nature reviews bioengineering

https://doi.org/10.1038/s44222-025-00326-1

Review article

Check for updates

Genome synthesis in plants

Tianlong Lan[1,2,3], Lian-Ge Chen[4], Ying Wang[5] ✉ & Yuling Jiao[1,3,6] ✉

Abstract

Owing to advances in genome sequencing and editing, a genome can now be redesigned, synthesized and introduced into cells as desired. The field of synthetic genomics not only aims to provide deeper understanding of how the genome functions but can also be harnessed for a wide range of synthetic biology and bioengineering applications, from rapid evolution and screening for favourable strains to biotechnological and bioproduction tool development. Although genome synthesis has been carried out mainly in simple unicellular organisms, plants and animals are now also being investigated. Compared with animals, plants have unique advantages, such as fewer ethical concerns, simpler experimental operations and easier regeneration from cells to organisms. In this Review, we focus on genome synthesis in plants, discuss the current research landscape and assess possible future directions.

Sections

[1]State Key Laboratory of Gene Function and Modulation Research, School of Life Sciences, Peking University, Beijing, China. [2]Peking University–Tsinghua University–National Institute of Biological Sciences Joint Graduate Program, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [3]Peking–Tsinghua Center for Life Sciences, Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [4]Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing, China. [5]College of Life Sciences, Capital Normal University, Beijing, China. [6]Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Weifang, China. ✉e-mail: ying.wang@cnu.edu.cn; yuling.jiao@pku.edu.cn

Key points

- Advances in genome sequencing, synthesis and editing technologies have enabled systematic redesign and chemical synthesis of plant genomes, establishing a new research discipline for functional genome exploration with potential for agricultural innovation and industrial bioproduction.
- Top-down chromosomal engineering strategies focus on the modification of endogenous chromosomes, and bottom-up strategies use de novo DNA assembly techniques to generate artificial chromosomes with distinct structural and functional properties.
- De novo synthesis of plant chromosome segments follows a stage-gated workflow involving computational design, modular DNA assembly, transformation, targeted integration and verification, plant regeneration and phenotypic validating, with post-analysis data driving iterative optimization cycles for genome design refinement.
- Accurate integration of large DNA fragments remains a technical bottleneck in seed plant synthetic genomics, requiring breakthroughs in homology-directed repair or nonhomologous end joining-mediated integration efficiency and engineered recombinase systems to improve plant genome engineering platforms.

Introduction

The term 'genome' refers to the entire genetic material of an organism, including all the genes and non-coding sequences. Genomics, that is, the study of genomes[1], is the cornerstone of modern life science research, and enormous efforts have been devoted to understanding the genome. However, as Richard Phillips Feynman once said: 'What I cannot create, I do not understand', therefore, the ability to synthesize and redesign genomes could propel genomic research to a new stage. Genome synthesis has now become possible owing to advances in genome sequencing, reduced DNA synthesis cost[2] and technologies to assemble large-fragment DNA, which is often referred to as 'chunk'[3,4]. As a subfield of synthetic biology, synthetic genomics is expected to enable the reconstruction and reshaping of the genome to better understand it and to serve as the foundation for other synthetic biology applications. For example, viruses were first synthesized in the early 2000s[5,6] and whole prokaryote genomes can now be synthesized routinely[7–9]. Chromosomal fragments and even intact yeast chromosomes have been synthesized and shown to function in vivo[10]. Furthermore, all budding yeast chromosomes have been fully redesigned and synthesized[11–13].

A logical next step is to achieve genome synthesis for multicellular organisms, which would facilitate the understanding and application of more complex genomes, including those of plants and animals[14–16]. Synthetic genomics is an 'enabling technology' that facilitates the development or implementation of other technologies or processes. For example, it enabled the rewiring of the chromosomal structure, the regulatory circuits or even the genetic code of bacteria and yeasts. Similar changes can be induced in multicellular plants, such as reorganization and simplification of chromosomes, redesigning signalling pathways and expanding the genetic codon to introduce non-canonical amino acids, which would establish transgenic biosecurity systems. Moreover, given the essential role of plants in human sustenance and industrial systems, along with their lower ethical and biosafety constraints compared with animal models, plant synthetic genome technology offers great potential for agricultural innovation and biopharmaceutical development (Box 1). There are numerous crop traits that need to be improved or even redesigned; for example, orthogonal pathways, such as C4 photosynthesis and nitrogen fixation, can be introduced into plants that lack these beneficial traits. Synthetic genomics also supports molecular pharming to produce small molecules, therapeutic proteins, vaccines and other pharmaceuticals. Synthetic genomics even supports molecular-level hybridization of species to generate beneficial new hybrid species or rescue endangered species. However, genome synthesis for multicellular organisms (including plants) presents considerable challenges compared with unicellular organisms[17], such as large genome sizes[18], inefficient DNA transformation and difficulties in cell regeneration and reconstruction of the epigenetic landscape.

In this Review, we first discuss milestones in synthetic genomics and then shift our focus to plants, starting with the top-down approach, in which native chromosomes are trimmed to be used as potential vectors for biotechnologies. We follow up by covering the bottom-up approach, which covers genome design, large-fragment synthesis and assembly, large-DNA-fragment transformation, site-directed targeting and elimination of targeted genome segments. Finally, we discuss current bottlenecks that need to be addressed to enable broader applications of plant synthetic genomics, including inefficient large-fragment transformation, difficulties in site-directed targeting and the low regeneration rate of seed plants.

Development of synthetic genome research

Because oligonucleotides can be chemically synthesized to 'create' life de novo, researchers are now exploring the possibility of synthesizing artificial genomes. A 77-bp alanine tRNA-encoding gene was the first gene to be artificially synthesized in 1970 (ref. 19), followed by virus, prokaryote, single-celled eukaryote genomes and even parts of plant genomes from the early 2000s and over the span of two decades (Fig. 1).

Viruses

Viruses lack cellular structures, and their genetic information is encapsulated solely by their own capsid proteins. Their genomes are usually simple and short, and single-stranded viruses often have genomes on the order of 1 kb (ref. 20), making them excellent starting points for synthetic genomic experiments. In 2002, the genome of infectious poliovirus was synthesized by assembling DNA oligonucleotides[5], resulting in a single-stranded RNA genome ~7.5 kb in length[21]. This synthetic poliovirus cDNA was then transcribed, translated and replicated in a cell-free extract, ultimately resulting in the de novo synthesis of infectious poliovirus. This was the first synthetic genome, confirming the feasibility of chemically synthesizing genomes de novo without relying on natural templates. In 2003, a 5,386 bp bacteriophage was synthesized with a φX174 DNA genome. Sequential ligation and polymerase cycling reactions enabled single-step construction of the entire >5 kb DNA without cloning[6]. This landmark research substantially increased the efficiency of DNA synthesis, achieving a tenfold increase in speed compared with the prior synthesis of the poliovirus[5] and laying the foundation for the subsequent synthesis of bacterial genomes[22]. In the following years, more and larger virus genomes were designed and synthesized[23–25], proving useful in the development of next-generation vaccines.

Box 1 | Potential applications of synthetic genomics in agriculture and bioproduction

As a transformative platform for large-scale genome writing, large-fragment replacement technology enables seamless integration of complex genetic cassettes. These cassettes range from multigene clusters to entire metabolic pathways comprising hundreds of genes into crop genomes while circumventing interference from endogenous regulatory networks. This capability enables radical rewiring of plant physiology, including photosynthetic apparatus optimization, nitrogenase system engineering and deployment of synthetic biosensors through integrated sensing-execution circuits. Moreover, genome synthesis allows for developmental reprogramming via synthetic gene circuits to enhance agronomic traits. In bioproduction contexts, targeted chromosomal integration of heterologous pathways ensures position-independent expression stability and enables genetic pyramiding through precise multilocus stacking.

Synthetic genomic tools can further revolutionize chromosomal engineering through synthetic centromere-embedded neochromosomes, which function as high-capacity vector platforms for trait stacking. These synthetic chromosomes surpass conventional plasmid-based methods by enabling tissue-specific expression control, stable inheritance and dynamic pathway optimization via modular swapping. Advanced chromosomal manipulation techniques, including interspecific chromosome fusion and alien chromosome addition, generate new hybrid species with introgressed beneficial traits at the whole-chromosome level.

Synthetic genomes also serve as evolutionary sandboxes for trait discovery and customized bioproduction chassis. The 'Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution' (SCRaMbLE) system drives accelerated chromosomal diversification, generating stress-resilient phenotypes. Integrated biocontainment strategies harness non-canonical amino-acid-dependent auxotrophy and viral resistance mechanism through codon reassignment, thereby blocking viral translation hijacking while preventing ecosystem escape. In optimized chassis plants, non-canonical amino acid incorporation expands proteome functionality, whereas endogenous pathway redirection through systematic genome refactoring maximizes resource allocation towards targeted compound synthesis.

Prokaryotes

Prokaryotic organisms possess a simple cell structure, with genomes on the order of 10^6 bp. Compared with viruses, which often have fewer than 10 genes, prokaryotes have several hundred genes. In 2008, the first complete 582,970-bp *Mycoplasma genitalium* genome was synthesized. However, owing to the extremely slow growth rate of *M. genitalium*, the synthesized genome was ultimately not transferred into *M. genitalium* cells[26]. To overcome this limitation, a 1.08-Mb genome of the closely related but faster growing *Mycoplasma mycoides* was synthesized. This genome, which closely resembled the original wild-type genome, supported cell viability and enabled the creation of JCVI-syn1.0, the first cell line fully controlled by a synthetic genome[7]. In the subsequent JCVI-syn3.0, many non-essential genes were removed, reducing the number of genes from 901 to 473 to generate a minimal synthetic genome cell[9]. Although JCVI-syn3.0 retained replication capacity and wild-type-like colony morphology, its doubling time tripled compared with JCVI-syn1.0, and daughter cells exhibited morphological heterogeneity post-division[9]. Subsequent reintroduction of seven essential genes involved in cell division morphology corrected these defects[27], demonstrating power of synthetic genomics for functional gene exploration. Furthermore, synthetic genomics enables the systematic replacement of naturally redundant codons genome-wide, facilitating biological containment or non-canonical amino acids incorporation. In 2016, a 3.97-Mb *Escherichia coli* synthetic genome was designed using only 57 codons, with 7 codons replaced by synonymous alternatives[28]. In 2019, REXER (Replicon EXcision for Enhanced genome engineering through programmed Recombination)[29] — enabled single-step replacement of >100 kb genome fragments via lambda-Red-mediated recombination — was used to generate the fully synthetic *E. coli* 'Syn61', which contained only 61 codons[8].

