Guo, S., Zhao, S., Sun, H., Wang, X., Wu, S., Lin, T., Ren, Y., Gao, L., Deng, Y., Zhang, J., et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. Nat. Genet. **51**:1616–1623.

Guo, W., Xin, M., Wang, Z., Yao, Y., Hu, Z., Song, W., Yu, K., Chen, Y., Wang, X., Guan, P., et al. (2020a). Origin and adaptation to high altitude of Tibetan semi-wild wheat. Nat. Commun. **11**:5085.

Guo, D., Jiang, H., Yan, W., Yang, L., Ye, J., Wang, Y., Yan, Q., Chen, J., Gao, Y., Duan, L., et al. (2019). Resequencing 200 Flax Cultivated Accessions Identifies Candidate Genes Related to Seed Size and Weight and Reveals Signatures of Artificial Selection. Front. Plant Sci. **10**:1682.

Han, Z., Crisp, P.A., Stelpflug, S., Kaeppler, S.M., Li, Q., and Springer, N.M. (2018). Heritable epigenomic changes to the maize methylome resulting from tissue culture. Genetics **209**:983–995.

Hao, C., Jiao, C., Hou, J., Li, T., Liu, H., Wang, Y., Zheng, J., Liu, H., Bi, Z., Xu, F., et al. (2020). Resequencing of 145 Landmark Cultivars Reveals Asymmetric Sub-genome Selection and Strong Founder Genotype Effects on Wheat Breeding in China. Mol. Plant **13**:1733–1751.

Harlan, J.R. (1971). Agricultural origins: centers and noncenters. Agriculture may originate in discrete centers or evolve over vast areas without definable centers. Science **174**:468–474.

Hartley, T.N., Thomas, A.S., and Maathuis, F.J. (2020). A role for the OsHKT 2; 1 sodium transporter in potassium use efficiency in rice. J. Exp. Bot. **71**:699–706.

Henchion, M., Hayes, M., Mullen, A.M., Fenelon, M., and Tiwari, B. (2017). Future protein supply and demand: Strategies and factors influencing a sustainable equilibrium. Foods **6**. 53–21.

Hendre, P.S., Muthemba, S., Kariba, R., Muchugi, A., Fu, Y., Chang, Y., Song, B., Liu, H., Liu, M., Liao, X., et al. (2019). African Orphan Crops

Consortium (AOCC): status of developing genomic resources for African orphan crops. Planta **250**:989–1003.

Hindu, V., Palacios-Rojas, N., Babu, R., Suwarno, W.B., Rashid, Z., Usha, R., Saykhedkar, G.R., and Nair, S.K. (2018). Identification and validation of genomic regions influencing kernel zinc and iron in maize. Theor. Appl. Genet. **131**:1443–1457.

Ho, S.S., Urban, A.E., and Mills, R.E. (2020). Structural variation in the sequencing era. Nat. Rev. Genet. **21**:171–189.

Hu, W., Ji, C., Liang, Z., Ye, J., Ou, W., Ding, Z., Zhou, G., Tie, W., Yan, Y., Yang, J., et al. (2021). Resequencing of 388 cassava accessions identifies valuable loci and selection for variation in heterozygosity. Genome Biol. **22**:316.

Hu, J., Chen, B., Zhao, J., Zhang, F., Xie, T., Xu, K., Gao, G., Yan, G., Li, H., Li, L., et al. (2022). Genomic selection and genetic architecture of agronomic traits during modern rapeseed breeding. Nat. Genet. **54**:694–704.

Huang, X., and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. Annu. Rev. Plant Biol. **65**:531–551.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. **42**:961–967.

Huang, X., Kurata, N., Wei, X., Wang, Z.X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., et al. (2012a). A map of rice genome variation reveals the origin of cultivated rice. Nature **490**:497–501.

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C., et al. (2012b). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat. Genet. **44**:32–39.

