nature plants

Perspective

https://doi.org/10.1038/s41477-024-01693-0

Designing a synthetic moss genome using GenoDesigner

Received: 12 October 2023

Accepted: 10 April 2024

Published online: 03 June 2024

Check for updates

Wenfei Yu[1,2,8], Shuo Zhang[2,3,8], Shijun Zhao[1,2], Lian-ge Chen[3], Jie Cao[4], Hao Ye[4], Jianbin Yan[4], Qiao Zhao[1], Beixin Mo[5], Ying Wang[2], Yuling Jiao[2,3,6,7], Yingxin Ma[1,2], Xiaoluo Huang[1,2] ✉, Wenfeng Qian[2,3] ✉ & Junbiao Dai[1,2,4,5] ✉

The de novo synthesis of genomes has made unprecedented progress and achieved milestones, particularly in bacteria and yeast. However, the process of synthesizing a multicellular plant genome has not progressed at the same pace, due to the complexity of multicellular plant genomes, technical difficulties associated with large genome size and structure, and the intricacies of gene regulation and expression in plants. Here we outline the bottom-up design principles for the de novo synthesis of the *Physcomitrium patens* (that is, earthmoss) genome. To facilitate international collaboration and accessibility, we have developed and launched a public online design platform called GenoDesigner. This platform offers an intuitive graphical interface enabling users to efficiently manipulate extensive genome sequences, even up to the gigabase level. This tool is poised to greatly expedite the synthesis of the *P. patens* genome, offering an essential reference and roadmap for the synthesis of plant genomes.

Synthetic biology has heralded a paradigm shift, marking a transition from an era characterized by reading (genome sequencing) to an emergent epoch centred on writing (genome synthesizing). The advent of recent advancements in DNA synthesis technologies, such as high-throughput DNA synthesis[1] and the assembly of large DNA fragments[2], coupled with the revolution of the robust and versatile genome editing technology CRISPR–Cas[3], has rendered the de novo synthesis of various organismal genomes as achievable reality[4–6]. The inception of whole-genome synthesis can be traced back to the pioneering work of Cello et al.[7], who assembled poliovirus complementary DNA from chemically synthesized oligonucleotides[7]. Concurrently, Smith et al. ventured into the synthesis of a ϕX174 bacteriophage by refining the assembly methods suitable to chemically synthesized oligonucleotides[8]. These efforts laid the groundwork for whole-genome synthesis.

The bacterial genomes of *Mycoplasma genitalium*, *M. mycoides* and *Escherichia coli* were subsequently designed and synthesized[4,6,9]. These works focused on prokaryotic organisms until the inception of the synthetic yeast genome project (Sc2.0)[5]. By 2017, this project had witnessed notable strides with the successful synthesis of six chromosomes of the *Saccharomyces cerevisiae* genome, complete with artificially introduced marker features[10–15]. Recently, all synthetic chromosomes in Sc2.0 have been accomplished, representing a milestone in synthetic genomics[16–25]. With the proposal of the Genome Project-Write and the increasing momentum of Sc3.0 in recent years[26,27], we are

[1]Shenzhen Key Laboratory of Synthetic Genomics, Guangdong Provincial Key Laboratory of Synthetic Genomics, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. [2]University of Chinese Academy of Sciences, Beijing, China. [3]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [4]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [5]College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China. [6]State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China. [7]Peking-Tsinghua Center for Life Sciences, Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [8]These authors contributed equally: Wenfei Yu, Shuo Zhang. ✉e-mail: huangxl@siat.ac.cn; wfqian@genetics.ac.cn; daijunbiao@caas.cn

entering what can be termed as the 'genome-write' era. This era beckons a deeper understanding and application of life, offering notable prospects for future genomic exploration and manipulation.

The expansive realm of genomics has thus far unveiled the sequences of numerous economically important crops and medicinal plants. These decoded sequences provide an invaluable 'genome 1.0' (that is, the original genome) template, serving as a cornerstone for driving further advancements in synthetic genome design. To accelerate the synthetic genomic revolution in plant biology, we have advocated for the initiation of genome synthesis in *Physcomitrium patens*, an earthmoss species that diverged from seed plant ancestors approximately ten million years ago[28]. The initial draft genome of *P. patens* was released in 2008. In 2017, Lang et al. further refined this genome, presenting a genome architecture consisting of 27 pseudochromosomes and leaving about 1% of sequences unresolved[29]. We recently reported a near telomere-to-telomere assembly confirming that *P. patens* comprises 26 chromosomes with a total length of 480 Mb[30].

The decision to select *P. patens* as our pioneer plant organism is based on several factors. First, it represents one of the earliest land plant lineages in a taxonomic group closely related to vascular plants and has been established as an exemplary model species for studying plant evolution, genetics and cell biology[31]. Second, its relatively short life cycle, involving one generational reproduction within approximately two months, makes it an attractive model for expediting the achievement of our synthetic genome goal. Third, *P. patens* possesses a high recombination rate and an inherent ability to efficiently incorporate synthetic large DNA fragments[28]. Fourth, the *P. patens* genome is notably composed of approximately 60% transposable elements (TEs)[29]. This composition not only raises intriguing questions regarding their role and persistence through evolution but also presents an opportunity for substantial genome reduction, should these elements be deemed non-essential. Lastly, its 26 chromosomes of variable sizes enable a wide range of chromosome reconstruction investigations, including chromosome fusion and circularization[30].

In a related study, we have successfully synthesized a portion of the left arm of chromosome 18 of *P. patens*, thereby strengthening the groundwork for future genome synthesis research[32]. We believe that the synthesis of the *P. patens* genome, termed SynMoss, not only is achievable but also can represent a key milestone in advancing genome synthesis research for other plants, including gymnosperms (for example, pine), eudicotyledons (for example, *Arabidopsis*, soybean and tomato) and monocotyledons (for example, maize and rice). Inspired by the Sc2.0 project, we propose that the synthesis of different chromosomes of *P. patens* could be concurrently pursued by multiple research groups to accelerate the pace of progress. We invite all scientists interested in this field to participate in the SynMoss project as a consortium.

Genome design stands as the cornerstone of genome synthesis. While earlier work on virus and bacteria genome synthesis adhered to a few simple design rules[4,7], recent progress in larger genome synthesis demands systematic modifications to the original genome, leading to a more comprehensive design process. As genome size increases, its design complexity grows substantially, thereby requiring an efficient and easy-to-use software to generate the synthetic sequences. Despite the development of the BioStudio program by the Sc2.0 project, enabling teams from the United States, China, the United Kingdom, Australia and Singapore to design their chromosomes[5,33], the lack of a user-friendly interface posed a challenge for wet-lab scientists. Consequently, many design-related tasks were manually and painstakingly executed by front-line scientists, such as the inspection of a missing fragment in the synthetic chromosome. Furthermore, BioStudio failed to address the requirements of genome design at the scale of hundreds of megabases to gigabases, making it unsuitable for *P. patens* genome synthesis design. In 2021, Schwartz and colleagues announced the initial development of a computer-aided design program capable of designing an organism and visualizing hundreds of megabases of a genome (https://spectrum.ieee.org/genome-editing-cad); however, no subsequent progress was reported.

