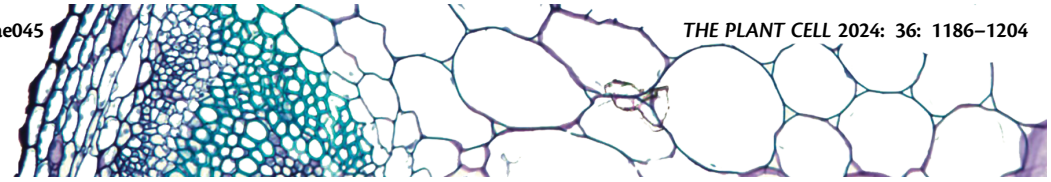
https://doi.org/10.1093/plcell/koae045 THE PLANT CELL 2024: 36: 1186–1204



Genome organization and botanical diversity

Andrew H. Paterson [1,*] **Christine Queitsch** [2]

1 Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, USA
2 Department of Genome Sciences, University of Washington, Seattle, WA, USA

*Author for correspondence: paterson@uga.edu

Review

Abstract

The rich diversity of angiosperms, both the planet's dominant flora and the cornerstone of agriculture, is integrally intertwined with a distinctive evolutionary history. Here, we explore the interplay between angiosperm genome organization and botanical diversity, empowered by genomic approaches ranging from genetic linkage mapping to analysis of gene regulation. Commonality in the genetic hardware of plants has enabled robust comparative genomics that has provided a broad picture of angiosperm evolution and implicated both general processes and specific elements in contributing to botanical diversity. We argue that the hardware of plant genomes—both in content and in dynamics—has been shaped by selection for rather substantial differences in gene regulation between plants and animals such as maize and human, organisms of comparable genome size and gene number. Their distinctive genome content and dynamics may reflect in part the indeterminate development of plants that puts strikingly different demands on gene regulation than in animals. Repeated polyploidization of plant genomes and multiplication of individual genes together with extensive rearrangement and differential retention provide rich raw material for selection of morphological and/or physiological variations conferring fitness in specific niches, whether natural or artificial. These findings exemplify the burgeoning information available to employ in increasing knowledge of plant biology and in modifying selected plants to better meet human needs.

Introduction

The principle that plants share many commonalities in their underlying genetic "hardware" traces to at least Vavilov, basing his law of homologous series in variation on observations that comparable variant forms tended to appear in different varieties of the same species, different species of the same genus, and different genera of the same family (Vavilov 1922).

How is underlying genetic commonality reconciled with the rich diversity of extant angiosperms? First, commonality is not identity. Most genes have recognizable homologs in most plant species, but with coding and/or regulatory sequence differences ranging from minute (single nucleotide polymorphisms [SNPs]) to massive. Likewise, gene order along the chromosomes ranges from near-identity within a population [although even subtle differences may be of evolutionary importance by reducing gene flow (Rieseberg 2001)] to barely discernible among distant taxa. The progression of such divergence is by no means linear—for example, the genome of sorghum more closely parallels that of rice, separated by ~100 million years, than that of maize, separated by approximately 12 million years (Swigonova *et al.* 2004). Often, striking differences in genome structure of closely related taxa reflect punctuational consequences of a genome duplication, for example in the maize lineage (see below).

Second, similar hardware may use different software. The regulatory cues that determine gene expression are often short and easily mutable. Indeed, regulatory mutations may create interdependence that is a basis for the retention of duplicated genes, for example by reciprocal loss of expression cues that make different members of a duplicated gene pair each essential in different tissues (Force *et al.* 1999, 2005). We elaborate on this important dimension of (plant) diversity below.

Third, polymorphism determines the subset of genes that contribute to standing genetic variation within a taxon. Both random and nonrandom factors determine the persistence of gene polymorphism with functional consequences, the

Received October 16, 2023. Accepted February 07, 2024. Advance access publication February 21, 2024.

former being heavily weighted in gene pools that have experienced evolutionarily recent genetic bottlenecks such as those of most major crops. Only the subset of genes in which functional polymorphism differentiates the progenitors of a study population are accessible by forward genetics, while reverse genetics scans most genes.

Thus, although botanical diversity is a focus of this paper, essential to its study has been commonality in the genetic hardware and software of plants. In 2 hypothetical genomes that differed by only a single nucleotide, the identical surroundings rather than the difference itself would inform its likely phenotypic consequences. Robust comparative genomics, distinctive in plants in accommodating recursive genome duplication (see below), enables information from facile models to be extrapolated across the plant phylogeny as at least an indicator of gene function. This principle has provided for tremendous leveraging of both genome sequence and reverse genetics resources, with deep annotation and hard-won functional data for botanical models accelerating hypothesis formation in diverse taxa. Forward genetics also benefits from comparative genomics, although with comparisons among taxa confounded by factors that determine the persistence of polymorphism with functional consequences.

In the final scientific paper of an illustrious career, G. Ledyard Stebbins articulated "*a plant evolutionist's point of view . . . based on profound differences . . . between the 2 kingdoms*." In particular, he highlighted the lack of a germline, propensity for inbreeding, and greater tolerance of diversity, particularly polyploidy, attributed to the fewer cellular structures of plants than animals (Stebbins 1999). Little more than a year after his paper was published, Stebbins had died when the initial sequence of the first higher plant genome (Arabidopsis Genome Initiative 2000), followed 2 months later by the human genome (Lander *et al*. 2001), kicked off a new era in comparative biology. This era has been enabled by technological improvements that made genome sequencing rapid and routine and empowered by CPUs sufficient to archive and sift voluminous digital data for grains of insight.

More than 2 decades hence, voluminous literature has accumulated on the hardware and software of genomes, how these differ between and among plants and animals, and how they relate to diversity among and polymorphism within plant taxa. Here, we present a plant-centric view, reviewing the broad state of the field based on findings of and connections among a selected subset of studies that focus on plant-specific features. The widespread application of genomic technologies, particularly in agriculture, make truly exhaustive coverage intractable. Overrepresentation herein of our own studies and others in our focal taxa reflects those we are best able to interpret and integrate, with no disrespect intended to the work of many others that would be similarly suitable. We have further stretched our writ to volunteer opinions about areas in which we see promise for future investigations, an act of commission intended to stimulate second-level thinking by our colleagues and successors.

The hardware: using genome structure and composition to probe diversity

Robust DNA sequences provide digital access to the genome of each new organism, filling gaps in knowledge of evolutionary history and the spectrum of botanical diversity. As more taxa are sequenced, diminishing returns are naturally realized in discovery of new genes and macroevolutionary events such as genome duplications. However, closely related taxa with less functional diversity may be favorable systems in which to identify causal genomic differences, just as subtle differences in beak shape related to diet of closely related finches contributed to the theory of natural selection (Darwin 1859).

An end in itself in that it comprises the entire hereditary information of an organism, a genome is also a means to other ends in providing improved tools, as summarized in Table 1. High-quality genome sequences have expedited forward genetics approaches to home in on the region of a causal gene and reverse genetics approaches to identify causative variants. Dramatic improvement in both cost and throughput of sequencing has empowered "evolutionary genetics" to achieve critical mass as a complement to reverse genetics toward determining causative variants and added the potential precision of genome-wide association study (GWAS) to the repertoire of forward genetics. Robust comparative genomics (section IV, below) enables hard-won functional information about specific genes from forward, reverse, evolutionary, and/or association approaches to be extrapolated to newly sequenced genomes as at least an early indicator of gene function.

Forward genetics remains central in the botany toolbox, especially in crop genomes for which target traits tend to be complex (quantitative trait loci). Robust genome sequences obviate the need for laborious "dart-throwing" approaches by which were produced early genetic maps that guided sequence assembly. A variety of reduced representation "genotyping by sequencing" approaches (Davey *et al*. 2011) now permit one to routinely undertake whole-genome scans for trait mapping, and reference genomes expedite digital identification of precisely targeted diagnostic tools such as Kompetitive Allele Specific PCR (Semagn *et al*. 2013) or simple-sequence repeat markers. Genetic maps based on voluminous SNPs permit anchoring of even small sequence contigs as well as orientation of those spanning a recombination event—for example, more than 2 million SNPs in 1,178 positions separated by recombinants anchored 57,270 scaffolds each containing 5 or more mapped SNPs and collectively comprising 14% of the subject genome (Bowers *et al*. 2016). Such rich data also empower GWAS (Ozaki *et al*. 2002) that use physically dense marker sets to take advantage of historical accumulations of recombinations in relating a phenotype to a gene(s) in a crop gene pool or natural population.

Reverse genetics has tremendously increased plant gene functional knowledge, in particular with many thousands

Table 1. Genome-enabled genetics approaches in plants. Most plant genetics studies utilize 1 or more of 4 broad approaches, each suitable in different contexts, with different strengths and limitations, and requiring different resources

Genetics approach	% (subset) of gene set accessed	Rate (source) of false positives?	Rate (source) of false negatives?	Resolution (level) of DNA characterization	Resources required
Forward	Low (polymorphic in single cross)	Low	Medium (small population size)	Low (10 cM marker spacing	Mapping populations (BC, F2, DH, RIL, NIL)
Reverse	High (mutable resulting in phenotype)	Low	Medium (subtle phenotypes)	High (gene space sequence)	Mutant collection
Evolutionary	Medium (polymorphic in species)	High (relatedness of genotypes)	Medium (rare alleles)	High (gene space sequence)	Diversity panel (core or mini-core collection)
GWAS	Medium (polymorphic in species subset)	High (relatedness of genotypes)	Medium (rare alleles)	Medium (deep SNP coverage)	Diversity panel (gene pool sample)

GWAS, genome-wide association study; SNP, single nucleotide polymorphism.

of genes identified using the *Agrobacterium* T-DNA as both a mutagen and a tag (Feldmann *et al.* 1989), taking advantage of high throughput methods (Alonso *et al.* 2003) to produce extensive collections of mutant alleles. Historically, the ability to generate phenotypically interesting mutations has greatly exceeded the ability to identify causative sequence variants. Easily identifiable "tags" such as T-DNA or various transposons (Fladung 2016) and methods for efficient searches for point mutations (McCallum *et al.* 2000; Colbert *et al.* 2001; Till *et al.* 2003; Henikoff *et al.* 2004) have mitigated constraints to identifying causative plant gene mutations. Knockout mutant populations generated by means such as fast neutron mediated mutagenesis that cause mostly deletions of DNA fragments ranging in size from a few base pairs to more than 30 kb (Bruggemann *et al.* 1996; Li *et al.* 2001, 2002) can readily be sequence tagged even by low-coverage methods, although tolerating far fewer mutations per genome (Belfield *et al.* 2012) than EMS populations in which it is more challenging to discern (single-nucleotide) mutations with confidence [e.g. (Addo-Quaye *et al.* 2018)].

Naturally occurring alleles, "biased" in informative ways by the action of selection, have become a powerful complement to both forward and reverse genetics. The ability to generate draft genome sequences enables comprehensive searches for naturally occurring alleles that are statistically associated with a trait, incorporating elements of both forward and reverse genetics. Demonstrated in 2005 (Ozaki *et al.* 2002), such GWAS have been widely used, taking advantage of the historical accumulation of recombination events since the evolution of an allele. Sequencing of large germplasm collections, such as many collected and phenotyped for numerous crops by CGIAR centers, reveals much of the spectrum of naturally occurring alleles in a gene pool, also evaluating associations of the more abundant ones with traits [e.g. (Morris *et al.* 2013)]. In addition to de novo searches for trait association, phenotypic and passport information for such collections can be of value in testing hypotheses, for example, regarding the geographic distribution of specific candidate alleles (Cuevas *et al.* 2016).

Germplasm drawn from cultivated gene pools derived recently from small numbers of progenitors, or prepared by crossing strategies that combine broad samples of diversity with appreciable linkage disequilibrium (Yu *et al.* 2008), harnesses naturally occurring alleles and recombinations to improve the precision of trait mapping while mitigating the propensity of GWAS for false-positive associations [e.g. (Buckler *et al.* 2009)]. Such a "nested association mapping" approach is an attractive means to investigate effects of naturally occurring alleles that are too rare to obtain a significant GWAS signal, in that biparental populations placing the allele in hundreds of progeny offer higher statistical power.

While commonality of plant gene content has permitted extensive leveraging of reverse genetics findings across taxa, extrapolation of forward genetics data is constrained by random and nonrandom factors that affect the persistence of functional polymorphism, together with statistical constraints that limit quantitative trait loci mapping by either biparental methods [e.g. (Lander and Botstein 1989)] or GWAS (Yu *et al.* 2008). For example, the notion that crosses between wild and elite forms of different crops may segregate for convergent alleles at corresponding loci (Paterson *et al.* 1995) has had both supporting (Lin *et al.* 2012) and conflicting evidence (Tang *et al.* 2013). Nonetheless, across numerous traits and taxa, meta-analysis of the collected literature empowered by robust comparative genomics is informative in revealing "hotspots" for discovery of quantitative trait loci in elite gene pools and in providing diagnostic tools for such hotspots.