Hufford, M.B., Xu, X., Van Heerwaarden, J., Pyhäjärvi, T., Chia, J.M., Cartwright, R.A., Elshire, R.J., Glaubitz, J.C., Guill, K.E., Kaeppler, S.M., et al. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. **44**:808–811.

Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V.S., Gundlach, H., Monat, C., Lux, T., Kamal, N., Lang, D., Himmelbach, A., et al. (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. Nature **588**:284–289.

Ji, L., Jordan, W.T., Shi, X., Hu, L., He, C., and Schmitz, R.J. (2018). TET-mediated epimutagenesis of the Arabidopsis thaliana methylome. Nat. Commun. **9**:895.

Ji, F., Ma, Q., Zhang, W., Liu, J., Feng, Y., Zhao, P., Song, X., Chen, J., Zhang, J., Wei, X., et al. (2021). A genome variation map provides insights into the genetics of walnut adaptation and agronomic traits. Genome Biol. **22**:300.

Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., Chai, Y., Yang, L., Liu, K., Lu, H., et al. (2013). A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (Setaria italica). Nat. Genet. **45**:957–961.

Jiang, M., Ning, W., Wu, S., Wang, X., Zhu, K., Li, A., Li, Y., Cheng, S., and Song, B. (2022). Three-nucleotide periodicity of nucleotide diversity in a population enables the identification of open reading frames. Brief. Bioinform. **23**:1–13.

Jupe, F., Witek, K., Verweij, W., Śliwka, J., Pritchard, L., Etherington, G.J., Maclean, D., Cock, P.J., Leggett, R.M., Bryan, G.J., et al. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. Plant J. **76**:530–544.

Kang, L., Qian, L., Zheng, M., Chen, L., Chen, H., Yang, L., You, L., Yang, B., Yan, M., Gu, Y., et al. (2021). Genomic insights into the origin, domestication and diversification of Brassica juncea. Nat. Genet. **53**:1392–1402.

Khan, A.W., Garg, V., Roorkiwal, M., Golicz, A.A., Edwards, D., and Varshney, R.K. (2020). Super-Pangenome by Integrating the Wild Side of a Species for Accelerated Crop Improvement. Trends Plant Sci. **25**:148–158.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37**:907–915.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. Genome Res. **27**:722–736.

Koren, S., Rhie, A., Walenz, B.P., Dilthey, A.T., Bickhart, D.M., Kingan, S.B., Hiendleder, S., Williams, J.L., Smith, T.P.L., and Phillippy, A.M. (2018). De novo assembly of haplotype-resolved genomes with trio binning. Nat. Biotechnol. **36**:1174–1182.

Kress, W.J., Soltis, D.E., Kersey, P.J., Wegrzyn, J.L., Leebens-Mack, J.H., Gostel, M.R., Liu, X., and Soltis, P.S. (2022). Green plant genomes: What we know in an era of rapidly expanding opportunities. Proc. Natl. Acad. Sci. USA **119**:1–9.

Li, W., Zhu, Z., Chern, M., Yin, J., Yang, C., Ran, L., Cheng, M., He, M., Wang, K., Wang, J., et al. (2017). A Natural Allele of a Transcription Factor in Rice Confers Broad-Spectrum Blast Resistance. Cell **170**:114–126.e15.

Li, Y., Cao, K., Zhu, G., Fang, W., Chen, C., Wang, X., Zhao, P., Guo, J., Ding, T., Guan, L., et al. (2019a). Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. Genome Biol. **20**:36.

Li, N., Lin, B., Wang, H., Li, X., Yang, F., Ding, X., Yan, J., and Chu, Z. (2019b). Natural variation in Zm FBL41 confers banded leaf and sheath blight resistance in maize. Nat. Genet. **51**:1540–1548.

Li, Y., Li, D., Jiao, Y., Schnable, J.C., Li, Y., Li, H., Chen, H., Hong, H., Zhang, T., Liu, B., et al. (2020). Identification of loci controlling adaptation in Chinese soya bean landraces via a combination of conventional and bioclimatic GWAS. Plant Biotechnol. J. **18**:389–401.