Here we developed GenoDesigner, a user-friendly and efficient online genome design tool. This tool is adept at handling genome designs up to the gigabase level, empowering researchers to manipulate target DNA sequences in each genome or chromosome. We anticipate that the principles, the experiences and the GenoDesigner tool outlined here will serve as invaluable resources, not only expediting the synthesis of the *P. patens* genome but also facilitating related biological research.

The fundamental rationale for designing a synthetic *P. patens* genome

To increase the likelihood of success in this pioneering endeavour, we have delineated three core rationales for the synthetic design: (1) it should constitute a wholly synthetic genome, not merely a partially edited one; (2) the *P. patens* plant harbouring a synthetic genome must be viable; and (3) the synthetic chromosomes should exhibit functionality beneficial to further biotechnological applications or biological studies.

Drawing from the insights gathered from the Sc2.0 project and combining them with our pilot plant genome synthesis experience, we propose the following design principles for the *P. patens* genome: (1) removal of all the TEs, (2) truncation of most intergenic regions, (3) introduction of PCRmarks, (4) unification of stop codons and (5) incorporation of the Synthetic Chromosome Recombination and Modification by LoxP-Mediated Evolution (SCRaMbLE) system (Fig. 1).

Removal of all the TEs

TEs, constituting approximately 57.2% of the *P. patens* genome, are predominantly composed of long terminal repeat retrotransposons. Unfortunately, the functional roles of these TEs remain largely unknown. The investigation of TEs in *P. patens* is impeded by their repetitive nature and seemingly random distribution; however, such exploration could greatly enhance our comprehension of plant genome evolution, spanning from unicellular plants to land plants. Most TEs in *P. patens* are located distantly from protein-coding genes (Supplementary Fig. 1), suggesting minimal direct impact on gene expression.

On the basis of the characteristics of TEs, we have adopted an assertive design strategy to completely eliminate all TEs in the synthetic *P. patens* genome, despite the inherent risk of potentially eradicating critical biological functions acquired by some TEs during evolution (Fig. 1a and Supplementary Figs. 2 and 3). During preliminary testing of the synthetic genome, we will meticulously verify the fitness of the synthetic plants, reintroducing any necessary TEs to maintain viability under specific environmental conditions.

Truncation of intergenic regions

Protein-coding genes are distributed throughout the genome, whereas repetitive sequences, notorious for inducing genome instability, are predominantly located in intergenic regions (Supplementary Fig. 1). Notably, the median length of intergenic regions in *P. patens* has been reduced from over 10 kbp to less than 6 kbp through the removal of all TEs (Supplementary Fig. 4). In an effort to further refine the genome, we have opted to excise the intergenic sequences. This approach also enables us to explore the functions of regulatory elements on a whole-genome scale. Previous studies have indicated that plant promoters typically range in length from 500 to 2,000 bp[34]. On the basis of RNA-seq analyses, we have observed that transcription start and end sites are approximately 500 bp upstream and downstream of genes, respectively (Supplementary Fig. 5). Consequently, in the synthetic *P. patens* genome, we have retained a sequence span of 3 kbp upstream and 2 kbp downstream of each annotated gene after the removal of TEs, serving as potential promoters and terminators, respectively (Fig. 1b). The remaining sequences, in conjunction with centromeres

https://doi.org/10.1038/s41477-024-01693-0

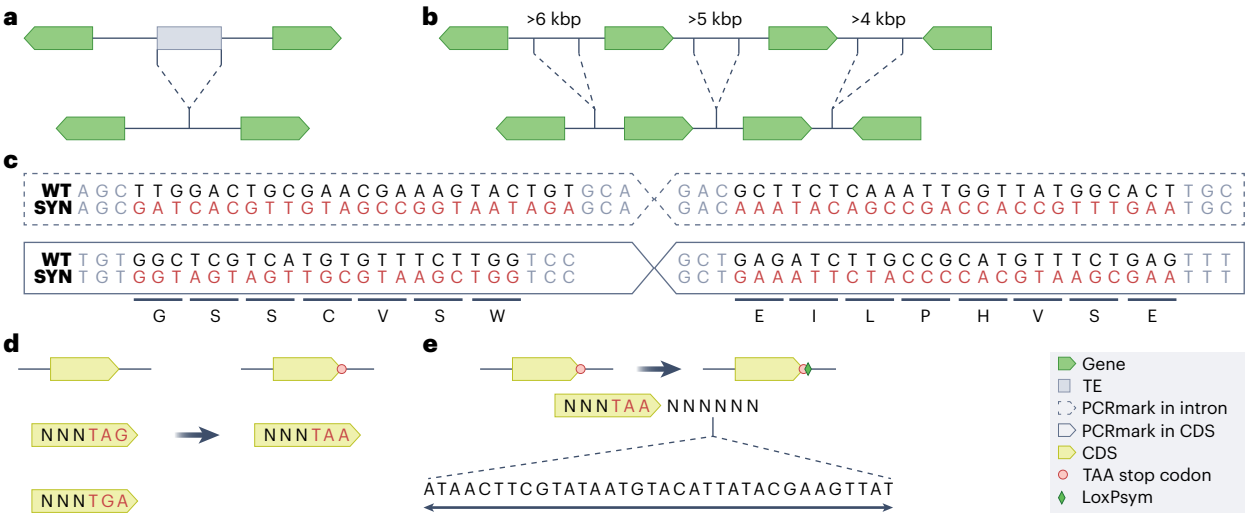

Fig. 1 | The design principles for the SynMoss genome. a, TEs are removed, and the adjacent sequences are joined together. **b**, For each gene, the regions 3 kbp upstream and 2 kbp downstream are retained as putative promoter and terminator sites, respectively, while intergenic regions beyond these limits are removed. **c**, Two examples of pairs of PCRmarks designed in introns and CDSs, respectively. SYN, synthetic; WT, wild type. Both WT and synthetic PCRmarks are designed for each gene to differentiate the native from the synthetic sequence. Synthetic PCRmarks in introns are made by shuffling the WT sequence. The synthetic PCRmarks in CDSs are made by codon switching, ensuring that the resulting amino acids remain unchanged. **d**, All stop codons are unified to TAA in SynMoss. **e**, LoxPsym sequences are inserted 3 bp downstream of the last stop codon for each gene. All sequences displayed are oriented on the positive strand.

and synthetic telomeres, constitute the synthetic chromosomes of *P. patens*.