Exceptions to the commonality of genetic hardware often offer hypotheses about potential causal agents of specific traits. Genome sequences abound with examples of copy number amplifications that are correlated with distinctive features of particular plant taxa. *Brassica napus*, or canola, has experienced massive expansion of oil biosynthesis genes in an oilseed plant (Chalhoub *et al.* 2014) at almost double the number annotated in soybean (Schmutz *et al.* 2010) and more than double those in oil palm (Singh *et al.* 2013). Striking differences in the cell walls of monocots and dicots

(McCann and Roberts 1991; Carpita and Gibeaut 1993) are correlated with abundance of CesA/Csl cell wall biogenesis gene superfamily members, *Arabidopsis* containing a single Group F GT31 gene, whereas sorghum and rice contain 6 and 10, respectively (Paterson *et al.* 2009). Contraction or complete loss in tomato of several cytochrome P450 subfamilies associated with toxic alkaloid biosynthesis (Sato *et al.* 2012) may be related to the importance of attracting vertebrate frugivores to disperse seeds via fleshy fruits (Howe and Smallwood 1982). Copy number amplifications contribute to the evolution of herbicide resistance in weeds of genetically engineered crops (Patterson *et al.* 2018), potentially with multiple origins of such resistance (Fernández *et al.* 2013), attesting to the speed at which this mechanism permits a plant to respond to new selective pressure. Some such rapid responses appear to be enabled by heritable extranuclear DNA (Molin *et al.* 2020).

While the commonality of genetic hardware among plants is a useful generalization, formation of new genes and loss of existing ones are each continuous, with the content of an extant genome ranging from taxonomically widespread, highly conserved genes to taxonomically restricted "orphan" genes. Percentages of orphan genes in a genome range widely, with 5% to 15% being typical (Arendsee *et al.* 2014) and often declining as additional closely related genomes are sequenced. Age stratification found 4% of *A. thaliana* genes to be species specific, 61.3% to date to the origin of eukaryotes (30.3%, an estimated 1.6 billion years ago) or cellular organisms (31.0%, 2.52 billion years ago), and varying percentages shared with ascending hierarchical taxonomic groups, for example 0.8% originating with the *Arabidopsis* genus and 3.3% with the Brassicaceae family (Arendsee *et al.* 2014).

In that their comparative analysis is constrained by definition, orphan genes are prone to artifactual annotation; however, various types of evidence attest to subsets having important functions. For example, sorghum genes containing 1 functional domain that was absent from rice encoded alpha kafirins that account for most sorghum seed storage protein (Paterson *et al.* 2009) and were found to correspond to maize zeins (Xu and Messing 2008). A recent analysis of 3,553 orthogroups (totaling 5,456 genes) found, with very few exceptions, that those comprised solely of species specific (in the example, *P. alba*) genes in which multiple genes showed molecular signatures of positive selection had features associated with disease resistance (Kong *et al.* 2023).

The software: plant gene regulation in light of genome structure and lifestyle

Gene expression translates the genome's hardware into messenger RNAs and noncoding RNAs and ultimately generates the diversity of plant cell types and tissues and their finely tuned responses to environmental cues and challenges. As in animals, epigenetic marks reinforce patterns of cell-type and condition-specific gene expression and contribute to appropriate transcriptional responses (Sullivan *et al.* 2014; Schmitz *et al.* 2022). Accessible chromatin regions (i.e. regulatory DNA, ACRs) are enriched for trait-associated genetic variation implicated in domestication and evolution (Maurano *et al.* 2012; Sullivan *et al.* 2014; Rodgers-Melnick *et al.* 2016). Although researchers have learned some of the rules that govern the transcription of genomic information in plants, there is much left to be discovered.

Plant gene regulation research lacked the unified, comprehensive, and well-funded ENCODE and ModENCODE projects that annotated transcripts, transcription factor binding sites, histone modifications, ACRs, and long-range regulatory interactions for many human and animal cell lines and tissues (ENCODE 2012; Sanyal *et al.* 2012; Thurman *et al.* 2012). However, during the past decade, the methods developed and employed in these projects and other genome-scale interrogations of animal genomes (Patwardhan *et al.* 2009; Arnold *et al.* 2013; Gasperini *et al.* 2019; Klein *et al.* 2020) have been adapted for use in plants in many studies (Rodríguez-Leal *et al.* 2017; Ricci *et al.* 2019; Jores *et al.* 2020, 2021, 2023; Deng *et al.* 2023; Tan *et al.* 2023). Taken together, these studies have identified ACRs, histone marks associated with various activity states, transcription factor binding sites, and candidate long-range regulatory interactions in diverse plant genomes (Schmitz *et al.* 2022; Jores *et al.* 2023). Too often, however, data interpretation and conclusions appear to be driven by the deep existing knowledge of gene regulation mechanisms in animals without sufficient consideration given to the distinct features of plant genomes and plant life that might be relevant to understand gene regulation in these sessile and largely autotrophic organisms.

Plant and animal genomes differ strikingly in gene size and prevalence of transposable elements, both features of likely importance for gene regulation. The maize and human genomes are of comparable size (2.4 and 3 Gb, respectively), yet the average protein-coding gene size is ~4,000 bp in maize and approximately 32,000 bp in human. This 8-fold difference is largely due to the enormous size of introns in human (human, mean length 6,174 bp, median 1,590 bp in protein coding transcripts; maize, mean length 699 bp, median 145 bp in protein coding transcripts). In human, the genomic "real estate" occupied by introns and intronic ACRs is far greater than in maize (Fig. 1) and impacts the distances that need to be spanned between intronic enhancer elements and their target transcription start sites (TSS; human, distance of 8,530 intronic ACRs to nearest TSS: mean 35,800 bp, median 11,501 bp; maize, distance of 350 intronic ACRs to nearest TSS: mean 8,074 bp, median 4,118 bp).

The first genome-wide plant regulatory landscapes established in *A. thaliana* (Zhang *et al.* 2012; Sullivan *et al.* 2014) show marked differences in the genomic distribution of ACRs compared with those in animals. In *A. thaliana*, less than 5% of ACRs reside in introns, about 37% reside within 400 bps of a TSS, and another 37% were classified as intergenic or distal (with the latter designation not well suited to this compact genome). In the equally compact and

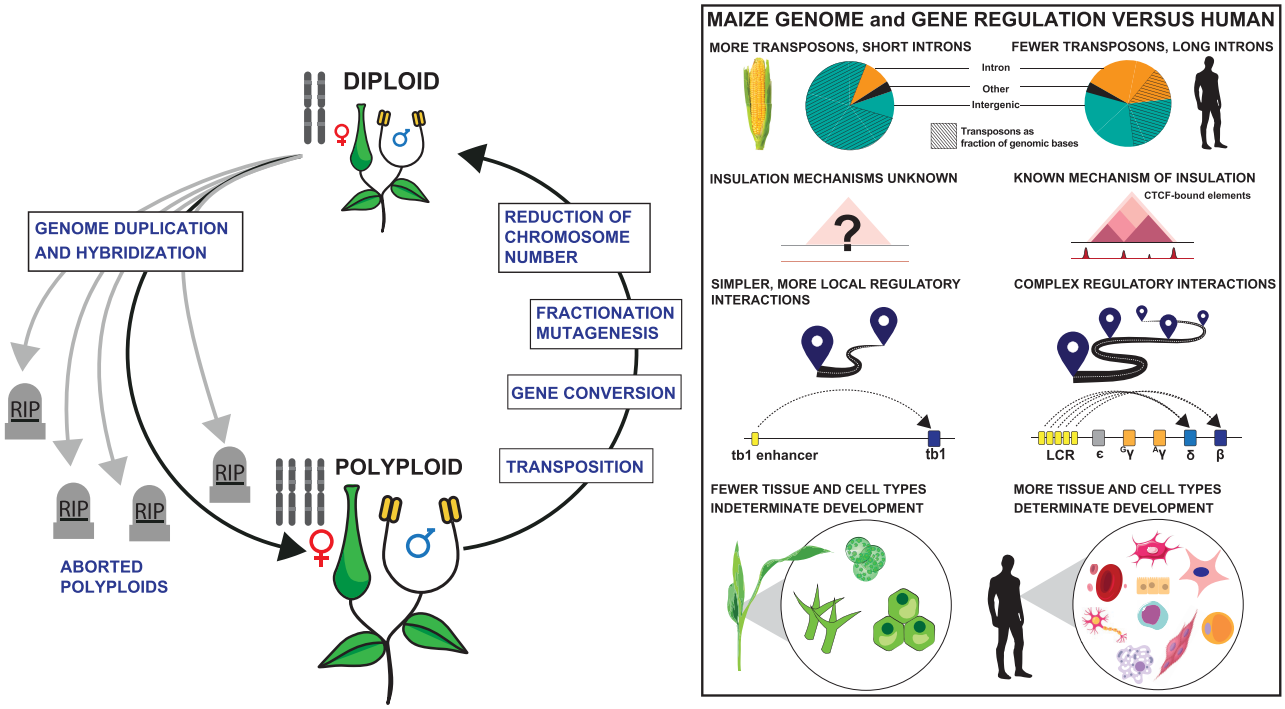


Figure 1. Plant genomes—the hardware—have been shaped by frequent polyploidization events (left figure panel) via hybridization or duplication. Newly formed polyploids face chromosome segregation and gene dosage challenges, leading to many aborted lineages. In a small subset of lineages, selection acts on the variation created via transposition, gene conversion, and fractionation mutagenesis, generally with reduction of chromosome number such that a diploid-like state is restored. Maize provides a particularly good example, now having the same chromosome number as sorghum despite having experienced genome duplication since their divergence approximately 20 million years ago. This cycle has consequences for plant genome structure and gene regulation—the software—as shown here for maize and human (right figure panel). Although both genomes are of comparable size and contain similar numbers of genes, human genes are much longer, largely because their introns are 10 times longer (pie charts, top). Compared with humans, maize genes and the maize regulatory landscape appear much more compact, possibly reducing the need for "expression domains" enabled by the human insulator protein CTCF (middle 2 panels). Although long-distance regulatory interactions exist in plants, including between the maize *tb1* gene and its transposon-derived enhancer, these interactions appear to be generally less complex than those observed in humans, as illustrated here with the beta-globin locus control region. The reduced complexity at this regulatory level is consistent with fewer cell types and the indeterminate, environmentally responsive mode of development in maize vs human (bottom panel). Shorter introns and a generally more compact and less complex regulatory landscape may render plant genomes less susceptible to disruption by the mutational processes following polyploidization.

gene-rich genome of the fly *Drosophila melanogaster*, 31% of ACRs reside in introns, 12% reside near a TSS, and 29% were classified as intergenic (Thomas *et al.* 2011). In human, the 2012 ENCODE study reported that 51% of ACRs reside in introns, 3% near a TSS, and 41% were classified as distal in 125 cell lines and tissue samples (Thurman *et al.* 2012). A 2020 ENCODE study improved on the latter work by vastly increasing the number of analyzed samples and broadening the coverage of primary cells and tissues (Moore *et al.* 2020). Moreover, the study integrated chromatin accessibility with histone modification marks on flanking nucleosomes and annotated CTCF-bound elements. This approach enables ACR annotation as promoters (high H3K4me3 signal, within 200 bp of a TSS), proximal (high H3K27ac, low H3K4me3, within 2 kb of a TSS) or distal enhancers (high H3K27ac, low H3K4me3, outside 2 kb of a TSS), insulator or looping elements (CTCF-bound), and elements of unknown function (high H3K4me3, not within 200 bp of a TSS). Using these annotations, ∼4% of ACRs are promoter-like, 15% are proximal enhancer-like elements within 2 kb of a TSS, and 71% are distal enhancer-like elements, many of which are residing in introns. In stark contrast to human, in maize, 29% of ACRs reside within 2 kb of a gene and only 45% are classified as distal ACRs residing outside of this interval (Ricci *et al.* 2019). Moreover, the majority of distal ACRs (51.2%) are depleted of flanking histone modifications, 10.2% show H3K9/K27/K56 acetylation, and 27.5% show histone modifications consistent with transcribed genes, possibly representing unannotated open reading frames.

The absence of the canonical flanking histone modifications marking active enhancers in animals (high H3K 27ac, high H3K4me1, low H3K4me3) is consistent with the absence of functional enhancer RNAs in plants. Enhancer RNAs are short (∼200 nt-2 kb) and short-lived transcripts that are pervasively and bidirectionally transcribed from active enhancers (Kim *et al.* 2010; Arner *et al.* 2015; Harrison and Bose 2022). They interact with the histone modifying enzyme complexes Polycomb Repressive Complex 2 and histone

acetyltransferases CREB binding protein/p300 to inhibit the deposition of repressive H3K27me3 histone marks and promote the deposition of activating H3K27ac histone marks, respectively, thereby maintaining enhancer chromatin accessibility and promoting further eRNA synthesis (Bose *et al.* 2017). Both enhancer RNAs and H3K27ac marks are used to annotate active animal enhancers (Andersson *et al.* 2014; Core *et al.* 2014; Wu *et al.* 2014; Moore *et al.* 2020). Analyses of nascent bi-directionally transcribed RNAs in *A. thaliana* and maize yielded conflicting results for (Lozano *et al.* 2021) and against the widespread existence of enhancer RNAs (Erhard *et al.* 2015; Hetzel *et al.* 2016; McDonald *et al.* 2023). A recent *A. thaliana* study revisited this controversy by inhibiting exosome-dependent degradation of unstable nascent RNAs and finding only 113 instances of bi-directionally transcription, of which 78 were intronic, and the remaining 35 were intergenic (Thieffry *et al.* 2020). Thus, plants appear to use different mechanisms than animals to maintain enhancer chromatin accessibility and activity or at least additional ones (McDonald *et al.* 2023).