Liang, Z., Duan, S., Sheng, J., Zhu, S., Ni, X., Shao, J., Liu, C., Nick, P., Du, F., Fan, P., et al. (2019). Whole-genome resequencing of 472 Vitis accessions for grapevine diversity and demographic history analyses. Nat. Commun. **10**:1190.

Liao, L., Zhang, W., Zhang, B., Fang, T., Wang, X.F., Cai, Y., Ogutu, C., Gao, L., Chen, G., Nie, X., et al. (2021). Unraveling a genetic roadmap for improved taste in the domesticated apple. Mol. Plant **14**:1454–1471.

Liu, J., Huang, L., Wang, C., Liu, Y., Yan, Z., Wang, Z., Xiang, L., Zhong, X., Gong, F., Zheng, Y., et al. (2019). Genome-wide association study reveals novel genomic regions associated with high grain protein content in wheat lines derived from wild emmer wheat. Front. Plant Sci. **10**:464.

Liu, Y., Wang, H., Jiang, Z., et al. (2021). Genomic basis of geographical adaptation to soil nitrogen in rice. Nature **590**:600–605.

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.A., Zhang, H., Liu, Z., Shi, M., et al. (2020a). Pan-Genome of Wild and Cultivated Soybeans. Cell **182**:162–176.e13.

Liu, H., Shi, J., Cai, Z., Huang, Y., Lv, M., Du, H., Gao, Q., Zuo, Y., Dong, Z., Huang, W., et al. (2020b). Evolution and Domestication Footprints Uncovered from the Genomes of Coix. Mol. Plant **13**:295–308.

Lovell, J.T., MacQueen, A.H., Mamidi, S., Bonnette, J., Jenkins, J., Napier, J.D., Sreedasyam, A., Healey, A., Session, A., Shu, S., et al. (2021). Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. Nature **590**:438–444.

Ma, X., Feng, F., Zhang, Y., et al. (2019). A novel rice grain size gene OsSNB was identified by genome-wide association study in natural population. PLoS Genet **15**, e1008191.

Ma, Y., Min, L., Wang, J., Li, Y., Wu, Y., Hu, Q., Ding, Y., Wang, M., Liang, Y., Gong, Z., et al. (2021a). A Combination of Genome-wide and Transcriptome-wide Association Studies Reveals Genetic Elements Leading to Male Sterility during High Temperature Stress in Cotton. https://doi.org/10.1111/nph.17325.

Ma, Z., Zhang, Y., Wu, L., Zhang, G., Sun, Z., Li, Z., Jiang, Y., Ke, H., Chen, B., Liu, Z., et al. (2021b). High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement. Nat. Genet. 53:1385–1391.

Mace, E.S., Tai, S., Gilding, E.K., Li, Y., Prentis, P.J., Bian, L., Campbell, B.C., Hu, W., Innes, D.J., Han, X., et al. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat. Commun. 4:2320.

Mamidi, S., Healey, A., Huang, P., Grimwood, J., Jenkins, J., Barry, K., Sreedasyam, A., Shu, S., Lovell, J.T., Feldman, M., et al. (2020). A genome resource for green millet Setaria viridis enables discovery of agronomically valuable loci. Nat. Biotechnol. 38:1203–1210.

Maphosa, Y., and Jideani, V.A. (2017). The Role of Legumes in Human Nutrition. Funct. Food - Improv. Heal. through Adequate Food 1:13. https://doi.org/10.5772/intechopen.69127.

Marks, R.A., Hotaling, S., Frandsen, P.B., and VanBuren, R. (2021). Representation and participation across 20 years of plant genome sequencing. Nat. Plants 7:1571–1578.

Marsh, J.I., Hu, H., Gill, M., Batley, J., and Edwards, D. (2021). Crop breeding for a changing climate: integrating phenomics and genomics with bioinformatics. Theor. Appl. Genet. 134:1677–1690.