Introduction of PCRmarks

When employing a bottom-up strategy for genome synthesis, it is crucial to include specific markers to differentiate the synthetic sequence from its wild-type counterpart. Leveraging insight from the Sc2.0 project, we integrated PCRmarks, spanning 20–30 nucleotides, into the synthetic chromosome, within either intron or coding sequence (CDS) regions (Fig. 1c). For introns, to avoid interfering with mRNA splicing, we placed the PCRmarks 10 bp downstream from the 5′ end and 50 bp upstream of the 3′ end, carefully avoiding the predicted 5′ splice site, 3′ splice site and branch point sequences. For CDSs, we adapted the approach outlined by Dymond et al.[5], with a few modifications. Specifically, to ensure the highest specificity of each primer pair towards the *P. patens* genome, we replaced each original codon with the most dissimilar synonymous codon, resulting in PCRmarks exhibiting at least 37.5% nucleotide divergence from their wild-type sequence counterparts. Furthermore, to minimize potential interference with translation initiation and termination, we confined the PCRmarks to positions beyond the first and last six bases of the coding region. To improve PCR efficiency, we limited the products generated by PCRmarks to less than 1,000 bp, regardless of whether the template sequence is located in introns or CDSs. We used Primer3 (ref. 35) to generate PCRmarks with annealing temperatures between 57 °C and 68 °C. Although it is preferable to introduce PCRmarks into the intron regions, if the design standards are not met, they are frequently positioned within CDSs.

Unification of stop codons

In the design of synthetic genomes, codon minimization can offer several advantages, including the potential for protein production with non-standard amino acids and the potential to confer antivirus properties[36,37]. It has been shown in synthetic *E. coli* that the substitution of the stop codon TAG with TAA as well as the removal of relevant release factors appears to have minimal impact on fitness[37]. In Sc2.0, similar changes to stop codons do not result in major defects[6,33], indicating the considerable plasticity of stop codon usage in eukaryotes too. Motivated by the mystery of stop codons and the ambition to construct a versatile cellular factory using synthetic *P. patens*, we have decided to recode stop codons, unifying them to TAA (Fig. 1d and Supplementary Fig. 7). This approach results in an extreme condition of stop codon usage that can be used to study translation termination in *P. patens* while also providing a foundation for the production of modified proteins with non-standard amino acids[38]. In cases where multiple transcripts are present (due to alternative splicing, for example), each stop codon is recoded.

Incorporation of the SCRaMbLE system

Alterations in the structure and function of the *P. patens* genome will provide a unique opportunity to elucidate the fundamental principles underlying moss genomes. The SCRaMbLE system[39–43], employed in the Sc2.0 project, uses Cre recombinase to catalyse recombination between loxPsym sites, thereby enabling the induction of various genome rearrangements at the whole-genome level. In SynMoss, loxPsym sequences will also be introduced 3 bp downstream of the stop codon of each gene to facilitate SCRaMbLE. In cases where a gene produces multiple transcripts, the loxPsym sequence will be positioned 3 bp downstream of the last CDS region of the gene (Fig. 1e). We intend to meticulously monitor the effects of inserting loxPsym sites into the 3′ untranslated regions (UTRs) of genes. If any gene function is adversely affected by these insertions, it will provide an opportunity to explore how modifying the 3′ UTR may alter gene expression and localization in plants. In such cases, we will use genome editing tools to reposition the loxPsym sites.

Manipulation of non-coding RNAs

Many non-coding RNAs (ncRNAs), such as ribosomal RNA, transfer RNA (tRNA), small nucleolar RNA, microRNA and others, primarily function as translation machineries and regulatory elements[44]. In the *P. patens* genome, ribosomal DNAs are found in tandem repeat arrays on chromosomes 6, 20 and 26 (Supplementary Fig. 1). Due to the potential impact on the genome's stability, these ribosomal DNA arrays are maintained in their original locations on synthetic chromosomes to prevent major defects. While it is theoretically possible to consolidate all tRNA genes into a newly synthesized artificial chromosome (as demonstrated in Sc2.0), given the current limitations in understanding

https://doi.org/10.1038/s41477-024-01693-0

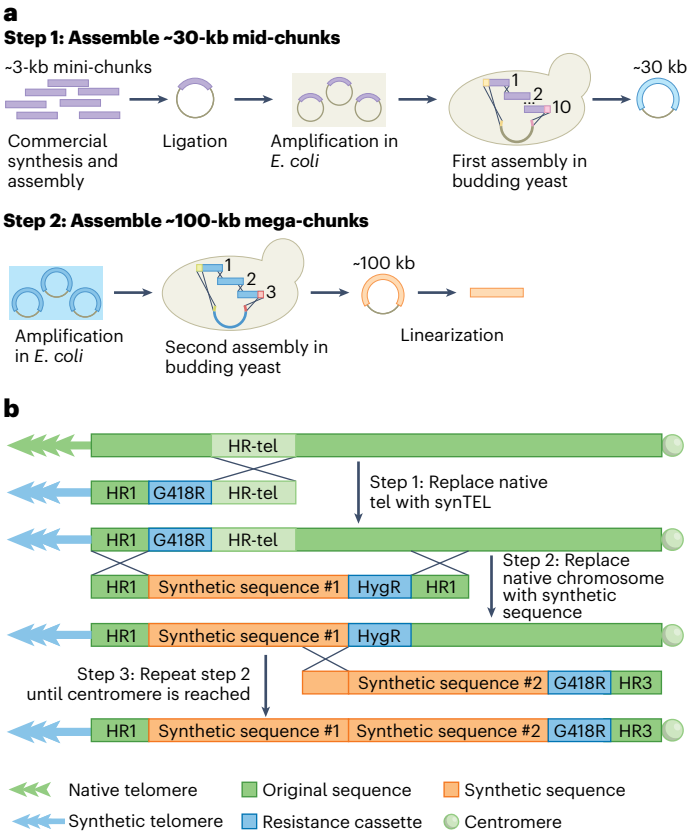

Fig. 2 | Strategy for constructing the SynMoss genome. a, Flowchart for the assembly of synthetic sequences. This procedure consists of two major steps. First, approximately 3-kbp commercially synthesized and assembled mini-chunks (purple) are assembled into about 30-kbp mid-chunks (blue) using homologous recombination in yeast. Second, these three mid-chunks are further assembled into approximately 100-kbp mega-chunks (orange) within the yeast system. These mega-chunks will be introduced into *P. patens* cells to replace the native sequence. **b**, Sequential replacement of native sequences by synthetic sequences from telomere to centromere. Starting from one telomere, each successive synthetic segment incorporates a resistance cassette (serving as a selective marker) positioned on the interior side of the adjoining homologous arm, oriented towards the centromere. The replacement is iteratively conducted until an entire chromosomal arm has been successfully replaced.

plant centromeres and the absence of robust artificial chromosomes, we decided to retain these tRNA genes in their original genomic locations. Besides ribosomal RNAs and tRNA genes, the *P. patens* genome houses over 500 other ncRNA genes, scattered indiscriminately across the genome (Supplementary Figs. 1 and 7). Due to limited data and knowledge about these ncRNAs, such as small nucleolar RNAs and microRNAs, these ncRNAs are not specifically targeted and may be deleted due to the aforementioned design criteria. Throughout the assembly and testing of the synthetic chromosomes, close monitoring of phenotypic changes will guide decisions on whether to reintroduce these ncRNA fragments.