Unicellular eukaryotes

Compared with prokaryotic organisms, eukaryotes possess chromosomal structures and more complex genomes. *Saccharomyces cerevisiae* was the first eukaryotic organism whose entire genome (~12 Mb with 16 chromosomes[30]) was sequenced owing to its strong recombination capability[3]. In 2011, the chromosomal arms IXR and VIL of *S. cerevisiae* were artificially synthesized and used to replace the endogenous arms. Subsequently, members of an international consortium comprising 21 institutions worldwide initiated the first eukaryotic genome synthesis project, Sc2.0 (ref. 11). After more than a decade of effort, in 2023, all 16 chromosomes of *S. cerevisiae*, along with an additional tRNA neochromosome, had been fully synthesized artificially and used to replace the natural genome[12,13,31–37]. In the Sc2.0 project, each chromosome was assembled via SwAP-In (switching auxotrophies progressively for integration), wherein iterative assembly and integration of synthetic DNA sequences with two selection markers enabled the assembly of large fragments. As all the chromosomes have been individually synthesized, the next step is to integrate all the synthetic chromosomes into a single cell to construct a yeast strain with a completely artificial genome[36]. Compared with prokaryotes, a major characteristic of the eukaryote genome is the presence of many intergenic regions and repetitive sequences. In the redesigned Sc2.0 synthetic chromosomes, many repetitive sequences, including transposons and introns known to be removable, have been deleted. This deletion not only provides a foundation for exploring the organizational mechanisms of yeast genome elements but also makes the stepwise replacement of DNA mega-chunk more efficient. Additionally, the design includes stop codon replacements to support the use of non-standard amino acids. Because tRNAs can induce genomic instability, including replication fork collapse[38,39] and Ty1 retrotransposon insertion upstream of tRNA loci[40], all native tRNA genes were deleted and their sequences redesigned in a dedicated tRNA neochromosome[37]. To enable inducible genome restructuring, multiple loxPsym sites were inserted near genes to construct the 'Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution' (SCRaMbLE) system. This system drives rapid, recombinase-dependent gene rearrangements or deletions upon induction[11]. The Sc3.0 project aims to synthesize a

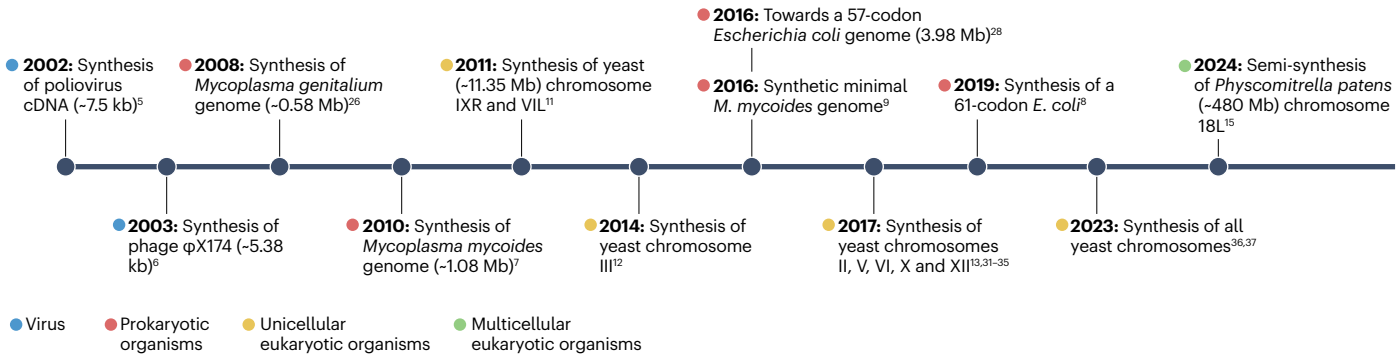

Fig. 1 | Timeline of milestones in synthetic genomics. Selected milestones in synthetic genomics over the past 20 years. The genome size of each species is shown in brackets. The blue, red, yellow and green points represent viruses, prokaryotes, unicellular eukaryotes and multicellular eukaryotes, respectively.

minimal yeast genome by using the SCRaMbLE system constructed in Sc2.0 to identify non-essential genes[41].

Multicellular eukaryotes

Compared with unicellular eukaryotes, multicellular eukaryotes typically have much larger genomes, usually ranging from 10^8 bp to 10^{10} bp, as well as more complex epigenetic regulation, which makes it challenging to perform transformation and directed recombination of large DNA fragments. Additionally, multicellular organisms often have long life cycles and are difficult to regenerate, which hinders genome synthesis. However, the synthesis of multicellular eukaryotes is highly desirable; on the one hand, it advances our understanding of the complex mechanisms of multicellular eukaryotic genomes, including heterochromatin formation, the establishment of epigenetic regulation, the role of intergenic regions and the evolution of genomic elements. On the other hand, large-scale tools for targeted assembly in multicellular eukaryotic genomes have research and industrial value. With the development of technologies for the synthesis, assembly and transformation of long nucleotide sequences into cells and targeting of these sequences to chromosomal loci, genome synthesis of multicellular organisms has become feasible. The Human Genome Project-Write was launched in 2016 with the goal of reducing the costs of engineering and testing large genomes within 10 years[16]. In 2023, a 'mammalian switching antibiotic resistance markers progressively for integration' (mSwAP-In) method was designed to facilitate the integration of large DNA constructs into mammalian genomes by progressively switching antibiotic resistance markers. mSwAP-In was used to rewrite up to 180 kb at three important disease loci in the mouse genome[14]. In 2024, the homologous recombination (HR) capability of *Physcomitrium patens* (a model moss species) was leveraged to achieve the artificial synthesis and replacement of genome fragments in plants for the first time[15]. By eliminating 55.8% of a 155 kb endogenous chromosomal region, the *P. patens* genome was substantially simplified without discernible phenotypic effects.

Artificial-chromosome research in plants

There are two approaches for synthesizing chromosomes in plants. The first involves using an organism's own chromosomal elements to generate new synthetic chromosomes via a 'top-down' approach. This strategy is typically geared towards practical applications, such as for generating stable vectors for transgenes, but also provides new insights into chromosomal organization. The second 'bottom-up' approach involves synthesizing chromosomes de novo (Fig. 2a). This method more closely aligns with synthetic genomic approaches for unicellular organisms and aims to understand life through synthesis. Owing to persistent technological challenges in synthesizing and precise targeting of large DNA fragments (>100 kb), the top-down engineering approach is often preferred.

Top-down approach

Currently, telomere-mediated chromosomal truncation (TMCT) is the primary method for generating top-down mini-chromosomes in plants, which provide a platform for stacking multiple genes[42] (Fig. 2b). Telomeres are specialized chromosomal structures typically composed of repetitive nucleotide sequences that prevent the degradation of chromosome ends and ensure proper chromosome segregation[43]. TMCT involves introducing telomere sequences into cells; the resulting insertion sites remain irreparable, leading to deletion of the distal chromosomal arm and generation of mini-chromosomes. TMCT was first applied to maize plants[44], which can harbour an extra chromosome known as the B chromosome. As an accessory element to the regular chromosome, a B chromosome is largely inert and features a nearly terminal centromere, rendering it an ideal substrate for constructing artificial mini-chromosomes[45]. By ligating telomere sequences, resistance selection markers and recombination sites on a plasmid, and introducing them into the genome via *Agrobacterium*-mediated transformation, lines containing truncated mini-chromosomes can be obtained through subsequent resistance screening and fluorescence in situ hybridization analysis[44]. Additional genes can be added stepwise via Cre-*lox* system, in which the P1 bacteriophage cyclization recombinase (Cre) mediates recombination between pairs of *loxP* sites or other recombinant systems[46]. In addition to maize, mini-chromosomes produced by TMCT have also been reported in other plants, such as *Arabidopsis*[47,48], rice[49] and barley[50]. Currently, the obtained mini-chromosomes are expected to serve as vectors for the transformation of large DNA fragments. However, many issues remain unaddressed, such as the concerns associated with the isolation and modification of the obtained mini-chromosomes.

Bottom-up approach

Compared with the top-down approach, the bottom-up approach aims at designing and synthesizing genomes from scratch. Although using human artificial chromosomes (HACs) with artificial centromeres can also be considered a bottom-up approach[51], here we discuss the

redesign and de novo synthesis of genomes on the basis of natural chromosomal frameworks. The bottom-up approach involves two distinct methods: the first is the assembly of the entire chromosome in vitro before introducing it into cells. The existing native chromosomes within the cells are then fragmented via methods such as Cas9 (Fig. 2c). However, owing to the large size of eukaryotic chromosomes, which typically span several tens of megabases, the current assembly and transformation technologies are inadequate to support this method. The second method is similar to the stepwise replacement used in Sc2.0, in which the synthetic genome sequence progressively replaces the native genome sequence (Fig. 2d). This method not only addresses the challenge of synthesizing excessively long nucleotide fragments but also enables real-time observation of the effects of replacing synthetic DNA on target cells.

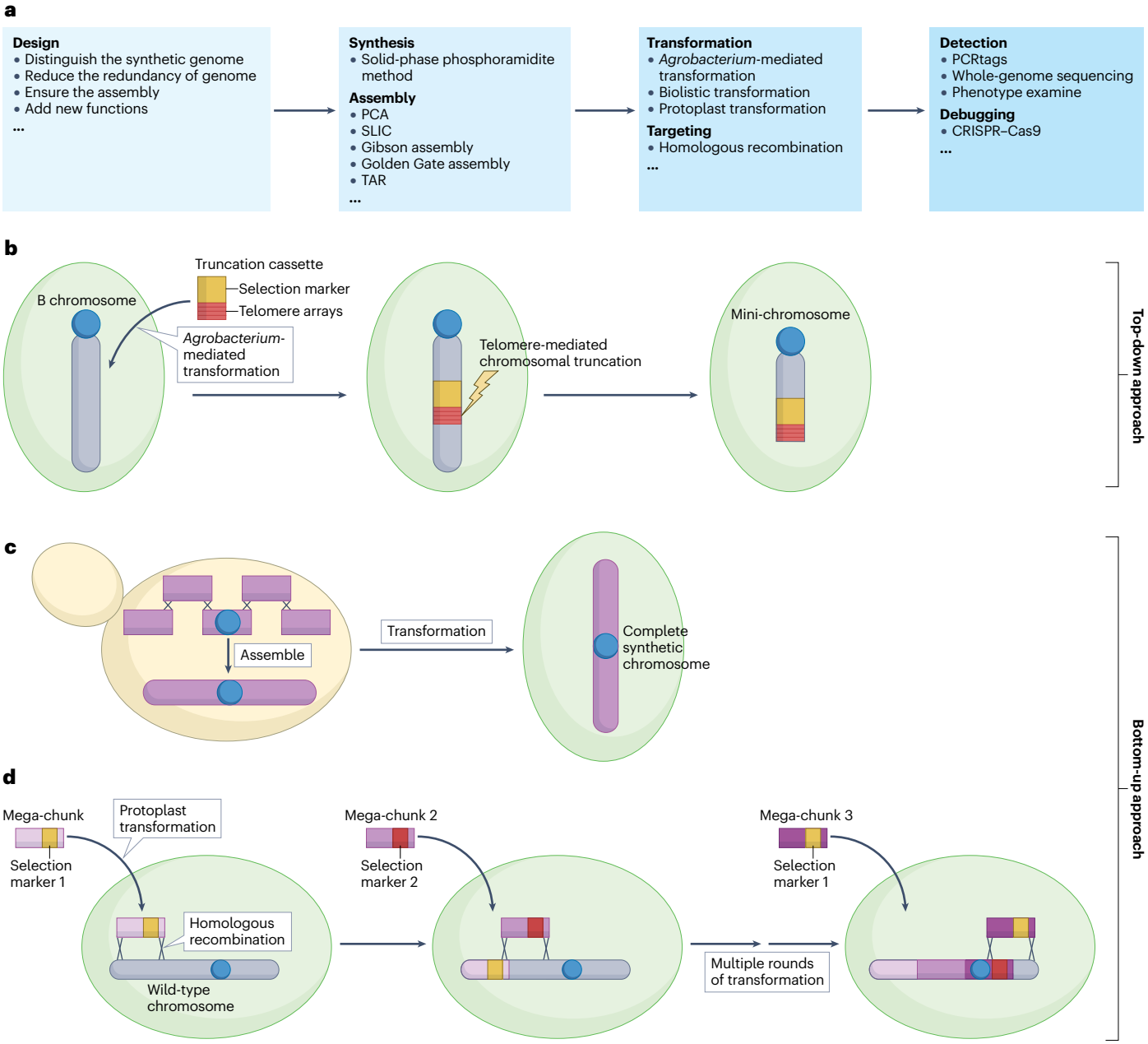

Fig. 2 | Strategies of chromosome synthesis for multicellular organisms. **a**, Technical workflow for bottom-up synthesis of plant genomes de novo. The process includes designing the synthetic genome, building the synthetic genome, transformation, targeting, detection and debugging. **b**, Creating new mini-chromosomes through telomere-mediated chromosomal truncation[44], that is, the top-down approach. The green symbols indicate in planta events. **c**, The complete chromosome is synthesized before being introduced into target cells (the bottom-up approach). Light-yellow graphics represent *Saccharomyces cerevisiae*. **d**, Stepwise replacement of the native genome, that is, the bottom–up approach. Mega-chunks of 10–100 kb are transformed at each step, with iterative selection of markers[13–15]. PCA, polymerase cycling assembly; SLIC, sequence-independent and ligation-independent cloning; TAR, transformation-associated recombination.