Marsh, J.I., Hu, H., Petereit, J., Bayer, P.E., Valliyodan, B., Batley, J., Nguyen, H.T., and Edwards, D. (2022). Haplotype mapping uncovers unexplored variation in wild and domesticated soybean at the major protein locus cqProt-003. Theor. Appl. Genet. 135:1443–1455.

Marsh, J.I., Nestor, B.J., Petereit, J., Tay Fernandez, C.G., Bayer, P.E., Batley, J., and Edwards, D. (2023). Legume-wide comparative analysis of pod shatter locus PDH1 reveals phaseoloid specificity, high cowpea expression, and stress responsive genomic context. Plant J. 115:68–80. https://doi.org/10.1111/tpj.16209.

Meyer, R.S., and Purugganan, M.D. (2013). Evolution of crop species: Genetics of domestication and diversification. Nat. Rev. Genet. 14:840–852.

Meyer, R.S., Choi, J.Y., Sanches, M., Plessis, A., Flowers, J.M., Amas, J., Dorph, K., Barretto, A., Gross, B., Fuller, D.Q., et al. (2016). Domestication history and geographical adaptation inferred from a SNP map of African rice. Nat. Genet. 48:1083–1088.

Michael, T.P., and VanBuren, R. (2020). Building near-complete plant genomes. Curr. Opin. Plant Biol. 54:26–33.

Millet, E.J., Kruijer, W., Coupel-Ledru, A., Alvarez Prado, S., Cabrera-Bosquet, L., Lacube, S., Charcosset, A., Welcker, C., van Eeuwijk, F., and Tardieu, F. (2019). Genomic prediction of maize yield across European environmental conditions. Nat. Genet. 51:952–956.

Ni, P., Nie, F., Zhong, Z., Xu, J., Huang, N., Zhang, J., Zhao, H., Zou, Y., Huang, Y., and Li, J. (2023). DNA 5-methylcytosine detection and methylation phasing using PacBio circular consensus sequencing. Nat. Commun. 14:4054.

Ong-Abdullah, M., Ordway, J.M., Jiang, N., Ooi, S.E., Kok, S.Y., Sarpan, N., Azimi, N., Hashim, A.T., Ishak, Z., Rosli, S.K., et al. (2015). Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. Nature 525:533–537.

Yang, M., Lu, K., Zhao, F.J., Xie, W., Ramakrishna, P., Wang, G., Du, Q., Liang, L., Sun, C., Zhao, H., et al. (2018). Genome-Wide Association Studies Reveal the Genetic Basis of Ionomic Variation in Rice. Plant Cell 30:2720–2740.

Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res. 18:2024–2033.

Parra, L., Nortman, K., Sah, A., Truco, M.J., Ochoa, O., and Michelmore, R. (2021). Identification and mapping of new genes for resistance to downy mildew in lettuce. Theor. Appl. Genet. 134:519–528.

Petereit, J., Bayer, P.E., Thomas, W.J.W., Tay Fernandez, C.G., Amas, J., Zhang, Y., Batley, J., and Edwards, D. (2022). Pangenomics and Crop Genome Adaptation in a Changing Climate. Plants 11:1949.

Poczai, P., Cernák, I., Varga, I., and Hyvönen, J. (2014). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Genet. Resour. Crop Evol. 61:796–815.

Purugganan, M.D., and Fuller, D.Q. (2009). The nature of selection during plant domestication. Nature 457:843–848.

Qin, P., Wang, L., Liu, K., Mao, S., Li, Z., Gao, S., Shi, H., and Liu, Y. (2015). Genomewide association study of Aegilops tauschii traits under seedling-stage cadmium stress. Crop J. 3:405–415.

Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., He, Q., Ou, S., Zhang, H., Li, X., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. Cell 184:3542–3558.e16.

Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., Van Ex, F., Pasha, A., et al. (2018). The transcriptional landscape of polyploid wheat. Science 361:eaar6089.

Rautiainen, M., and Marschall, T. (2020). GraphAligner: rapid and versatile sequence-to-graph alignment. Genome Biol. 21:253.