Strategy to construct a synthetic *P. patens* chromosome

Although *P. patens* demonstrates a higher recombination efficiency than other model plants such as *Arabidopsis thaliana*[45], incorporating DNA fragments over 10 kbp into the genome has remained a challenge for many years[46]. During our pilot study concerning the synthesis of the short arm of *P. patens* chromosome 18 (ref. 32), we successfully demonstrated the feasibility of incorporating synthetic DNA chunks with a length of about 100 kbp in *P. patens*. As a result, we propose a general pipeline as a standard strategy for *P. patens* genome synthesis (Fig. 2).

The process begins by initially dividing the synthetic chromosome into mega-chunks of approximately 100 kbp. These mega-chunks are then further divided into mid-chunks of around 30 kbp, which are subsequently segmented into mini-chunks of approximately 3 kbp. Each mini-chunk is designed to have 150-bp overlaps with adjacent fragments, allowing chemical synthesis and assembly in yeast by homologous recombination. Both free ends of these mega-chunks contain 1-kbp regions that are identical to endogenous sequences in *P. patens*. Adjacent to one of the two regions for homologous recombination, a marker gene for selection (for example, a hygromycin resistance cassette) is inserted (Fig. 2b).

Ten mini-chunks are assembled into a mid-chunk through homologous recombination in the budding yeast. Similarly, three mid-chunks can be assembled into a mega-chunk using the YCp/BAC vector (Fig. 2a); these are subsequently prepared from *E. coli* and released by restriction digestion. The resulting linearized mega-chunk is then integrated into the host genome via PEG-mediated transformation, replacing the corresponding original segment (Fig. 2b). The recovered antibiotic-resistant lines will be screened using both synthetic and wild-type PCRmarks to identify candidates containing only the synthetic sequences (Supplementary Fig. 8).

Open-source software for synthetic genome design

GenoDesigner consists of four crucial modules: genome visualization, genome manipulation, biological parts management and task management. The main functions and relevant interfaces are shown in Fig. 3a. The detailed instructions for GenoDesigner can be found on GitHub (https://github.com/SynMoss/GenoDesigner) and the GenoDesigner server (http://49.51.34.198:3000/client/chromosome/en/?#/).

Genome visualization

This capability allows researchers to swiftly pinpoint the genomic location of genes or functional sequences, evaluate their potential upstream or downstream effects and analyse unexpected DNA losses or mutagenesis occurring in synthesized genomes. While traditional molecular biology tools such as SnapGene (www.snapgene.com) can rapidly visualize plasmid genomes several kilobases in size, they fall short when it comes to loading genomes at the scale of hundreds of megabases. GenoDesigner, through systematic optimization of the back-end algorithm, can now upload gigabase-level genomes within minutes on a computer with a quad-core processor. It supports genome files in the .gb, .gbk and .gbff formats from local computers, which can be loaded either separately or together. Furthermore, GenoDesigner can load .gtf, .gff, .fa, .fasta and .fna genome files separately from a local computer. GenoDesigner supports four types of views: linear map, circular map, table map (feature list) and sequence map. Users can easily browse sequences and switch from these views.

Genome manipulation

GenoDesigner empowers users to cut, copy, paste, insert and replace any sequence graphically on the genome. The platform also allows users to create, edit and delete features on the current genome.

Besides basic editing functions, GenoDesigner supports many global operations. The Delete by Feature Type function allows users to remove a designated type of featured sequences at the chromosome or whole-genome level. This function can be incredibly useful as current genome synthesis designs tend to delete numerous repetitive regions such as TEs from the genome. Considering that most genome simplification efforts aim to concatenate the remaining genome, GenoDesigner allows users to achieve this with the function Truncate Genome. The replacement of stop codons and global insertions can also be achieved by GenoDesigner.