The bottom-up approach for chromosome synthesis typically involves several steps: designing and building the synthetic genome, transformation, targeting, detection and debugging (Fig. 2a). First, the design of the target genome is completed based on research needs, assembly requirements and application demands. Next, long fragments are synthesized according to the design. Typically, oligonucleotide segments longer than 100 bp are synthesized first, followed by multiple rounds of in vitro and yeast assembly to obtain nucleotide mega-chunks for transformation[52]. The length of the synthetic fragments used for each round of transformation can range from 10 kb to 100 kb. This length needs to be determined experimentally based on transformation and recombination requirements for the target species. Each round of transformation involves iterating on the resistance gene cassette of the previous segment until target replacement is completed. After each round of transformation, resistance screening, sequencing and phenotypic analysis are required to ensure that the synthetic fragment is fully replaced, that the expression levels of the synthetic genes are normal and that the synthetic genome does not cause substantial phenotypic changes.

The early land plant *P. patens* is a mature model organism for non-seed plants and is highly suitable for research on bottom-up plant chromosome synthesis[53,54]. As a widely used model for research on evolutionary development and cell biology[54,55], the genome of *P. patens* has been extensively sequenced and annotated[56]. *P. patens* possesses high-efficiency HR capabilities, enabling integration of exogenous fragments via homologous arms, thus providing a starting point for exploring large-scale fragment assembly[15,57]. *P. patens* also has high protoplast regeneration capacity, enabling the rapid acquisition of transgenic lines through protoplast transformation[58]. *P. patens* is involved primarily in haploid gametophyte generation, allowing gene manipulation of only one set of chromosomes. Additionally, plant genomes typically harbour a greater proportion of transposable elements compared with animals or microorganisms. *P. patens* exemplifies this trait, with transposable elements constituting ~60% of its genome[56], establishing it as a model system for studying genome simplification mechanisms. Moreover, it possesses a complex epigenetic landscape comparable to that of seed plants[59,60] and is an excellent chassis for the bioproduction of proteins and metabolites of research or industrial importance[61–63]. As an example, a 155 kb sequence was redesigned by standardizing all stop codons to TAA and removing partial intergenic regions, reducing the sequence to 68 kb. The complete designed sequence was assembled through polymerase cycling assembly (PCA) and two rounds of yeast-based assembly. Following protoplast transformation and HR, the designed sequence replaced the original genome, generating the synthetic line semi-syn 18L. Semi-syn 18L and the wild-type exhibited comparable phenotypes across all developmental stages, with no substantial differences in tolerance to high salinity or drought. Epigenetic landscapes, including histone modifications, were re-established de novo in the synthetic regions. This study established the foundation for the synthetic moss genome (SynMoss) project. Compared with that of *P. patens*, synthesizing the genomes of seed plants requires overcoming technical challenges related to the transformation and recombination of large DNA fragments.

Box 2 | Recommendations for special cases in plant genome design

Special situations may arise during genome design. For instance, if short tandem repeats appear within the coding sequence, recoding the sequence to reduce repetitiveness can facilitate synthesis and assembly. If the stop codon of the first gene overlaps with the coding sequence region of the second gene, it is advisable not to alter the stop codon or add loxPsym sites. Alternatively, splitting the two genes to better support the subsequent addition of non-canonical amino acids is another viable approach. If the two loxPsym sites are too close to each other, their contribution to recombination may be minimal. It is advisable to retain only one loxPsym site to ensure smooth synthesis and assembly. The genome contains a substantial amount of non-coding RNA and their redesign is also worth considering. tRNAs might influence genome stability and can be integrated into a new chromosome (similar to the approach used in Sc2.0), provided that synthetic centromere technology can support stable replication and transmission of the neochromosome. For rRNA tandem repeat arrays, the safest approach is to leave them unaltered. However, on the basis of the experience from Sc2.0, relocating rRNA tandem repeat arrays is feasible. Current understanding of other non-coding RNAs in plants is limited, and unless there is evidence indicating that a specific non-coding RNA gene has a function, it is generally not a primary focus during the design phase. However, it is important to consider any altered non-coding RNAs during subsequent testing and debugging if issues arise.

Designing synthetic genomes

Genome structures are complex, harbouring coding sequences, regulatory elements and various types of repetitive sequences. Currently, genome design relies on natural genome sequences as templates. Thus, high-quality genome sequences, ideally telomere-to-telomere sequences, need to be obtained. For the design of synthetic genomes, consistent design principles must be followed to obtain a standardized genome architecture for genome engineering (Box 2). Generally, the following aspects should be considered: how to distinguish between synthetic and natural genomes, how to reduce redundancy in the native genome, how to ensure successful assembly and how to insert new elements for new functions.

To distinguish the synthetic from the wild-type genome, PCRtags can be introduced without influencing gene expression (Fig. 3a). The designed PCRtag can replace a segment of 20–30 nucleotides in the coding region with synonymous codons, as used in Sc2.0 (ref. 64). Primers are designed to distinguish these synonymous replacements[11]. PCRtags can also be designed in intronic regions, named PCRmarks. To avoid disrupting mRNA splicing, PCRtags should be carefully positioned to avoid disrupting 5′ splice sites, 3′ splice sites or branch point sequences[64]. Standardizing all stop codons to 'TAA' can help distinguish between synthetic and wild-type genomes when the synthetic sequences are resequenced. Additionally, this standardization helps in studying translation termination in plant genes and supports the expansion of non-standard amino acid codons[28].

To reduce redundancy in the natural genome, considerations may include the removal of transposable elements and intergenic regions (Fig. 3b), which are abundant in plant genomes[65]. Transposable elements are genomic sequences that can change their location within the host genome. There are two different perspectives on the widespread presence of transposable elements in multicellular eukaryotic genomes. One hypothesis suggests that transposable elements

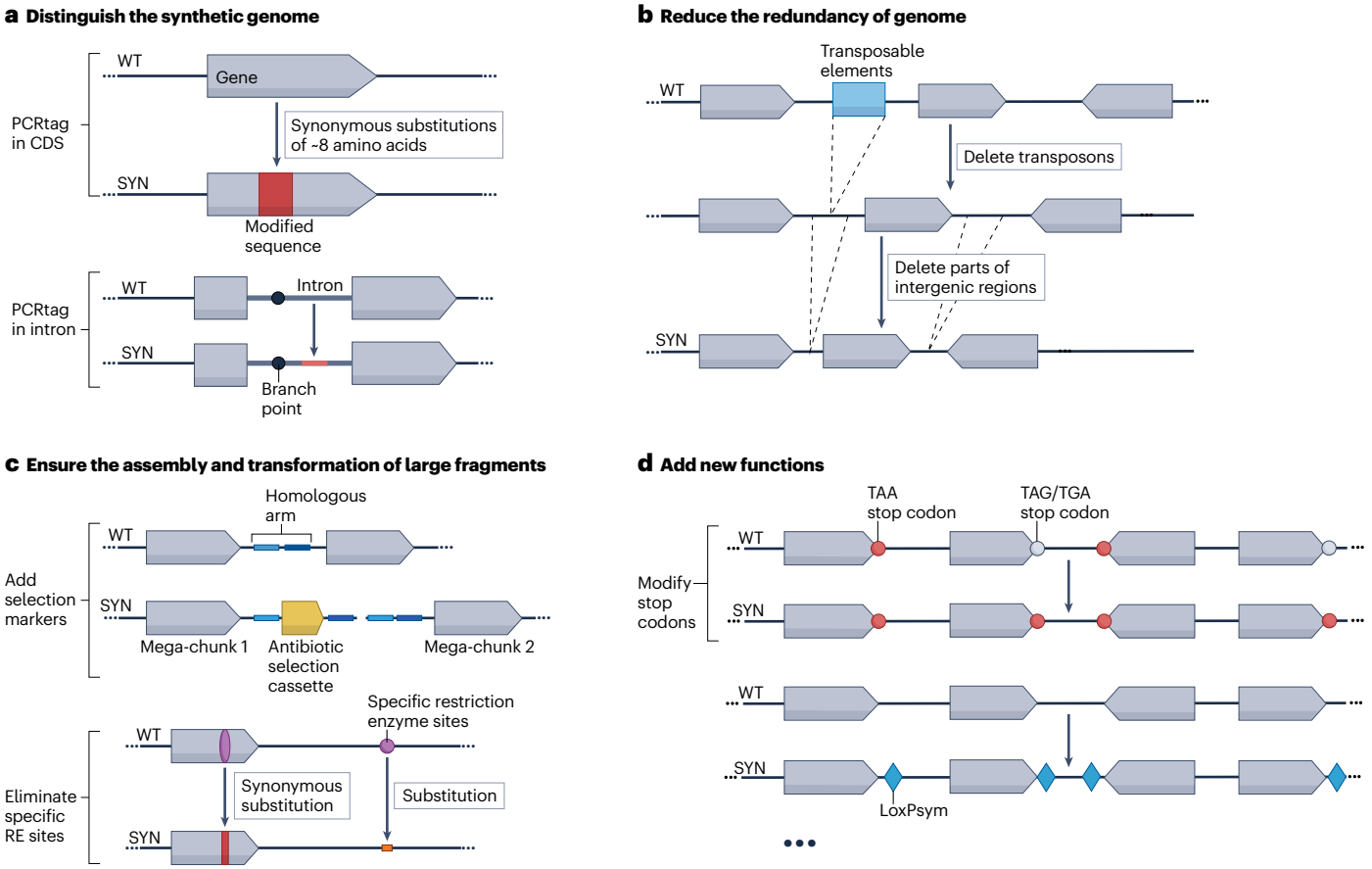

Fig. 3 | Designing synthetic plant genomes. a, To distinguish the synthetic genome, PCRtags can be introduced through synonymous codon substitutions in the coding sequence (CDS) or base substitutions in the intron regions. **b**, To reduce genome redundancy, transposable elements and non-conserved intergenic spacers can be removed, whereas essential *cis*-regulatory elements flanking code sequences need to be preserved. **c**, To ensure the assembly and transformation of large fragments, antibiotic selection cassettes are inserted between alternating homologous arms. Moreover, restriction enzyme (RE) sites used during the construction process are removed. **d**, New functions for the synthetic genome are added. All stop codons are standardized to TAA to facilitate the future incorporation of non-standard amino acids. LoxPsym sites are added downstream of genes to construct the Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution (SCRaMbLE) system. SYN, synthetic sequence; WT, wild-type sequence.

have minor detrimental effects, potentially causing genomic instability, but not to a degree sufficient for natural selection to effectively eliminate them[66,67]. By contrast, other studies have argued that transposable elements are essential for maintaining chromosomal integrity and for the survival of the organism[68]. Using synthetic genomic approaches to remove transposable elements at the chromosomal level in plants has shown that many of them are non-essential[15]. After they are removed, many repetitive sequences may still exist in intergenic regions, potentially influencing genome stability. In synthetic genome design, non-functional and redundant intergenic regions can be selectively deleted, preserving only coding sequences and essential upstream and downstream regulatory elements. Nevertheless, adaptability or even viability might become a problem after the removal of certain transposable elements or intergenic regions, which will need to be reintroduced into the synthetic genome. In Sc2.0, most of the introns in the yeast genome were removed[13]. However, many examples in which introns have regulatory roles in multicellular eukaryotes exist. For instance, intron retention can post-transcriptionally regulate gene expression, thereby enhancing transcriptome diversity and modulating diverse biological processes in mammals[69,70]. Moreover, introns improve gene expression by influencing transcription elongation or translation efficiency in eukaryotic systems[71]. To avoid introducing excessive variables, introns may be retained when simplifying plant genomes. Additionally, as demonstrated by JCVI-syn3.0 (ref. 9) and planned for Sc3.0 (ref. 41), removing non-essential genes can lead to the creation of a 'minimal genome'. However, the current understanding of non-essential genes in plants is limited, and there remains a long road ahead to attempt the synthesis of a minimal plant genome.