Ren, G., Zhang, X., Li, Y., Ridout, K., Serrano-Serrano, M.L., Yang, Y., Liu, A., Ravikanth, G., Nawaz, M.A., Mumtaz, A.S., et al. (2021). Large-scale whole-genome resequencing unravels the domestication history of Cannabis sativa. Sci. Adv. 7:eabg2286.

Rijzaani, H., Bayer, P.E., Rouard, M., Doležel, J., Batley, J., and Edwards, D. (2022). The pangenome of banana highlights differences between genera and genomes. Plant Genome 15:e20100.

Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. Nat. Methods 17:155–158.

Ruperao, P., Thirunavukkarasu, N., Gandham, P., Selvanayagam, S., Govindaraj, M., Nebie, B., Manyasa, E., Gupta, R., Das, R.R., Odeny, D.A., et al. (2021). Sorghum Pan-Genome Explores the Functional Utility for Genomic-Assisted Breeding to Accelerate the Genetic Gain. Front. Plant Sci. 12:666342.

Sadhukhan, A., Kobayashi, Y., Nakano, Y., Iuchi, S., Kobayashi, M., Sahoo, L., and Koyama, H. (2017). Genome-wide Association Study Reveals that the Aquaporin NIP1 ; 1 Contributes to Variation in Hydrogen Peroxide Sensitivity in Arabidopsis thaliana. Mol. Plant 10:1082–1094.

Shang, Y., Ma, Y., Zhou, Y., et al. (2014). Biosynthesis, regulation, and domestication of bitterness in cucumber. Science 346:1084–1088.

Shi, A., Bhattarai, G., Xiong, H., et al. (2022). Genome-wide association study and genomic prediction of white rust resistance in USDA GRIN spinach germplasm. Hortic. Res 9:uhac069.

Si, L., Chen, J., Huang, X., et al. (2016). OsSPL13 controls grain size in cultivated rice. Nat. Genet. 48:447–456.

Sirén, J., Monlong, J., Chang, X., et al. (2021). Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. Science 374:1461.

Slatko, B.E., Gardner, A.F., and Ausubel, F.M. (2018). Overview of Next-Generation Sequencing Technologies. Curr. Protoc. Mol. Biol. 122:1–11.

Song, Q., Zhang, T., Stelly, D.M., and Chen, Z.J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biol. **18**:99.

Song, J.M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. Nat. Plants **6**:34–45.

Staniak, M., Ksiak, J., and Bojarszczuk, J. (2014). Mixtures of Legumes with Cereals as a Source of Feed for Animals. Org. Agric. Towar. Sustain. 123–145. https://doi.org/10.5772/58358.

Steuernagel, B., Periyannan, S.K., Hernández-Pinzón, I., Witek, K., Rouse, M.N., Yu, G., Hatta, A., Ayliffe, M., Bariana, H., Jones, J.D.G., et al. (2016). Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. Nat. Biotechnol. **34**:652–655.

Strachan, S.M., Armstrong, M.R., Kaur, A., Wright, K.M., Lim, T.Y., Baker, K., Jones, J., Bryan, G., Blok, V., and Hein, I. (2019). Mapping the H2 resistance effective against Globodera pallida pathotype Pa1 in tetraploid potato. Theor. Appl. Genet. **132**:1283–1294.

Sun, D., Cen, H., Weng, H., Wan, L., Abdalla, A., El-Manawy, A.I., Zhu, Y., Zhao, N., Fu, H., Tang, J., et al. (2019). Using hyperspectral analysis as a potential high throughput phenotyping tool in GWAS for protein content of rice quality. Plant Methods **15**:54.

Sun, Y., Shang, L., Zhu, Q.-H., Fan, L., and Guo, L. (2022). Twenty years of plant genome sequencing: achievements and challenges. Trends Plant Sci. **27**:391–401. https://doi.org/10.1016/j.tplants.2021.10.006.