https://doi.org/10.1038/s41477-024-01693-0

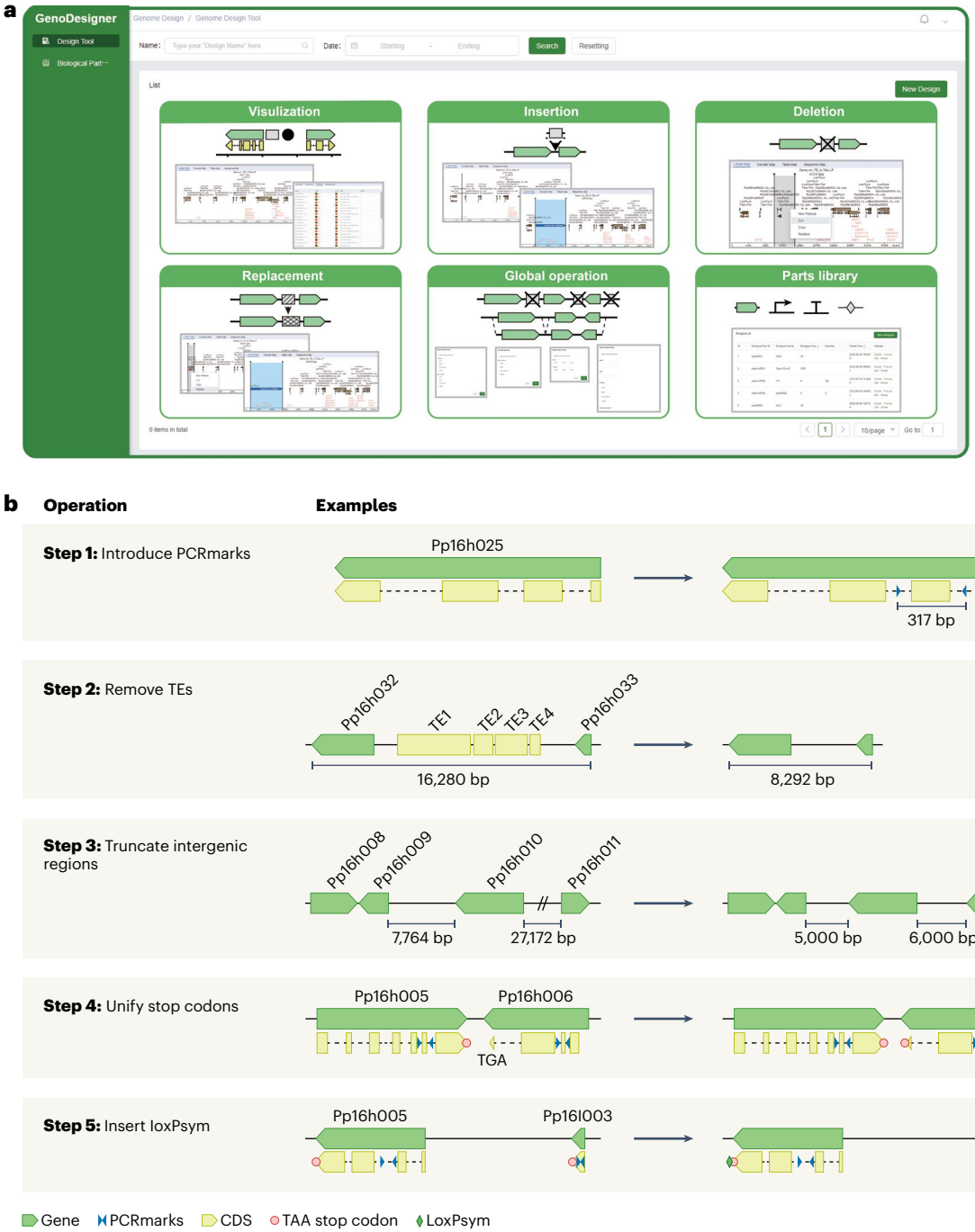

Fig. 3 | Overview of the GenoDesigner software. a, GenoDesigner provides a series of functions and interfaces. It facilitates genome visualization; enables genome editing such as insertion, deletion and replacement; and streamlines genome design through global operation functions and systematic management of biological parts. **b**, Synthetic *P. patens* Chr16L generated by GenoDesigner. The process of generating synthetic sequences from the wild type involves five designed steps: introduce PCRmarks, remove TEs, truncate intergenic regions, unify stop codons and insert loxPsym. Examples of each step are illustrated in the middle, and the output modifications made in each step are listed on the right.

Additionally, GenoDesigner provides the History function, enabling users to revert to a previous version. With these functionalities, GenoDesigner provides a comprehensive workflow for graphically designing a synthetic genome (see the file Example_of_GenoDesigner.pdf on GitHub).

Management of biological parts

The concept of a parts-based design has been integral to synthetic biology since its inception[47]. Several libraries of biological parts have been constructed, making a substantial contribution to the establishment of functional gene pathways or the engineering of protein expression cassettes. Integrating biological parts into synthetic genome design at specific genomic locations or across the entire genome has the potential to enrich the diversity of genome design. Consequently, GenoDesigner provides a valuable section for biological parts management. Users can define biological parts and add them into the biological parts library. These parts are displayed with their ID, name, length and functional description. GenoDesigner allows users to edit or delete these parts.

https://doi.org/10.1038/s41477-024-01693-0

Table 1 | Modifications in the SynMoss genome

Chromosome	Length of WT chromosome (bp)	Length of SYN chromosome (bp)	Length of TEs (bp)	No. of PCRmarks introduced[a]	Length of TEs removed (bp)	Length of intergenic regions removed (bp)	No. of stop codons replaced[b]	No. of LoxPsym inserted[c]
Chr01	31,557,254	10,993,845	18,832,377	1,856	18,669,469	1,957,452	1,393	1,868
Chr02	26,289,282	10,320,165	14,445,971	1,704	14,165,842	1,861,993	1,256	1,727
Chr03	25,956,603	9,907,049	14,320,292	1,672	14,154,323	1,953,201	1,296	1,705
Chr04	22,798,254	8,236,368	13,096,561	1,339	12,911,542	1,696,924	972	1,370
Chr05	21,028,244	7,285,440	12,514,804	1,195	12,266,318	1,517,830	878	1,216
Chr06	21,236,458	9,017,048	10,746,959	1,271	10,493,097	1,770,445	975	1,298
Chr07	18,499,901	7,332,577	10,135,618	1,205	9,955,939	1,253,103	915	1,227
Chr08	18,448,947	6,412,237	11,002,486	1,070	10,839,613	1,234,395	791	1,097
Chr09	18,012,735	6,702,321	10,073,388	1,070	9,913,536	1,433,904	832	1,089
Chr10	17,932,518	6,769,360	10,145,039	1,101	9,860,980	1,340,122	838	1,116
Chr11	17,862,078	7,132,877	9,825,983	1,199	9,606,392	1,164,425	902	1,224
Chr12	17,639,930	6,222,313	10,450,116	1,035	10,288,322	1,165,267	772	1,058
Chr13	17,710,172	5,540,906	11,297,421	983	11,091,090	1,111,054	714	967
Chr14	17,463,378	7,148,419	9,260,384	1,198	9,078,030	1,278,341	888	1,218
Chr15	17,178,229	6,601,468	9,390,789	1,070	9,253,174	1,360,103	793	1,074
Chr16	16,868,143	6,686,151	9,180,035	1,142	9,003,051	1,218,143	876	1,153
Chr17	16,300,214	6,309,073	8,972,048	1,028	8,780,766	1,245,565	791	1,035
Chr18	16,375,218	5,361,312	10,026,345	882	9,832,537	1,211,493	673	886
Chr19	16,049,963	5,160,013	9,973,812	905	9,736,634	1,183,950	656	901
Chr20	17,006,457	7,220,312	8,887,472	946	8,664,610	1,154,175	686	960
Chr21	15,939,076	5,424,899	9,721,228	938	9,441,880	1,104,529	674	948
Chr22	15,267,569	5,322,327	9,070,563	891	8,883,791	1,091,847	675	894
Chr23	15,418,210	5,586,410	9,205,948	922	9,044,774	818,952	688	939
Chr24	14,190,027	5,311,491	8,166,570	920	7,964,950	945,240	664	931
Chr25	16,757,958	5,961,977	9,949,145	1,004	9,757,145	1,073,176	742	1,010
Chr26	11,963,395	5,372,705	5,968,580	680	5,777,739	836,513	523	693
Total	**481,750,213**	**179,339,063**	**274,659,934**	**29,226**	**269,435,544**	**33,982,142**	**21,863**	**29,604**

[a]A PCRmark consists of a pair of forward and reverse primers. [b]Stop codons of different CDSs at the same position are counted as one. [c]Insertion sites are not within other CDSs.

Task management

Current genome synthesis projects often involve collaboration among multiple labs worldwide. To meet this demand, GenoDesigner allows users to maintain tasks under their own accounts by typing in their username. Built on this, users can create a new task for further genome editing or rewriting. Completed tasks can be packed and downloaded, including .fasta files, GenBank files and historical records. The packed task can also be cloned for further editing or saved as a template for further use. Users can quickly locate their tasks by searching the task name and updating time.