To ensure the successful assembly and replacement of large DNA segments, it is essential to include resistance gene cassettes and appropriate homologous arms when HR is used (Fig. 3c). Currently, synthetic genomes are often divided into chunks of varying sizes, with different chunks assembled through homologous sequences. Only by rationally dividing the synthetic genome, rapid and accurate genome assembly can be achieved. Generally, the boundaries of chunks should be placed in intergenic regions to avoid interrupting gene sequences

and to facilitate debugging[35]. Repetitive sequences in the homologous arms should be avoided to reduce the difficulty of assembly and the off-target rate during large-fragment recombination. Additionally, rotating plant antibiotic selection cassettes could be added to the ends of the mega-chunks used for transformation to facilitate selection in each round[15]. Alternatively, conditional marker genes that allow positive and negative selection in plants can also be applied. For example, D-amino acid oxidase, which detoxifies D-serine and catalyses the conversion of the non-toxic D-valine to a cytotoxic product, can serve as a dual-function selection marker in plants[72]. Restriction enzyme sites that might interfere with experimental workflows, such as those used for post-assembly plasmid linearization or Type IIS sites critical for Golden Gate assembly, should be systematically removed during construct design.

When designing a genome, the addition of new sequences might endow the genome with new functions (Fig. 3d). In Sc2.0, the loxPsym sequence was introduced 3 bp downstream of the stop codon of each gene to facilitate the activity of the SCRaMbLE system. The use of Cre recombinase to catalyse recombination between loxPsym sites enables the induction of various genomic rearrangements or gene deletions at the whole-genome level, achieving rapid evolution[73]. The *P. patens* synthetic chromosome project SynMoss also incorporates similar designs[64]. Another example involves the incorporation of an inducible Cas9 gene as a component of the CRISPR–Cas9 system. This addition can facilitate subsequent large-scale recombination or targeted gene editing.

The genomes of multicellular organisms are typically in the 0.1–10 Gb range and consist of many chromosomes. Synthesizing the complete genome of a multicellular organism requires a unified standard in chassis genome design and often involves collaboration among different groups. The Sc2.0 project developed the BioStudio program[13], which enables teams from different countries to design their chromosomes using the same criteria. For plant synthetic genomics, a public design platform called GenoDesigner exists[64], which offers an intuitive graphical interface, enabling users to manipulate a wide range of genomic sequences. The aim is to accelerate the synthesis of the *P. patens* genome, providing a test run and roadmap for the synthesis of other plant genomes[64].

Design and synthesis of functional centromeres and telomeres

In plant chromosome synthesis, synthesizing functional centromeres that support mitosis and meiosis is an important step to ensure chromosome replication and transmission to daughter cells. During mitosis and meiosis, spindle fibres bind to the centromeric regions of chromosomes, facilitating the even distribution of chromosomes to daughter cells[74]. Most plants possess monocentric chromosomes, with centromeres typically surrounded by pericentromeric heterochromatin regions that are devoid of genes. Generally, the DNA sequences in the centromere core domain of plants contain extensive tandem repeats, consisting primarily of satellite DNA and retrotransposons[75]. Centromeric regions possess specific epigenetic markers, particularly CENH3 (also referred to as CENP-A). CENH3 is a centromere-specific histone H3 variant that primarily resides in centromeric nucleosomes, partially replacing the canonical histone H3 (refs. 74,76). CENH3 is crucial for centromere function and assembly, making it the primary epigenetic marker for determining centromere identity[77].

Unlike most other eukaryotes, *S. cerevisiae* has unique point centromeres, which consist of an essential and conserved 125-bp sequence and whose function fully depends on primary sequences. In Sc2.0, synthetic centromeres could use the original yeast centromere sequences to obtain functional centromeres. For example, the new tRNA neochromosome uses the centromere sequence from yeast chromosome VI, which is sufficient to support the segregation of synthetic yeast chromosomes[37]. HACs are formed de novo by introducing large centromeric sequences consisting of highly repeated 171-bp alpha satellite (alphoid) DNA into HT1080 cells[78]. However, owing to the difficulty of synthesizing and assembling megabase-level tandem repeat sequences, new methods have been developed to obtain synthetic centromeres for HACs de novo. For example, a rolling circle amplification transformation-associated recombination (RCA-TAR) method has been used for the rapid synthesis and modification of alphoid sequences[79,80]. When artificial centromeres are constructed via the RCA-TAR method, tetracycline operator (tetO) is used to replace the CENP-B binding box, enabling the tetR fusion protein to be targeted to the alphoid array containing tetO sequences[51,81,82]. HAC centromeres can also bypass the alphoid array and directly use LacO sequences to recruit LacI-HJURP, in which HJURP is a centromeric chromatin assembly protein, to form artificial centromeres[83]. HACs formed via this method are stably inherited in subsequent lineages.

In plants, simply copying natural plant centromere repeats to synthesize plant centromere sequences seems to be ineffective[42,84]. The formation of centromere sequences might not require centromere DNA sequences, as the mere presence of these does not result in the formation of a functional centromere[85]. The function of centromeres primarily depends on the binding of CENH3 and other centromere proteins[86]. For example, LexA/LexO or LacI/LacO can be used to recruit CENH3 or other centromere proteins to construct synthetic centromeres in plants[87,88]. In *Arabidopsis thaliana*, recombinant proteins comprising the kinetochore component CENH3 fused with LacI can recruit other kinetochore components to non-centromeric LacO insertion sites[87]. In maize, the LexA–CENH3 fusion protein was used to recruit native CENH3 to long arrays of LexO repeats on a chromosome arm. The newly recruited CENH3 generated functional centromeres, causing chromosome breakage, and the released chromosome fragments could be transmitted to progeny through meiosis[88] (Fig. 4a). Although small chromosomes with synthetic centromeres were ultimately produced in maize, these newly generated chromosomes were not completely stable in somatic cells. This instability might be due to centromere-mediated genome elimination in the embryo, which leads to the removal of small or defective centromeres. In the future, the use of large-fragment replacement technology to replace the original centromere sequences with synthetic LexO repeats is expected to yield fully stable chromosomes with artificial centromeres (Fig. 4b). The length of centromeric regions in plants is often correlated with the length of the chromosome[89,90], and different numbers of LexO repeats need to be tested to find an appropriate length for functional synthetic centromeres.

The functional telomere is also an important part of plant chromosome synthesis and can prevent the degradation of chromosome ends and ensure proper chromosome segregation. Similar to those in other eukaryotes, plant telomeres consist of tandem repeat sequences that are generally conserved[91], typically $(TTTAGGG)_n$. Telomerase is a ribonucleoprotein complex composed of RNA and protein. It is a reverse transcriptase that can synthesize telomere sequences at chromosome ends using its own RNA as a template[92]. A segment of telomere repeat sequences at the chromosome end is required, and sufficient, as a 'seed', enabling telomerase to form full-length telomeres (Fig. 4c).

For example, top-down artificial mini-chromosomes in maize were generated by inserting a 2.6 kb telomere repeat from *A. thaliana*, which resulted in chromosome truncation. These artificial chromosomes have telomere seeds at their ends, which can facilitate the de novo establishment of stable new telomeres with the assistance of telomerase[44]. When designing synthetic plant chromosomes, choosing telomere seeds of appropriate length is a key consideration. These sequences must be neither excessively long (exceeding 10 kb, which complicates assembly) nor overly short — for instance, the shortest functional centromere in *A. thaliana* cannot be shorter than 0.3 kb (ref. 93) — to ensure stable neotelomere formation.

Synthesis and assembly of large fragments

After the overall blueprint of the synthetic genome is obtained, the next step is to synthesize and assemble the designed genome fragments. Given the substantial costs and time investments required for large-fragment assembly and transformation, iterative transformation of ~100-kb-level mega-chunks is the preferred strategy in *P. patens*. Regardless of the assembly method used, the initial step involves synthesizing oligonucleotide fragments via chemical methods such as solid-phase phosphoramidite[94,95]. This synthesis method extends the DNA chain from the 3′ end to the 5′ end, adding nucleotides with dimethoxytrityl protecting groups each time. Through a cyclical series of steps, including deprotection (detritylation), activation, coupling, oxidation and capping, oligonucleotide fragments longer than 150 bp can be synthesized. However, the throughput of this method remains relatively low (typically limited to several hundred sequences) compared with high-throughput oligonucleotide synthesis platforms. The latter can increase the yield of oligonucleotides and reduce synthesis costs, especially when using microarrays to prepare tens of thousands of oligonucleotides simultaneously in a mixed format. However, array-based synthesis exhibits inherent limitations, including high error rates (typically >1%) and low oligonucleotide yields owing to suboptimal monomer concentrations. Furthermore, partitioning the resulting oligonucleotide pools, which contain great sequence diversity, into functional subpools remains technically challenging. These limitations must be addressed to enable broader adoption of this technology[10,96,97]. In parallel, template-free enzymatic oligonucleotide synthesis using terminal deoxynucleotidyl transferase (TdT) shows promise for replacing conventional chemical methods. TdT, a DNA polymerase that sequentially adds deoxynucleotides (dNTPs) to oligonucleotide 3′ ends without requiring a template, enables rapid synthesis, with TdT–dNTP conjugates achieving quantitative single-nucleotide primer extension within 10–20 s (refs. 98,99).

The next step is to assemble the ~150-bp oligonucleotide fragments obtained by chemical synthesis into larger fragments, either in vitro or in vivo, to obtain hundred kilobase-level fragments. Typically, small DNA fragments are first assembled seamlessly via in vitro assembly methods, including PCA, sequence-independent and ligation-independent cloning (SLIC), Gibson assembly and Golden Gate assembly. PCA, which is based on PCR, can rapidly connect dozens of partially overlapping oligonucleotide chains together and is especially useful in the first-stage assembly. Kilobase-scale fragments generated via PCA assembly can be further assembled into larger constructs using in vitro seamless cloning methods. The SLIC assembly method uses the exonuclease activity of T4 DNA polymerase to generate single-stranded DNA ends. These single strands are then annealed to connect homologous sequences, enabling 5–10 DNA fragments to be joined in a single reaction[100]. Although SLIC requires only T4 DNA polymerase, which reduces costs, the gaps introduced during the in vivo annealing repair process in *E. coli* complicates the workflow. Gibson assembly is a technique for in vitro, one-step joining of multiple DNA fragments that requires 15–25 bp of overlap between two neighbouring fragments. Through the concerted actions of a 5′ exonuclease, a DNA polymerase and a DNA ligase, multiple fragments can be seamlessly connected[101]. This method is operationally simple, typically joining 2–6 fragments per reaction. Golden Gate assembly uses Type IIS restriction enzymes (such as BsaI and AarI) that cleave outside their recognition sites to generate unique sticky ends, enabling simultaneous, seamless assembly of up to 20 DNA fragments[102]. However, Golden Gate requires careful avoidance of endogenous Type IIS