Taken together, there appear to be profound differences in ACR spacing, ACR genomic context, and ACR histone modifications between human and maize. These differences may relate to another notable difference between animal and plant genomes: the absence of the transcriptional repressor CTCF in plant genomes (Fig. 1). In animals, CTCF is involved in regulating the 3D structure of chromatin and forming loops and marking TAD (Topologically Associating Domain) boundaries. CTCF-bound TAD boundaries function as insulators, separating regulatory units within which enhancers and promoters interact (Lupiáñez *et al.* 2015). In animals and human, disruption of TAD boundaries can lead to inappropriate interactions and profound misregulation (Okhovat *et al.* 2023). Although large plant genomes exhibit TAD-like structures, long-range chromatin interactions can span these structures (Dong *et al.* 2017; Liu *et al.* 2017; Mascher *et al.* 2017; Doğan and Liu 2018). As neither the protein(s) nor any sequence motif(s) underpinning boundary activity are known in any plant (Heger and Wiehe 2014), it remains unresolved how plant TAD-like structures arise and how relevant they are for gene regulation (Domb *et al.* 2022; Schmitz *et al.* 2022). Given the differences in spacing and genomic context of maize and human ACRs, the need for strictly enforcing insulation of long-distance enhancer-promoter interactions might be less profound in plants.

It is tempting to speculate that some of the described differences arise in part from the indeterminate development of plants that puts strikingly different demands on gene regulation. Throughout their life, plants continue to form new organs in response to environmental cues, which requires close and constant integration of environmental response and developmental gene expression pathways in ways simply not present in fly, mouse, and human. This constant integration might not be compatible with the strict expression boundaries observed in these animals; instead, it likely requires rapid rewiring across both long- and short-distance regulatory interactions. In human, of the nearly 3 million union ACRs detected in 125 diverse samples, nearly 1 million were specific to 1 sample, nearly 2 million in 2 or more samples, and only 3,692 ACRs were detected in all samples (Thurman *et al.* 2012). ACRs in human are so specific to cell and tissue type that their changing patterns and decreasing numbers along developmental trajectories can be used to reconstruct cell fate and lineage relationships from embryonic stem cells to terminal fates (Stergachis *et al.* 2013). Although direct comparisons are challenging because far fewer plant cell and tissue types have been studied in this way, ACRs in *A. thaliana* and maize appear to be far less dynamic—i.e. cell-type or condition-specific—than those in human. Across 13 diverse *A. thaliana* samples that detected nearly 47,000 union ACRs, only about 2,000 ACRs were specific to 1 sample while nearly 45,000 ACRs were detected in 2 or more samples (Sullivan *et al.* 2019). In maize, only 15% to 21% of the distal ACRs accessible in leaf tissue were inaccessible in inflorescence tissue (Ricci *et al.* 2019). Although single-cell genomics studies in both *A. thaliana* and maize have yielded evidence for larger numbers of dynamic, cell-type-specific ACR (Dorrity *et al.* 2021; Marand *et al.* 2021) cell and tissue identity appears less rigidly engrained in the regulatory landscape of plants than in animals.

However, rigid cell fate determination through altered patterns of ACRs might be less required in plants. Unlike animal development, plant development does not involve movement of cells because cell identity and lineage are established by cell division and position (van den Berg *et al.* 1995). This developmental mode likely minimizes the role of cell-autonomous gene regulation and cell-type-specific epigenomes and allows for scenarios in which transcription factors and other molecules expressed in a particular cell type can act in a concerted fashion with those in neighboring cells. It is tempting to speculate that the less rigid (i.e. less cell and tissue-specific) regulatory landscapes found in plants contribute to their capacity to regenerate fully functional plants from excised tissue or protoplasts (Gaillochet and Lohmann 2015) because there are fewer ACRs that have to be reopened to turn terminally differentiated plant cells into stem-cell-like ones. Moreover, coming back to Stebbin's observations about animal and plant kingdom differences, there are likely many fewer distinct cell types in maize than in human or mouse (Fig. 1).

The relative stasis of plant regulatory landscapes may also reflect another phenomenon: the constant integration of environmental cues with developmental gene expression may require that a large number of ACRs are poised for activation, i.e., occupied by trans-acting factors, without active transcription occurring until a signal is perceived. This mode of transcriptional regulation, occurring at enhancers but also immediately downstream of the transcription start site with paused, transcriptionally engaged Pol II molecules awaiting activation, allows for fast response times and is often observed for developmental and environmental response genes (Boehm *et al.* 2003; Creyghton *et al.* 2010; Xi *et al.*

2011). For example, in *A. thaliana*, treatment with the growth hormone auxin results in differential regulation of over 1,200 genes (Lewis *et al.* 2013); however, accessibility increases for only 155 ACRs and decreases for another 167 ACRs (Sullivan *et al.* 2019), in part because auxin-responsive factors are DNA-bound irrespective of auxin treatment. A greater fraction of poised ACRs in plants than in human would result in a comparatively more static regulatory landscape and a weaker correlation between chromatin accessibility and nearby gene expression. Alternatively, the relative stasis of plant regulatory landscapes might be explained by the large expansion of plant transcription factor (TF) families (Shiu *et al.* 2005), with their repressive and activating family members recognizing very similar motifs. While repressive and activating TFs of a family may trade places upon an environmental or developmental stimulus, this dynamic change would be largely invisible at the level of ACRs.

Stebbins highlighted another plant-specific phenomenon —greater tolerance of polyploidy—that is of consequence for the genome structure of angiosperms and gene expression regulation. Polyploidy is well tolerated across maturing tissues of individual plants (endoreduplication) (Lang and Schnittger 2020), and ploidy changes occur frequently along plant lineages (Song and Chen 2015), as discussed in detail below. There is very little accounting in most gene expression studies for the fact that commonly assayed tissues like leaves and roots are highly divergent in ploidy (Bhosale *et al.* 2018; Lang and Schnittger 2020). Single-cell expression studies in *A. thaliana* roots suggest that gene expression overall drops with increasing ploidy and maturity of root hair cells, but the expression of cell-type-specific genes increases (Jean-Baptiste *et al.* 2019). It is unknown to what extent the many additional copies contribute to gene expression in endoreduplicated cells and how regulatory landscapes and interactions among regulatory elements may change in response to endoreduplication.

The frequent changes of ploidy across plant lineages may contribute to shorter genes, shorter introns, less pronounced TAD structures, fewer distal enhancers, and higher transposon content and activity in plants (Fig. 1). Increases in ploidy require rapid dosage compensation and rapid divergence (Conant and Wolfe 2008). Plants have mastered both through RNA-mediated DNA methylation (Song and Chen 2015) and through high transposon activity (Flagel and Wendel 2009). Of course, transposon activity has remodeled both plant and animal genomes and contributed to evolutionary novelty (Flagel and Wendel 2009; Lynch *et al.* 2015) and domestication traits (Studer *et al.* 2011). Nevertheless, the frequent reshuffling of plant genomes during the course of their deep evolutionary history through cycles of polyploidization and returns to a diploid state might contribute to their shorter genes, their more compact distance distribution of regulatory elements, their less cell type–specific landscapes, and their higher genomic transposon content, all which favor local gene regulation environments and disfavor the highly intricate enhancer architecture and large distances observed in the animal *HOX* gene clusters (Montavon and Duboule 2013) or the beta-globin locus control region (Levings and Bungert 2002) (Fig. 1).

In summary, we argue that the hardware of plant genomes —both in content and in dynamics—has contributed to rather substantial differences in gene regulation from that in animals as exemplified by maize and human, organisms of comparable genome size and gene number. The mechanistic underpinnings of several of the phenomena discussed here such as insulation and TAD formation, maintenance of enhancer accessibility, ACR presence, and gene expression in endoreduplicated cells remain unknown and will require technological innovation to be resolved. Specifically, we need to move beyond averaging large numbers of small fragments to infer chromatin states and element activity as we do in the current chromosome conformation and accessibility assays. We favor recently developed single-molecule, long-read assays like Fiber-seq and others (Abdulhay *et al.* 2020; Stergachis *et al.* 2020), which can capture ACRs, TF binding sites, nucleosome position, Pol II footprints, and cytosine methylation along 20-kb fibers. The cited methods use a DNA N^6-adenine methyltransferase to methylate accessible adenines and long-read sequencing to detect this modification at nucleotide resolution, in addition to cytosine methylation. Adenine methylation is extremely sparse in plants (Kong *et al.* 2022), contrary to earlier reports. Assessing gene regulatory landscapes in this way eliminates the need to infer activity states and co-regulation of neighboring regulatory elements (Pliner *et al.* 2018) and eliminates bias from interpreting data across different assay types. Single-molecule regulatory landscapes will give insight into the regulatory activity of formerly inaccessible, highly repetitive regions, including the large regions of plant genomes occupied by transposons, and they will capture the stochasticity of TF occupancy and nucleosome positioning across many fibers corresponding to a large genomic interval. As long-read sequencing becomes more efficient (and thus cheaper) and capable of even longer reads, the promise of visualizing and understanding gene regulation at the level of single molecules across diverse plants and diverse plant cell types cannot be overstated.

A singular history of polyploidization

Stebbins' appreciation of the importance of polyploidy in plant evolution, albeit considerable, was an underestimate. Building on early clues from genetic mapping (Kowalski *et al.* 1994; Paterson *et al.* 1996), its genome sequence revealed that even the relatively small 5-chromosome genome of *A. thaliana*, chosen as the first plant genome to be sequenced in part for its simplicity, retained traces of 2 duplications and 1 triplication (Blanc *et al.* 2000; Arabidopsis Genome Initiative 2000; Lynch and Conery 2000; Paterson *et al.* 2000; Vision *et al.* 2000; Bowers *et al.* 2003). With the further discovery that the first monocot genome sequence, rice, also confirmed clues from cytology (Lawrence 1931)

and genetic mapping (Kishimoto *et al.* 1994; Nagamura *et al.* 1995) of genome duplication (Goff *et al.* 2002), the study of duplicated gene relationships advanced from an isolated specialization to an essential element of genetic, evolutionary, and developmental investigations in higher plants.

In contrast to algorithms refined in microbes and/or animals to discern often-faint signal of ancient genome duplications (Bailey *et al.* 2002, 2004; Cheung *et al.* 2003a, 2003b; Tuzun *et al.* 2004), consideration of multiple colinearity was necessary to take full advantage of the relatively strong signals of, often recursive, plant genome duplications. For example, if a chromosomal region is preserved in 3 genomes (A, B, and C), then pair-wise predictions would combinatorially yield 3 inferences about ancestral gene repertoire and order (A-B, B-C, A-C). Polyploidy in 1 or more of the 3 genomes multiply the number of comparisons accordingly. To retrieve maximal information, both consequences of paleopolyploidy within a genome and comprehensive alignments of multiple paleopolyploid genomes to one another, require related pairwise colinear segment to be combined into one inferred order (A-B-C). For example, multi-alignment of *A. thaliana*, *Populus trichocarpa*, and *Carica papaya* genomes revealed evidence of previously unknown "triplicated" structure that was validated empirically by comparison to the *Vitis* genome (Tang *et al.* 2008), a phylogenetic outgroup that had not experienced any more recent polyploidy.

The more than 1,600 green plant genomes now available, representing over 800 species, have yielded a relatively good (though by no means complete) history of plant genome duplications (Fig. 2) and knowledge of their consequences, as follows:

1) Phylogenetic tree topologies for hundreds of putatively orthologous expressed sequence tagged (EST) sequences support the occurrence of 1 genome duplication in the common ancestor of seed plants, and another in the common ancestor of angiosperms (Jiao *et al.* 2011), with the lineage of only 1 extant angiosperm, *Amborella trichopoda*, not known to have preserved additional paleopolyploidy events (Amborela Genome Project 2013);
2) In the monocots, phylogenetic analysis of nested synteny blocks indicates 1 genome duplication in a common ancestor of grasses and commelinids, upon which additional duplications have been superimposed in many lineages (e.g. 1 in oil palm, 2 in grasses, possibly 3 in banana) (Jiao *et al.* 2014) and with a host of still more recent duplications in some, for example, allopolyploidy in a common *Miscanthus-Saccharum* ancestor ~3.8 to 4.6 million years ago (Kim *et al.* 2014);
3) In a common ancestor of eudicots, colinearity information revealed an ancient event first discerned in *Arabidopsis* (Bowers *et al.* 2003) and clearly determined to be a genome triplication in the grape genome (Jaillon *et al.* 2007), upon which a host of additional lineage specific events have been superimposed;
4) While most paleopolyploidizations are duplications (i.e. forming tetraploids), several triplications have been inferred to form hexaploids, for example, in a eudicot ancestor (Jaillon *et al.* 2007) and twice consecutively in the tomato lineage (Sato *et al.* 2012); and a decaploid may have been formed by a complex event in the cotton lineage involving multiple polyploidizations plus hybridization in short succession (Wang *et al.* 2016);
5) Plant genome duplication may not be merely episodic but cyclic in the sense that some fitness benefits gradually deteriorate and favor recursive polyploidization (Chapman *et al.* 2006). However, the interval between successive events varies widely even in closely-related taxa, for example, with nearly 100 million years between a pan-grass duplication and the formation of *Sorghum halepense* (Paterson *et al.* 2020), but only ~1 million years between an allopolyploidy shared by Miscanthus and Saccharum, and a Saccharum-specific autoploidy (Kim *et al.* 2014);
6) Genome duplication is a punctuational event in the evolution of a lineage, triggering changes such that closely related genomes differing by a duplication may be less similar to one another than much more ancient genomes of common ploidy. Study of recently formed natural polyploids (Paterson *et al.* 2012; Chalhoub *et al.* 2014; Liu *et al.* 2014; Chen *et al.* 2016; Zhuang *et al.* 2019) and 1 synthetic polyploid (Zhuang *et al.* 2019) reveals macromolecular processes similar to those well known in synthetic polyploids, including loss and restructuring of low-copy DNA sequences (Song et al. 1995; Feldman et al. 1997; Liu et al. 1998; Ozkan et al. 2001; Shaked et al. 2001; Kashkush et al. 2002; Ozkan et al. 2002; Pires et al. 2004), activation of genes and retrotransposons (O'Neill *et al.* 2002; Kashkush *et al.* 2003; Fontdevila 2005), gene silencing (Chen and Pikaard 1997a, 1997b; Comai *et al.* 2000; Lee and Chen 2001) and outright loss (Langham *et al.* 2004; Freeling 2009; Schnable *et al.* 2010), and subfunctionalization of gene expression patterns (Adams et al. 2003; Adams et al. 2004; Adams and Wendel 2005).
7) Modern chromosome numbers provide little information about the history of polyploidy in a lineage, as chromosome numbers tend to return to a narrow range following polyploidization, often via joining of homoeologous chromosomes near their termini (Wang *et al.* 2014). For example, the 5 chromosomes of *A. thaliana* are thought to trace to a total of 84 ancestral chromosomes (Wang *et al.* 2014). Reciprocal gene loss in different polyploid individuals or subpopulations, leading to a special case of Bateson–Dobzhansky–Muller incompatibility (Werth and Windham 1991), may favor survival of lineages with low chromosome numbers (Bowers and Paterson 2021).