Design of SynMoss using GenoDesigner

Here we present the design of the left arm of Chr16 (Chr16L, 1.44 Mbp) to illustrate the design process of SynMoss using GenoDesigner (Fig. 3b). First, 49 PCRmarks were designed using a customized R script and integrated using the Open Editing File function of GenoDesigner, leading to a sequence change of 2,392 bp. Second, 1.04 Mbp of TEs were removed by applying the Delete by Feature Type function, resulting in a remaining chromosome arm of 398.5 kbp. Third, intergenic regions exceeding 3 kbp upstream and 2 kbp downstream of each protein-coding gene were deleted using the Truncate Genome function, totalling 109.8 kbp. Fourth, using the Replace Stop Codons function, all stop codons were identified, and 39 TAG and TGA stop codons were replaced by TAA. However, two stop codons that overlapped with other coding regions remained unchanged and required manual inspection. Lastly, 49 loxPsym sites were introduced 3 bp downstream of the last stop codon of each gene using the Insert by Feature Type function. The structure of synthetic Chr16L, termed Ppsyn16L, is shown in Supplementary Fig. 9. With a similar workflow, the whole SynMoss genome was generated. The details of the modifications in SynMoss are listed in Table 1.

Discussion

Evoking the wisdom of renowned physicist and Nobel laureate Richard Feynman—"What I cannot create, I do not understand"—we emphasize the transformative power of genome synthesis. Since the inaugural synthesis of the poliovirus in 2002[7], genome synthesis has emerged as a pivotal area of exploration within the scientific community, deepening our comprehension of the intricacies of life. Embarking on a new frontier, we advocate for the synthesis of the *P. patens* genome as a preliminary step in plant genome synthesis, facilitated by accessible design principles and cutting-edge tools. The synthesis of the *P. patens* genome is anticipated to enrich our understanding of plant evolution and development.

To expedite the synthesis of the *P. patens* genome, we have introduced GenoDesigner, a tool designed to address the need for universal genome design software. GenoDesigner offers comprehensive genome visualization capacities, crucial for researchers requiring immediate insights into gene locations and overall genome structure. Notably, GenoDesigner supports visualization at the gigabase level, potentially benefiting not only the synthesis of the *P. patens* genome but also the exploration of other plant and animal genomes. Furthermore, GenoDesigner provides full genome manipulation capacities, including Delete by Feature Type, Truncate Genome, Replace Stop Codons, Insert by Feature Type and others, thereby facilitating intuitive genome design. While GenoDesigner is currently tailored to specific purposes, future enhancements are anticipated as more scientists use the tool and contribute feedback, thereby bolstering the development of robust engineering software for genome synthesis. We also provide Python and R scripts here (https://github.com/WenfeiY/P.patens_geno_design) for genome design as a supplement to GenoDesigner to mitigate GenoDesigner's potential drawbacks such as PCRmark design. These scripts can design PCRmarks for most genes and generate drafts of synthetic sequences of chromosome arms in a command-line manner.

The *P. patens* genome, comprising 26 chromosomes spanning approximately 480 Mb[30], represents a formidable challenge for scientific enquiry. Even with the complete elimination of potentially redundant TEs, a minimum of approximately 200 Mb of the genome requires synthesis. This immense task necessitates a collective effort on a much larger scale than that of the Sc2.0 project. We propose an open global collaborative approach for *P. patens* genome synthesis, providing detailed experimental protocols and expertise to interested scientists. Furthermore, the SynMoss project is expected to encounter various unforeseen challenges, such as biological variations among individuals, difficulties in integrating multiple synthetic chromosomes into a single plant and inevitable large-fragment losses or base substitutions during genome synthesis and manipulation. We propose that strict adherence to the design sequences may not be crucial, as long as the overall genome functions effectively. This approach could expedite the synthesis process. The acceptable degree of deviation should be a subject for future discussion within the *P. patens* genome synthesis scientific community, especially as the project progresses and more data become available. Our goal is not to create a genetically precise artefact with 100% sequence accuracy, but to use synthetic biology to deepen our understanding and utility of organisms.

The ability to incorporate large synthetic modules into the *P. patens* genome positions this model plant as a robust platform for plant synthetic biology applications. Moreover, this endeavour has the potential to revolutionize our understanding of plant biology. By leveraging the synthetic genome platform and technologies developed during this project, *P. patens* could become a pivotal host for the production of plant-derived medicinal and biotechnological products. Potential applications are expansive. For example, the production of plant-derived natural products or medicines could be considerably enhanced through the evolution of the synthetic organism, specifically using the introduced SCRaMbLE system.

We welcome the participation of DNA synthesis companies, especially those harnessing innovative technologies such as enzymatic synthesis and chip-based DNA synthesis, in this endeavour. Recognizing the trend of diminishing costs for commercial DNA synthesis, we anticipate that the per-base cost for the complete synthesis of the *P. patens* genome will be substantially lower than for the Sc2.0 project[33]. With such advancement, the next stage of the *P. patens* genome synthesis could focus more on gene rearrangement or the introduction of exogenous genes. This would provide a broader lens to probe the evolution of life and find applications across a wider spectrum. Thus, through this project, we aim to go beyond merely understanding life—we aspire to reshape it and to optimize its potential for the betterment of humankind. With that, we strongly advocate for greater financial support from both government and non-governmental organizations towards this ground-breaking project.

Methods

Genome data analysis

All genome data used in the design of SynMoss are from our recent work (https://doi.org/10.6084/m9.figshare.22975925.v5), including genome sequences and annotations. The file Physcomitrium_patens_V4_genome.fasta records a new version of the genome sequences of *P. patens*. The file Physcomitrium_patens_V4_rename.gff3 contains annotations of genes, mRNAs, CDSs and UTRs. The file Pp.genome.repeat.out contains predicted repetitive sequences in *P. patens* generated by RepeatMasker (https://www.repeatmasker.org/). The tRNA annotations came from the file Pp.tRNA.gff3, while the other ncRNA annotations came from the file Pp.rfam.gff3.

All TE annotations were extracted from Pp.genome.repeat.out by filtering out simple repeats and ribosomal RNAs. The total length of TEs was computed by summing up the lengths of all TE regions, excluding the overlapping areas of TEs, which should not be counted repeatedly. Stop codons at the same positions were counted once. The circus plot displaying genome features was generated by the R package circlize[48].

Details of genome visualization by GenoDesigner

The code, instructions and example data for GenoDesigner can be found on GitHub (https://github.com/SynMoss/GenoDesigner). While uploading sequences in GenoDesigner, the percentage of genome uploading is shown in a progress window. After uploading, users can choose to select one chromosome from the genome file for viewing the genome as a linear map, circular map, table map or sequence map. The features such as genes and CDSs that are annotated in the original loaded file will be graphically shown on the map. In general, different styles of the genome map are shown in separate windows, but users can drag two types of maps into one window to compare the sequence and its structure. Notably, to accelerate genome manipulation, only 100,000-bp sequences from the chromosome or genome are shown in the window at one time by default, but users can drag the sliding bar and adjust the display settings to see the whole genome. Moreover, GenoDesigner supports showing multiple chromosomes in one window to compare their sequence structure. Overall, GenoDesigner provides a quick and robust way to visualize small to enormous genomes.