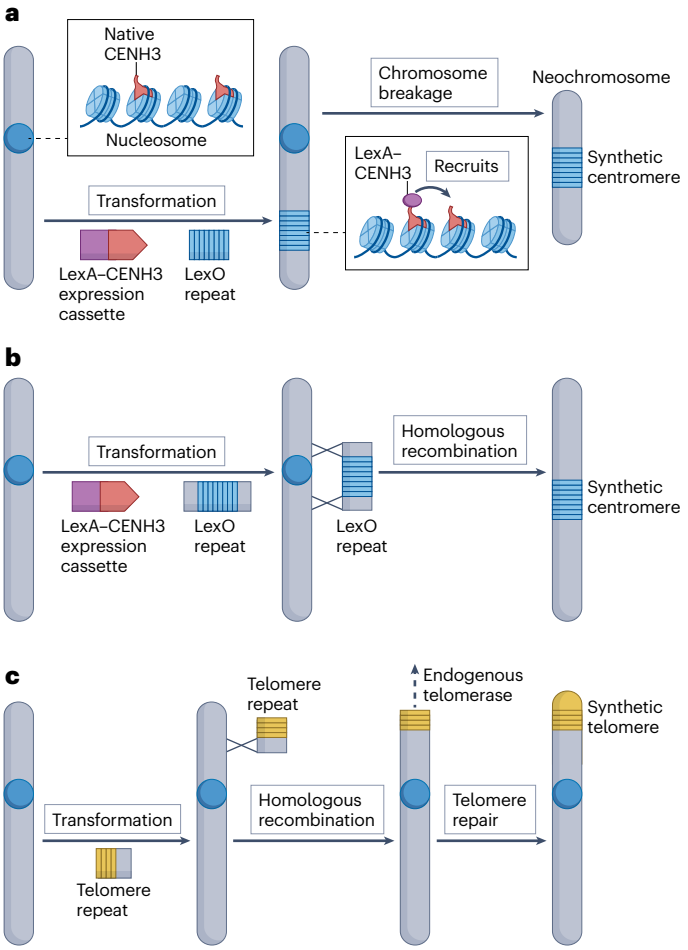

Fig. 4 | Artificial plant centromeres. a, Insertion of LexO repeats into the genome with *trans*-acting LexA–CENH3 to create synthetic centromeres. Epigenetic landscape of native centromeres and synthetic centromeres are illustrated. The LexO repeat array recruits LexA–CENH3, further recruiting native CENH3 to form centromeres. Dicentric chromosomes can cause chromosome breakage, leading to the formation of neochromosomes. **b**, Large-fragment replacement technology replaces the native centromere with a LexO repeat array, in which chromosome breakage can be avoided. **c**, Synthesis of plant telomeres by replacing the native chromosome ends with large fragments containing telomere repeats of appropriate length. The telomere repeat sequences act as 'seeds' for telomere formation, which are then extended into functional telomeres under the action of endogenous telomerase.

sites during sequence design. Selection among these in vitro assembly strategies depends on experimental requirements, including fragment number, size and cost constraints.

When the sequence to be assembled is greater than 10 kb in size, in vitro assembly becomes exceptionally cumbersome and inefficient. *S. cerevisiae* has strong HR capabilities, which enables it to connect multiple DNA fragments with overlapping sequences (at least 40 bp) through HR. This ability makes budding yeast an ideal host for the assembly of long DNA fragments[103]. TAR in *S. cerevisiae* is widely used in various synthetic biology and synthetic genomics studies[26,104]. However, yeast-based large-scale multiround assembly remains time-consuming. Each round of assembly requires the extraction of plasmids from yeasts and their transformation to *E. coli* for amplification, and multiple rounds of assembly are often required to obtain the desired length[52]. To overcome some of these limitations, a yeast life cycle-based assembly method was developed that enables in vivo iterative assembly of large DNA by controlled transfer of assembled sequences in the cycle of yeast mating and sporulation[4]. This method avoids extensive in vitro handling of large DNA fragments and is expected to reduce the operational and time costs of large-fragment assembly. In summary, improving the efficiency of TAR assembly holds promise for lowering the time costs of genome synthesis.

Transformation of long DNA fragments

Traditional methods for transferring large fragments into plant cells rely on *Agrobacterium*-mediated transformation[105,106] or biolistic transformation[107,108]. *Agrobacterium* is a plant pathogen capable of infecting wounded sites of most seed plants and integrating transfer DNA randomly into the plant genome. *Agrobacterium*-mediated transformation is convenient but the integration sites are usually random and is often limited to up to 150 kb in size for transfer DNA[109].

Biolistic transformation involves the binding of naked DNA to tungsten microparticles, which are only a few micrometres in diameter, and then shooting them into plant tissue via a particle gun[107]. Biolistic transformation is nearly unrestricted by plant species; however, it carries the risk of interrupting endogenous and exogenous genes[110], which is detrimental to the large-scale fragment transformations required in synthetic genome research. The transformed cells need to regenerate into new plants, and both *Agrobacterium*-mediated transformation and biolistic transformation commonly target calluses derived from explant-based regeneration, in which vasculature stem cells transdifferentiate into root-like calluses through the adventitious root formation pathways. The ability and efficiency of regeneration are often species-specific and genotype-specific and can be a limiting factor for transformation[111].

Given the random-insertion nature and length constraints of *Agrobacterium*-mediated transformation and the substantial risk of genomic damage associated with biolistic transformation, these approaches are suboptimal for iterative large-fragment delivery in planta. Owing to its capacity for large fragments and compatibility with controllable HR, protoplast transformation might be appropriate for in planta assembly of synthetic sequences. Plant protoplast isolation can be achieved by enzymatic digestion of cell walls using a cellulase-pectinase solution isotonic or slightly hypertonic relative to cellular osmolarity (typically supplemented with sorbitol or mannitol). Polyethylene glycol (PEG)-mediated protoplast transformation is a simple and commonly used transfection method that leverages membrane permeability-enhancing properties of PEG to deliver exogenous DNA into protoplasts with high efficiency (>30%)[112]. Notably, protoplast transformation exhibits compatibility with large DNA fragments, evidenced by the delivery of ~69 kb fragments in *P. patens*[15]. In addition to PEG-mediated protoplast transformation, protoplasts might also be amenable to yeast cell fusion for megabase-scale large-fragment delivery while circumventing the laborious purification of large DNA fragments, as used in mammalian cells[113,114]. Despite being a developed technology, protoplast regeneration is challenging in seed plants; unlike *P. patens*, whose protoplasts regenerate directly into protonema and almost completely bypass the callus stage, seed plant protoplasts regenerate into calli and then into plantlets or somatic embryos, which is time-consuming and inefficient[115–117]. Thus, protoplast regeneration is likely to be a challenge for genome synthesis in seed plants.

To overcome the low regeneration efficiency of protoplasts in most seed plants, implementing developmental regulators, which are critical genes governing embryogenesis, and meristem functions, can substantially enhance regeneration capacity. This approach, previously validated in explant-based regeneration systems[118–121], shows similar potential for improving protoplast regeneration. For example, shoot meristem regulatory genes, including *DORNRÖSCHEN* and *WUSCHEL*, promote protoplast regeneration in the model plant *Arabidopsis*[122]. Moreover, REGENERATION FACTOR 1 (REF1), a local wound signal peptide, promotes explant-based regeneration in multiple species[123], raising the question of whether this peptide also enhances protoplast regeneration. Chemical stimulation can also induce reprogramming; adding the histone deacetylase inhibitor trichostatin A, which increases chromatin accessibility, enhances protoplast regeneration efficiency in *Arabidopsis*[122]. Accelerating protoplast regeneration is also highly desirable to facilitate multiround transformations required for in vivo genome replacement or targeting.

Targeting of synthetic DNA fragments

After large fragments are transformed into plant cells, these fragments must also be integrated at designed locations, which are often accompanied by the replacement of endogenous natural genome sequences with the synthetic sequences. Given the challenge of targeting synthetic large fragments in plants and based on experiences with genome synthesis in yeast and *P. patens*, HR is an effective approach. *P. patens* has high HR efficiency (>90% gene targeting rates[57]), enabling direct integration of synthetic fragments using 1 kb homologous arms[15] (Fig. 5a). Nevertheless, HR-mediated replacement of large (~100 kb) fragments remains inefficient (<1%)[124,125].

Notably, site-specific double-stranded breaks (DSBs) generated by endonucleases can substantially increase targeting efficiency[126,127]. For example, the CRISPR–Cas9 system uses single-guide RNA to direct Cas nucleases to specific genomic loci for targeted DNA cleavage[128,129]. Following Cas9-induced targeted DSBs, homology-directed repair (HDR) using exogenous repair templates facilitates precise mutations or short sequence integration, albeit with low efficiency: 0.2% in rice[125] and 0.8% in *Arabidopsis*[130]. By contrast, Cas9-mediated DSBs near homologous arms enabled efficient HDR of ~100 kb fragments in mice, achieving integration rates exceeding 15% (ref. 14). However, HDR is often constrained by competitive nonhomologous end joining (NHEJ); therefore, variants of Cas9, such as Cas9 (D10A), have been used to generate single-strand breaks only[131], or vCas9 to create misaligned sticky ends[132] to reduce NHEJ. Fusing 5′ exonucleases with Cas endonucleases facilitates the generation of long single-stranded 3′ overhangs at DSB sites, thereby enhancing HDR. This approach has enabled scar-free insertion of several kilobases in *Arabidopsis*[133].

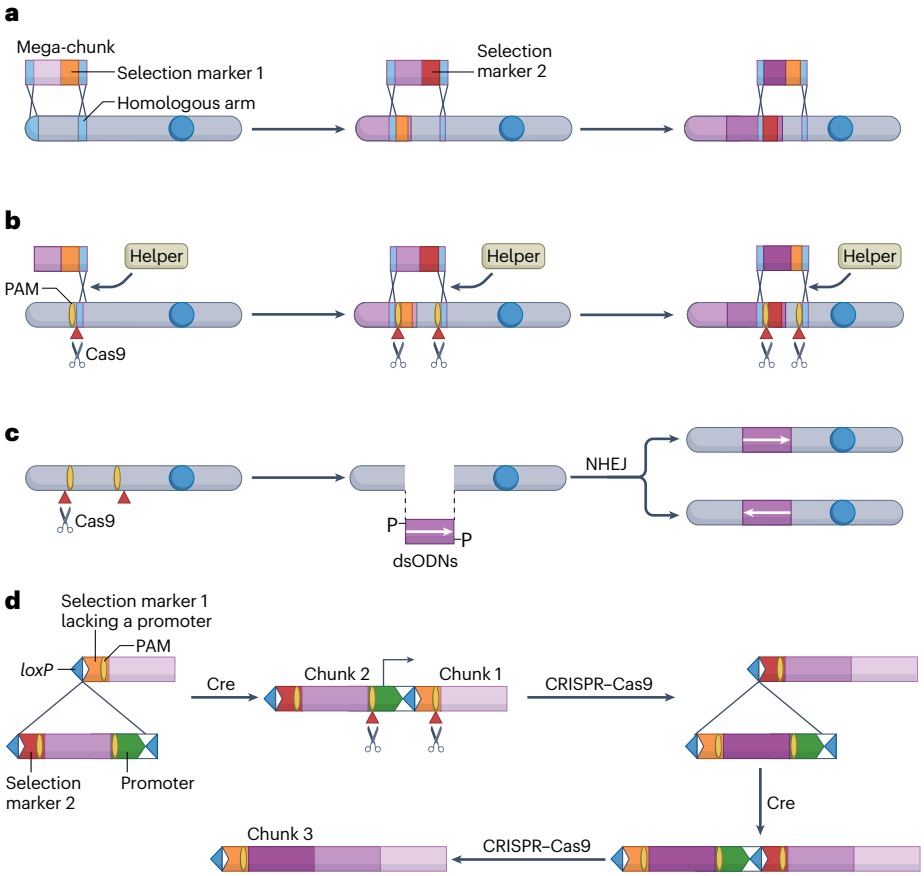

Fig. 5 | Targeted replacement of large DNA fragments. a, Multiple rounds of targeted large-fragment replacement can be achieved through endogenous homologous recombination with two selection markers (red and orange elements) replacing each other in each round of replacement. **b**, Large-fragment homologous recombination can be improved by Cas9 or Cas9 variants. The yellow ovals refer to the protospacer-adjacent motif (PAM) sequences, which are required for Cas9 recognition and are designed near homologous arms. The scissors refer to Cas9 or Cas9 variants. The beige element 'Helper' refers to homology-directed repair-promoting factors, such as RS-1, RAD51 and Scr7. **c**, Targeted large-fragment replacement can be achieved via the nonhomologous end joining (NHEJ) pathway. CRISPR–Cas9 is used to generate two distal double-stranded breaks, enabling exogenous double-stranded oligodeoxynucleotides (dsODNs) to replace the original sequence via the NHEJ pathway. The arrows in purple boxes refer to dsODNs. **d**, An example of fragment stacking using recombination sites[152]; the blue triangles refer to *loxP* sites, the green arrows refer to promoters and the red and orange elements refer to selection markers lacking promoters. Each time a new chunk is recombined into the target location by Cre, the proximity of the promoter and the selection marker lacking a promoter activate the selection marker. By using endonucleases such as CRISPR–Cas9 to eliminate the activated selection gene and a *loxP* site, the next round of recombination is then allowed to proceed. After each round of recombination, gene editing techniques are used to specifically delete the corresponding native sequences, resulting in large-fragment replacement.