Tang, D., Jia, Y., Zhang, J., Li, H., Cheng, L., Wang, P., Bao, Z., Liu, Z., Feng, S., Zhu, X., et al. (2022). Genome evolution and diversity of wild and cultivated potatoes. Nature **606**:535–541. https://doi.org/10.1038/s41586-022-04822-x.

Tieman, D., Zhu, G., Resende, M.F., Jr., et al. (2017). A chemical genetic roadmap to improved tomato flavor, 355 (Science), pp. 391–394.

Tirnaz, S., Zandberg, J., Thomas, W.J.W., Marsh, J., Edwards, D., and Batley, J. (2022). Application of crop wild relatives in modern breeding: An overview of resources, experimental and computational methodologies. Front. Plant Sci. **13**:1008904.

Todesco, M., Owens, G.L., Bercovich, N., Légaré, J.S., Soudi, S., Burge, D.O., Huang, K., Ostevik, K.L., Drummond, E.B.M., Imerovski, I., et al. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature **584**:602–607.

Tse, O.Y.O., Jiang, P., Cheng, S.H., Peng, W., Shang, H., Wong, J., Chan, S.L., Poon, L.C.Y., Leung, T.Y., Chan, K.C.A., et al. (2021). Genome-wide detection of cytosine methylation by single molecule real-time sequencing. Proc. Natl. Acad. Sci. USA **118**. e2019768118.

Turner, T.L., Bourne, E.C., Von Wettberg, E.J., Hu, T.T., and Nuzhdin, S.V. (2010). Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. Nat. Genet. **42**:260–263.

Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., and Posthuma, D. (2021). Genome-wide association studies. Nat. Rev. Methods Primers **1**:59.

Van de Weyer, A.L., Monteiro, F., Furzer, O.J., Nishimura, M.T., Cevik, V., Witek, K., Jones, J.D.G., Dangl, J.L., Weigel, D., and Bemm, F. (2019). A Species-Wide Inventory of NLR Genes and Alleles in Arabidopsis thaliana. Cell **178**:1260–1272.e14.

Varshney, R.K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., Zhang, H., Zhao, Y., Wang, X., Rathore, A., et al. (2017). Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nat. Biotechnol. **35**:969–976.

Varshney, R.K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H.D., Yang, W., Bajaj, P., Cubry, P., Rathore, A., Jian, J., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. Nat. Genet. **51**:857–864.

Varshney, R.K., Roorkiwal, M., Sun, S., Bajaj, P., Chitikineni, A., Thudi, M., Singh, N.P., Du, X., Upadhyaya, H.D., Khan, A.W., et al. (2021). A chickpea genetic variation map based on the sequencing of 3,366 genomes. Nature **599**:622–627. https://doi.org/10.1038/s41586-021-04066-1.

Vasilevsky, N.A., Minnier, J., Haendel, M.A., and Champieux, R.E. (2017). Reproducible and reusable research: Are journal data sharing policies meeting the mark? PeerJ **5**:e3208.

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J., Ramirez-Gonzalez, R.H., Kolodziej, M.C., Delorean, E., Thambugala, D., et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. Nature **588**:277–283.

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R., Zhang, F., et al. (2018a). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature **557**:43–49.

Wang, X., Zou, B., Shao, Q., Cui, Y., Lu, S., Zhang, Y., and Huang, Q. (2018b). Natural Variation Reveals that OsSAP16 Controls Low-Temperature Germination in Rice, 69, pp. 413–421.

Wang, X., Liu, S., Zuo, H., Zheng, W., Zhang, S., Huang, Y., Pingcuo, G., Ying, H., Zhao, F., Li, Y., et al. (2021). Genomic basis of high-altitude adaptation in Tibetan Prunus fruit trees. Curr. Biol. **31**:3848–3860.e8.

Wang, J., Hu, Z., Liao, X., Wang, Z., Li, W., Zhang, P., Cheng, H., Wang, Q., Bhat, J.A., Wang, H., et al. (2022). Whole-genome resequencing reveals signature of local adaptation and divergence in wild soybean. Evol. Appl. **15**:1820–1833. https://doi.org/10.1111/eva.13480.