Details of genome manipulation functions

The Cut function allows users to cut sequences from a selected area or defined region of the genome and temporarily save the sequence for further paste manipulation. The Copy function allows users to copy a specific region of the genome sequence or the whole chromosome and insert or paste this part later at other genomic locations. The Paste function allows users to paste the cut or copied sequence into a specific genomic location. The Insert function allows users to insert a fragment of a new sequence into a specific region. The Replace function allows users to replace a specific genomic region with other sequences or replace several nucleotides or a defined feature of the sequence at the whole-genome level. The Delete by Feature Type function enables users to delete sequences of a certain feature type. The Truncate Genome function allows users to preserve certain types of features, along with their upstream and downstream regions, and delete the rest of the genome. The Replace Stop Codons function can detect stop codons and allows users to replace one kind of stop codon with another (for example, TAG to TAA) or even unify them. The Insert by Feature Type function allows users to insert any element globally on the basis of feature types and relevant positions.

Implementation of GenoDesigner

GenoDesigner is written in Python and Java using a B/S system architecture that can be accessed using a browser. It can be implemented

on the server with a four-core CPU, 16 GB memory and 500 GB hard drive. GenoDesigner creates its own logic for genome-scaled sequence editing while integrating the visual style of the free Open Vector Editor (https://github.com/TeselaGen/openVectorEditor), a plasmid editing tool. The overall software adopts a front-end and back-end separation development mode. The front end uses Vue.js and React.js as visual rendering and human–machine interaction frameworks and supports parsing and rendering of large files (genome files above 200 M). The back end adopts a microservices architecture and uses SpringCloud as the framework, ensuring software stability while being fully compatible with heterogeneous language integration.

Design workflow of SynMoss

The workflow of generating SynMoss sequences comprises inserting PCRmarks, unifying stop codons, removing TEs, reducing intergenic areas and inserting loxPsym sites. The files containing genome sequences and annotations (genes, CDSs, TEs and so on) of *P. patens* are required during the construction of SynMoss. The example data and instructions for additional Python and R scripts can be found on GitHub (https://github.com/WenfeiY/P.patens_geno_design).

The R script PCRmark_design.R was used earlier to design PCRmarks for each gene. We employed Primer3 (ref. 35) to calculate the melting temperatures of candidate PCRmarks, guaranteeing that the generated PCRmarks can work between 57 and 68 degrees Celsius. Furthermore, to ensure the uniqueness of all PCRmarks within their respective wild-type or synthetic sequences, we performed in silico PCR on the PCRmarks using the AmplifyDNA function from the DECIPHER package[49].

While designing SynMoss through GenoDesigner, a semi-synthetic sequence with PCRmarks can be generated by opening the genome and the PCRmark files such as PpChr16L_PCRmark_editing.txt in GenoDesigner. The TE sequences can then be eliminated by using the Delete by Feature Type function, choosing the feature type of TEs and selecting 'gene', 'tRNA', 'centromere' and so on to be protected. The intergenic regions will next be shortened by using Truncate Genome, selecting 'gene', 'tRNA', 'centromere' and so on to preserve, and entering the required lengths to be retained upstream and downstream (3,000 and 2,000, 500 and 200, and 0 and 0 in this case, respectively). TAG and TGA stop codons can then be unified to TAA using the Replace Stop Codons function. Finally, loxPsym sites will be inserted 3 bp downstream of the last CDS region of each gene by pressing the Insert by Feature Type button and inputting the information for loxPsym (name, LoxPsym; type, protein_bind; sequence, ATAACTTCGTATAATGTACATTATACGAAGTTAT). Synthetic sequences generated by GenoDesigner can be further customized by changing centromeres, telomeres and other components of interest. The finished synthetic sequences can be download and subjected to synthesis.

The Python script Pp.syn.generate.py can be used for generating synthetic sequences of chromosome arms of SynMoss. BioPython[50] is employed in this script to manipulate sequence objects. The script allows for the importing of PCRmarks, the removal of all 'misc_feature' except the regions inside genes, the elimination of intergenic regions exceeding 3 kbp upstream and 2 kbp downstream of each gene, the replacement of all TAG and TGA stop codons with TAA, and the insertion of loxPsym sites 3 bp downstream of the last stop codon of each gene.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All genome data used in the design of SynMoss are available via Figshare at https://doi.org/10.6084/m9.figshare.22975925.v5 (ref. 51).

Code availability

The source code and instructions for GenoDesigner are available at https://github.com/SynMoss/GenoDesigner. The server is accessible at http://49.51.34.198:3000/client/chromosome/en/#/. Related Python and R programs are available at https://github.com/WenfeiY/P.patens_geno_design.

References

1. Sinyakov, A. N., Ryabinin, V. A. & Kostina, E. V. Application of array-based oligonucleotides for synthesis of genetic designs. *Mol. Biol. (Mosk.)* **55**, 562–577 (2021).
2. He, B. et al. YLC-assembly: large DNA assembly via yeast life cycle. *Nucleic Acids Res.* **51**, 8283–8292 (2023).
3. Hanna, R. E. & Doench, J. G. Design and analysis of CRISPR–Cas experiments. *Nat. Biotechnol.* **38**, 813–823 (2020).
4. Gibson, D. G. et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
5. Dymond, J. S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).
6. Fredens, J. et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019).
7. 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).
8. Smith, H. O., Hutchison, C. A. 3rd, Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl Acad. Sci. USA* **100**, 15440–15445 (2003).
9. Gibson, D. G. et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
10. Annaluru, N. et al. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–58 (2014).
11. Mitchell, L. A. et al. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* https://doi.org/10.1126/science.aaf4831 (2017).
12. Shen, Y. et al. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* https://doi.org/10.1126/science.aaf4791 (2017).
13. Vashee, S. et al. Cloning, assembly, and modification of the primary human cytomegalovirus isolate toledo by yeast-based transformation-associated recombination. *mSphere* **2**, e00331–00317 (2017).
14. Wu, Y. et al. Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* https://doi.org/10.1126/science.aaf4706 (2017).
15. Xie, Z. X. et al. 'Perfect' designer chromosome V and behavior of a ring derivative. *Science* https://doi.org/10.1126/science.aaf4704 (2017).
16. Blount, B. A. et al. Synthetic yeast chromosome XI design provides a testbed for the study of extrachromosomal circular DNA dynamics. *Cell Genom.* https://doi.org/10.1016/j.xgen.2023.100418 (2023).
17. Foo, J. L. et al. Establishing chromosomal design–build–test–learn through a synthetic chromosome and its combinatorial reconfiguration. *Cell Genom.* https://doi.org/10.1016/j.xgen.2023.100435 (2023).
18. Lauer, S. et al. Context-dependent neocentromere activity in synthetic yeast chromosome VIII. *Cell Genom.* https://doi.org/10.1016/j.xgen.2023.100437 (2023).
19. Luo, J. et al. Synthetic chromosome fusion: effects on mitotic and meiotic genome structure and function. *Cell Genom.* https://doi.org/10.1016/j.xgen.2023.100439 (2023).