Fusion of the Cas9 endonuclease with the *Agrobacterium* VirD2 relaxase can deliver the repair template to the vicinity of the DSB, thereby enhancing HDR efficiency by fivefold to sixfold in rice[134].

Another approach to enhance HDR is to improve endogenous repair pathways. For example, suppressing the NHEJ pathway by depleting key NHEJ factors, such as DNA ligase IV (inhibited by Scr7)[135–137] and *KU70* (ref. 137), has been explored. Overexpressing DNA repair proteins (such as RAD51) or supplementing HDR enhancers (such as RS-1 (ref. 138)) can tune the repair pathways favourable to HDR. Knockout of genes that inhibit HDR, such as the nuclease gene *TREX1*, also has the potential to enhance HDR efficiency[139]. The application of these methods in plants is expected to improve the efficiency of large-fragment replacement based on homologous arms (Fig. 5b).

In contexts in which replacement fragment boundaries lie outside genic or regulatory regions, imprecise target-site integration might be tolerable. Although HDR enables precise and scar-free modifications, its efficiency remains orders of magnitude lower than error-prone NHEJ in seed plants[126] (Fig. 5c). For example, 5′-phosphorylated, double-stranded oligodeoxynucleotides and CRISPR–Cas9-induced DSBs were used to insert a 2 kb fragment in rice via the NHEJ pathway, achieving an efficiency of up to 25% (ref. 140). Similarly, fusing *Pong* transposase with Cas9 endonucleases to direct transposase-excised donor fragments to the vicinity of DSBs improves the efficiency to 35.5% of targeted insertion via NHEJ in *Arabidopsis*[141]. However, NHEJ frequently results in erroneous insertions, deletions, inversions or off-targets, necessitating subsequent detailed screening and verification.

Recombinase-mediated gene targeting, a method using site-specific recombinases to recognize defined recombination loci and to enable precise DNA fragment insertion or replacement, can bypass

the low efficiency of HR[142]. For example, PrimeRoot, a method that harnesses prime editing whereby Cas9 is fused with reverse transcriptase to efficiently and specifically insert a recombination site into the plant genome, has been used to target >10 kb DNA insertions in rice[143]. Plasmids containing the desired DNA fragment flanked by two recombination sites were then transformed into the cells. Under the action of site-specific recombinases, recombination occurred between two recombination sites, resulting in the insertion of a large DNA fragment at the recombination site location in the genome with an efficiency of 6.3% (ref. 143). Comparable approaches can be combined with synthetic sequences to enable precise targeting into the genome (Fig. 5d). However, recombination sites might interfere with subsequent genome engineering and might need to be removed from the genome.

Alternatively, chromosome engineering in *Arabidopsis* and other seed plant species has enabled megabase-scale chromosomal inversions and rearrangements[144–147]. For instance, simultaneous Cas9-mediated DSBs on chromosomes 1 and 2 in *Arabidopsis*, using *Staphylococcus aureus*-derived SaCas9, induced reciprocal megabase-scale translocations between heterologous chromosomes, albeit with low efficiency (0.01%)[148]. Although these methods are not directly applicable to large-fragment integration, they provide valuable frameworks for optimizing targeting and debugging strategies.

Detecting synthetic sequences

The screening and detection of synthetic lines are necessary steps in genome synthesis. Initial detection involves performing PCR on positive lines obtained after resistance screening. Typically, two methods can be used for PCR identification: the three-primer method and the use of PCRtags introduced during genome design. The three-primer method involves designing a primer within the deleted region, such as in transposable elements or removed intergenic regions, and on both sides of the deleted region. Amplicons corresponding to the deleted regions are not detected in lines with the wild-type sequences fully replaced by synthetic sequences[15]. In long synthetic sequences without any deletions, PCRtags are introduced through synonymous substitutions in exons or modifications in introns[13,64] to enable PCR-based separation of synthetic and wild-type sequences. After preliminary PCR, whole-genome sequencing can be used to check the replacement of synthetic sequences, to identify possible off-targets or insertions at other genome locations and to reveal single-nucleotide mutations, as well as other small insertions or deletions.

Testing and debugging

It is essential to test whether lines with partially or fully synthetic genomes exhibit the same phenotypes as the wild type, including not only viability but also stress resistance and sexual reproduction at each developmental stage[13]. Some changes in gene expression may not manifest phenotypically immediately but could exert latent effects through epistatic interactions or under untested stress conditions. For instance, disruptions to genetic redundancy (for example, loss of paralogues) may compromise phenotypic robustness, whereas stress-resistant genes might only become functionally relevant under specific environmental pressures. Thus, using RNA sequencing to assess gene expression in synthetic lines is recommended to identify unintended transcriptional perturbations.

Epigenetics also regulates gene expression in plants as in other multicellular organisms[149,150]. In a pilot study conducted in moss, the correct epigenetic landscape was reconstructed de novo in the synthetic region, except for H3K9me and DNA methylation, which are associated with the most transposable elements removed. Nevertheless, chromatin loops in the synthetic region of semi-synthetic moss are altered, which might influence expression levels but do not result in strong phenotypic changes[15]. If the epigenetic landscape cannot be established by the DNA sequence, epigenetic editing[151] techniques can be used for debugging. Debugging not only corrects unexpected genome synthesis and assembly errors but might also deepen our understanding of the plant genome, including the identification of unknown genes, discovery of non-coding region functions and identification of high-order genome organizations.

Outlook

Research on the de novo synthesis of plant genomes is still in its infancy but holds promise for revolutionizing our understanding of plant genomics and supporting sustainable agriculture, bioproduction and pharmaceutical applications. Several steps during the synthesis process, including genome design, large-fragment synthesis and assembly, DNA transformation and targeting, detection and debugging, can benefit from further optimization.

For genome design, the limited understanding of multicellular genomes restricts the extent of possible modifications. Conversely, synthetic genomics can be used to investigate uncharted properties of multicellular genomes, such as whether all epigenetic modifications are essential or whether primary DNA sequences alone can fully reconstitute epigenetic patterns, particularly in specialized regions such as centromeres. Replacing native centromeric regions with engineered repeats or CENH3-recruiting arrays could help answer these fundamental questions. Similarly, the plasticity of chromatin interactions, including topologically associating domains and chromatin loops in plants, remains poorly understood. Synthetic genomics enables targeted alteration of 3D genome architecture, thereby facilitating functional analysis of these structural features in planta. The systematic removal of genomic 'dark matter', such as transposons and heterochromatin, during redesign processes presents unique opportunities to uncover latent functions of these elements through subsequent debugging. Methodical manipulation of intronic sequences could yield comprehensive evaluation of the importance of splicing, whereas implementation of SCRaMbLE systems could help accelerate functional gene discovery. Emerging capabilities in genome engineering now permit ambitious rewriting strategies, including elimination of non-essential genes, modular reorganization of genetic elements and even pathway optimization. Notably, current plant synthetic genome designs predominantly rely on reference genome frameworks and annotation data. Future advancements could integrate multidimensional data sets encompassing gene expression profiles, functional information, histone modification patterns, chromatin accessibility maps and 3D chromatin architecture to inform more sophisticated synthetic genome blueprints.

A pressing challenge during the large-scale synthesis and assembly phase is to reduce the cost of oligonucleotide synthesis and the time required for large-fragment assembly. Using terminal TdT for oligonucleotide synthesis could improve synthesis efficiency and reduce costs[98,99]. Optimizing TAR in yeast is a critical step for accelerating synthetic genomics research. For example, assembly methods based on the yeast life cycle may help avoid extensive in vitro handling[4]. Moreover, the multistage workflow spanning from oligonucleotide synthesis to final assembly of 100 kb-scale large fragments remains both laborious and redundant. Establishing a standardized and automated platform for DNA assembly would address critical scalability challenges in next-generation genome foundry operations.

In the large-fragment transformation phase, compared with *Agrobacterium*-mediated transformation or biolistic transformation, protoplast transformation might be more suitable as it is compatible with large-fragment transformation and targeted replacement. Moreover, protoplast transformation provides more starting material to obtain enough transformants. The main technical bottleneck lies in the inherently low protoplast regeneration efficiency observed across most plant species. Potential solutions include transient co-overexpression of developmental regulators or stress-responsive genes, coupled with targeted suppression of differentiation-promoting genes, possibly augmented with pharmacological induction. Moreover, fusing plant and yeast cells presents a strategy for synthetic DNA payload transfer. This approach circumvents the labour-intensive and technically demanding processes of yeast-derived megabase plasmid purification and subsequent transformation into *E. coli*, while expanding the theoretical payload capacity beyond current physical delivery limitations.

In the 100-kb-level large-fragment targeting phase, a major challenge is the lack of efficient in planta targeting and replacement techniques. Although the first plant synthetic genome study was conducted in the model moss *P. patens* using endogenous HR, the efficiency was low (<1%) and the procedure was time-consuming, requiring several months for screening. Further optimization of existing HDR-based techniques, such as using Cas variants, HDR-promoting factors or NHEJ inhibitors to enhance HDR, or knocking out genes that inhibit HDR, could enable targeting and replacement of larger fragments. A major priority involves mechanistically validating HDR-mediated precision replacement efficiency in *P. patens*, particularly their capacity to enable megabase-range chromosomal substitutions. Successful demonstration would establish foundational capabilities for genome-scale synthesis initiatives targeting this model bryophyte. Moreover, recombinase-mediated targeting schemes could achieve the targeting and replacement of large fragments through orthogonal recombinase. The design of the targeting schemes to reduce rounds of transformation events is key to seed plants because their life cycle is much longer than that of unicellular organisms.

Building on these technological advancements, the field of plant synthetic genomics is now expected to achieve the full synthesis of a chromosome arm, an entire chromosome and the full genome of *P. patens*. The next milestone would be the de novo synthesis of chromosomal arms with functional telomeres and centromeres in moss. Achieving this goal will help unravel the mysteries of plant centromeres and also pave the way to further de novo synthesis and complete replacement of plant chromosomes. With the development of *P. patens* genome synthesis technology, moss could also be harnessed as a platform to construct diverse, complex synthetic pathways found only in seed plants. An initial test run can be the design and replacement of genome fragments at the scale of tens of kilobases. Introducing an entire biosynthetic pathway as a single module at a target site can be highly useful for bioproduction and other bioengineering applications.