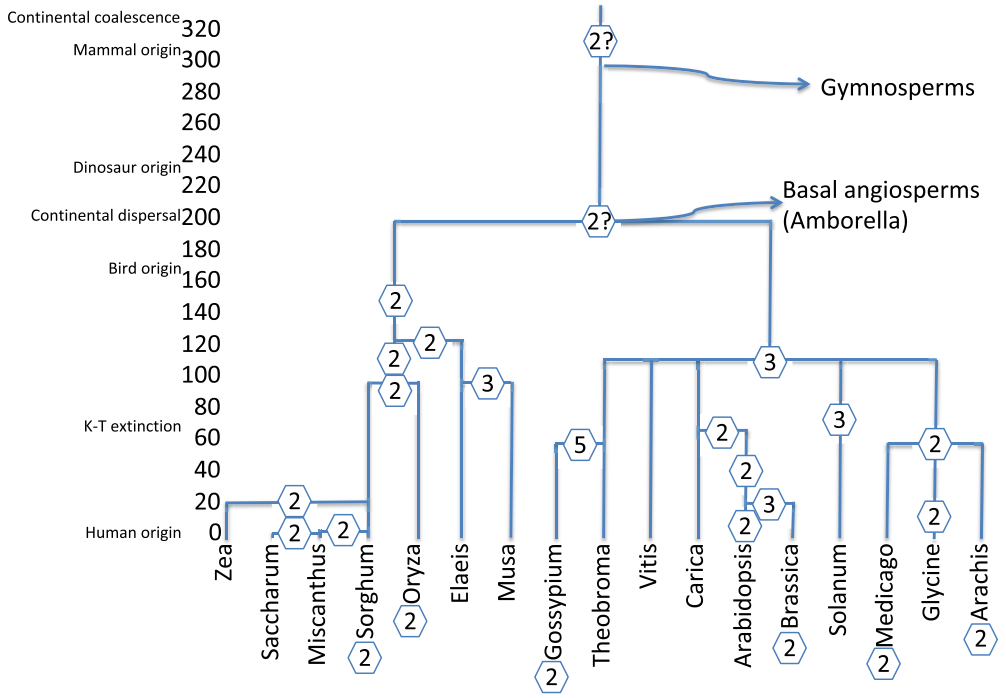


Figure 2. A brief history of plant genome duplication, illustrated in selected taxa. Genome duplication has been central to evolution in plants, arguably to a greater degree than in any other taxon. Duplication of regulatory genes important to seed and flower development appears to be concentrated around 319 and 192 million years ago (MYA) (Jiao *et al.* 2011): the former, roughly coincident with continental coalescence (Rogers and Santosh 2004), suggests a genome duplication in a common ancestor of seed plants; the latter, roughly coincident with continental dispersal (Rogers and Santosh 2004), preceded the diversification of angiosperms. While 1 basal angiosperm, *Amborella trichopoda*, has not preserved further duplications (Amborela Genome Project 2013), others have [e.g. (Chaw *et al.* 2019; Chen *et al.* 2019)]. Early duplication events also shaped the eudicot (Bowers *et al.* 2003; Jaillon *et al.* 2007) and monocot (Jiao *et al.* 2014) lineages. Genomic analyses have revealed many additional prehistoric duplications (Goff *et al.* 2002; Paterson *et al.* 2004; Schmutz *et al.* 2010; D'Hont *et al.* 2012; Sato *et al.* 2012; Kim *et al.* 2014), the most affected lineage being that of Brassica with 36-fold multiplication relative to the angiosperm common ancestor, plus an additional post-Neolithic polyploid formation between extant diploids (Chalhoub *et al.* 2014). Indeed, genus-specific polyploid formations led to some primary cultigens, for example in Arachis (Zhuang *et al.* 2019), Gossypium (Paterson *et al.* 2012), and Medicago (Shen *et al.* 2020); widespread invasives, for example, in sorghum (Paterson *et al.* 2020), and abundant species, for example, in Oryza (Zou *et al.* 2015). Selected outgroups are depicted to delineate timing of salient events. Timing of all events is based on indicated citations, and variation in estimation methods together with resolution of the figure make these timings approximate—readers should refer to original citations for precise estimates.

A natural question is whether the propensity of plants for genome duplication may have contributed to their evolutionary success. Duplication of existing genes is thought to be a primary source of genetic material available for evolution of genes with new functions (Taylor and Raes 2004), and a long-held hypothesis has been that polyploidy is an engine for the evolution of genetic diversity by facilitating "functional divergence" of duplicated genes. Duplication of an entire genome at once retains stoichiometric balance between its constituents (Birchler and Veitia 2007) and has long been recognized as less disruptive to the phenotype and viability than duplication of only a subset such as a single chromosome (Blakeslee *et al.* 1920). Following duplication, 1 member of a gene pair may be free to experience functional divergence, acquiring unique functionality [neofunctionalization, (Stephens 1951; Ohno 1970)] or the 2 copies subdividing ancestral functions, for example, by reciprocal loss of regulatory cues that render expression of different copies specific to different tissues [subfunctionalization —(Lynch and Force 2000)], with the fitness of the organism insulated by the homeolog. However, in that endoreduplication is well tolerated across maturing tissues of individual plants as noted above, it remains an open question whether expression-based neo- or subfunctionalization associated with polyploidy [e.g. (Adams *et al.* 2003, 2004; Adams and Wendel 2005)] is a striking response to a newly duplicated nucleus or merely a previously undetected adaptation to the frequent occurrence of many additional gene copies.

While diverse examples associate polyploidy with broader zones of adaptation (Kiedrzynski *et al.* 2021) and striking phenotypic changes such as the seedborne epidermal fibers of cotton (Jiang *et al.* 1998; Paterson *et al.* 2012), genomic data have also raised questions about the classical "functional divergence" model. For example, if a primary advantage of polyploidy is the opportunity for the evolution of genes with new functions, then patterns of genetic diversity among strains within paleopolyploid taxa might reveal footprints of selection that are consistent with duplicated genes being

relatively free to acquire unique functionality. Contrary to this prediction, SNPs encode less radical amino acid changes in genes for which there exists a duplicated copy at a "paleologous" locus, than in "singleton" genes among both *Arabidopsis* ecotypes and *Oryza* subspecies (Chapman *et al.* 2006). While this does not preclude the possibility that important adaptations have occurred by functional divergence of duplicated genes, it suggests that there may also be a population of genes in which fitness benefits accrue to functional buffering, consistent with remarkably long duration that duplicated yeast genes continue to compensate one another (Gu *et al.* 2003).

Organisms continuously require genetic variation to adapt to constantly fluctuating environments, yet genome duplications are episodic—in lineages that do not preserve whole genome duplications for long time periods, diverse mechanisms have provided the raw genetic material for adaptation. Sorghum and rice, which have not experienced whole genome duplication in an estimated 98 million years (Wang *et al.* 2015), preserve relatively more SNPs in tandem than paralogous duplicated genes (Guo *et al.* 2019). However, maize, which experienced genome duplication shortly after divergence from a common ancestor shared with sorghum as recently as 12 million years ago (Swigonova *et al.* 2004), shows SNP enrichment in its large supply of paralogous duplicates. The proportion of genes showing signatures of recent positive selection is higher in small-scale (tandem and transposed) than genome-scale duplicates in sorghum, but the opposite in maize (Guo *et al.* 2019).

Even complex biochemical pathways, which might be favored by stoichiometric balance resulting from whole-genome duplications, in at least 1 case did not evolve from available whole-genome duplicates but came together subsequently from a series of single-gene duplicates. For example, genome duplication in a common ancestor of grasses putatively provided "spare" copies of genes for the entire photosynthetic pathway; however, the evolution of C4 photosynthesis from C3 ancestors used few if any of these, instead involving single-gene duplications (Wang *et al.* 2009).

Paleopolyploidy results in islands of conserved sequence duplicated in parallel along stretches of homoeologous chromosomes, providing a genomic environment conducive to genetic exchanges between non-allelic loci. Similar DNA sequences may transiently form heteroduplex DNA (Holliday 1964, 1966), in which repair of unmatched bases may allow "gene conversion," rendering similar DNA sequences identical (Galtier 2003). Often implicated in homogenization of small tracts of paralogous DNA sequences, usually between several and several hundred base pairs (Petes and Symington 1991), gene conversion has long been thought to account for the evolution of various multigene families (Sawyer 1989; White and Crowther 2000; Mondragon-Palomino and Gaut 2005) and proximal gene clusters such as rRNA (Brown *et al.* 1972) and histone genes (Ohta 1984). Gene conversion may explain low divergence rates between paralogs produced by ancient large-scale duplication events in yeast (Gao and Innan 2004), appearing to have affected 2% of *Caenorhabditis elegans* duplicated genes (Semple and Wolfe 1999) and 18% of homeologs duplicated before the mouse–rat divergence (Ezawa *et al.* 2006).

Plant whole-genome duplications and associated genome-wide sets of simultaneously duplicated homoeologous genes have facilitated surveys for gene conversion, finding rich evidence of the action of this mechanism with clear temporal and spatial patterns. Complete genome sequences enable searches for footprints of gene conversion using elegant approaches such as phylogenetic incongruity in "gene trees," for example, finding that ancient duplicates within the same genome are more similar to one another than they are to their alleles in a different genome. Gene conversion begins virtually immediately after polyploid formation, being found at appreciable frequency in "synthetic" polyploid *Arachis* a few years after being produced by humans (Zhuang *et al.* 2019) and neopolyploid canola formed in nature a few thousand years ago (Chalhoub *et al.* 2014), with gradually declining frequency over a few million years of evolution of a new lineage but still discernible within the elite gene pools of crops such as cotton (Guo *et al.* 2014).

Striking spatial patterns of gene evolution are evident along plant chromosomes, with some general principles that appear to have persisted for millennia, and intriguing special cases. "Euchromatin" and "heterochromatin," long recognizable cytologically by differential staining, have profoundly different composition and evolution, effectively providing each plant genome with 2 qualitatively different compartments that respectively facilitate different types and rates of evolution. Euchromatin tends to be terminal to the chromosomes, accounts for most reciprocal exchange between orthologs ("conventional" recombination) and is gene-rich with gene orders persisting over long time periods (Bowers *et al.* 2005). Heterochromatin tends to be the physically larger compartment, accounting for much of the difference in DNA content between closely related taxa (Bowers *et al.* 2005), with relative paucity of genes but enrichment for recent single-gene duplications [or perhaps merely less effective removal than in euchromatin (Paterson *et al.* 2009)] and extensive migration of DNA between pericentromeric regions of different chromosomes. On paleo-duplicated chromosomes, regions of euchromatin and heterochromatin closely correspond (Paterson *et al.* 2009), indicating that these respective states have persisted for long time periods, for example, nearly 100 million years in the cereals (Wang *et al.* 2015).

Intriguing and perplexing are the case of rice chromosomes 11 and 12 and their homologs and homeologs in other cereals. Derived from a single ancestral chromosome in the pan-cereal genome duplication (Wang *et al.* 2007), these chromosomes experienced illegitimate recombination that has been temporally restricted in a stepwise manner, independently forming "strata" in different grass lineages with perplexing properties (Wang *et al.* 2011). The pericentromeric region accounts for two-thirds of the gene content

differences between this homeologous chromosome pair, yet there is generally low sequence divergence between paleoduplicated genes. Indeed, a distal region has the greatest DNA similarity between surviving duplicated genes found anywhere in the genome but also the highest concentration of lineage-specific gene pairs found anywhere in these genomes and with a significantly elevated gene evolutionary rate.