Wei, T., van Treuren, R., Liu, X., Zhang, Z., Chen, J., Liu, Y., Dong, S., Sun, P., Yang, T., Lan, T., et al. (2021a). Whole-genome resequencing of 445 Lactuca accessions reveals the domestication history of cultivated lettuce. Nat. Genet. **53**:752–760. https://doi.org/10.1038/s41588-021-00831-0.

Wei, X., Qiu, J., Yong, K., Fan, J., Zhang, Q., Hua, H., Liu, J., Wang, Q., Olsen, K.M., Han, B., and Huang, X. (2021b). A quantitative genomics map of rice provides genetic insights and guides breeding. Nat. Genet. **53**:243–253.

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. **37**:1155–1162.

Wibowo, A., Becker, C., Durr, J., Price, J., Spaepen, S., Hilton, S., Putra, H., Papareddy, R., Saintain, Q., Harvey, S., et al. (2018). Partial maintenance of organ-specific epigenetic marks during plant asexual reproduction leads to heritable phenotypic variation. Proc. Natl. Acad. Sci. USA **115**:E9145–E9152.

Witek, K., Jupe, F., Witek, A.I., Baker, D., Clark, M.D., and Jones, J.D.G. (2016). Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. Nat. Biotechnol. **34**:656–660.

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., Zhou, G., Lohwasser, U., Hua, S., Wang, H., et al. (2019). Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype Divergence. Mol. Plant **12**:30–43.

Wu, J., Wang, L., Fu, J., Chen, J., Wei, S., Zhang, S., Zhang, J., Tang, Y., Chen, M., Zhu, J., et al. (2020). Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. Nat. Genet. **52**:118–125.

Wu, X., Feng, H., Wu, D., et al. (2021). Using high-throughput multiple optical phenotyping to decipher the genetic architecture of maize drought tolerance. Genome Biol 22:1–26.

Wu, X., Liu, Y., Luo, H., Shang, L., Leng, C., Liu, Z., Li, Z., Lu, X., Cai, H., Hao, H., and Jing, H.C. (2022). Genomic footprints of sorghum domestication and breeding selection for multiple end uses. Mol. Plant 15:537–551.

Xiao, N., Pan, C., Li, Y., et al. (2021). Genomic insight into balancing high yield, good quality, and blast resistance of japonica rice. Genome Biol. 22:1–22.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.C., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H., Hirano, K., and Matsuoka, M. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat. Genet. 48:927–934.

Yano, K., Morinaka, Y., Wang, F., et al. (2019). GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. Proc. Natl. Acad. Sci. U S A 116:21262–21267.

Yu, J., Xiong, H., Zhu, X., et al. (2017). OsLG3 contributing to rice grain length and yield was mined by Ho-LAMap. BMC Biol. 15:1–18.

Yuan, Y., Chung, C.Y.L., and Chan, T.F. (2020). Advances in optical mapping for genomic research. Comput. Struct. Biotechnol. J. 18:2051–2062.

Yuan, Y., Bayer, P.E., Batley, J., and Edwards, D. (2021). Current status of structural variation studies in plants. Plant Biotechnol. J. 19:2153–2163.

Zanini, S.F., Bayer, P.E., Wells, R., Snowdon, R.J., Batley, J., Varshney, R.K., Nguyen, H.T., Edwards, D., and Golicz, A.A. (2022). Pangenomics in crop improvement—from coding structural variations to finding regulatory variants with pangenome graphs. Plant Genome 15:e20177.

Zeng, X., Guo, Y., Xu, Q., Mascher, M., Guo, G., Li, S., Mao, L., Liu, Q., Xia, Z., Zhou, J., et al. (2018). Origin and evolution of qingke barley in Tibet. Nat. Commun. 9:5433.