Published online: 25 June 2025

References

1. McKusick, V. A. & Ruddle, F. H. Toward a complete map of the human genome. *Genomics* **1**, 103–106 (1987).
2. Sinyakov, A. N., Ryabinin, V. A. & Kostina, E. V. Application of array-based oligonucleotides for synthesis of genetic designs. *Mol. Biol.* **55**, 487–500 (2021).
3. Casini, A., Storch, M., Baldwin, G. S. & Ellis, T. Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.* **16**, 568–576 (2015).
4. He, B. et al. YLC-assembly: large DNA assembly via yeast life cycle. *Nucleic Acids Res.* **51**, 8283–8292 (2023).
5. Cello, J., Paul, A. V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**, 1016–1018 (2002).
   This study established the first fully synthetic genome, validating template-independent de novo genome construction through chemical synthesis methodologies.
6. Smith, H. O., Hutchison, C. A., Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: φX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl Acad. Sci. USA* **100**, 15440–15445 (2003).
7. Gibson, D. G. et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
8. Fredens, J. et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019).
9. Hutchison, C. A. et al. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
   This study reports the excision of dispensable genes (47% reduction from 901 to 473) through systematic genome streamlining, resulting in the first functional minimal synthetic cell prototype.
10. Zhang, W., Mitchell, L. A., Bader, J. S. & Boeke, J. D. Synthetic genomes. *Annu. Rev. Biochem.* **89**, 77–101 (2020).
11. Dymond, J. S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).
12. Annaluru, N. et al. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–58 (2014).
13. Richardson, S. M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
    This study describes the Sc2.0 project, which delivered a definitive blueprint for synthetic eukaryotic genomics, incorporating sequence-editing strategies, synthetic telomere design and inducible evolution systems.
14. Zhang, W. et al. Mouse genome rewriting and tailoring of three important disease loci. *Nature* **623**, 423–431 (2023).
15. Chen, L.-G. et al. A designer synthetic chromosome fragment functions in moss. *Nat. Plants* **10**, 228–239 (2024).
    This study is the first demonstration of partial chromosome arm synthesis in plants.
16. Boeke, J. D. et al. The genome project-write. *Science* **353**, 126–127 (2016).
17. Jiao, Y. & Wang, Y. Towards plant synthetic genomics. *Biodes. Res.* **5**, 0020 (2023).
18. Blommaert, J. Genome size evolution: towards new model systems for old questions. *Proc. Biol. Sci.* **287**, 20201441 (2020).
19. AGARWAL, K. L. et al. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* **227**, 27–34 (1970).
20. Koonin, E. V., Dolja, V. V. & Krupovic, M. The logic of virus evolution. *Cell Host Microbe* **30**, 917–929 (2022).
21. Racaniello, V. R. & Baltimore, D. Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proc. Natl Acad. Sci. USA* **78**, 4887–4891 (1981).
22. Venter, J. C., Glass, J. I., Hutchison, C. A. & Vashee, S. Synthetic chromosomes, genomes, viruses, and cells. *Cell* **185**, 2708–2724 (2022).
23. Chan, L. Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1**, 2005.0018 (2005).
24. Thi Nhu Thao, T. et al. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565 (2020).
25. Dormitzer, P. R. et al. Synthetic generation of influenza vaccine viruses for rapid response to pandemics. *Sci. Transl. Med.* **5**, 185ra68 (2013).
26. Gibson, D. G. et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
27. Pelletier, J. F. et al. Genetic requirements for cell division in a genomically minimal cell. *Cell* **184**, 2430–2440.e16 (2021).
28. Ostrov, N. et al. Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
29. Wang, K. et al. Defining synonymous codon compression schemes by genome recoding. *Nature* **539**, 59–64 (2016).
30. Goffeau, A. et al. Life with 6000 genes. *Science* **274**, 546–567 (1996).
31. Mitchell, L. A. et al. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* **355**, eaaf4831 (2017).
32. Shen, Y. et al. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355**, eaaf4791 (2017).
33. Wu, Y. et al. Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* **355**, eaaf4706 (2017).
34. Xie, Z.-X. et al. 'Perfect' designer chromosome V and behavior of a ring derivative. *Science* **355**, eaaf4704 (2017).
35. Zhang, W. et al. Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* **355**, eaaf3981 (2017).
36. Zhao, Y. et al. Debugging and consolidating multiple synthetic chromosomes reveals combinatorial genetic interactions. *Cell* **186**, 5220–5236.e16 (2023).
37. Schindler, D. et al. Design, construction, and functional characterization of a tRNA neochromosome in yeast. *Cell* **186**, 5237–5253.e22 (2023).
38. Admire, A. et al. Cycles of chromosome instability are associated with a fragile site and are increased by defects in DNA replication and checkpoint controls in yeast. *Genes Dev.* **20**, 159–173 (2006).

39. Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. Transcription–replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* **170**, 774–786.e19 (2017).
40. Mularoni, L. et al. Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. *Genome Res.* **22**, 693–703 (2012).
41. Dai, J., Boeke, J. D., Luo, Z., Jiang, S. & Cai, Y. Sc3.0: revamping and minimizing the yeast genome. *Genome Biol.* **21**, 205 (2020).
42. Birchler, J. A., Graham, N. D., Swyers, N. C., Cody, J. P. & McCaw, M. E. Plant minichromosomes. *Curr. Opin. Biotechnol.* **37**, 135–142 (2016).
43. Shay, J. W. & Wright, W. E. Telomeres and telomerase: three decades of progress. *Nat. Rev. Genet.* **20**, 299–309 (2019).
44. Yu, W., Lamb, J. C., Han, F. & Birchler, J. A. Telomere-mediated chromosomal truncation in maize. *Proc. Natl Acad. Sci. USA* **103**, 17331–17336 (2006).
45. Carlson, W. R. & Phillips, R. L. The B chromosome of maize. *Crit. Rev. Plant Sci.* **3**, 201–226 (1986).
46. Gaeta, R. T. et al. In vivo modification of a maize engineered minichromosome. *Chromosoma* **122**, 221–232 (2013).
47. Teo, C. H. et al. Induction of telomere-mediated chromosomal truncation and stability of truncated chromosomes in *Arabidopsis thaliana*. *Plant J.* **68**, 28–39 (2011).
48. Murata, M. Minichromosomes and artificial chromosomes in *Arabidopsis*. *Chr. Res.* **22**, 167–178 (2014).
49. Xu, C., Cheng, Z. & Yu, W. Construction of rice mini-chromosomes by telomere-mediated chromosomal truncation. *Plant J. Cell Mol. Biol.* **70**, 1070–1079 (2012).
50. Kapusi, E. et al. Telomere-mediated truncation of barley chromosomes. *Chromosoma* **121**, 181–190 (2012).
51. Kouprina, N. et al. Human artificial chromosome with regulated centromere: a tool for genome and cancer studies. *ACS Synth. Biol.* **7**, 1974–1989 (2018).
52. Jiang, S. et al. Efficient de novo assembly and modification of large DNA fragments. *Sci. China Life Sci.* **65**, 1445–1455 (2022).
53. Cove, D. The moss *Physcomitrella patens*. *Annu. Rev. Genet.* **39**, 339–358 (2005).
54. Rensing, S. A., Goffinet, B., Meyberg, R., Wu, S.-Z. & Bezanilla, M. The moss *Physcomitrium* (*Physcomitrella*) *patens*: a model organism for non-seed plants. *Plant Cell* **32**, 1361–1376 (2020).
55. de Keijzer, J., Freire Rios, A. & Willemsen, V. *Physcomitrium* patens: a single model to study oriented cell divisions in 1D to 3D patterning. *Int. J. Mol. Sci.* **22**, 2626 (2021).
56. Bi, G. et al. Near telomere-to-telomere genome of the model plant *Physcomitrium* patens. *Nat. Plants* **10**, 327–343 (2024).
57. Schaefer, D. G. & Zrÿd, J.-P. Efficient gene targeting in the moss *Physcomitrella patens*. *Plant J.* **11**, 1195–1206 (1997).
58. Charlot, F., Goudounet, G., Nogué, F. & Perroud, P.-F. in *Protoplast Technology*: *Methods and Protocols* (eds Wang, K. & Zhang, F.) 3–19 (Humana Press, 2022).
59. Widiez, T. et al. The chromatin landscape of the moss *Physcomitrella patens* and its dynamics during development and drought stress. *Plant J.* **79**, 67–81 (2014).
60. Lang, D. et al. The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–533 (2018).
61. Khairul Ikram, N. K., Zakariya, A. M., Saiman, M. Z., Kashkooli, A. B. & Simonsen, H. T. Heterologous production of artemisinin in *Physcomitrium patens* by direct in vivo assembly of multiple DNA fragments. *Bio-Protocol* **13**, e4719 (2023).
62. Reski, R., Parsons, J. & Decker, E. L. Moss-made pharmaceuticals: from bench to bedside. *Plant Biotechnol. J.* **13**, 1191–1198 (2015).
63. Decker, E. L., Parsons, J. & Reski, R. Glyco-engineering for biopharmaceutical production in moss bioreactors. *Front. Plant Sci.* **5**, 346 (2014).
64. Yu, W. et al. Designing a synthetic moss genome using GenoDesigner. *Nat. Plants* **10**, 848–856 (2024).
This study reports a systematic framework to govern synthesized chromosome in *Physcomitrella patens*, integrating sequence modularization and orthogonal design constraints.
65. Gao, D. Introduction of plant transposon annotation for beginners. *Biology* **12**, 1468 (2023).
66. Elliott, T. A., Linquist, S. & Gregory, T. R. Conceptual and empirical challenges of ascribing functions to transposable elements. *Am. Nat.* **184**, 14–24 (2014).
67. Doolittle, W. F., Brunet, T. D. P., Linquist, S. & Gregory, T. R. Distinguishing between 'function' and 'effect' in genome biology. *Genome Biol. Evol.* **6**, 1234–1237 (2014).
68. Gemmell, N. J. Repetitive DNA: genomic dark matter matters. *Nat. Rev. Genet.* **22**, 342 (2021).
69. Buckley, P. T., Khaladkar, M., Kim, J. & Eberwine, J. Cytoplasmic intron retention, function, splicing, and the sentinel RNA hypothesis. *Wiley Interdiscip. Rev. RNA* **5**, 223–230 (2014).
70. Schmitz, U. et al. Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol.* **18**, 216 (2017).
71. Shaul, O. How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* **91**, 145–155 (2017).
72. Erikson, O., Hertzberg, M. & Näsholm, T. A conditional marker gene allowing both positive and negative selection in plants. *Nat. Biotechnol.* **22**, 455–458 (2004).
73. Dymond, J. & Boeke, J. The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioeng. Bugs* **3**, 168–171 (2012).
74. Comai, L., Maheshwari, S. & Marimuthu, M. P. A. Plant centromeres. *Curr. Opin. Plant Biol.* **36**, 158–167 (2017).
75. Melters, D. P. et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **14**, R10 (2013).
76. McKinley, K. L. & Cheeseman, I. M. The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* **17**, 16–29 (2016).
77. Carroll, C. W., Milks, K. J. & Straight, A. F. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J. Cell Biol.* **189**, 1143–1155 (2010).
78. Harrington, J. J., Bokkelen, G. V., Mays, R. W., Gustashaw, K. & Willard, H. F. Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat. Genet.* **15**, 345–355 (1997).
79. Ebersole, T. et al. Rapid generation of long synthetic tandem repeats and its application for analysis in human artificial chromosome formation. *Nucleic Acids Res.* **33**, e130 (2005).
80. Noskov, V. N., Lee, N. C., Larionov, V. & Kouprina, N. Rapid generation of long tandem DNA repeat arrays by homologous recombination in yeast to study their function in mammalian genomes. *Biol. Proc. Online* **13**, 8 (2011).
81. Nakano, M. et al. Inactivation of a human kinetochore by specific targeting of chromatin modifiers. *Dev. Cell* **14**, 507–522 (2008).
82. Kouprina, N. et al. Organization of synthetic alphoid DNA array in human artificial chromosome (HAC) with a conditional centromere. *ACS Synth. Biol.* **1**, 590–601 (2012).
83. Logsdon, G. A. et al. Human artificial chromosomes that bypass centromeric DNA. *Cell* **178**, 624–639.e19 (2019).
84. Phan, B. H. et al. Transformation of rice with long DNA-segments consisting of random genomic DNA or centromere-specific DNA. *Transgenic Res.* **16**, 341–351 (2007).
85. Dawe, R. K. Charting the path to fully synthetic plant chromosomes. *Exp. Cell Res.* **390**, 111951 (2020).
86. Gent, J. I., Dong, Y., Jiang, J. & Dawe, R. K. Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. *Nucleic Acids Res.* **40**, 1550–1560 (2012).
87. Teo, C. H., Lermontova, I., Houben, A., Mette, M. F. & Schubert, I. De novo generation of plant centromeres at tandem repeats. *Chromosoma* **122**, 233–241 (2013).
88. Dawe, R. K. et al. Synthetic maize centromeres transmit chromosomes across generations. *Nat. Plants* **9**, 433–441 (2023).
This article describes the engineering of LexA–CENH3 chimaeras to mediate accurate recruitment of centromeric histones to LexO repeats, establishing de novo kinetochore complexes that triggered programmable chromosome breakage.
89. Zhang, H. & Dawe, R. K. Total centromere size and genome size are strongly correlated in ten grass species. *Chromosome Res.* **20**, 403–412 (2012).
90. Plačková, K., Bureš, P. & Zedek, F. Centromere size scales with genome size across eukaryotes. *Sci. Rep.* **11**, 19811 (2021).
91. Kumawat, S. & Choi, J. Y. No end in sight: mysteries of the telomeric variation in plants. *Am. J. Bot.* **110**, e16244 (2023).
92. Zvereva, M. I., Shcherbakova, D. M. & Dontsova, O. A. Telomerase: structure, functions, and activity regulation. *Biochemistry (Mosc)* **75**, 1563–1583 (2010).
93. Heacock, M., Spangler, E., Riha, K., Puizina, J. & Shippen, D. E. Molecular analysis of telomere fusions in *Arabidopsis*: multiple pathways for chromosome end-joining. *EMBO J.* **23**, 2304–2313 (2004).
94. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507 (2014).
95. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites — a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862 (1981).
96. LeProust, E. M. et al. Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res.* **38**, 2522–2540 (2010).
97. Matzas, M. et al. High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing. *Nat. Biotechnol.* **28**, 1291–1294 (2010).
98. Palluk, S. et al. De novo DNA synthesis using polymerase–nucleotide conjugates. *Nat. Biotechnol.* **36**, 645–650 (2018).
99. Barthel, S., Palluk, S., Hillson, N. J., Keasling, J. D. & Arlow, D. H. Enhancing terminal deoxynucleotidyl transferase activity on substrates with 3′ terminal structures for enzymatic de novo DNA synthesis. *Genes* **11**, 102 (2020).
100. Benoit, R. M. et al. Seamless insert-plasmid assembly at high efficiency and low cost. *PLoS ONE* **11**, e0153158 (2016).
101. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
102. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLOS ONE* **4**, e5553 (2009).
103. Gibson, D. G. Gene and genome construction in yeast. *Curr. Protoc. Mol. Biol.* **Ch. 3**, Unit3.22 (2011).
104. Muller, H. et al. Assembling large DNA segments in yeast. *Methods Mol. Biol.* **852**, 133–150 (2012).
105. Gelvin, S. B. Integration of *Agrobacterium* T-DNA into the plant genome. *Annu. Rev. Genet.* **51**, 195–217 (2017).
106. Hwang, H.-H., Yu, M. & Lai, E.-M. *Agrobacterium*-mediated plant transformation: biology and applications. *Arab. Book* **15**, e0186 (2017).
107. Altpeter, F. et al. Particle bombardment and the genetic enhancement of crops: myths and realities. *Mol. Breed.* **15**, 305–327 (2005).
108. Lowe, B. A. et al. Enhanced single copy integration events in corn via particle bombardment using low quantities of DNA. *Transgenic Res.* **18**, 831–840 (2009).