In partial summary, the neopolyploidy that has long been evident from classical genetics and cytology [e.g. (Stebbins 1966)] is merely the tip of the iceberg—plants have been experiencing recursive whole-genome duplications throughout their evolutionary history, followed by gene and chromosome losses and rearrangements that obscure cytological evidence but leave signatures of these ancient events in genome sequences. The classical notion of polyploidy as a generator of diversity may be somewhat simplistic in that endoreduplication of some tissues may "preadapt" plants to having multiple functioning gene copies in a cell, that signatures of selection consistent with functional buffering are also evident, and that the availability of duplicated genes for entire pathways was sometimes insufficient to catalyze major events.

Toward harnessing botanical diversity

With a broad range of plant genome sequences now available and the ability to routinely deepen the sample as needed to address fundamental questions and/or applied goals, better knowledge of plant evolutionary history informs and guides utilization of botanical diversity to improve human lives.

Building on early forays showing that plant genomes were not the staid and stable environments we once envisioned but in fact were highly fluid [e.g. (McClintock 1984)], detailed assemblies have highlighted heterogeneity across the genomic landscape. Plant genomes are dynamic and variable environments, broadly comprised of 2 qualitatively different compartments that respectively facilitate different types and rates of evolution, corresponding to classical "euchromatin" and "heterochromatin" (Bowers *et al.* 2005). Particularly striking chromosomal regions show both the most extensive gene loss and the most striking conservation of the remaining genes, putatively reflecting high levels of nonreciprocal exchanges but driven by selective forces that are not yet understood (Wang *et al.* 2011). In sum, a plant genome provides a range of options that may permit the same adaptive need to be met by different means at different times and/or in different lineages.

Much like the genome as a whole, the gene space also includes qualitatively different components that may respectively facilitate different types and rates of evolution. Polyploids have been thought to acquire capabilities that are "more than the sum" of those of their diploid progenitors, with gene duplication (Maere *et al.* 2005) providing material available for divergence to new function (Stephens 1951; Ohno 1970; Force *et al.* 1999; Lynch and Conery 2000; Taylor and Raes 2004), while normal function by the homeolog insulates the fitness of the organism. However, in contrast with the classical "functional divergence" model, some groups of gene families show signatures of selection consistent with "functional buffering" (conservation of ancestral function) (Chapman *et al.* 2006; Paterson *et al.* 2006), with duplicated copies continuing to functionally compensate one another for far longer periods (Gu *et al.* 2003) than are thought to be necessary for "functional divergence" to occur (Lynch and Conery 2000).

In nature, heterogeneity and flexibility, rather than a specific pattern of genome organization, may confer evolutionary success. Evolution occurs at the intersection of genetics and ecology, each the outcome of hosts of multifactorial interactions, and this complexity may be reflected in outcomes incongruous with evolutionary expectations rooted in genomics. For example, the evolution of a complex pathway such as C4 photosynthesis intuitively would have been greatly simplified by stoichiometric balance between its constituent parts (Birchler and Veitia 2007)—the inference that it evolved *not* from available whole-genome duplicates but from single-gene duplications (Wang *et al.* 2009) suggests a series of adaptive steps toward its greater efficiency at ambient CO_2 levels and elevated temperatures (Heckmann *et al.* 2013; Schluter and Weber 2020). Likewise, in lineages that survived long periods during which no whole-genome duplications survived, adaptation utilized raw material provided by other mechanisms (Guo *et al.* 2019).

The spectrum of adaptations observed in nature informs approaches by which plants might be improved to better meet human needs. For example, cotton domestication was associated with conversion of a few dozen genes in the "D" genome (from a progenitor that does not produce spinnable fibers) to the "A" genome sequence, thus doubling copy number of the allele from the progenitor that does produce spinnable fibers (Guo *et al.* 2014). This raises several questions. First, did this doubling confer variation in fiber yield or quality for which domesticates were selected? Second, would doubling of other cotton genes permit further improvements—or does a similar phenomenon contribute to phenotypes of other crops? Genome editing now permits empirical testing of such questions by engineering of conversions not found in nature.

A natural means by which large numbers of single gene-sized chromosomal tracts could be transferred between exotic and elite genotypes for empirical phenotypic evaluation would be of high value. Transfer of favorable traits from wild or exotic relatives by recurrent "backcrossing" to an elite crop cultivar typically includes 10% to 20% of the chromosome carrying the target gene, including hundreds of nearby genes that often impose "linkage drag" from associated undesirable phenotypes. DNA markers expedite this process, especially for target genes with recessive or subtle phenotypes otherwise requiring replicated progeny testing and identifying recombinants that minimized linkage drag. However, better still would be a means to induce such single-allele transfers, for example, in F1 hybrids between an exotic

donor and an elite line engineered with appropriate molecular machinery—followed by selfing or backcrossing to permit "screening" large numbers of individual exotic alleles for useful effects in progeny in much the same manner as reverse genetics screens while dramatically reducing the impact of linkage drag. This would especially facilitate evaluation of allelic variation in recombinationally recalcitrant heterochromatin, usually the physically larger compartment of a plant genome. Albeit not involving transfer or gene-sized tracts, methods by which hundreds of novel regulatory alleles can be generated (Rodríguez-Leal *et al.* 2017) foretell similar new opportunities to screen large numbers of candidates for alleles conferring novel functionality.

Polyploidy itself has long been employed to introgress botanical diversity from wild species into cultivated gene pools, but many species combinations fail to produce viable hybrids. Moreover, hybrids from such crosses often form novel alleles per se, by loss and restructuring of low-copy DNA sequences (cited above), which in some cases are associated with phenotypes (Schranz and Osborn 2000; Pires et al. 2004; Schranz and Osborn 2004). Convergent loss of duplicated copies of specific genes following independent duplications that are separated by hundreds of millions of years of evolution may reflect an underlying set of principles of molecular evolution that contribute to the fates of genome duplications (Paterson *et al.* 2006), raising the intriguing hypothesis that persistence of artificial polyploids might be facilitated by silencing 1 copy of these genes. Methods for simultaneously editing dozens of genes at once (Campa *et al.* 2019; Yuan and Gao 2022) or in orthogonal combinations (Cetin *et al.* 2023) may permit empirical tests of this hypothesis, potentially accessing rich additional diversity.

While the creation of "synthetic" tetraploids by humans has been widely attempted to exploit otherwise-inaccessible botanical variation in crop improvement, a recent example reverses the process. The formation of tetraploid *S. halepense* ~1 to 2 million years ago from a naturally occurring event merging the genomes of African *S. bicolor* (sorghum) and Asian *S. propinquum* was the first surviving genome duplication in the sorghum lineage in nearly 100 million years (Wang *et al.* 2015), and its spread across 6 continents has exposed its rich variation to diverse selective forces (Paterson *et al.* 2020). Recent discovery that diploid sorghums can be obtained from certain crosses with tetraploid *S. halepense*-derived materials (Cox *et al.* 2017) permits genetic novelty from *S. halepense* to be investigated for contributions to the conventional sorghum gene pool.

In closing, the world has changed since the first author walked the wheat fields as a graduate student, learning empirical whole organism–level methods of plant breeding. It is hard to imagine an alternative that evaluates the voluminous number of interactions between and among hardware, software, and environment that are reflected in the whole organism–level phenotyping essential to making plant breeding decisions. However, the potential inputs into crop improvement programs have snowballed, genomics transcending transgenesis with enriched knowledge of the innate hardware and software of plant genomes and their function, and with new abilities to track genome transmission and alter the hardware and software in designed ways that leave behind no exogenous DNA. Moreover, we have far better understanding of the heterogeneity and flexibility of plant genomes and the forces that have acted to permit them to adapt to habitats from the tropics to near the poles and from sea level to at least 20,100 feet in altitude (Younghusband 1926), forming ramets ranging from floating *Wolffia* spp. plants of 1 mm in length (Raven *et al.* 1992) to *Eucalyptus regnans* trees of 100 m in height and 10 m in trunk diameter. Writing this article was especially interesting—perhaps as was envisioned by the editors who suggested such a partnership, the second ("junior") author challenged the first to reach well beyond the original assignment, with the end result presenting a broad and, we hope, integrative picture of the state of knowledge of botanical diversity at the molecular level. We emphasize once again that this breadth made truly exhaustive coverage intractable, and the overrepresentation of our own studies and others in our focal taxa is with no disrespect intended to the excellent work of many esteemed colleagues that would be similarly suitable.

Acknowledgments

Paterson thanks numerous past and current PGML and other colleagues for influencing his thinking, empowering, and conducting research salient to this paper. Queitsch is grateful to Kerry Bubb for inspiration, discussion, and analyses concerning the "software" of plant genomes. She thanks Sayeh Gorjifard for assistance with creating Fig. 1.

Author contributions

Both authors contributed significantly to the preparation of this review. A.H.P. was the primary author of the "hardware", "polyploidization" and "harnessing" sections. C.Q. was the primary author of the "software" section. A.H.P. and C.Q. were the primary authors of the abstract and introduction sections, also each reviewing and revising the sections written by the other.

Funding

Salient work in the laboratory of A.H.P. was partially funded by many past grants from the US National Science Foundation, US Department of Agriculture, US Agency for International Development, US Department of Energy Joint Genome Institute Community Sequencing Program, Cotton Incorporated, and a University System of Georgia Regents Professorship; and for C.Q., from the US National Science Foundation (RESEARCH-PGR grant no. 1748843, PlantSynBio grant no. 2240888) and the US National Institute of Health (NIGMS MIRA grant no. 1R35GM13953).

Conflict of interest statement. None declared.

Data availability

No new data were generated or analysed in support of this research.

References

Abdulhay NJ, McNally CP, Hsieh LJ, Kasinathan S, Keith A, Estes LS, Karimzadeh M, Underwood JG, Goodarzi H, Narlikar GJ, et al. Massively multiplex single-molecule oligonucleosome footprinting. Elife. 2020:**9**:e59404. https://doi.org/10.7554/eLife.59404

Adams KL, Cronn R, Percifield R, Wendel JF. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci U S A. 2003:**100**(8):4649–4654. https://doi.org/10.1073/pnas.0630618100

Adams KL, Percifield R, Wendel JF. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics. 2004:**168**(4):2217–2226. https://doi.org/10.1534/genetics.104.033522

Adams KL, Wendel JF. Novel patterns of gene expression in polyploid plants. Trends Genet. 2005:**21**(10):539–543. https://doi.org/10.1016/j.tig.2005.07.009

Addo-Quaye C, Tuinstra M, Carraro N, Weil C, Dilkes BP. Whole-genome sequence accuracy is improved by replication in a population of mutagenized sorghum. 2018:**8**(3):1079–1094. https://doi.org/10.1534/g3.117.300301

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al. Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science. 2003:**301**(5633):653–657. https://doi.org/10.1126/science.1086391

Amborela Genome Project. The amborella genome and the evolution of flowering plants. Science. 2013:**342**:1241089. https://doi.org/10.1126/science.1241089

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014:**507**(7493):455–461. https://doi.org/10.1038/nature12787

Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature. 2000:**408**(6814):796–815. https://doi.org/10.1038/35048692

Arendsee ZW, Li L, Wurtele ES. Coming of age: orphan genes in plants. Trends Plant Sci. 2014:**19**(11):698–708. https://doi.org/10.1016/j.tplants.2014.07.003

Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drabløs F, Lennartsson A, Rönnerblad M, Hrydziuszko O, Vitezic M, et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science. 2015:**347**(6225):1010–1014. https://doi.org/10.1126/science.1259418

Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science. 2013:**339**(6123):1074–1077. https://doi.org/10.1126/science.1232542

Bailey JA, Church DM, Ventura M, Rocchi M, Eichler EE. Analysis of segmental duplications and genome assembly in the mouse. Genome Res. 2004:**14**(5):789–801. https://doi.org/10.1101/gr.2238404

Bailey JA, Gu ZP, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. Recent segmental duplications in the human genome. Science. 2002:**297**(5583):1003–1007. https://doi.org/10.1126/science.1072047

Belfield EJ, Gan X, Mithani A, Brown C, Jiang C, Franklin K, Alvey E, Wibowo A, Jung M, Bailey K, et al. Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of Arabidopsis thaliana. Genome Res. 2012:**22**(7):1306–1315. https://doi.org/10.1101/gr.131474.111

Bhosale R, Boudolf V, Cuevas F, Lu R, Eekhout T, Hu Z, Van Isterdael G, Lambert GM, Xu F, Nowack MK, et al. A spatiotemporal DNA endoploidy map of the arabidopsis root reveals roles for the endocycle in root development and stress adaptation. Plant Cell. 2018:**30**(10):2330–2351. https://doi.org/10.1105/tpc.17.00983

Birchler JA, Veitia RA. The gene balance hypothesis: FROM classical genetics to modern genomics. Plant Cell. 2007:**19**(2):395–402. https://doi.org/10.1105/tpc.106.049338

Blakeslee AF, Belling J, Farnham ME. Chromosomal duplication and Mendelian phenomena in datura mutants. Science. 1920:**52**(1347):388–390. https://doi.org/10.1126/science.52.1347.388

Blanc G, Barakat A, Guyot R, Cooke R, Delseny I. Extensive duplication and reshuffling in the arabidopsis genome. Plant Cell. 2000:**12**(7):1093–1101. https://doi.org/10.1105/tpc.12.7.1093

Boehm AK, Saunders A, Werner J, Lis JT. Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. Mol Cell Biol. 2003:**23**(21):7628–7637. https://doi.org/10.1128/MCB.23.21.7628-7637.2003

Bose DA, Donahue G, Reinberg D, Shiekhattar R, Bonasio R, Berger SL. RNA binding to CBP stimulates histone acetylation and transcription. Cell. 2017:**168**(1–2):135–149.e122. https://doi.org/10.1016/j.cell.2016.12.020