20. McCulloch, L. H. et al. Consequences of a telomerase-related fitness defect and chromosome substitution technology in yeast synIX strains. *Cell Genom*. https://doi.org/10.1016/j.xgen.2023.100419 (2023).
21. Schindler, D. et al. Design, construction, and functional characterization of a tRNA neochromosome in yeast. *Cell* https://doi.org/10.1016/j.cell.2023.10.015 (2023).
22. Shen, Y. et al. Dissecting aneuploidy phenotypes by constructing Sc2.0 chromosome VII and SCRaMbLEing synthetic disomic yeast. *Cell Genom*. https://doi.org/10.1016/j.xgen.2023.100364 (2023).
23. Williams, T. C. et al. Parallel laboratory evolution and rational debugging reveal genomic plasticity to *S. cerevisiae* synthetic chromosome XIV defects. *Cell Genom*. https://doi.org/10.1016/j.xgen.2023.100379 (2023).
24. Zhang, W. et al. Manipulating the 3D organization of the largest synthetic yeast chromosome. *Mol. Cell* **83**, 4424–4437.e5 (2023).
25. Zhao, Y. et al. Debugging and consolidating multiple synthetic chromosomes reveals combinatorial genetic interactions. *Cell* **186**, 5220–5236.e16 (2023).
26. Boeke, J. D. et al. The Genome Project-Write. *Science* **353**, 126–127 (2016).
27. Dai, J., Boeke, J. D., Luo, Z., Jiang, S. & Cai, Y. Sc3.0: revamping and minimizing the yeast genome. *Genome Biol.* **21**, 205 (2020).
28. 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).
29. Lang, D. et al. The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–533 (2018).
30. Bi, G. et al. Near telomere-to-telomere genome of the model plant *Physcomitrium patens*. *Nat. Plants* **10**, 327–343 (2024).
31. Ye, Z.-H. & Zhong, R. Cell wall biology of the moss *Physcomitrium patens*. *J. Exp. Bot.* **73**, 4440–4453 (2022).
32. Chen, L.-G. et al. A designer synthetic chromosome fragment functions in moss. *Nat. Plants* **10**, 228–239 (2024).
33. Richardson, S. M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
34. Liu, W. & Stewart, C. N. Plant synthetic promoters and transcription factors. *Curr. Opin. Biotechnol.* **37**, 36–44 (2016).
35. Untergasser, A. et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
36. Ostrov, N. et al. Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
37. Robertson, W. E. et al. Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**, 1057–1062 (2021).
38. Venter, J. C., Glass, J. I., Hutchison, C. A. & Vashee, S. Synthetic chromosomes, genomes, viruses, and cells. *Cell* **185**, 2708–2724 (2022).
39. Dymond, J. & Boeke, J. The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioeng. Bugs* **3**, 168–171 (2012).
40. Shen, Y. et al. SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res.* **26**, 36–49 (2016).
41. Mercy, G. et al. 3D organization of synthetic and scrambled chromosomes. *Science* https://doi.org/10.1126/science.aaf4597 (2017).
42. Blount, B. A. et al. Rapid host strain improvement by in vivo rearrangement of a synthetic yeast chromosome. *Nat. Commun.* **9**, 1932 (2018).
43. Jia, B. et al. Precise control of SCRaMbLE in synthetic haploid and diploid yeast. *Nat. Commun.* https://doi.org/10.1038/s41467-018-03084-4 (2018).
44. Anastasiadou, E., Jacob, L. S. & Slack, F. J. Non-coding RNA networks in cancer. *Nat. Rev. Cancer* **18**, 5–18 (2018).
45. Kobayashi, W. et al. Homologous pairing activities of *Arabidopsis thaliana* RAD51 and DMC1. *J. Biochem.* **165**, 289–295 (2019).
46. Kamisugi, Y. et al. The mechanism of gene targeting in *Physcomitrella patens*: homologous recombination, concatenation and multiple integration. *Nucleic Acids Res.* **34**, 6205–6214 (2006).
47. Galdzicki, M., Rodriguez, C., Chandran, D., Sauro, H. M. & Gennari, J. H. Standard Biological Parts Knowledgebase. *PLoS ONE* **6**, e17005 (2011).
48. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
49. Wright, E. S. Using DECIPHER v2.0 to analyze big biological sequence data in R. *R J.* **8**, 352–359 (2016).
50. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
51. Zhao, S. Telomere-to-telomere (T2T) genome of the model plant *Physcomitrium patens*. *Figshare* https://doi.org/10.6084/m9.figshare.22975925.v5 (2023).

Acknowledgements

This study was supported by the National Key R&D Program of China (grant nos 2022YFF1201800, 2021YFF1201700 and 2019YFA0903900), the National Natural Science Foundation of China (grant no. 32201207), the Innovation Program of the Chinese Academy of Agricultural Sciences, the Shenzhen Science and Technology Program (grant nos KQTD20180413181837372 and RCYX20221008092950122), the Shenzhen Outstanding Talents Training Fund, the CAS Project for Young Scientists in Basic Research (grant no. YSBR-078) and the Strategic Priority Research Program (Precision Seed Design and Breeding, grant no. XDA24020103).

Author contributions

J.Y., Q.Z., B.M., Y.W., Y.J., Y.M., X.H., W.Q. and J.D. conceived and designed the project. X.H., W.Q. and J.D. supervised the research. W.Y., S. Zhang and X.H. completed the programming. S. Zhao analysed the RNA-seq data. L.-g.C., J.C. and H.Y. performed the experiments. W.Y., S. Zhang, X.H., W.Q. and J.D. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-024-01693-0.

Correspondence should be addressed to Xiaoluo Huang, Wenfeng Qian or Junbiao Dai.

Peer review information *Nature Plants* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Xiaoluo Huang, Wenfeng Qian & Junbiao Dai

Last updated by author(s): Mar 25, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect data in this study.
Data analysis	Deeptools was used for transcription start & stop sites analysis (https://deeptools.readthedocs.io/en/latest/index.html, Extended Data Fig. 5). The source code and instructions of GenoDesigner is available on Github repository (https://github.com/SynMoss/GenoDesigner). The GenoDesigner server is freely available at http://49.51.34.198:3000/client/chromosome/en/#/.The custom Python & R scripts used for genome design are available on Github repository (https://github.com/WenfeiY/P.patens_geno_design).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Genome sequence and annotations are available at Figshare repository (https://doi.org/10.6084/m9.figshare.22975925.v4). The source code and instructions of

GenoDesigner are freely available at https://github.com/SynMoss/GenoDesigner. GenoDesigner server is freely available at http://49.51.34.198:3000/client/chromosome/en/#/. The additional Python and R programs attached to GenoDesigner are freely available at https://github.com/WenfeiY/P.patens_geno_design.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	N/A
Data exclusions	N/A
Replication	N/A
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	The wild-type strain used in this study is the Gransden ecotype.
Novel plant genotypes	N/A
Authentication	N/A