Zhang, M., Liang, X., Wang, L., Cao, Y., Song, W., Shi, J., Lai, J., and Jiang, C. (2019a). variation of salt tolerance in maize. Nat. Plants 5:1297–1308.

Zhang, T., Wu, T., Wang, L., Jiang, B., Zhen, C., Yuan, S., Hou, W., Wu, C., Han, T., and Sun, S. (2019b). A combined linkage and GWAS analysis identifies QTLs linked to soybean seed protein and oil content. Int. J. Mol. Sci. 20:5915.

Zhang, K., He, M., Fan, Y., Zhao, H., Gao, B., Yang, K., Li, F., Tang, Y., Gao, Q., Lin, T., et al. (2021a). Resequencing of global Tartary buckwheat accessions reveals multiple domestication events and key loci associated with agronomic traits. Genome Biol. 22:23.

Zhang, M.Y., Xue, C., Hu, H., Li, J., Xue, Y., Wang, R., Fan, J., Zou, C., Tao, S., Qin, M., et al. (2021b). Genome-wide association studies provide insights into the genetic determination of fruit traits of pear. Nat. Commun. 12:1144.

Zhang, Y., Fu, Y., Xian, W., et al. (2023). Comparative Phylogenomics and Phylotranscriptomics Provide Insights into the Genetic Complexity of Nitrogen Fixing Root Nodule Symbiosis. Plant Commu.

Zhao, G., Lian, Q., Zhang, Z., Fu, Q., He, Y., Ma, S., Ruggieri, V., Monforte, A.J., Wang, P., Julca, I., et al. (2019a). A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. Nat. Genet. 51:1607–1615.

Zhao, Y.P., Fan, G., Yin, P.P., Sun, S., Li, N., Hong, X., Hu, G., Zhang, H., Zhang, F.M., Han, J.D., et al. (2019b). Resequencing 545 ginkgo genomes across the world reveals the evolutionary history of the living fossil. Nat. Commun. 10:4201.

Zhao, J., Sauvage, C., Zhao, J., Bitton, F., Bauchet, G., Liu, D., Huang, S., Tieman, D.M., Klee, H.J., and Causse, M. (2019c). Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavor. Nat. Commun. 10:1534.

Zhao, J., Bayer, P.E., Ruperao, P., Saxena, R.K., Khan, A.W., Golicz, A.A., Nguyen, H.T., Batley, J., Edwards, D., and Varshney, R.K. (2020). Trait associations in the pangenome of pigeon pea (Cajanus cajan). Plant Biotechnol. J. 18:1946–1954.

Zheng, C., Zhou, J., Zhang, F., Yin, J., Zhou, G., Li, Y., Chen, F., and Xie, X. (2020a). OsABAR1, a novel GRAM domain-containing protein, confers drought and salt tolerance via an ABA-dependent pathway in rice. Plant Physiol. Biochem. 152:138–146.

Zheng, C., Tan, L., Sang, M., Ye, M., and Wu, R. (2020b). Genetic adaptation of Tibetan poplar (Populus szechuanica var. tibetica) to high altitudes on the Qinghai–Tibetan Plateau. Ecol. Evol. 10:10974–10985.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol. 33:408–414.

Zhou, Y., Zhao, X., Li, Y., Xu, J., Bi, A., Kang, L., Xu, D., Chen, H., Wang, Y., Wang, Y.G., et al. (2020). Triticum population sequencing provides insights into wheat adaptation. Nat. Genet. 52:1412–1422.

Zhou, Y., Zhang, Z., Bao, Z., Li, H., Lyu, Y., Zan, Y., Wu, Y., Cheng, L., Fang, Y., Wu, K., et al. (2022). Graph pangenome captures missing heritability and empowers tomato breeding. Nature 606:527–534. https://doi.org/10.1038/s41586-022-04808-9.

Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., Lin, T., Qin, M., Peng, M., Yang, C., et al. (2018). Rewiring of the Fruit Metabolome in Tomato Breeding. Cell 172:249–261.e12.