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC, et al. Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. Proc Natl Acad Sci U S A. 2005:**102**(37):13206–13211. https://doi.org/10.1073/pnas.0502365102

Bowers JE, Chapman BA, Rong J, Paterson AH. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature. 2003:**422**(6930):433–438. https://doi.org/10.1038/nature01521

Bowers JE, Paterson AH. Chromosome number is key to longevity of polyploid lineages. New phytol. 2021:**231**(1):19–28. https://doi.org/10.1111/nph.17361

Bowers JE, Pearl SA, Burke JM. Genetic mapping of millions of SNPs in safflower (Carthamus tinctorius L.) via whole-genome resequencing. G3 (Bethesda). 2016:**3**(6):2203–2211. https://doi.org/10.1534/g3.115.026690

Brown DD, Wensink PC, Jordan E. A comparison of the ribosomal DNA's of Xenopus laevis and Xenopus mulleri: the evolution of tandem genes. J Mol Biol. 1972:**63**(1):57–73. https://doi.org/10.1016/0022-2836(72)90521-9

Bruggemann E, Handwerger K, Essex C, Storz G. Analysis of fast neutron-generated mutants at the Arabidopsis thaliana HY4 locus. Plant J. 1996:**10**(4):755–760. https://doi.org/10.1046/j.1365-313X.1996.10040755.x

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, et al. The genetic architecture of maize flowering time. Science. 2009:**325**(5941):714–718. https://doi.org/10.1126/science.1174276

Campa CC, Weisbach NR, Santinha AJ, Incarnato D, Platt RJ. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. Nat Methods. 2019:**16**(9):887–893. https://doi.org/10.1038/s41592-019-0508-6

Carpita NC, Gibeaut DM. Structural models of primary-cell walls in flowering plants—consistency of molecular-structure with the physical-properties of the walls during growth. Plant J. 1993:**3**(1):1–30. https://doi.org/10.1111/j.1365-313X.1993.tb00007.x

Cetin R, Wegner M, Luwisch L, Saud S, Achmedov T, Susser S, Vera-Gupti A, Muller K, Matthess Y, Quandt E, et al. Optimized metrics for orthogonal combinatorial CRISPR screens. Sci Rep. 2023:**13**(1):7405. https://doi.org/10.1038/s41598-023-34597-8

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al. Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome. Science. 2014:**345**(6199):950–953. https://doi.org/10.1126/science.1253435

Chapman BA, Bowers JE, Feltus FA, Paterson AH. Buffering crucial functions by paleologous duplicated genes may impart

cyclicality to angiosperm genome duplication. Proc Natl Acad Sci U S A. 2006:**103**(8):2730–2735. https://doi.org/10.1073/pnas.0507782103

Chaw SM, Liu YC, Wu YW, Wang HY, Lin CI, Wu CS, Ke HM, Chang LY, Hsu CY, Yang HT, et al. Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. Nat Plants. 2019:**5**(1):63–73. https://doi.org/10.1038/s41477-018-0337-0

Chen J, Hao Z, Guang X, Zhao C, Wang P, Xue L, Zhu Q, Yang L, Sheng Y, Zhou Y, et al. Liriodendron genome sheds light on angiosperm phylogeny and species-pair differentiation. Nat Plants. 2019:**5**(1):18–25. https://doi.org/10.1038/s41477-018-0323-6

Chen X, Li H, Pandey MK, Yang Q, Wang X, Garg V, Li H, Chi X, Doddamani D, Hong Y, et al. Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis and allergens. Proc Natl Acad Sci USA. 2016:**113**(24):6785–6790. https://doi.org/10.1073/pnas.1600899113

Chen ZJ, Pikaard CS. Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in Brassica. Proc Natl Acad Sci U S A. 1997a:**94**(7):3442–3447. https://doi.org/10.1073/pnas.94.7.3442

Chen ZJ, Pikaard CS. Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. Genes Dev. 1997b:**11**(16):2124–2136. https://doi.org/10.1101/gad.11.16.2124

Cheung J, Estivill X, Khaja R, MacDonald JR, Lau K, Tsui LC, Scherer SW. Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. Genome Biol. 2003a:**4**(4):R25. https://doi.org/10.1186/gb-2003-4-4-r25

Cheung J, Wilson MD, Zhang JJ, Khaja R, MacDonald JR, Heng HHQ, Koop BF, Scherer SW. Recent segmental and gene duplications in the mouse genome. Genome Biol. 2003b:**4**(8):R47. https://doi.org/10.1186/gb-2003-4-8-r47

Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S. High-throughput screening for induced point mutations. Plant Physiol. 2001:**126**(2):480–484. https://doi.org/10.1104/pp.126.2.480

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. Phenotypic instability and rapid gene silencing in newly formed arabidopsis allotetraploids. Plant Cell. 2000:**12**(9):1551–1567. https://doi.org/10.1105/tpc.12.9.1551

Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet. 2008:**9**(12):938–950. https://doi.org/10.1038/nrg2482

Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet. 2014:**46**(12):1311–1320. https://doi.org/10.1038/ng.3142

Cox S, Nabukalu P, Paterson AH, Kong W, Auckland S, Rainville L, Cox S, Wang S. High proportion of diploid hybrids produced by interspecific diploid × tetraploid Sorghum hybridization. Genet Resour Crop Evol. 2017:**65**:645–657. https://doi.org/10.1007/s10722-017-0562-9

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010:**107**(50):21931–21936. https://doi.org/10.1073/pnas.1016071107

Cuevas HE, Zhou C, Tang H, Khadke PP, Das SK, Lin YR, Ge Z, Clemente T, Upadhyaya HD, Hash CT, et al. The evolution of photoperiod-insensitive flowering in sorghum, a genomic model for panicoid grasses. Mol Biol Evol. 2016:**33**(9):2417–2428. https://doi.org/10.1093/molbev/msw120

Darwin C. On the origin of species. London: Murray; 1859.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011:**12**(7):499–510. https://doi.org/10.1038/nrg3012

Deng L, Zhou Q, Zhou J, Zhang Q, Jia Z, Zhu G, Cheng S, Cheng L, Yin C, Yang C, et al. 3D organization of regulatory elements for transcriptional regulation in Arabidopsis. Genome Biol. 2023:**24**(1):181. https://doi.org/10.1186/s13059-023-03018-4

D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, et al. The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. Nature. 2012:**488**(7410):213–217. https://doi.org/10.1038/nature11241

Doğan ES, Liu C. Three-dimensional chromatin packing and positioning of plant genomes. Nat Plants. 2018:**4**(8):521–529. https://doi.org/10.1038/s41477-018-0199-5

Domb K, Wang N, Hummel G, Liu C. Spatial features and functional implications of plant 3D genome organization. Annu Rev Plant Biol. 2022:**73**(1):173–200. https://doi.org/10.1146/annurev-arplant-102720-022810

Dong P, Tu X, Chu PY, Lü P, Zhu N, Grierson D, Du B, Li P, Zhong S. 3D chromatin architecture of large plant genomes determined by local A/B compartments. Mol Plant. 2017:**10**(12):1497–1509. https://doi.org/10.1016/j.molp.2017.11.005

Dorrity MW, Alexandre CM, Hamm MO, Vigil AL, Fields S, Queitsch C, Cuperus JT. The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. Nat Commun. 2021:**12**(1):3334. https://doi.org/10.1038/s41467-021-23675-y

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012:**489**(7414):57–74. https://doi.org/10.1038/nature11247

Erhard KF Jr., Talbot JE, Deans NC, McClish AE, Hollick JB. Nascent transcription affected by RNA polymerase IV in Zea mays. Genetics. 2015:**199**(4):1107–1125. https://doi.org/10.1534/genetics.115.174714

Ezawa K, Oota S, Saitou N. Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. Genome-wide search of gene conversions in duplicated genes of mouse and rat. Mol Biol Evol. 2006:**23**(5):927–940. https://doi.org/10.1093/molbev/msj093

Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM. Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics. 1997:**147**(3):1381–1387. https://doi.org/10.1093/genetics/147.3.1381

Feldmann KA, Marks MD, Christianson ML, Quatrano RS. A dwarf mutant of arabidopsis generated by T-Dna insertion mutagenesis. Science. 1989:**243**(4896):1351–1354. https://doi.org/10.1126/science.243.4896.1351

Fernández L, de Haro LA, Distefano AJ, Carolina Martínez M, Lía V, Papa JC, Olea I, Tosto D, Esteban Hopp H. Population genetics structure of glyphosate-resistant Johnsongrass (Sorghum halepense L. Pers) does not support a single origin of the resistance. Ecol Evol. 2013:**3**:3388–3400. https://doi.org/10.1002/ece3.67

Fladung M. Transposon activation tagging in plants for gene function discovery. In: **Botany PI, Luttge U, Canovas F, Matyssek R**, editors. Progress in Botany. vol. 77. Switzerland: Springer Cham; 2016. p. 265–289.

Flagel LE, Wendel JF. Gene duplication and evolutionary novelty in plants. New Phytol. 2009:**183**(3):557–564. https://doi.org/10.1111/j.1469-8137.2009.02923.x

Fontdevila A. Hybrid genome evolution by transposition. Cytogenet Genome Res. 2005:**110**(1–4):49–55. https://doi.org/10.1159/000084937

Force A, Cresko WA, Pickett FB, Proulx SR, Amemiya C, Lynch M. The origin of subfunctions and modular gene regulation. Genetics. 2005:**170**(1):433–446. https://doi.org/10.1534/genetics.104.027607

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999:**151**(4):1531–1545. https://doi.org/10.1093/genetics/151.4.1531

Freeling M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Ann Rev Plant Biol. 2009:**60**(1):433–453. https://doi.org/10.1146/annurev.arplant.043008.092122

Gaillochet C, Lohmann JU. The never-ending story: from pluripotency to plant developmental plasticity. Development. 2015:**142**(13): 2237–2249. https://doi.org/10.1242/dev.117614

Galtier N. Gene conversion drives GC content evolution in mammalian histones. Trends Genet. 2003:**19**(2):65–68. https://doi.org/10.1016/S0168-9525(02)00002-1

Gao LZ, Innan H. Very low gene duplication rate in the yeast genome. Science. 2004:**306**(5700):1367–1370. https://doi.org/10.1126/science.1102033

Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD, Jackson D, Leith A, Schreiber J, Noble WS, et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. Cell. 2019:**176**(1–2):377–390.e319. https://doi.org/10.1016/j.cell.2018.11.029

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. A draft sequence of the rice genome (Oryza sativa L. ssp japonica). Science. 2002:**296**(5565): 92–100. https://doi.org/10.1126/science.1068275

Gu ZL, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH. Role of duplicate genes in genetic robustness against null mutations. Nature. 2003:**421**(6918):63–66. https://doi.org/10.1038/nature01198

Guo H, Jiao Y, Tan X, Wang X, Huang X, Jin H, Paterson AH. Gene duplication and genetic innovation in cereal genomes. Genome Res. 2019:**29**(2):261–269. https://doi.org/10.1101/gr.237511.118

Guo H, Wang X, Gundlach H, Mayer KFX, Peterson DG, Scheffler BE, Chee PW, Paterson AH. Extensive and biased intergenomic non-reciprocal DNA exchanges shaped a nascent polyploid genome, gossypium (cotton). Genetics. 2014:**197**(4):1153–1163. https://doi.org/10.1534/genetics.114.166124

Harrison LJ, Bose D. Enhancer RNAs step forward: new insights into enhancer function. Development. 2022:**149**(16):dev200398. https://doi.org/10.1242/dev.200398

Heckmann D, Schulze S, Denton A, Westhoff P, Weber APM, Lercher MJ. Predicting C4 photosynthesis evolution: modular, individually adaptive steps on a mount fuji fitness landscape. Cell. 2013:**153**(7):1579–1588. https://doi.org/10.1016/j.cell.2013.04.058

Heger P, Wiehe T. New tools in the box: an evolutionary synopsis of chromatin insulators. Trends Genet. 2014:**30**(5):161–171. https://doi.org/10.1016/j.tig.2014.03.004

Henikoff S, Till BJ, Comai L. TILLING. Traditional mutagenesis meets functional genomics. Plant Physiol. 2004:**135**(2):630–636. https://doi.org/10.1104/pp.104.041061

Hetzel J, Duttke SH, Benner C, Chory J. Nascent RNA sequencing reveals distinct features in plant transcription. Proc Natl Acad Sci U S A. 2016:**113**(43):12316–12321. https://doi.org/10.1073/pnas.1603217113

Holliday R. Mechanism for gene conversion in fungi. Genet Res. 1964:**5**(2):282–304. https://doi.org/10.1017/S0016672300001233

Holliday R. Studies on mitotic gene conversion in ustilago. Genet Res. 1966:**8**(3):323–337. https://doi.org/10.1017/S0016672300010181

Howe HF, Smallwood J. Ecology of seed dispersal. Annu Rev Ecol Syst. 1982:**13**(1):201–228. https://doi.org/10.1146/annurev.es.13.110182.001221

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 2007:**449**(7161):463–467. https://doi.org/10.1038/nature06148

Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperus JT. Dynamics of gene expression in single root cells of Arabidopsis thaliana. Plant Cell. 2019:**31**(5):993–1011. https://doi.org/10.1105/tpc.18.00785

Jiang CX, Wright RJ, El-Zik KM, Paterson AH. Polyploid formation created unique avenues for response to selection in Gossypium (cotton). Proc Natl Acad Sci U S A. 1998:**95**(8):4419–4424. https://doi.org/10.1073/pnas.95.8.4419