109. Hamilton, C. M., Frary, A., Lewis, C. & Tanksley, S. D. Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proc. Natl Acad. Sci. USA* **93**, 9975–9979 (1996).
110. Liu, J. et al. Genome-scale sequence disruption following biolistic transformation in rice and maize. *Plant Cell* **31**, 368–383 (2019).
111. Ikeuchi, M., Sugimoto, K. & Iwase, A. Plant callus: mechanisms of induction and repression. *Plant Cell* **25**, 3159–3173 (2013).
112. Duarte, P., Ribeiro, D., Carqueijeiro, I., Bettencourt, S. & Sottomayor, M. in *Biotechnology of Plant Secondary Metabolism: Methods and Protocols* (ed. Fett-Neto, A. G.) 137–148 (Humana Press, 2016).
113. Li, L. & Blankenstein, T. Generation of transgenic mice with megabase-sized human yeast artificial chromosomes by yeast spheroplast–embryonic stem cell fusion. *Nat. Protoc.* **8**, 1567–1582 (2013).
114. Brown, D. M. et al. Efficient size-independent chromosome delivery from yeast to cultured cell lines. *Nucleic Acids Res.* **45**, e50–e50 (2017).
115. Wu, F.-H., Yuan, Y.-H., Hsu, C.-T., Cheng, Q.-W. & Lin, C.-S. in *Protoplast Technology: Methods and Protocols* (eds Wang, K. & Zhang, F.) 49–64 (Humana Press, 2022).
116. Jones, A. M. P., Shukla, M. R., Biswas, G. C. G. & Saxena, P. K. Protoplast-to-plant regeneration of American elm (*Ulmus americana*). *Protoplasma* **252**, 925–931 (2015).
117. Jeong, Y. Y., Lee, H.-Y., Kim, S. W., Noh, Y.-S. & Seo, P. J. Optimization of protoplast regeneration in the model plant *Arabidopsis thaliana*. *Plant Methods* **17**, 21 (2021).
118. Lowe, K. et al. Morphogenic regulators *Baby boom* and *Wuschel* improve monocot transformation. *Plant Cell* **28**, 1998–2015 (2016).
119. Debernardi, J. M. et al. A GRF–GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nat. Biotechnol.* **38**, 1274–1279 (2020).
120. Yu, Y. et al. Enhancing wheat regeneration and genetic transformation through overexpression of TaLAX1. *Focus Issue Wheat Biol.* **5**, 100738 (2024).
121. Liu, X. et al. Uncovering the transcriptional regulatory network involved in boosting wheat regeneration and transformation. *Nat. Plants* **9**, 908–925 (2023).
122. Xu, M., Du, Q., Tian, C., Wang, Y. & Jiao, Y. Stochastic gene expression drives mesophyll protoplast regeneration. *Sci. Adv.* **7**, eabg8466 (2021).
123. Yang, W. et al. Peptide REF1 is a local wound signal promoting plant regeneration. *Cell* **187**, 3024–3038.e14 (2024).
124. Wang, N. et al. Leaf transformation for efficient random integration and targeted genome modification in maize and sorghum. *Nat. Plants* **9**, 255–270 (2023).
125. Sun, Y. et al. Engineering herbicide-resistant rice plants through CRISPR/Cas9-mediated homologous recombination of acetolactate synthase. *Mol. Plant* **9**, 628–631 (2016).
126. Schmidt, C., Pacher, M. & Puchta, H. in *Transgenic Plants: Methods and Protocols* (eds Kumar, S. et al.) 237–266 (Humana Press, 2019).
127. Fauser, F. et al. In planta gene targeting. *Proc. Natl Acad. Sci. USA* **109**, 7535–7540 (2012).
128. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
129. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
130. Zhao, Y. et al. An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Sci. Rep.* **6**, 23890 (2016).
131. Rees, H. A., Yeh, W.-H. & Liu, D. R. Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat. Commun.* **10**, 2212 (2019).
132. Chauhan, V. P., Sharp, P. A. & Langer, R. Altered DNA repair pathway engagement by engineered CRISPR–Cas9 nucleases. *Proc. Natl Acad. Sci. USA* **120**, e2300605120 (2023).
133. Schreiber, T. et al. Efficient scar-free knock-ins of several kilobases in plants by engineered CRISPR–Cas endonucleases. *Mol. Plant* **17**, 824–837 (2024).
134. Ali, Z. et al. Fusion of the Cas9 endonuclease and the VirD2 relaxase facilitates homology-directed repair for precise genome engineering in rice. *Commun. Biol.* **3**, 44 (2020).
135. Morton, J., Davis, M. W., Jorgensen, E. M. & Carroll, D. Induction and repair of zinc-finger nuclease-targeted double-strand breaks in *Caenorhabditis elegans* somatic cells. *Proc. Natl Acad. Sci. USA* **103**, 16370–16375 (2006).
136. Maruyama, T. et al. Increasing the efficiency of precise genome editing with CRISPR–Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).
137. Qi, Y. et al. Increasing frequencies of site-specific mutagenesis and gene targeting in *Arabidopsis* by manipulating DNA repair pathways. *Genome Res.* **23**, 547–554 (2013).
138. Song, J. et al. RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.* **7**, 10548 (2016).
139. Karasu, M. E. et al. Removal of TREX1 activity enhances CRISPR–Cas9-mediated homologous recombination. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-024-02356-3 (2024).
140. Lu, Y. et al. Targeted, efficient sequence insertion and replacement in rice. *Nat. Biotechnol.* **38**, 1402–1407 (2020).
141. Liu, P. et al. Transposase-assisted target-site integration for efficient plant genome engineering. *Nature* **631**, 593–600 (2024).
142. Kühn, R. & Torres, R. M. in *Transgenesis Techniques: Principles and Protocols* (ed. Clarke, A. R.) 175–204 (Springer, 2002).
143. Sun, C. et al. Precise integration of large DNA sequences in plant genomes using PrimeRoot editors. *Nat. Biotechnol.* **42**, 316–327 (2024).
144. Rönspies, M., Dorn, A., Schindele, P. & Puchta, H. CRISPR–Cas-mediated chromosome engineering for crop improvement and synthetic biology. *Nat. Plants* **7**, 566–573 (2021).
145. Rönspies, M., Schindele, P., Wetzel, R. & Puchta, H. CRISPR–Cas9-mediated chromosome engineering in *Arabidopsis thaliana*. *Nat. Protoc.* **17**, 1332–1358 (2022).
146. Schmidt, C. et al. Changing local recombination patterns in *Arabidopsis* by CRISPR/Cas mediated chromosome engineering. *Nat. Commun.* **11**, 4418 (2020).
147. Rönspies, M. et al. Massive crossover suppression by CRISPR–Cas-mediated plant chromosome engineering. *Nat. Plants* **8**, 1153–1159 (2022).
148. Beying, N., Schmidt, C., Pacher, M., Houben, A. & Puchta, H. CRISPR–Cas9-mediated induction of heritable chromosomal translocations in *Arabidopsis*. *Nat. Plants* **6**, 638–645 (2020).
149. Godwin, J. & Farrona, S. Plant epigenetic stress memory induced by drought: a physiological and molecular perspective. *Methods Mol. Biol.* **2093**, 243–259 (2020).
150. Hemenway, E. A. & Gehring, M. Epigenetic regulation during plant development and the capacity for epigenetic memory. *Annu. Rev. Plant Biol.* **74**, 87–109 (2023).
151. Gjaltema, R. A. F. & Rots, M. G. Advances of epigenetic editing. *Curr. Opin. Chem. Biol.* **57**, 75–81 (2020).
152. Srivastava, V. & Thomson, J. Gene stacking by recombinases. *Plant Biotechnol. J.* **14**, 471–482 (2016).

Acknowledgements

Work in the authors' laboratory is funded by the National Key R&D Program of China (2023YFE0101100 and 2019YFA0903900), the Natural Science Foundation of China (32230010 and 32270345), the Qidong-SLS Innovation Fund, and the Guangdong Provincial Key Laboratory of Synthetic Genomics (2023B1212060054).

Author contributions

The manuscript was written by T.L. All the authors contributed to discussion of the content and review and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Citation diversity statement

We acknowledge that papers authored by scholars from historically excluded groups are systematically under-cited. Here, we have made every attempt to reference relevant papers in a manner that is equitable in terms of racial, ethnic, gender and geographical representation.

Additional information

Peer review information *Nature Reviews Bioengineering* thanks Ilha Lee, Yiping Qi, who co-reviewed with Rushil Mandlik, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.