Jiao Y, Li J, Tang H, Paterson AH. Integrated syntenic and phylogenomic analyses reveal a new ancient genome duplication in monocots. Plant Cell. 2014:**26**(7):2792–2802. https://doi.org/10.1105/tpc.114.127597

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. Ancestral polyploidy in seed plants and angiosperms. Nature. 2011:**473**(7345): 97–100. https://doi.org/10.1038/nature09916

Jores T, Hamm M, Cuperus JT, Queitsch C. Frontiers and techniques in plant gene regulation. Curr Opin Plant Biol. 2023:**75**:102403. https://doi.org/10.1016/j.pbi.2023.102403

Jores T, Tonnies J, Dorrity MW, Cuperus JT, Fields S, Queitsch C. Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves. Plant Cell. 2020:**32**(7):2120–2131. https://doi.org/10.1105/tpc.20.00155

Jores T, Tonnies J, Wrightsman T, Buckler ES, Cuperus JT, Fields S, Queitsch C. Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. Nat Plants. 2021:**7**(6):842–855. https://doi.org/10.1038/s41477-021-00932-y

Kashkush K, Feldman M, Levy AA. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics. 2002:**160**(4): 1651–1659. https://doi.org/10.1093/genetics/160.4.1651

Kashkush K, Feldman M, Levy AA. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet. 2003:**33**(1):102–106. https://doi.org/10.1038/ng1063

Kiedrzynski M, Zielinska KM, Jedrzecczyk I, Kiedrzynska E, Toomczyk PP, Rewicz A, Rewers M, Indreica A, Bednarska I, Stupar V, et al. Tetraploids expanded beyond the mountain niche of their diploid ancestors in the mixed-ploidy grass Festuca amethystina L. Sci Rep. 2021:**11**(1):18735. https://doi.org/10.1038/s41598-021-97767-6

Kim C, Wang X, Lee T-H, Jakob K, Lee G-J, Paterson AH. Comparative analysis of two saccharinae genera, miscanthus and saccharum reveals a shared whole-genome duplication but different evolutionary fate. Plant Cell. 2014:**26**(6):2420–2429. https://doi.org/10.1105/tpc.114.125583

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010:**465**(7295):182–187. https://doi.org/10.1038/nature09033

Kishimoto N, Higo H, Abe K, Arai S, Saito A, Higo K. Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of cDNA markers of known functions. Theor Appl Genet. 1994:**88**(6–7):722–726. https://doi.org/10.1007/BF01253976

Klein JC, Agarwal V, Inoue F, Keith A, Martin B, Kircher M, Ahituv N, Shendure J. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. Nat Methods. 2020:**17**(11):1083–1091. https://doi.org/10.1038/s41592-020-0965-y

Kong W, Liu M, Felker P, Ewens M, Bessega C, Pometti C, Wang J, Peng X, Teng J, Wang J, et al. Genome and evolution of Prosopis alba Griseb., a drought and salinity tolerant tree legume crop for arid climates. Plants People Planet. 2023:**5**(6):933–947. https://doi.org/10.1002/ppp3.10404

Kong Y, Cao L, Deikus G, Fan Y, Mead EA, Lai W, Zhang Y, Yong R, Sebra R, Wang H, et al. Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. Science. 2022:**375**(6580):515–522. https://doi.org/10.1126/science.abe7489

Kowalski S, Lan T-H, Feldmann K, Paterson A. Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved gene order. Genetics. 1994:**138**(2):499–510. https://doi.org/10.1093/genetics/138.2.499

Lander E, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989:**121**(1): 185–199. https://doi.org/10.1093/genetics/121.1.185

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing

and analysis of the human genome. Nature. 2001:**409**(6822):860–921. https://doi.org/10.1038/35057062

Lang L, Schnittger A. Endoreplication—a means to an end in cell growth and stress response. Curr Opin Plant Biol. 2020:**54**:85–92. https://doi.org/10.1016/j.pbi.2020.02.006

Langham RJ, Walsh J, Dunn M, Ko C, Goff SA, Freeling M. Genomic duplication, fractionation and the origin of regulatory novelty. Genetics. 2004:**166**(2):935–945. https://doi.org/10.1093/genetics/166.2.935

Lawrence WJC. The secondary association of chromosomes. Cytologia (Tokyo). 1931:**2**(4):352–384. https://doi.org/10.1508/cytologia.2.352

Lee HS, Chen ZJ. Protein-coding genes are epigenetically regulated in Arabidopsis polyploids. Proc Natl Acad Sci U S A. 2001:**98**(12):6753–6758. https://doi.org/10.1073/pnas.121064698

Levings PP, Bungert J. The human beta-globin locus control region. Eur J Biochem. 2002:**269**(6):1589–1599. https://doi.org/10.1046/j.1432-1327.2002.02797.x

Lewis DR, Olex AL, Lundy SR, Turkett WH, Fetrow JS, Muday GK. A kinetic analysis of the auxin transcriptome reveals cell wall remodeling proteins that modulate lateral root development in Arabidopsis. Plant Cell. 2013:**25**(9):3329–3346. https://doi.org/10.1105/tpc.113.114868

Li X, Lassner M, Zhang Y. Deleteagene: a fast neutron deletion mutagenesis-based gene knockout system for plants. Comp Funct Genomics. 2002:**3**(2):158–160. https://doi.org/10.1002/cfg.148

Li X, Song Y, Century K, Straight S, Ronald PC, Dong X, Lassner M, Zhang Y. A fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J. 2001:**27**(3):235–242. https://doi.org/10.1046/j.1365-313x.2001.01084.x

Lin Z, Li X, Shannon LM, Yeh C-T, Wang ML, Bai G, Peng Z, Li J, Trick HN, Clemente TE, et al. Parallel domestication of the Shattering1 genes in cereals. Nat Genet. 2012:**44**(6):720–724. https://doi.org/10.1038/ng.2281

Liu B, Vega JM, Feldman M. Rapid genomic changes in newly synthesized amphiploids of triticum and aegilops. II. changes in low-copy coding DNA sequences. Genome. 1998:**41**(4):535–542. https://doi.org/10.1139/g98-052

Liu C, Cheng YJ, Wang JW, Weigel D. Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. Nat Plants. 2017:**3**(9):742–748. https://doi.org/10.1038/s41477-017-0005-9

Liu S, Liu Y, Yang Y, Tong C, Edwards D, Parkin I, Zhao M, Ma J, Yu J, Huang S, et al. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun. 2014:**5**(1):3930. https://doi.org/10.1038/ncomms4930

Lozano R, Booth GT, Omar BY, Li B, Buckler ES, Lis JT, Del Carpio DP, Jannink JL. RNA polymerase mapping in plants identifies intergenic regulatory elements enriched in causal variants. G3 (Bethesda). 2021:**11**(11):jkab273. https://doi.org/10.1093/g3journal/jkab273

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015:**161**(5):1012–1025. https://doi.org/10.1016/j.cell.2015.04.004

Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000:**290**(5494):1151–1155. https://doi.org/10.1126/science.290.5494.1151

Lynch M, Force A. The probability of duplicate gene preservation by subfunctionalization. Genetics. 2000:**154**(1):459–473. https://doi.org/10.1093/genetics/154.1.459

Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza SL, Mazur EC, Emera D, Sheikh SZ, Grützner F, Bauersachs S, et al. Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. Cell Rep. 2015:**10**(4):551–561. https://doi.org/10.1016/j.celrep.2014.12.052

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci U S A. 2005:**102**(15):5454–5459. https://doi.org/10.1073/pnas.0501102102

Marand AP, Chen Z, Gallavotti A, Schmitz RJ. A cis-regulatory atlas in maize at single-cell resolution. Cell. 2021:**184**(11):3041–3055.e3021. https://doi.org/10.1016/j.cell.2021.04.014

Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, et al. A chromosome conformation capture ordered sequence of the barley genome. Nature. 2017:**544**(7651):427–433. https://doi.org/10.1038/nature22043

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012:**337**(6099):1190–1195. https://doi.org/10.1126/science.1222794

McCallum CM, Comai L, Greene EA, Henikoff S. Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol. 2000:**123**(2):439–442. https://doi.org/10.1104/pp.123.2.439

McCann MC, Roberts K. Architecture of the primary cell wall. In: **Lloyd CW**, editor. The cytoskeletal basis of plant growth and form. New York: Academic Press; 1991. p. 109–129.

McClintock B. The significance of responses of the genome to challenge. Science. 1984:**226**(4676):792–801. https://doi.org/10.1126/science.15739260

McDonald BR, Picard C, Brabb IM, Savenkova MI, Schmitz RJ, Jacobsen SE, and Duttke SH. Enhancers associated with unstable RNAs are rare in plants. bioRxiv 2023.09.25.559415. https://doi.org/10.1101/2023.09.25.559415, 2023, preprint: not peer reviewed.

Molin WT, Yaguchi A, Blenner M, Saski CA. The EccDNA replicon: a heritable, extranuclear vehicle that enables gene amplification and glyphosate resistance in amaranthus palmeri. Plant Cell. 2020:**32**(7):2132–2140. https://doi.org/10.1105/tpc.20.00099

Mondragon-Palomino M, Gaut BS. Gene conversion and the evolution of three leucine-rich repeat gene families in Arabidopsis thaliana. Mol Biol Evol. 2005:**22**(12):2444–2456. https://doi.org/10.1093/molbev/msi241

Montavon T, Duboule D. Chromatin organization and global regulation of Hox gene clusters. Philos Trans R Soc Lond B Biol Sci. 2013:**368**(1620):20120367. https://doi.org/10.1098/rstb.2012.0367

Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A, Kaul R, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020:**583**(7818):699–710. https://doi.org/10.1038/s41586-020-2493-4

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci U S A. 2013:**110**(2):453–458. https://doi.org/10.1073/pnas.1215985110

Nagamura Y, Inoue T, Antonio B, Shimano T, Kajiya H, Shomura A, Lin S, Kuboki Y, Harushima Y, Kurata N, et al. Conservation of duplicated segments between rice chromosomes 11 and 12. Breed Sci. 1995:**45**:373–376. https://doi.org/10.1270/JSBBS1951.45.373

Ohno S. Evolution by gene duplication. Berlin: Springer; 1970.

Ohta T. Some models of gene conversion for treating the evolution of multigene families. Genetics. 1984:**106**(3):517–528. https://doi.org/10.1093/genetics/106.3.517

Okhovat M, VanCampen J, Nevonen KA, Harshman L, Li W, Layman CE, Ward S, Herrera J, Wells J, Sheng RR, et al. TAD evolutionary and functional characterization reveals diversity in mammalian TAD boundary properties and function. Nat Commun. 2023:**14**(1):8111. https://doi.org/10.1038/s41467-023-43841-8

O'Neill RJW, O'Neill MJ, Graves JAM. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature. 2002:**420**:106–106. https://doi.org/10.1038/nature01162

Ozaki K, Ohnishi Y, Iida A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T. Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. Nat Genet. 2002:**32**(4):650–654. https://doi.org/10.1038/ng1047

Ozkan H, Levy AA, Feldman M. Allopolyploidy-induced rapid genome evolution in the wheat (Aegilops-Triticum) group. Plant Cell. 2001:**13**(8):1735–1747. https://doi.org/10.1105/tpc.010082

Ozkan H, Levy AA, Feldman M. Rapid differentiation of homeologous chromosomes in newly-formed allopolyploid wheat. Isr J Plant Sci. 2002:**50**(1):S65–S76. https://doi.org/10.1560/E282-PV55-G4XT-DRWJ

Paterson A, Bowers J, Burow M, Draye X, Elsik C, Jiang C, Katsar C, Lan T, Lin Y, Ming R, et al. Comparative genomics of plant chromosomes. Plant Cell. 2000:**12**(9):1523–1539. https://doi.org/10.1105/tpc.12.9.1523

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al. The Sorghum bicolor genome and the diversification of grasses. Nature. 2009:**457**(7229):551–556. https://doi.org/10.1038/nature07723

Paterson AH, Bowers JE, Chapman BA. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci U S A. 2004:**101**(26):9903–9908. https://doi.org/10.1073/pnas.0307901101

Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC. Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. Trends Genet. 2006:**22**(11):597–602. https://doi.org/10.1016/j.tig.2006.09.003

Paterson AH, Kong W, Johnston RM, Nabukalu P, Wu G, Poehlman W, Goff VH, Isaacs K, Lee T-H, Guo H, et al. The evolution of an invasive plant, Sorghum halepense L. ('johnsongrass'). Front Genet. 2020:**11**:317. https://doi.org/10.3389/fgene.2020.00317

Paterson AH, Lan TH, Reischmann KP, Chang C, Lin YR, Liu SC, Burow MD, Kowalski SP, Katsar CS, DelMonte TA, et al. Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. Nat Genet. 1996:**14**(4):380–382. https://doi.org/10.1038/ng1296-380

Paterson AH, Lin YR, Li ZK, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science. 1995:**269**(5231):1714–1718. https://doi.org/10.1126/science.269.5231.1714

Paterson AH, Wendel J, Gundlach F, Guo H, Jenkins H, Jin J, Llewellyn D, Showmaker D, Shu KC, Udall S, et al. Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature. 2012:**492**(7429):423–427. https://doi.org/10.1038/nature11798

Patterson EL, Pettinga DJ, Ravet K, Neve P, Gaines TA. Glyphosate resistance and EPSPS gene duplication: convergent evolution in multiple plant species. J Hered. 2018:**109**(2):117–125. https://doi.org/10.1093/jhered/esx087

Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat Biotechnol. 2009:**27**(12):1173–1175. https://doi.org/10.1038/nbt.1589

Petes TD, Symington LS. The molecular and cellular biology of the yeast Saccharomyces: genome dynamics, protein synthesis and energetics. In: **Broach JJ, Pringle J**, editors. Recombination in yeast. Cold Spring Harbor (NY): Cold Spring Harbor Press; 1991.

Pires JC, Zhao JW, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC. Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids (Brassicaceae). Biol J Linn Soc. 2004:**82**(4):675–688. https://doi.org/10.1111/j.1095-8312.2004.00350.x

Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. Mol Cell. 2018:**71**(5):858–871.e858. https://doi.org/10.1016/j.molcel.2018.06.044

Raven P, Evert R, Eichhorn S. Biology of plants. New York: Worth Publishers, Inc.; 1992.

Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-Tatché M, et al. Widespread long-range cis-regulatory elements in the maize genome. Nat Plants. 2019:**5**(12):1237–1249. https://doi.org/10.1038/s41477-019-0547-0

Rieseberg LH. Chromosomal rearrangements and speciation. Trends Ecol Evol. 2001:**16**(7):351–358. https://doi.org/10.1016/S0169-5347(01)02187-5

Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. Open chromatin reveals the functional maize genome. Proc Natl Acad Sci U S A. 2016:**113**(22):E3177-3184. https://doi.org/10.1073/pnas.1525244113

Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB. Engineering quantitative trait variation for crop improvement by genome editing. Cell. 2017:**171**(2):470–480.e478. https://doi.org/10.1016/j.cell.2017.08.030

Rogers JJW, Santosh M. Continents and supercontinents. Oxford: Oxford University Press; 2004.

Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012:**489**(7414):109–113. https://doi.org/10.1038/nature11279

Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isobe S, Kaneko T, Nakamura Y, Shibata D, Aoki K, et al. The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 2012:**485**(7400):635–641. https://doi.org/10.1038/nature11119

Sawyer S. Statistical tests for detecting gene conversion. Mol Biol Evol. 1989:**6**(5):526–538. https://doi.org/10.1093/oxfordjournals.molbev.a040567

Schluter U, Weber APM. Regulation and evolution of C4 photosynthesis. Annu Rev Plant Biol. 2020:**71**(1):183–215. https://doi.org/10.1146/annurev-arplant-042916-040915

Schmitz RJ, Grotewold E, Stam M. Cis-regulatory sequences in plants: their importance, discovery, and future challenges. Plant Cell. 2022:**34**(2):718–741. https://doi.org/10.1093/plcell/koab281

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2010:**463**(7278):178–183. https://doi.org/10.1038/nature08670

Schnable JC, Springer N, Freeling M. Biased gene loss following tetraploidy in maize reflects genome dominance and ongoing selection. Proc Natl Acad Sci U S A. 2010:**108**(10):4069–4074. https://doi.org/10.1073/pnas.1101368108

Schranz ME, Osborn TC. Novel flowering time variation in the resynthesized polyploid Brassica napus. J Hered. 2000:**91**(3):242–246. https://doi.org/10.1093/jhered/91.3.242

Schranz ME, Osborn TC. De novo variation in life-history traits and responses to growth conditions of resynthesized polyploid Brassica napus (Brassicaceae). Am J Bot. 2004:**91**(2):174–183. https://doi.org/10.3732/ajb.91.2.174

Semagn K, Babu R, Hearne S, Olsen M. Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed. 2013:**33**(1):1–14. https://doi.org/10.1007/s11032-013-9917-x

Semple C, Wolfe KH. Gene duplication and gene conversion in the caenorhabditis elegans genome. J Mol Evol. 1999:**48**(5):555–564. https://doi.org/10.1007/PL00006498

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell. 2001:**13**(8):1749–1759. https://doi.org/10.1105/TPC.010083

Shen C, Du H, Chen Z, Lu H, Zhu F, Chen H, Meng X, Liu Q, Liu P, Zheng L, et al. The chromosome-level genome sequence of the

autotetraploid alfalfa and resequencing of core germplasms provide genomic resources for alfalfa research. Mol Plant. 2020:**13**(9): 1250–1261. https://doi.org/10.1016/j.molp.2020.07.003

Shiu SH, Shih MC, Li WH. Transcription factor families have much higher expansion rates in plants than in animals. Plant Physiol. 2005:**139**(1):18–26. https://doi.org/10.1104/pp.105.065110

Singh R, Ong-Abdullah M, Low ET, Manaf MA, Rosli R, Nookiah R, Ooi LC, Ooi SE, Chan KL, Halim MA, et al. Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. Nature. 2013:**500**(7462):335–339. https://doi.org/10.1038/nature12309

Song KM, Lu P, Tang KL, Osborn TC. Rapid genome change in synthetic polyploids of brassica and its implications for polyploid evolution. Proc Natl Acad Sci U S A. 1995:**92**(17):7719–7723. https://doi.org/10.1073/pnas.92.17.7719

Song Q, Chen ZJ. Epigenetic and developmental regulation in plant polyploids. Curr Opin Plant Biol. 2015:**24**:101–109. https://doi.org/10.1016/j.pbi.2015.02.007

Stebbins G. Chromosomal variation and evolution; polyploidy and chromosome size and number shed light on evolutionary processes in higher plants. Science. 1966:**152**(3728):1463–1469. https://doi.org/10.1126/science.152.3728.1463

Stebbins GL. A brief summary of my ideas on evolution. Am J Bot. 1999:**86**(8):1207–1208. https://doi.org/10.2307/2656985

Stephens S. Possible significance of duplications in evolution. Adv Genet. 1951:**4**:247–265. https://doi.org/10.1016/S0065-2660(08)60237-0

Stergachis AB, Debo BM, Haugen E, Churchman LS, Stamatoyannopoulos JA. Single-molecule regulatory architectures captured by chromatin fiber sequencing. Science. 2020:**368**(6498): 1449–1454. https://doi.org/10.1126/science.aaz1646

Stergachis AB, Neph S, Reynolds A, Humbert R, Miller B, Paige SL, Vernot B, Cheng JB, Thurman RE, Sandstrom R, et al. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell. 2013:**154**(4):888–903. https://doi.org/10.1016/j.cell.2013.07.020

Studer A, Zhao Q, Ross-Ibarra J, Doebley J. Identification of a functional transposon insertion in the maize domestication gene tb1. Nat Genet. 2011:**43**(11):1160–1163. https://doi.org/10.1038/ng.942

Sullivan AM, Arsovski AA, Lempe J, Bubb KL, Weirauch MT, Sabo PJ, Sandstrom R, Thurman RE, Neph S, Reynolds AP, et al. Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. Cell Rep. 2014:**8**(6):2015–2030. https://doi.org/10.1016/j.celrep.2014.08.019

Sullivan AM, Arsovski AA, Thompson A, Sandstrom R, Thurman RE, Neph S, Johnson AK, Sullivan ST, Sabo PJ, Neri FV, et al. Mapping and dynamics of regulatory DNA in maturing arabidopsis thaliana siliques. Front Plant Sci. 2019:**10**:1434. https://doi.org/10.3389/fpls.2019.01434

Swigonova Z, Lai JS, Ma JX, Ramakrishna W, Llaca V, Bennetzen JL, Messing J. Close split of sorghum and maize genome progenitors. Genome Res. 2004:**14**(10a):1916–1923. https://doi.org/10.1101/gr.2332504

Tan Y, Yan X, Sun J, Wan J, Li X, Huang Y, Li L, Niu L, Hou C. Genome-wide enhancer identification by massively parallel reporter assay in Arabidopsis. Plant J. 2023:**116**(1):234–250. https://doi.org/10.1111/tpj.16373

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and colinearity in plant genomes. Science. 2008:**320**(5875):486–488. https://doi.org/10.1126/science.1153917

Tang H, Cuevas HE, Das S, Sezen UU, Zhou C, Guo H, Goff VH, Ge Z, Clemente TE, Paterson AH. Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication. Proc Natl Acad Sci U S A. 2013:**110**(39):15824–15829. https://doi.org/10.1073/pnas.1305213110

Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. Ann Rev Genet. 2004:**38**(1):615–643. https://doi.org/10.1146/annurev.genet.38.072902.092831

Thieffry A, Vigh ML, Bornholdt J, Ivanov M, Brodersen P, Sandelin A. Characterization of Arabidopsis thaliana promoter bidirectionality and antisense RNAs by inactivation of nuclear RNA decay pathways. Plant Cell. 2020:**32**(6):1845–1867. https://doi.org/10.1105/tpc.19.00815

Thomas S, Li XY, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE, et al. Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. Genome Biol. 2011:**12**(5):R43. https://doi.org/10.1186/gb-2011-12-5-r43

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. The accessible chromatin landscape of the human genome. Nature. 2012:**489**(7414):75–82. https://doi.org/10.1038/nature11232

Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Young K, Taylor NE, et al. Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res. 2003:**13**(3):524–530. https://doi.org/10.1101/gr.977903

Tuzun E, Bailey JA, Eichler EE. Recent segmental duplications in the working draft assembly of the brown Norway rat. Genome Res. 2004:**14**(4):493–506. https://doi.org/10.1101/gr.1907504

van den Berg C, Willemsen V, Hage W, Weisbeek P, Scheres B. Cell fate in the Arabidopsis root meristem determined by directional signalling. Nature. 1995:**378**(6552):62–65. https://doi.org/10.1038/378062a0

Vavilov N. The law of homologous series in variation. J Genet. 1922:**12**(1):1. https://doi.org/10.1007/BF02983073

Vision TJ, Brown DG, Tanksley SD. The origins of genomic duplications in Arabidopsis. Science. 2000:**290**(5499):2114–2117. https://doi.org/10.1126/science.290.5499.2114

Wang X, Gowik U, Tang H, Bowers JE, Westhoff P, Paterson AH. Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. Genome Biol. 2009:**10**(6):R68. https://doi.org/10.1186/gb-2009-10-6-r68

Wang X, Guo H, Wang J, Liu T, Lei T, Lee T-H, Li J, Tang H, Paterson AH. Comparative genomic de-convolution of the cotton genome revealed a decaploid ancestor and widespread chromosomal fractionation. New Phytol. 2016:**209**(3):1252–1263. https://doi.org/10.1111/nph.13689

Wang X, Tang H, Bowers JE, Feltus FA, Paterson AH. Extensive concerted evolution of rice paralogs and the road to regaining independence. Genetics. 2007:**177**(3):1753–1763. https://doi.org/10.1534/genetics.107.073197

Wang X, Tang H, Paterson AH. Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel in major Poaceae lineages. Plant Cell. 2011:**23**(1):27–37. https://doi.org/10.1105/tpc.110.080622

Wang X, Wang J, Guo H, Jin D, Lee T-H, Liu T, Paterson AH. Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. Mol Plant. 2015:**8**(6):885–898. https://doi.org/10.1016/j.molp.2015.04.004

Wang X, Wang Z, Guo H, Zhang L, Wang L, Li J, Jin D, Paterson AH. Telomere-centric genome repatterning determines recurring chromosome number reductions during the evolution of eukaryotes. New Phytol. 2014:**205**(1):378–389. https://doi.org/10.1111/nph.12985

Werth CR, Windham MD. A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. Am Nat. 1991:**137**(4):515–526. https://doi.org/10.1086/285180

White ME, Crowther BI. Gene conversions may obscure actin gene family relationships. J Mol Evol. 2000:**50**(2):170–174. https://doi.org/10.1007/s002399910018

Wu H, Nord AS, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA, Visel A. Tissue-specific RNA expression marks distant-acting developmental enhancers. PLoS Genet. 2014:**10**(9): e1004610. https://doi.org/10.1371/journal.pgen.1004610

Xi Q, Wang Z, Zaromytidou AI, Zhang XH, Chow-Tsang LF, Liu JX, Kim H, Barlas A, Manova-Todorova K, Kaartinen V, et al. A poised chromatin platform for TGF-β access to master regulators. Cell. 2011:**147**(7):1511–1524. https://doi.org/10.1016/j.cell.2011.11.032

Xu J-H, Messing J. Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. Proc Natl Acad Sci USA. 2008:**105**(38):14330–14335. https://doi.org/10.1073/pnas.0807026105

Younghusband F. The epic of Mount Everest. London: E.P. Publishing; 1926.

Yu JM, Holland JB, McMullen MD, Buckler ES. Genetic design and statistical power of nested association mapping in maize. Genetics. 2008:**178**(1):539–551. https://doi.org/10.1534/genetics.107.074245

Yuan Q, Gao X. Multiplex base- and prime-editing with drive-and-process CRISPR arrays. Nat Commun. 2022:**13**(1):2771. https://doi.org/10.1038/s41467-022-30514-1

Zhang W, Zhang T, Wu Y, Jiang J. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. Plant Cell. 2012:**24**(7):2719–2731. https://doi.org/10.1105/tpc.112.098061

Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang WC, Zhang L, Zhang X, Tang R, et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet. 2019:**51**(5):865–876. https://doi.org/10.1038/s41588-019-0402-2

Zou XH, Du YS, Tang L, Xu XW, Doyle JJ, Sang T, Ge S. Multiple origins of BBCC allopolyploid species in the rice genus (Oryza). Sci Rep. 2015:**5**(1):14876. https://doi.org/10.1038/srep14876